BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF TELECOMMUNICATIONS
ÚSTAV TELEKOMUNIKACÍ

COMPUTERISED ASSESSMENT OF GRAPHOMOTOR AND HANDWRITING DISABILITIES IN SUBJECTS WITH NEURODEGENERATIVE AND NEURODEVELOPMENTAL DISORDERS

POČÍTAČOVÉ HODNOCENÍ GRAFOMOTORICKÝCH OBTÍŽÍ A OBTÍŽÍ S PSANÍM U OSOB S NEURODEGENERATIVNÍMI ONEMOCNĚNÍMI A NEUROVÝVOJOVÝMI PORUCHAMI

HABILITATION THESIS
HABILITAČNÍ PRÁCE

AUTHOR — Ing. Jiří Mekyska, Ph.D.
AUTOR PRÁCE

BRNO 2022

ABSTRACT

People with neurodevelopmental (e.g. developmental dysgraphia) or neurodegenerative (e.g. Parkinson's disease) disorders are very likely to exhibit graphomotor disabilities (GD) such as motor-memory dysfunction, graphomotor production deficits, motor feedback difficulties, etc., leading to various drawing and handwriting difficulties (HD). These alterations have a very detrimental impact on the quality of life. Unfortunately, current diagnostic methods have many limitations, which results in underdiagnosis or incorrect diagnosis of GD/HD and consequently in ineffective therapy. In recent years, online handwriting processing proved to be a promising approach to an objective and accurate assessment of GD/HD. Nevertheless, the field is still relatively unexplored and has several knowledge gaps. The main goal of this habilitation thesis is to progress beyond the state of the art and to research new approaches to the computerised assessment of GD/HD that would facilitate objective diagnosis and monitoring of neurodegenerative and neurodevelopmental disorders. The thesis summarises 34 peer-reviewed works that bridge the main knowledge gaps, and provides new directions in the field.

KEYWORDS

graphomotor disabilities, handwriting disabilities, neurodegenerative disorders, neurodevelopmental disorders, dysgraphia, Parkinson's disease, online handwriting, assessment, diagnosis

ABSTRAKT

U osob s neurovývojovými poruchami (např. s neurovývojovou dysgrafií) nebo neurodegenerativními onemocněními (např. s Parkinsonovou nemocí) je velká pravděpodobnost výskytu grafomotorických obtíží (GD), jako např. motoricko-paměťové dysfunkce, poruchy grafomotoriky, potíže s motorickou zpětnou vazbou atd., což vede dále k různým potížím s kreslením a psaním (HD). Tyto poruchy mají velmi negativní dopad na kvalitu života. Bohužel, aktuální diagnostické metody manifestují mnoho nedostatků, což vede ke špatné diagnóze GD/HD, a následně k neefektivní terapii. V posledních letech se prokázalo, že je zpracování online písma slibným přístupem k objektivnímu a přesnému hodnocení GD/HD. Nicméně tato oblast je stále poměrně neprozkoumána a obsahuje mnoho mezer ve znalostech. Hlavním cílem této habilitační práce je postoupit za aktuální stav vědění a zkoumat nové přístupy hodnocení GD/HD, které by ulehčily objektivní diagnózu a monitorování neurodegenerativních onemocnění a neurovývojových poruch. Práce shrnuje 34 recenzovaných článků, které překlenují zmíněné mezery, a poskytují nové směry v oblasti.

KLÍČOVÁ SLOVA

grafomotorické obtíže, potíže s psaním, neurodegenerativní onemocnění, neurovývojové poruchy, dysgrafie, Parkinsonova nemoc, online písmo, hodnocení, diagnóza

MEKYSKA, Jiří. *Computerised assessment of graphomotor and handwriting disabilities in subjects with neurodegenerative and neurodevelopmental disorders*. Brno: Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Telecommunications, 2022, 422 p. Habilitation thesis.

ACKNOWLEDGEMENT

I would like to thank my beloved wife Simona for being a great support during my career and for being a wonderful mother to our three sons (Jiri, Maty, and Miki), who always fill me with happiness and joy. I would also like to acknowledge the great support from the rest of my family and from my close friends.

Big thanks go to Marcos Faundez-Zanuy, my colleague, my friend, my gastronomical guide, and the father of my research, who identified my research potential and helped me to shape my career. He taught me a lot in both my professional and personal life. Simultaneously, I would like to thank Prof. Zdenek Smekal, the second person from a couple of men important to my career. He supervised my bachelor's, master's, and doctoral thesis, enabled me to become a researcher, always provided me good advice, and (most importantly) motivated me to write this thesis. I also humbly admire his knowledge in the field of signal processing and appreciate the many philosophical discussions we had.

Generally, my research directions were strongly influenced by two amazing women, one of them is Irena Rektorova. I am lucky to say that my work is my hobby. More than ten years ago, Irena enabled me to enter the field of paraclinical analysis of neurodegenerative disorders, which became my passion. I love this multidisciplinary research, i.e. the combination of biomedical signal processing, data science and neuroscience. The biggest satisfaction is that the outcomes of my research are not just papers but also the knowledge that helps to understand and fight the neurodegenerative disorders. Thus my heartful thanks go to Irena and her team.

Next, I would like to thank a lot Katarina Safarova, who I met in 2015. She was the one who opened another very interesting field in my career, and that is the computerised assessment of neurodevelopmental dysgraphia. Since this field has many knowledge gaps and challenges, I really enjoy discussions with Katarina and appreciate her drive, interest, and innovativeness. Besides, as a father, I am very happy I could work on something that has a positive impact on children's quality of life.

I would not achieve my scientific results without a great team of colleagues and friends that I have in the Brain Diseases Analysis Laboratory. I would like to thank Zolo, Jano, Vojta, Palo, Dan, Marek, and Michal. Although everyone is completely different, we built a unique organism, which is full of synergy, which is pushing research beyond the state of the art, and which brings fun to our work.

Besides, I would like to thank all other colleagues who participated in our research and who were the important pieces in the final mosaic. Finally, I would like to thank the Department of Telecommunications for being my second home and for providing me with great support.

FUNDING

My work conducted in the frame of this thesis was supported by the following projects:

- Czech Ministry of Health, project no. NU20-04-00294, Diagnostics of Lewy body diseases in prodromal stage based on multimodal data analysis
- Technology Agency of the Czech Republic, project no. TL03000287, Software for advanced diagnosis of graphomotor disabilities
- Czech Ministry of Health, project no. 16-30805A, Effects of non-invasive brain stimulation on hypokinetic dysarthria, micrographia, and brain plasticity in patients with Parkinson's disease
- Czech Science Foundation, project no. 18-16835S, Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing
- Spanish Ministry of Science and Innovation, project no. PID2020-113242RB-I00, Advanced Online Writing Preprocessing Techniques for Device Interoperability
- Horizon 2020, Marie Sklodowska-Curie project CoBeN, Novel Network-Based Approaches for Studying Cognitive Dysfunction in Behavioral Neuology
- COST, project no. IC1206, De-Identification for Privacy Protection in Multimedia Content
- Czech Ministry of Health, project no. NT13499, Speech, its Impairment and Cognitive Performance in Parkinson's Disease
- Czech Ministry of Education, Youth and Sports, project no. CZ.1.07/2.3.00/20.0094, Support for Incorporating R&D Teams in International Cooperation in the Area of Image and Audio Signal Processing
- Czech Ministry of Education, Youth and Sports, project no. ED2.1.00/03.0072, Centre of Sensor, Information and Communication Systems

Contents

Introduction

Handwriting is a complex perceptual-motor skill composed of a coordinated combination of fine graphomotor movements, visual–perceptual abilities, visual–motor coordination, motor planning and execution, kinesthetic feedback, and orthographic coding [1]. These underpinnings of drawing and consequently handwriting abilities are usually considered as graphomotor skills (GS) [2], and should be mastered at the age of 8–9. When a person suffers from a neurodevelopmental (e.g. developmental dysgraphia) or neurodegenerative (e.g. Parkinson's disease) disorder, she/he is very likely to exhibit graphomotor disabilities (GD) such as graphomotor production deficits, motor-memory dysfunction, motor feedback difficulties, etc. [3], leading to various drawing and handwriting difficulties (HD). Such difficulties can have serious consequences, and can greatly affect a person's every-day life starting with slow and less-legible handwriting, lower self-esteem, poor emotional well-being, as well as problematic communication and social interaction [4]. To be able to introduce a timely and effective treatment/therapy and to improve a person's quality of life as much as possible, neurologists, psychologists, special education counselors, and other experts need a theory-based, proven and robust framework that will enable them to diagnose GD and HD in an objective and complex way with minimum manual intervention, cost and time constraints [5].

Nowadays, the most promising approach into robust, objective, and computerised assessment of GD/HD utilises various signals describing the process/product of handwriting/drawing acquired by a digitizing tablet [6]–[9]. Such signals represent movement of a digitizing stylus (pen) on horizontal and vertical axis, pressure exert on the surface of a digitizer, tilt and azimuth, acquired with respect to a specific series of timestamps forming a collection of time-series describing the process of handwriting/drawing from its beginning to the end (referred to as online handwriting) [10]. In addition, modern digitizers have the ability to record not only the movement of a pen on the surface of a digitizer, but also the movement above the surface (in-air movement) [11]. As shown in a variety of studies [12], [13], online handwriting provides us with the capability of going beyond the limitations of human perception and to characterize the handwriting/drawing process in terms of its temporal, spatial, kinematic, and dynamic features.

In recent years, online handwriting has been advantageously used in a variety of research studies focusing on identification and assessment of GD/HD in children experiencing developmental dysgraphia (DD) [9], [14], or in adults suffering from Parkinson's disease (PD), Alzheimer's disease (AD), essential tremor [6]–[8], [15], etc. Although the significant potential of this technique has been proved, the field is still relatively unexplored and has many knowledge gaps that should be bridged before we adopt this technology in practice. This compilation thesis summarises 34 works ([10]–[13], [15]–[44]) with the aim to go beyond the state of the art and to introduce new knowledge and directions in the computerised assessment of GD/HD facilitating objective diagnosis and monitoring of neurodegenerative and neurodevelopmental disorders.

1 Knowledge gaps

1.1 Graphomotor and handwriting disabilities in patients with PD

Parkinson's disease (PD) is a chronic idiopathic disorder characterized by a pathophysiological process of α-synuclein accumulation leading to the formation of Lewy bodies and Lewy neurites resulting in loss/degeneration of dopaminergic neurons in the substancia nigra pars compacta [45]–[47]. It is the second most frequent neurodegenerative disorder, with the prevalence rate estimated to be approximately 2.0 % for people aged over 65 years [48]. To date, the gradual deficiency of dopaminergic neurons in the basal ganglia has been recognized as a major cause of parkinsonian symptoms [49]. In addition to a large variety of other motor symptoms, such as tremor at rest [50], progressive bradykinesia [51], muscular rigidity [50], postural instability [52] and hypokinetic dysarthria [53], one of the prominent and early markers of PD is so-called Parkinson's disease dysgraphia (PDD) [3], [54]–[56].

PDD is a term describing a spectrum of neuromuscular difficulties, including motor-memory dysfunction (problems combining memory input with motor output), graphomotor production deficits (poor muscle coordination), motor feedback difficulties (overactivation of certain muscles and joints during handwriting as well as problems tracking the location of the pen's tip) and others. These cause a variety of HD manifesting as dysfluent, shaky, slow, and less readable handwriting; a progressive decrease in letter amplitude or width, namely, micrographia [3], [57], [58]; etc. Hence, PDD has serious consequences that significantly affect a person's everyday life, starting with slow and less legible handwriting and often progressing to lower self-esteem, poor emotional well-being, problematic communication and social interaction, and many others.

To introduce a timely and effective treatment to improve a patient's quality of life as much as possible, neurologists and other experts could benefit from the computer-aided assessment of PDD. We entered this field of science more than ten years ago when we reviewed some pioneering studies and identified the following knowledge gaps:

Knowledge gap 1 Although neurologists considered micrographia to be the main alteration of PD handwriting, just a few studies explored the presence of other manifestations.

Knowledge gap 2 The spiral drawing and spring task was used as a gold standard for the assessment of PDD. However, the potential of more complex (handwriting) tasks was not fully investigated.

Knowledge gap 3 Although the computer-aided diagnosis of PDD was not a new technology, it was still in its beginning, classification models had poor performance, there were almost no studies dealing with the rating of PD severity, no studies utilising multilingual datasets, and no studies focusing on the prodromal diagnosis.

1.2 Graphomotor and handwriting disabilities in children with DD

A child starts to develop GS [59], [60] and form the foundation of drawing [61] and consequently, handwriting abilities [62] around the age of 6. These skills should be mastered at the age of 8–9 and should result in automated, legible, well-coordinated and fast-paced handwriting [5], [63], which is used for quantification of a child's timely maturation and integration of linguistic, psycho-motor and mental abilities, and readiness for education [64]. Although a child is intensively exposed to modern technologies that bring new ways of communication, education and self-expression, handwriting takes 30–60 % of a child's school-time [65] and is still an important part of her/his life [60].

Proper acquisition of handwriting is crucial for a child's academic success and self-esteem [66]. However, 10–30 % of children experience an impairment of the neuro-muscular system manifested in GD, such as graphomotor production deficits (poor muscle coordination, less precise graphomotor movements, and unusual pen-grip), motor-memory dysfunction (problems combining memory input with motor output), motor feedback difficulties (problems tracking the location of the pen's tip and over-activation of certain muscles and joints during handwriting), etc. [59], [60] GD/HD are tightly linked with the developmental dysgraphia (DD), which belongs to the category of specific learning disabilities according to DSM V [67], and to the category of specific developmental disorders of scholastic skills according to ICD-10 [68]. DD could have serious pedagogical and psychological consequences such as lack of motivation to write, poor emotional well-being, bad attitude and behaviour, communication and social interaction problems, etc. [60], [69]–[71]

Nowadays, GD/HD in children with DD are diagnosed by occupational therapists and/or special educational counsellors, who visually assess the handwriting product and process, and score it in several domains using a questionnaire (rating scale). Some representatives of these questionnaires could be the Concise Assessment Scale for Children's Handwriting (Brave Handwriting Kinder) (BHK) [72], Handwriting Proficiency Screening Questionnaire (HPSQ) [73] or Handwriting Proficiency Screening Questionnaire for Children (HPSQ–C) [74]. Unfortunately, assessment based on these questionnaires is very subjective, depends on the rater's experience, perceptual abilities, and is subject to inter-rater variability [75], [76]. Due to the above-mentioned limitations, many children are undiagnosed or badly diagnosed, which has a detrimental impact on their quality of life.

The limitations could be effectively addressed by the computerised analysis of online handwriting. We entered this field of science in 2016 and identified the following knowledge gaps:

Knowledge gap 4 Most of the available studies reported some conclusions based on the quantitative analysis, but almost no studies investigated, whether mathematical modelling (e.g. employing machine learning) of handwriting features could support the diagnosis or rating of GD/HD.

Knowledge gap 5 We have not identified any study comparing different graphomotor tasks in supportive GD diagnosis.

Knowledge gap 6 There was no scale enabling objective and fully automatic assessment of manifestations associated with GD/HD.

1.3 Computerised assessment of GD/HD

In the concept of computerised assessment (see Section 3.1), GD/HD are usually quantified in terms of features (measures) that could be split into several categories: temporal (e.g. duration), spatial (e.g. width and length of the product of handwriting), kinematic (e.g. velocity and acceleration), dynamic (e.g. pressure or pen tilt), and other (e.g. the number of pen stops) [6]–[9], [77]. For a more detailed review of these parameters, we refer to Section 3.3. The advantage of the features is that they are usually easily interpretable, and they could be linked with specific manifestations of GD/HD.

Nonetheless, from the signal processing point of view, handwriting is time series that is the result of several interacting physiological mechanisms. This kind of signal contains complex fluctuations, which could provide information related to underlying processes and states of the physiological system. Disfluent movement, irregular muscle contractions, and cognitive deficits introduce randomness to handwriting and increase its complexity (e.g., add tremor, more handwriting interruptions, sudden changes in velocity, etc.). However, this complexity is difficult to be analysed using only conventional parameters. To better quantify the hidden complexities, an advanced and more sophisticated apparatus is needed.

Knowledge gap 7 Most of the existing online handwriting parameterisation algorithms were adopted from the field of biometrics and were not designed to quantify GD/HD.

Finally, until 2012, almost no attention was paid to the potential of in-air movement analysis. In that year, Sesa-Nogueras et al. observed, that information contained in the in-air movement could be advantageously used in biometric recognition [78]. This study opened new questions related to the utilisation of the in-air movement in the field of GD/HD assessment.

Knowledge gap 8 There was no research exploring how is the in-air movement linked with physiological processes and whether it contributes to more accurate diagnosis of neurodegenerative and neurodevelopmental disorders.

2 Aims of the thesis

Concerning the knowledge gaps mentioned in Section 1, the main goal of this habilitation thesis is to progress beyond the state of the art, and to **research new approaches to the computerised assessment of GD/HD that would facilitate objective diagnosis and monitoring of neurodegenerative/neurodevelopmental disorders**, more specifically, the thesis has the following aims.

Aim 1 Explore the impact of in-air movement analysis on diagnostic accuracy.

Aim 2 Introduce new online handwriting parameterisation techniques enabling advanced quantification of GD/HD.

Aim 3 Identify what tasks are suitable for assessment of drawing/handwriting alterations in PD/DD.

Aim 4 Evaluate the researched methodology in the computerised assessment of PDD.

Aim 5 Evaluate the researched methodology in the computerised assessment of DD.

3 General methods

3.1 Concept of the computerised assessment of GD/HD

The general concept of the computerised assessment of GD/HD in subjects with neurodegenerative/neurodevelopmental disorders is illustrated in Figure 3.1, and described below:

1. Based on an acquisition protocol (templates and instructions) a subject performs a set of drawing or handwriting tasks using a stylus and a digitizer (tablet).
2. Signals (time series) recorded by the tablet are consequently parametrized, and the resulting vector/matrix of features is extended by demographic/clinical data such as age, gender, information about medication, etc.
3. To get some first insight into data, we perform visualisations (e.g. kernel density estimation, violin graphs) and exploratory statistical analysis (e.g. correlation analysis or parametric/non-parametric tests). Usually, we also model the data employing machine learning algorithms, e.g. logistic regression of XGBoost.
4. The machine learning models could be used for supportive diagnosis (e.g. diagnosis of PDD), or for a rating of severity of GD/HD. The performance of a subject could be also followed in time (e.g. to monitor the effect of a therapy).

Each of the steps is in more detail explained in the following sections.

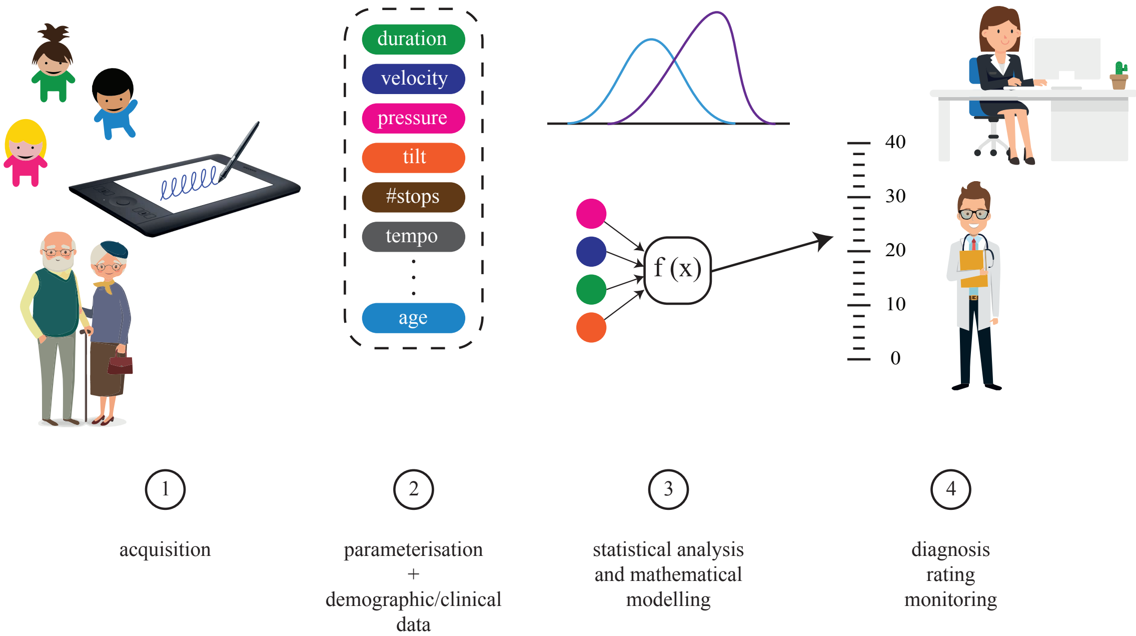


Fig. 3.1: Concept of the computerised assessment of GD/HD

3.2 Datasets and acquisition protocols

Through our research, we enrolled several hundreds of participants who performed specifically designed drawing/handwriting tasks on an A4 paper that was laid down and fixed to a digitizing tablet Wacom Intuos 4 M or Wacom Intuos Pro L (sampling frequency $f_s = 130\,\text{Hz}$). A special Wacom inking pen was used to provide immediate visual feedback, i.e., simulating classical pen-and-paper writing/drawing. Before the acquisition, the participants were seated in a comfortable way and had some time to get familiar with the hardware. The following datasets were acquired:

- **PaHaW** – The dataset contains 75 Czech subjects (37 PD patients and 38 age- and gender-matched healthy controls – HC) [A.23]. The participants were enrolled at the First Department of Neurology, St. Anne's University Hospital in Brno, Czech Republic. They are associated with information about gender, age, PD duration, UPDRS (Unified Parkinson's Disease Rating Scale), part V – modified Hoehn and Yahr staging score [79], and levodopa equivalent daily dose (LED) [80].
  The participants performed 9 tasks following the template available in Figure 3.2: TSK1 – Archimedean spiral; TSK2 – overlapped circles; TSK3 – five graphemes "l"; TSK4 – five bigrams "le"; TSK5 – five trigrams "les"; TSK6 – two words "lektorka"; TSK7 – two words "porovnat"; TSK8 – one word "nepopadnout"; and TSK9 – one sentence "Tramvaj dnes už nepojede." The dataset is freely available for a scientific community [A.23], and until now, it is probably the most popular database of online handwriting collected in PD patients.
- **CoBeN** – This is a multilingual dataset containing 59 Czech participants (19 PD patients and 40 HC) enrolled at the Central European Institute of Technology, 21 US participants (9 PD patients and 12 HC) enrolled at the University of Arizona, and 21 Hungarian participants (9 PD patients and 12 HC) enrolled at the University of Szeged [A.1]. They are associated with information about gender, age, PD duration, UPDRS part III – motor part score [79], and LED.
  The Czech participants performed 8 tasks following the template available in Figure 3.3: TSK1 – five graphemes "l"; TSK2 – a task, where a participant has to horizontally connect two dots; TSK3 – a signature performed with opened eyes; TSK4 – a signature performed with closed eyes; TSK5 – one sentence "Tramvaj dnes už nepojede."; TSK6 – one sentence "Máma a táta jeli dvakrát na dovolenou."; TSK7 – Archimedean spiral; and TSK8 – the pentagon copy test [A.4]. Except for the TSK5 and TSK6, the US and Hungarian participants performed the same tasks. Sentences were not the same, nevertheless, they contained the same number of letters.
- **preLBD** – We enrolled 39 subjects diagnosed with possible or probable MCI (based on the scores of the MoCA – Montreal Cognitive Assessment [81] and based on the CCB – Complex Cognitive Battery (see the explanation below), who were simultaneously diagnosed with possible or probable MCI-LB (i.e. mild cognitive impairment with Lewy bodies) based on the criteria published by McKeith et al. [82]. In this group, 21 subjects also had more than 50 % probability of developing PD (calculated following the MDS criteria published in [48]). In addition, we enrolled 7 subjects without possible/probable MCI-LB, but still with more than 50 % probability of developing PD. The participants performed the same protocol as in the CoBeN dataset (see Figure 3.3).
  CCB was used to evaluate four cognitive domains: 1) memory (The Brief Visuospatial memory test–revised [83], Philadelphia Verbal Learning Test [84]); 2) attention (Wechsler Adult Intelligence Scale-III: Letter-Number Sequencing, Digit Symbol Substitution [85]); 3) executive functions (Semantic and phonemic verbal fluency [86], Picture arrangement test [85]); and 4) visuospatial functions (Judgment of Line Orientation [87]). The cognitive domain z-scores were computed as the average z-scores of the tests included in the particular domain.

- **DYS_CZ_001** – The database contains 65 Czech children (33 diagnosed with DD and 32 intact) attending the 3rd or 4th grade of an elementary school [A.14]. The children performed three tasks using cursive letters: TSK1 – the children wrote all letters of the Czech alphabet [A.16]; TSK2 – a copy of a paragraph [A.14]; and TSK3 – the children wrote several sentences on a random topic. Besides the gender, age, and grade, the children were also associated with the scores of HPSQ and HPSQ–C.
- **DYS_CZ_002** – The database consists of 353 children from the final grade of kindergarten to the fourth grade of elementary schools. Participants were enrolled in 8 kindergartens, 14 elementary schools, and 2 counselling centres in the Czech Republic, covering 62 children with GD/HD and 291 intact children. The database includes socio-demographic data, several diagnostic scores, and the HPSQ–C score.
  All children were asked to perform a protocol (see Figure 3.4) consisting of 7 elementary graphomotor tasks (TSK1 – Archimedean spiral (approximately 15 cm in height); TSK2 – half-sized version of TSK1; TSK3 – connected loops (the spring task); TSK4 – flipped version of TSK3; TSK5 – saw; TSK6 – rainbow; TSK7 – a combination of TSK3 and TSK4) [A.7, A.9, A.13], and one paragraph copy task, whose content was depending on the grade of a child. Regarding the graphomotor part of the protocol, it was designed in a way so that the tasks cover the building blocks of letters used in the Latin alphabet.

Besides the above-mentioned datasets, in some studies, we also analysed databases of our partners, e.g. the Colombian HWUDEA [A.1], a database of patients with AD [A.21, A.32], or the BIOSECURID database [A.28].

Regarding our databases, all subjects used their dominant hand. None of the participants had a history or presence of any psychiatric symptoms or any disease affecting the central nervous system (other than PD in the PD cohort). All PD patients were well compensated on their stable dopaminergic medication and without major motor fluctuations or dyskinesias (they were examined while on their regular dopaminergic medication (ON state) approximately 1 h after the L-dopa dose). All subjects signed an informed consent form (in the datasets of children, the form was signed by parents). All studies were approved by the relevant local ethics committees.

3.3 Baseline parameters

The tablets, that we used, recorded the following time series/signals: x and y position ($x[n]$ and $y[n]$); timestamp ($t[n]$); a binary variable ($b[n]$), being 0 for in-air movement (i.e. movement of pen tip up to 1.5 cm above the tablet's surface) and 1 for on-surface movement (i.e. movement of pen tip on the paper), respectively; pressure exerted on the tablet's surface during writing ($p[n]$); pen tilt ($a[n]$); azimuth ($az[n]$).

In most studies, the signals were parameterised by employing baseline features that are frequently used in the field of neurodegenerative and neurodevelopmental disorders [6]–[9], [77], [88], [89]. These features could be split into several groups:

1. temporal – duration of writing, ratio of the on-surface/in-air duration, duration of strokes, and ratio of the on-surface/in-air stroke duration,
2. spatial – width, height, and length of the whole product as well as those of its individual strokes, i.e., stroke width, height, and length,
3. kinematic – velocity, angular velocity, acceleration, and jerk,

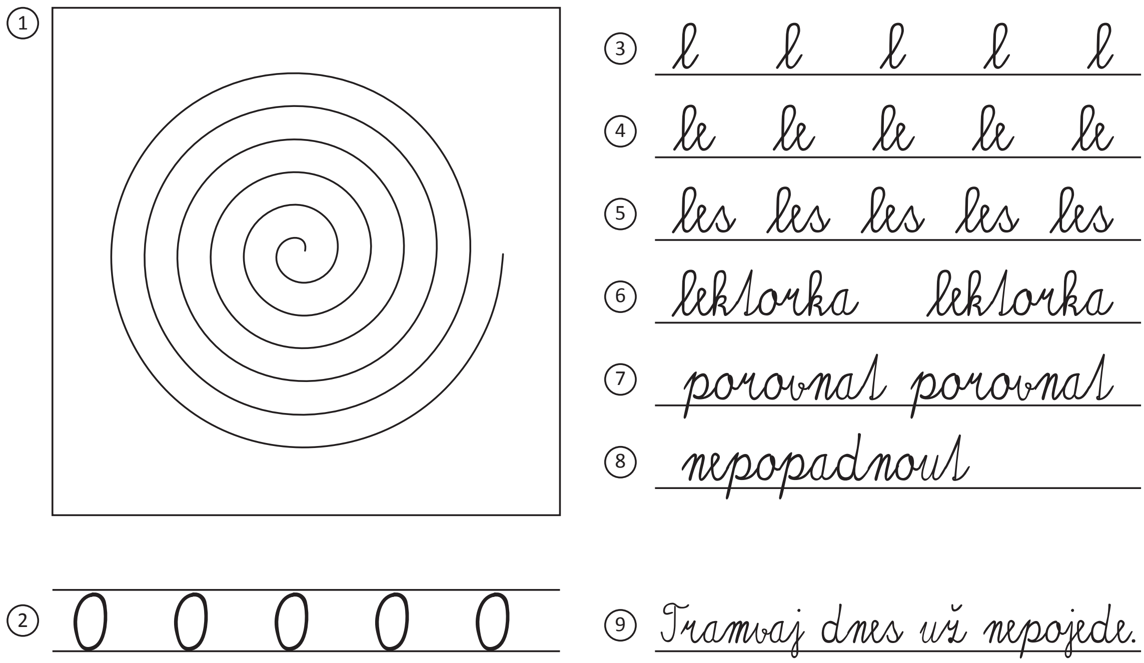


Fig. 3.2: Acquisition protocol of the PaHaW database

4. dynamic – pressure, tilt, and azimuth,
5. spiral-specific – degree of spiral drawing severity, mean drawing speed of spiral, second-order smoothness of spiral, spiral precision index, spiral tightness, variability of spiral width, and first-order zero-crossing rate of spiral,
6. loops/saw/rainbow-specific – local minima, local maxima, distance between neighbouring local maxima, velocity at local maxima, width of teeth (on a horizontal line going through 95 % of a particular tooth height), normalised width of teeth (normalised by a mean distance between local minima), and distance between neighbour bows (on a horizontal line going through 50 % of the first of them),
7. other – number of interruptions or pen elevations, relative number of interruptions, number of pen stops, tempo (number of strokes normalised by duration), number of on-surface interstroke intersections, relative number of on-surface interstroke intersections, number of on-surface intrastroke intersections, relative number of on-surface intrastroke intersections, total number of on-surface intrastroke intersections, relative total number of on-surface intrastroke intersections, relative number of changes in velocity profile, relative number of changes in pressure profile, relative number of changes in tilt profile, and relative number of changes in azimuth profile.

The spatial, temporal and kinematic features were extracted from both the on-surface and in-air movements. In addition, the kinematic features were also analyzed for the horizontal and vertical projections of the movements. Features that were represented by time series were transformed into scalar values using statistics such as median, interquartile range (iqr), nonparametric coefficient of variation (defined as iqr/median), 95th percentile, slope by applying the Theil–Sen estimator, etc.

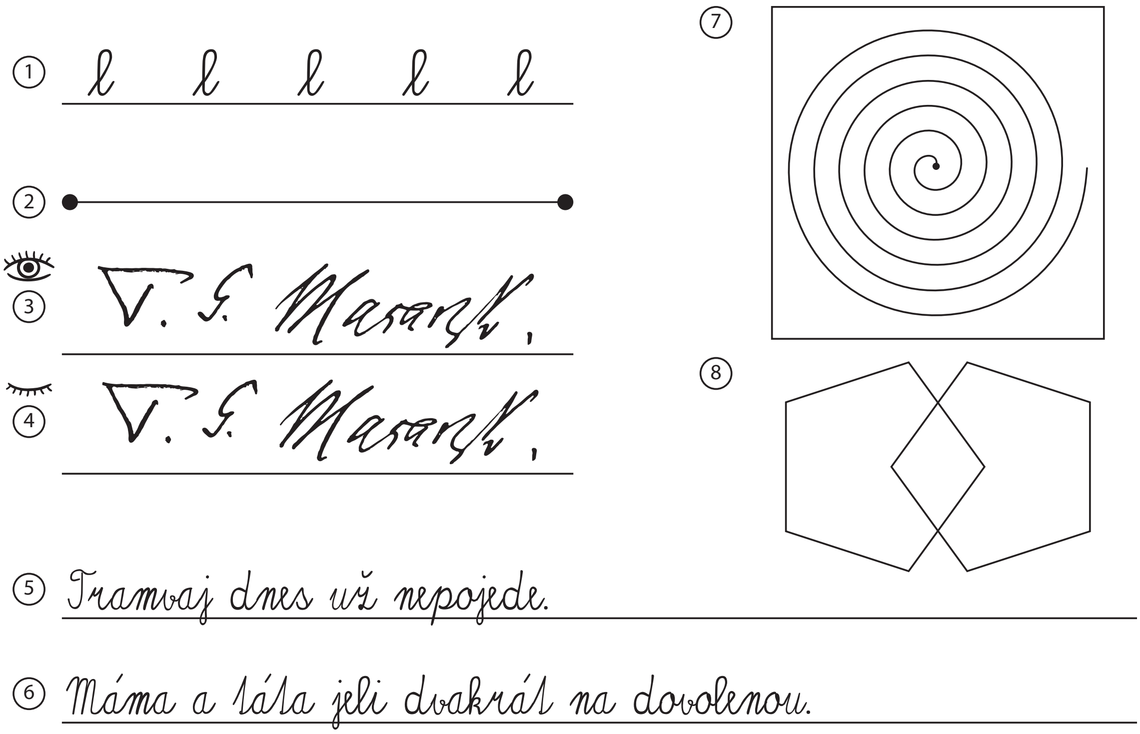


Fig. 3.3: Czech version of the acquisition protocol of the CoBeN database

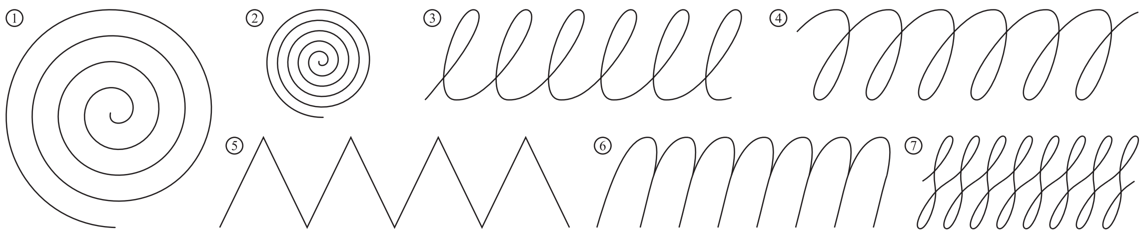


Fig. 3.4: A part of the acquisition protocol of the DYS_CZ_002 database

3.4 Statistical analysis and mathematical modelling

To get some first insight into the data, we typically plotted kernel density estimation, violin graphs, or correlation matrices. When necessary, we regressed out the effect of confounding factors (e.g. age or level of medication). Since the features usually did not have a normal distribution (as assessed by e.g. the Kolmogorov-Smirnov test), during the exploratory analysis, we usually applied the Mann-Whitney U test, Wilcoxon signed-rank test, and/or Pearson's correlation with the level of significance $\alpha = 0.05$. In the case of a higher number of features, we also applied the false discovery rate correction.

Depending on a specific application, we modelled the feature space utilising e.g. logistic regression, classification and regression trees, bagging and gradient boosting algorithms, or artificial neural networks. To prevent overfitting, and to get robust results, we usually followed the cross-validation strategy with several repetitions. Hyperparameters were optimised based on the random- or grid-search algorithm. Performance of classifiers was most frequently evaluated by sensitivity, specificity, balanced accuracy and Matthews correlation coefficient. The performance of regressors was evaluated by mean absolute error, mean squared error, root mean squared error, estimation error rate, etc. To get some intuition about the robustness of the models, we employed the permutation test. Finally, to interpret the models, we used feature importances. In some cases, we visualised the performance of a model using the ROC (Receiver Operating Characteristic) curve.

4 Discussion of main findings

4.1 Impact of the in-air movement analysis

More than ten years ago, the added value of the in-air movement analysis was not fully explored. In our first work dealing with this topic, we compared the on-surface and in-air movement from the information theory point of view [A.34], i.e. in samples of the BIOSECURID dataset, we investigated how much information each movement contains (based on the Shannon entropy) and how much information the movements share (based on the mutual information). Our experiments proved that the amount of information is similar in both movements, moreover, both trajectories appear to be notably non-redundant. After this finding, we moved further and explored, whether the quantitative analysis of in-air movement could improve the computer-aided diagnosis of PDD [A.1, A.11, A.24, A.27, A.31]. For this purpose, we processed the sentence copy task of the PaHaW database. Based on the results, we observed that the quantification of the in-air movement not only improves the classification accuracy but when considering both movements separately, the in-air one provides higher discrimination power [A.27, A.31]. Proficient handwriters without any disease affecting the central nervous system have the so-called “open-loop” handwriting performance, i.e. their handwriting is automatic (they do not concentrate on the process of handwriting). Vice versa PD patients experience the so-called “closed-loop” handwriting performance, i.e. they pay more attention to the process and thus also manifest increased in-air movement [A.27]. We hypothesise that the in-air movement is tightly linked with cognitive processes. E.g. in [A.32] we measured the in-air time of AD patients and HC performing a 3D house copy task. AD patients spent significantly higher time in-air, probably having issues with the visuospatial and memory functions. Similarly, in [A.20] we observed that (in a cohort of AD and MCI patients) this time is significantly higher even at a long distance, i.e. in a distance from the tablet’s surface where we are not able to monitor displacement of a pen. Next, in [A.4] we also proved that entropy-based features extracted from the in-air movement could be used to identify early cognitive changes in PD patients performing the pentagon copy test. In addition, we observed that the features are closely linked to attention levels and to the grey matter volume variability of the posterior cortical region engaged in both visual attention and visual-spatial processing. Besides the cognitive deficits, we assume that the in-air movement could be used to quantify fatigue. E.g. in [A.2] we identified progressively increasing duration of in-air strokes in a cohort of subjects with a high risk of developing Lewy body diseases (LBDs), thus suggesting, that this parameter could be used as a prodromal marker.

We observed that the in-air movement plays a significant role in the assessment of HD in children with neurodevelopmental disorders as well [A.7, A.14, A.16]. Similarly to the PPD, children with DD spend more time in-air and make more pauses/interruptions [A.7]. On the other hand, the in-air movement was less important in the quantification of GD [A.13]. However, this finding is expectable, because all the graphomotor elements we considered in our protocol (see Figure 3.4]) could be theoretically performed without pen elevation.

4.2 Advanced online handwriting parameterisation techniques

In [A.25] we introduced a new set of entropy (Shannon and Rényi) and energy (squared and Teager-Kaiser energy operator) based features extracted from raw signals and intrinsic mode functions of the empirical mode decomposition (EMD). The new feature set significantly outperformed the conventional one in supportive PDD diagnosis. In the following studies [A.23, A.24], we further extended the set by new pressure-based parameters extracted from different parts of pressure trajectories (raising, sustained, and falling) and further improved the classification accuracy.

Next, we introduced very successful features based on the theory of fractional calculus (the theory of integrals and derivatives of arbitrary order). In the first two studies dealing with this topic, we established new handwriting parametrization techniques utilizing fractional-order derivatives (FD) as a substitution of the conventionally used differential derivatives in the kinematic handwriting features extraction [A.17, A.19]. The newly proposed features improved the classification accuracy in absolute value by approximately 10 %. In [A.11] we additionally confirmed the potential of FD-based features to assess the severity of PD (as measured by the UPDRS V). In the following study [A.15], we optimized the order of FD so that we significantly reduced computational costs, moreover, we explored whether FD-based parameterisation of pressure, azimuth, and tilt time series brings some advancement. Next, we found out the features could be easily adjusted to the diagnosis of GD [A.9] and HD [A.16] in the children population. In addition, we observed that when applied to the in-air movement, they outperform the conventional ones [A.16]. Finally, in [A.3, A.10] we shed light on the impact of different FD approximations, namely on the Grünwald-Letnikov's, Riemann–Liouville's, and Caputo's.

In [A.14] we introduced features based on the tunable Q-factor wavelet transform (TQWT) and showed that HD manifest themselves in higher energies of the residual component of the decomposed signal computed by the transform. Following this research, in [A.9] we investigated the potential of TQWT to describe limited motor skills, poor dexterity and muscle tone or unspecified motor clumsiness in school-aged children suffering from GD. Although the TQWT-based parameters were comparable to other advanced ones, their limitation lies in a need for apriori knowledge about the analysed signal that is required for the optimisation of the transform parameters.

On top of the above-mentioned parameters, we also introduced measures based on the modulation spectra (quantifying the ratio between the low and high-frequency movements present in a given handwriting signal) [A.9] or based on real cepstrum [A.5]. Finally, in our recent publication [A.1] we paid attention to the increasing popularity of convolutional neural networks (CNNs) as feature extractors. We compared the discrimination power of the baseline handcrafted parameters with the ones learned by CNN in a multilingual dataset of PD patients performing Archimedean spiral and a sentence. We found that the two approaches are competitive, especially for the spiral drawing task, which is independent of language. Handcrafted features (especially kinematic measures) proved to be the better choice for the sentence writing task. This is expected since CNN-based features are extracted only from offline handwriting samples, from which temporal information is not available. In addition, the orthography of a sentence is strongly affected by the language of a writer.

4.3 Most discriminative tasks

In a review published in [A.6], we identified a wide range of tasks that could be used to quantify different pathologies associated with drawing/handwriting. Regarding the assessment of PDD, the most frequent ones are the Archimedean spiral, the spring task, and the sentence copy task. The first version of our protocol (see Figure 3.2) contains the spiral and the sentence copy task. In addition, we included overlapped circles (to quantify continuous kinematics), graphemes and some words. Unfortunately, at that time, we did not know the spring task has a good potential to quantify micrographia, therefore this important task is missing. In a couple of studies [A.24, A.33], we investigated which task of the protocol provides the best discrimination power. Although the Archimedean spiral is still considered a gold standard in the assessment of PDD [A.1], we found out that the sentence copy task significantly outperforms it. We assume that this task requires higher cognitive effort and accents the effect of rigidity and bradikinesia [A.27]. In terms of projection, the deficits mainly dominated in the vertical projection [A.1, A.2]. The finger system (which is mainly involved in vertical movement) is more affected by muscular fatigue than the wrist system (which controls the horizontal movement). From the anatomical point of view, the vertical movement requires coordinated movement and finer flexions/extensions of more joints (interphalangeal and metacarpophalangeal), i.e., it is more complex than ulnar abductions of the wrist, and we assume this to be the reason why kinematic deficits were more strongly observed in this direction. This finding could also be somehow linked with progressive/consistent vertical micrographia, i.e., progressive/consistent reduction in letter amplitude. However, this hypothesis requires further research because some studies suggest that the horizontal version of micrographia is even more common than the vertical version. Regarding the sentence copy task, we further confirmed its importance in a study, where we employed FD-based measures [A.17], and in a study, where we compared the performance of handcrafted and CNN-learned features in a multilingual dataset [A.1]. In [A.2] we also observed that handwriting (represented by a sentence), in comparison to a graphomotor task (the Archimedean spiral) or cognitive task (the pentagon copy test), enabled the highest classification accuracy when performing the prodromal diagnosis of LBDs.

In terms of DD, we intensively cooperated with psychologists and remedial teachers, and identified a complex set of graphomotor tasks that could be used for the assessment of GD [A.7]. Consequently, in [A.13], we noticed that the most discriminative one is the combined loop task (see TSK7 in Figure 3.4), i.e. the most complex task in our protocol, which requires coordinated movement of fingers, wrist, elbow and shoulder. In addition, the task is demanding in terms of visuospatial cognitive functions. The results also suggest that the task, where children draw a saw-tooth (TSK5), can also work well during the differential analysis. This task requires a precise change in direction when hitting the top of each tooth. Children with GD were associated with higher instability of acceleration when performing this task. We assume that the children were unstable especially in acceleration between upward and downward strokes, which is, again, linked with the vertical movement of the finger system.

4.4 Computerised assessment of PDD

During the last decade, we published more than 15 works dealing with the computerised assessment of PDD [A.1, A.2, A.3, A.4, A.5, A.11, A.12, A.15, A.17, A.19, A.23, A.24, A.25, A.26, A.27, A.31, A.33]. Depending on specific objectives, we followed different statistical and machine learning pipelines (see Section 3.4), processed different tasks (see Section 3.2) and employed different features (see Sections 3.3 and 4.2). When considering the computer-aided diagnosis of PD, in our first work [A.33], we reached 79 % accuracy (ACC), 80 % sensitivity (SEN), and 79 % specificity (SPE) in the PaHaW database. In [A.23] we made the dataset available for the scientific community, it became very popular, and was used by many teams around the world (the article has more than 100 citations without self-citations on the Web of Science), who published classification accuracies beyond 90 % (in some cases with questionable methodology). Regarding our own research, in [A.11], we reached 97 % ACC (SEN = 96 %, SPE = 100 %) when processing the whole protocol by the FD-based features. This and other recent publications suggest that the supportive diagnosis utilising online handwriting could provide very high accuracies. Nevertheless, majority of them were conducted in a single-language cohort. We were the first who explored the impact of language in a big dataset containg 143 PD patients and 151 HC enrolled in the Czech Republic, Hungary, Colombia and the United States of America [A.1]. We observed that the classification accuracies in multi-language scenarios droped to approximately 70 %, thus concluding that this field is still highly challenging and requiring further research. Besides PD, in [A.2], we focused on a more general task, i.e. supportive diagnosis of LBDs (including DLB), moreover in their prodromal state. Employing baseline parameters, we were able to differentiate LBDs and HC with ACC = 74 %, SEN = 80 %, and SPE = 67 %.

Regarding the assessment of PD severity, since the PaHaW database contains clinical data such as the duration of PD, or UPDRS V, we introduced several regression models estimating these variables. In [A.11] we predicted UPDRS V and PD duration with 13 % and 24 % error, respectively. The latter one was further improved in [A.15], where we reached 22 % error. The errors are still high, i.e. it is another challenging field. On the other hand, two patients could experience different severity in, e.g., ten years of PD, which is differently manifested in drawing/handwriting. Therefore, it makes sense to focus more on the computer-aided estimation of UPDRS scores and other metrics evaluating the progress of the disease.

4.5 Computerised assessment of DD

In comparison to PD, DD does not have any unified diagnostic criteria that could be used independently from a language. Nowadays, the diagnosis is usually done subjectively, based on several scales with poor psychometric properties [A.7]. This fact accents the need to introduce an objective approach. On the other hand, since most of external validation criteria are less reliable, it is even more challenging to establish a good classification model (we can observe a wide range of classification accuracies). In our first work dealing with this topic [A.22], we reached ACC = 96 % (SEN = 96 %, SPE = 97 %) when diagnosing DD in Israeli children performing a graphomotor tasks similar to a rainbow. In a work, where we processed the paragraph copy task of the DYS_CZ_001 dataset, we reached ACC = 85 % (SEN = 89 %, SPE = 83 %) [A.14]. When quantifying the combined loop task of the DYS_CZ_002 cohort, we were able to diagnose GD with ACC = 82 %, but with very imbalanced SEN (47 %) and SPE (90 %) [A.13]. We further improved the results in [A.9] (ACC = 84 %, SEN = 83 %, SPE = 81 %), however, we had to extend our pipeline by advanced features (see Section 4.2) and process all the tasks of the protocol together.

Regarding the severity of DD, we were the first in the world who defined and evaluated the concept of computer-aided rating [A.22]. In the cohort of the Israeli children, we were able to estimate the total score of HPSQ with 8 % error. Nevertheless, in a cohort of Czech children (DYS_CZ_001), the minimum error we were able to reach was 18 % [A.18]. In [A.8], we found out that since adults are usually influenced by their point of view, children could better evaluate their own performance using the HPSQ–C scale than teachers using the HPSQ one. This could also be one of the explanations why the estimation error of HPSQ–C was lower than the estimation error of HPSQ [A.18]. Nonetheless, assessment based on HPSQ–C is still subjective, and its mathematical modelling is challenging, e.g., in the DYS_CZ_002 cohort, we were not able to get the estimation error lower than 30 %.

5 Conclusion

5.1 General contributions of the thesis

Nowadays, the field of computerised assessment of GD/HD is very dynamic and perspective. Its benefit for the computer-aided diagnosis of neurodegenerative and neurodevelopmental diseases was proved in many studies. We were the pioneers in many paths in this field of science, bridged several gaps, and brought the community new knowledge and directions with an indirect impact on the quality of life of children experiencing DD and adults suffering from PD. Our specific contributions are structured according to the aims of this thesis:

Aim 1 Explore the impact of in-air movement analysis on diagnostic accuracy.
Progress beyond the state of the art: We were the first who quantitatively showed that the in-air movement contains almost the same amount of information as the on-surface one, moreover, that this information is not redundant. We were one of the first who successfully utilised the in-air movement in the computerised assessment of GD/HD and proved that it could be employed to quantify cognitive processes and fatigue. The in-air movement analysis is already well established in the domain, and plays a significant role in the supportive diagnosis.

Aim 2 Introduce new online handwriting parameterisation techniques enabling advanced quantification of GD/HD.
Progress beyond the state of the art: We introduced a set of new features specifically designed to quantify GD/HD. We observed that these parameters could improve classification accuracies, as well as improve the performance of regression models. Although a clinical interpretation or connection with physiological processes was sometimes very challenging, in all our studies, we tried to avoid black boxes and provide neurologists, neuroscientists, psychologists, and other experts with meaningful and understandable outcomes.

Aim 3 Identify what tasks are suitable for assessment of drawing/handwriting alterations in PD/DD.
Progress beyond the state of the art: We helped to fight the stigma of gold standards in the field of PDD, and proved that handwriting tasks could be much better candidates in terms of quantitative analysis (e.g. they support better quantification of cognitive processes) of GD/HD, and computer-aided diagnosis of PD or prodromal diagnosis of LBDs in general. Regarding the DD, we identified tasks that accent GD during the performance of graphomotor elements, and that improve their classification accuracies.

Aim 4 Evaluate the researched methodology in the computerised assessment of PDD.
Progress beyond the state of the art: We confirmed that micrographia is just one of the PDD manifestations, and that PD is associated with far more complex alterations that could be identified in the temporal, kinematic and dynamic aspects of handwriting. We helped the community to advance the computer-aided diagnosis of PD by providing it with the PaHaW database, and by pushing the research in the field of machine-learning-based diagnosis beyond the state of the art. Although it can look that the computer-aided diagnosis of PD (based on the online handwriting processing) is an almost solved task, in our recent work focusing on the multilingual dataset, we demonstrated that it is still a very

challenging field with several knowledge gaps. Finally, we were the first who revealed that the computerised assessment of GD/HD could support the early diagnosis of LBDs.

Aim 5 Evaluate the researched methodology in the computerised assessment of DD.
Progress beyond the state of the art: We were the first who defined and evaluated the concept of DD rating by employing online handwriting processing. Our study served as a building block and a baseline for other works published in the community. We proved that the machine-learning-based approach could bridge the limitations of current assessment methods, but on the other hand, we also observed how unreliable the current external validation criteria are – this is a big obstacle in training a model with good psychometric properties.

Besides the above-mentioned achievements, the conducted research and works have also several secondary contributions:

- Our team laid the foundations of online handwriting processing in the Czech Republic and brought it to the world-class level.
- From the educational point of view, the research was part of two defended PhD theses ("Advanced parameterisation of online handwriting in writers with graphomotor disabilities" defended by Ing. Ján Mucha, Ph.D., and "Research of advanced online handwriting analysis methods with a special focus on assessment of graphomotor disabilities in school-aged children" defended by Ing. Vojtěch Zvončák, Ph.D.), and is part of one ongoing ("Research of online handwriting parameterisation in subjects with graphomotor difficulties" being solved by Ing. Michal Gavenčiak).
- This multidisciplinary research helped us to strengthen or establish new cooperation with the Central European Institute of Technology, St. Anne's University Hospital in Brno, Masaryk University, Czech Academy of Sciences, Escola Superior Politecnica (TecnoCampus Mataro-Maresme), University of Haifa, University of the Basque Country, University of Las Palmas de Gran Canaria, University of Vic, Technical University of Košice, University of Antioquia, University of Arizona, University of Szeged, University of Bari Aldo Moro, Taipei Veterans General Hospital, and Wacom Co., Ltd.
- We raised awareness about the computerised assessment of GD/HD during several workshops with students held at the Masaryk University. We also raised the awareness among the general public, e.g. during the Wacom Connected Ink event.[1]
- Besides the PaHaW database that is freely available for research purposes, we also made available a Python library for online handwriting processing [90], [91], and user-friendly software for online handwriting acquisition [92], [93].

Regarding the neurodegenerative disorders, from a practical point of view, the outcomes of our research were used to better understand their pathophysiology. Even though we proposed some models supporting diagnosis, there is still a long path until they are used in clinical practice. On the other hand, concerning the DD, in the frame of project no. TL03000287 (Software for advanced diagnosis of graphomotor disabilities) supported by the Technology Agency of the Czech Republic, we cooperate with the company Propsyco s. r. o. and prepare software, that will help psychologists and remedial teachers to objectively diagnose GD/HD in school-aged children.

[1] `https://www.youtube.com/watch?v=04G5ksvNFBY`

5.2 Future directions

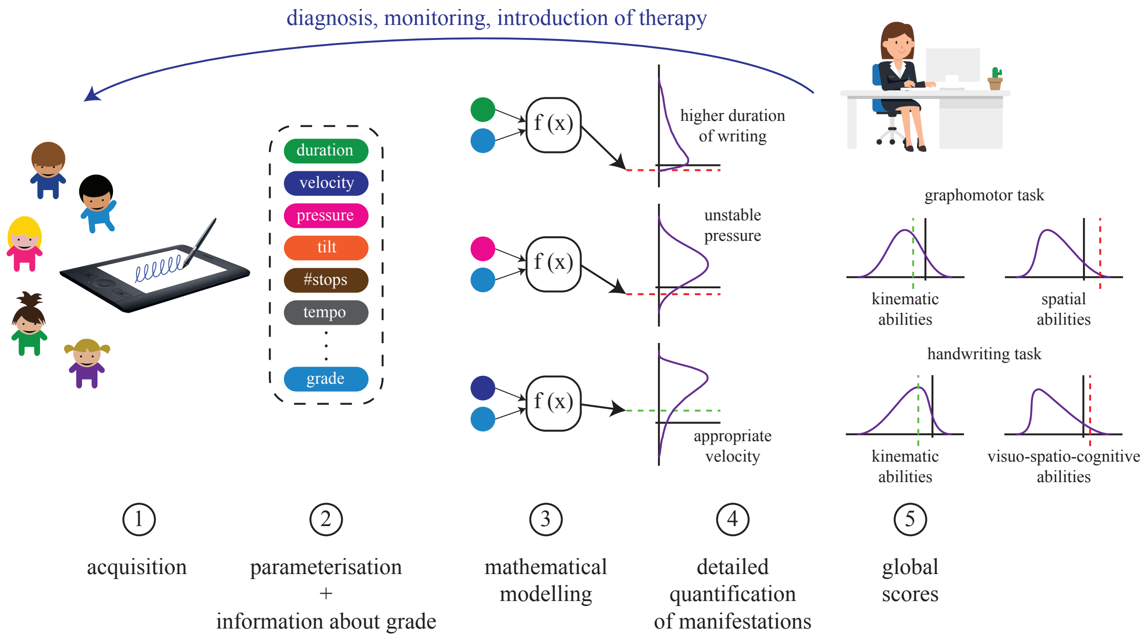


Fig. 5.1: Assessment based on the GHDRS scale

As already mentioned, the field of the computerised assessment of GD/HD is very dynamic, opens new questions and brings new challenges. We have identified the following future directions:

- Since external validation criteria in studies dealing with DD are very unreliable, we propose to introduce diagnostic models based on semi-supervised and data-driven approaches. The only work following (at least partially) this approach was published by Asselborn et al. in 2020 [94]. Our team is just finishing a study, where we go much beyond the state of the art, and which we believe will revolutionise the computer-aided diagnosis of GD/HD. The process could be summarised as follows: 1) based on discussions with well recognised special educational counsellors, and based on a very comprehensive review of symptoms associated with GD/HD, we pre-identified manifestations and related handwriting measures that could quantify them; 2) based on some simulations, we performed a finer selection of features that were consequently used to create a scale for each manifestation (e.g. dysfluent handwriting), and normative values that are used during diagnosis/rating. The concept is visualised in Figure 5.1. As can be seen, the concept enables us not only to diagnose GD/HD but also to assess specific manifestations (moreover, it could be scaled to pre-school children who are not able to write). This is very important because we observed that several children, all diagnosed with DD, could actually have different difficulties and require different therapy. Figure 5.2 displays our newly proposed GHDRS scale (Graphomotor and Handwriting Disabilities Scale) of three children attending 3rd grade of a primary school. As can be seen, the first girl has no GD/HD (she is intact). The second girl has GD/HD, more specifically, she has the impaired process of drawing and handwriting. Vice versa, the boy has impaired product of drawing and handwriting, which could be also seen in Figure 5.3 (he was not able to maintain the loops in a line and was not able to keep a stable tilt – this probably explains the different orientation of loops) and in Figure 5.4 (frequent overwriting, disability to perform longer strokes, all letters tended to have the same amplitude). The methodology (maximally transparent so that it could be used in practice) will be published in an upcoming article.

Fig. 5.2: Three children attending 3rd grade of a primary school assessed based on the GHDRS scale (the first top block contains the global scores; the next four blocks contain specific manifestations, i.e. they represent a detailed profile associated with the global scores; all scores are transformed by a sigmoid function so that the minimum is 0, maximum 1 and the threshold determining disability has a value 0.5): a) an intact girl without any GD/HD; b) a girl with the affected process of handwriting (too high duration of writing, lower variability of velocity) and affected process of drawing the loops (low velocity, low acceleration); also, she is not able to perform longer strokes during writing; c) a boy whose handwriting is characteristic by frequent overwriting (see Figure 5.4), disability to perform longer strokes, moreover, all letters tended to have the same amplitude; in addition, he was not able to maintain the loops in a line (see Figure 5.3) and was not able to keep a stable tilt.

- In 2021, Jan Rusz et al. published a work, where the authors identified distinct speech phenotypes in de novo PD patients [95]. This finding and the finding mentioned above (supported by Figure 5.2) led us to postulate that some phenotypes could be observed even in PDD. We plan to investigate it in the next few years.
- Although there has been a body of research focusing on PDD, studies dealing with GD/HD in patients with atypical Parkinsonian syndromes (e.g. multiple system atrophy or progressive supranuclear palsy) are missing. One year ago, we teamed up with the neuroscientific group of the Taipei Veterans General Hospital and currently collect a handwriting dataset of patients with these syndromes.
- As in the other fields of science, the popularity of deep neural networks (DNNs) in the computer-aided diagnosis of neurodegenerative disorders is rapidly increasing. The conven-

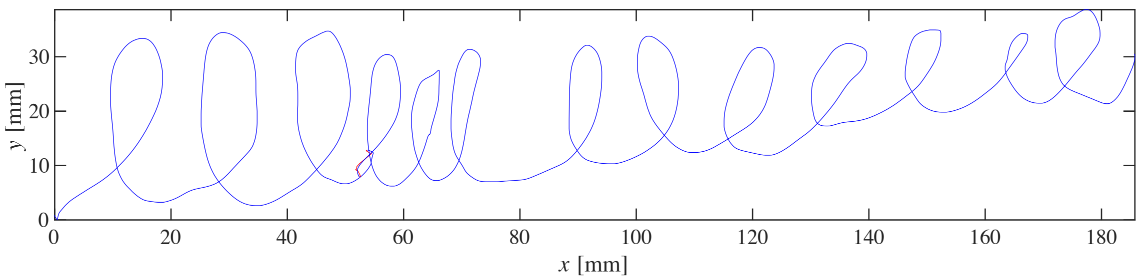


Fig. 5.3: Connected loops (spring task) performed by a boy, whose GHDRS is depicted in Figure 5.2c (the blue line represents the on-surface movement, the red line represents the in-air one).

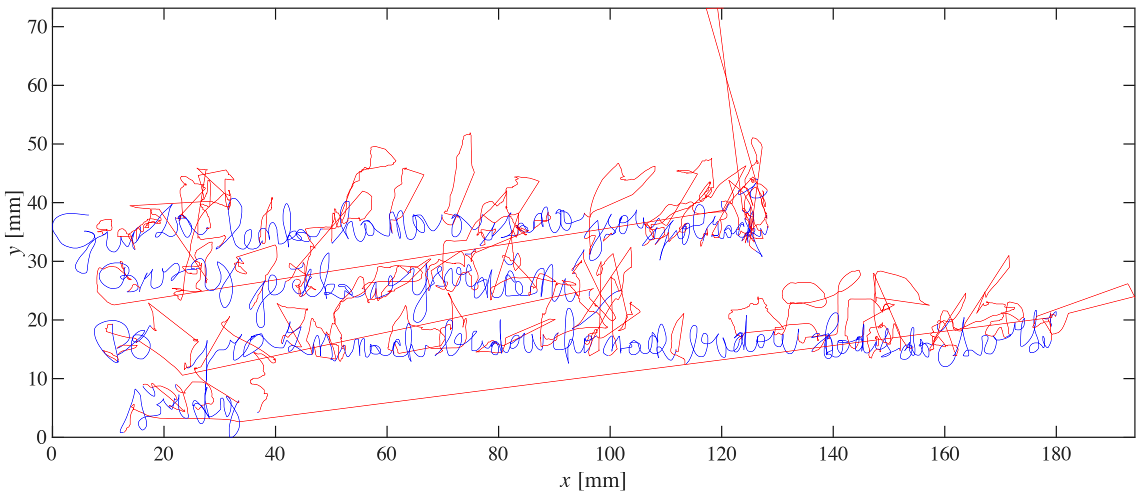


Fig. 5.4: Paragraph copy task performed by a boy, whose GHDRS is depicted in Figure 5.2c (the blue line represents the on-surface movement, the red line represents the in-air one).

tional methodology combining handcrafted features and shallow machine learning algorithms is still outperforming DNN-based approaches (and facilitating clinical interpretation). On the other hand, with the increasing size of datasets, new data augmentation techniques, model interpretation methodologies, and approaches such as transfer learning, we believe that DNNs will play a significant role. Moreover, some recent works put in place networks specifically designed for online handwriting processing (i.e. they are able to combine the spatial and temporal information) [96], [97].

- To adjust therapy or modify a treatment policy, children visit pedagogical-psychological counselling centres (and adults hospitals) a few times per year. Nevertheless, the frequency is insufficient, moreover, patients under clinical examination could be subjected to the Hawthorne effect [98]. To monitor the therapy effectively and intervene when necessary, neurologists, psychologists, educational counsellors, and other experts could take the advantage of mobile devices. In spite of the fact that some remote monitoring applications already exist [2], this field is still at its beginning.

[2] `https://dynamilis.com/en/`

Bibliography

[1] S. Rosenblum, S. Parush, and P. L. Weiss, "Computerized temporal handwriting characteristics of proficient and non-proficient handwriters", *American Journal of Occupational Therapy*, vol. 57, no. 2, pp. 129–138, 2003.

[2] J. Ziviani, "The development of graphomotor skills", in *Hand Function in the Child*, Elsevier, 2006, pp. 217–236.

[3] A. Letanneux, J. Danna, J.-L. Velay, F. Viallet, and S. Pinto, "From micrographia to Parkinson's disease dysgraphia", *Movement Disorders*, vol. 29, no. 12, pp. 1467–1475, 2014.

[4] B. Engel-Yeger, L. Nagauker-Yanuv, and S. Rosenblum, "Handwriting performance, self-reports, and perceived self-efficacy among children with dysgraphia", *American Journal of Occupational Therapy*, vol. 63, no. 2, pp. 182–192, 2009.

[5] S. Rosenblum, "Inter-relationships between objective handwriting features and executive control among children with developmental dysgraphia", *PLOS ONE*, vol. 13, no. 4, E. Manalo, Ed., e0196098, 2018.

[6] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. S. di Freca, "Handwriting analysis to support neurodegenerative diseases diagnosis: A review", *Pattern Recognition Letters*, vol. 121, pp. 37–45, 2019.

[7] D. Impedovo and G. Pirlo, "Dynamic handwriting analysis for the assessment of neurodegenerative diseases: A pattern recognition perspective", *IEEE reviews in biomedical engineering*, vol. 12, pp. 209–220, 2018.

[8] G. Vessio, "Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review", *Applied Sciences*, vol. 9, no. 21, p. 4666, 2019.

[9] J. Danna, V. Paz-Villagrán, and J.-L. Velay, "Signal-to-noise velocity peaks difference: A new method for evaluating the handwriting movement fluency in children with dysgraphia", *Research in developmental disabilities*, vol. 34, no. 12, pp. 4375–4384, 2013.

[10] M. Faundez-Zanuy and J. Mekyska, "Privacy of online handwriting biometrics related to biomedical analysis", in *User-Centric Privacy and Security in Biometrics*, Institution of Engineering and Technology, 2017, pp. 17–39.

[11] C. Alonso-Martinez, M. Faundez-Zanuy, and J. Mekyska, "A comparative study of in-air trajectories at short and long distances in online handwriting", *Cognitive Computation*, vol. 9, no. 5, pp. 712–720, 2017.

[12] Z. Galaz, J. Mucha, V. Zvoncak, J. Mekyska, Z. Smekal, K. Safarova, A. Ondrackova, T. Urbanek, J. M. Havigerova, J. Bednarova, and M. Faundez-Zanuy, "Advanced parametrization of graphomotor difficulties in school-aged children", *IEEE Access*, vol. 8, pp. 112 883–112 897, 2020.

[13] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease", *Computer Methods and Programs in Biomedicine*, vol. 117, no. 3, pp. 405–411, 2014.

[14] T. Asselborn, T. Gargot, Ł. Kidziński, W. Johal, D. Cohen, C. Jolly, and P. Dillenbourg, "Automated human-level diagnosis of dysgraphia using a consumer tablet", *npj Digital Medicine*, vol. 1, no. 1, p. 42, 2018.

[15] M. Faundez-Zanuy, J. Mekyska, and D. Impedovo, “Online handwriting, signature and touch dynamics: Tasks and potential applications in the field of security and health”, *Cognitive Computation*, vol. 13, no. 5, pp. 1406–1421, 2021.

[16] Z. Galaz, P. Drotar, J. Mekyska, M. Gazda, J. Mucha, V. Zvoncak, Z. Smekal, M. Faundez-Zanuy, R. Castrillon, J. R. Orozco-Arroyave, S. Rapcsak, T. Kincses, L. Brabenec, and I. Rektorova, “Comparison of CNN-learned vs. handcrafted features for detection of Parkinson’s disease dysgraphia in a multilingual dataset”, *Frontiers in Neuroinformatics*, vol. 1, no. 1, pp. 1–26, 2022.

[17] Z. Galaz, J. Mekyska, J. Mucha, V. Zvoncak, Z. Smekal, M. Faundez-Zanuy, L. Brabenec, I. Moravkova, and I. Rektorova, “Prodromal diagnosis of Lewy body diseases based on the assessment of graphomotor and handwriting difficulties”, in *20th Conference of the International Graphonomics Society (IGS)*, (in press), Springer, 2022, pp. 1–16.

[18] J. Mucha, Z. Galaz, J. Mekyska, M. Faundez-Zanuy, V. Zvoncak, Z. Smekal, L. Brabenec, and I. Rektorova, “Exploration of various fractional order derivatives in Parkinson’s disease dysgraphi analysis”, in *20th Conference of the International Graphonomics Society (IGS)*, (in press), Springer, 2022, pp. 1–15.

[19] L. Brabenec, P. Klobusiakova, J. Mekyska, and I. Rektorova, “Shannon entropy: A novel parameter for quantifying pentagon copying performance in non-demented Parkinson’s disease patients”, *Parkinsonism & Related Disorders*, vol. 94, pp. 45–48, 2022.

[20] J. A. Nolazco-Flores, M. Faundez-Zanuy, V. De La Cueva, and J. Mekyska, “Exploiting spectral and cepstral handwriting features on diagnosing Parkinson’s disease”, *IEEE Access*, vol. 9, pp. 141 599–141 610, 2021.

[21] K. Safarova, J. Mekyska, and V. Zvoncak, “Developmental dysgraphia: A new approach to diagnosis.”, *The International Journal of Assessment and Evaluation*, no. 1, 2021.

[22] K. Safarova, J. Mekyska, V. Zvoncak, Z. Galaz, P. Francova, B. Cechova, B. Losenicka, Z. Smekal, T. Urbanek, J. M. Havigerova, and S. Rosenblum, “Psychometric properties of screening questionnaires for children with handwriting issues”, *Frontiers in Psychology*, p. 2937, 2020.

[23] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, V. Zvoncak, K. Safarova, T. Urbanek, J. M. Havigerova, J. Bednarova, and Z. Smekal, “Analysis of various fractional order derivatives approaches in assessment of graphomotor difficulties”, *IEEE Access*, vol. 8, pp. 218 234–218 244, 2020.

[24] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, K. Lopez-de-Ipina, V. Zvoncak, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, “Identification and monitoring of Parkinson’s disease dysgraphia based on fractional-order derivatives of online handwriting”, *Applied Sciences*, vol. 8, no. 12, p. 2566, 2018.

[25] J. Mucha, J. Mekyska, M. Faundez-Zanuy, P. S.-C. Z. Galaz, V. Zvoncak, T. Kiska, Z. Smekal, K. Lopez-de-Ipina, and I. Rektorova, “Advanced analysis of online handwriting in a multilingual cohort of patients with Parkinson’s disease”, in *Advances in Signal Processing and Artificial Intelligence: Proceedings of the 1st International Conference on Advances in Signal Processing and Artificial Intelligence*, IFSA, 2019, pp. 144–147.

[26] J. Mekyska, Z. Galaz, K. Safarova, V. Zvoncak, J. Mucha, Z. Smekal, A. Ondrackova, T. Urbanek, J. M. Havigerova, J. Bednarova, and M. Faundez-Zanuy, “Computerised assessment of graphomotor difficulties in a cohort of school-aged children”, in *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2019, pp. 1–6.

[27] V. Zvoncak, J. Mekyska, K. Safarova, Z. Smekal, and P. Brezany, “New approach of dysgraphic handwriting analysis based on the tunable Q-factor wavelet transform”, in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, 2019, pp. 289–294.

[28] J. Mucha, M. Faundez-Zanuy, J. Mekyska, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de-Ipina, “Analysis of Parkinson’s disease dysgraphia based on optimized fractional order derivative features”, in *2019 27th European Signal Processing Conference (EUSIPCO)*, IEEE, 2019, pp. 1–5.

[29] V. Zvoncak, J. Mucha, Z. Galaz, J. Mekyska, K. Safarova, M. Faundez-Zanuy, and Z. Smekal, “Fractional order derivatives evaluation in computerized assessment of handwriting difficulties in school-aged children”, in *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2019, pp. 1–6.

[30] J. Mucha, J. Mekyska, M. Faundez-Zanuy, K. Lopez-De-Ipina, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, “Advanced Parkinson’s disease dysgraphia analysis based on fractional derivatives of online handwriting”, in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2018, pp. 1–6.

[31] V. Zvoncak, J. Mekyska, K. Safarova, Z. Galaz, J. Mucha, T. Kiska, Z. Smekal, B. Losenicka, B. Cechova, P. Francova, M. Faundez-Zanuy, and S. Rosenblum, “Effect of stroke-level intra-writer normalization on computerized assessment of developmental dysgraphia”, in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2018, pp. 1–5.

[32] J. Mucha, V. Zvoncak, Z. Galaz, M. Faundez-Zanuy, J. Mekyska, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de-Ipina, “Fractional derivatives of online handwriting: A new approach of parkinsonic dysgraphia analysis”, in *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 2018, pp. 1–4.

[33] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, “Identification and rating of developmental dysgraphia by handwriting analysis”, *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, 2016.

[34] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, “Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson’s disease”, *Artificial intelligence in Medicine*, vol. 67, pp. 39–46, 2016.

[35] P. Drotar, J. Mekyska, Z. Smekal, I. Rektorova, L. Masarova, and M. Faundez-Zanuy, “Contribution of different handwriting modalities to differential diagnosis of Parkinson’s disease”, in *2015 IEEE international symposium on medical measurements and applications (MeMeA) proceedings*, IEEE, 2015, pp. 344–348.

[36] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, “Decision support framework for Parkinson’s disease based on novel handwriting markers”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 508–516, 2014.

[37] L. Masarova, P. Drotar, J. Mekyska, Z. Smekal, and I. Rektorova, "Hodnocení písma u pacientů s Parkinsonovou nemocí", *Cesk Slov Neurol N*, vol. 77, no. 110.4, pp. 456–462, 2014.

[38] M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure-Alcobe, A. Esposito, J. Mekyska, and K. Lopez-de-Ipina, "A preliminary study on aging examining online handwriting", in *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)*, IEEE, 2014, pp. 221–224.

[39] M. Faundez-Zanuy, A. Hussain, J. Mekyska, E. Sesa-Nogueras, E. Monte-Moreno, A. Esposito, M. Chetouani, J. Garre-Olmo, A. Abel, Z. Smekal, *et al.*, "Biometric applications related to human beings: There is life beyond security", *Cognitive Computation*, vol. 5, no. 1, pp. 136–151, 2013.

[40] Z. Smekal, J. Mekyska, I. Rektorova, and M. Faundez-Zanuy, "Analysis of neurological disorders based on digital processing of speech and handwritten text", in *International symposium on signals, circuits and systems ISSCS2013*, IEEE, 2013, pp. 1–6.

[41] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "A new modality for quantitative evaluation of Parkinson's disease: In-air movement", in *13th IEEE international conference on bioinformatics and bioengineering*, IEEE, 2013, pp. 1–4.

[42] M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure-Alcobe, J. Garre-Olmo, J. Mekyska, K. Lopez-de-Ipina, and A. Esposito, "A preliminary study of online drawings and dementia diagnose", in *Neural Nets and Surroundings*, Springer, 2013, pp. 367–374.

[43] P. Drotar, J. Mekyska, Z. Smekal, I. Rektorova, L. Masarova, and M. Faundez-Zanuy, "Prediction potential of different handwriting tasks for diagnosis of Parkinson's", in *2013 E-Health and Bioengineering Conference (EHB)*, IEEE, 2013, pp. 1–4.

[44] E. Sesa-Nogueras, M. Faundez-Zanuy, and J. Mekyska, "An information analysis of in-air and on-surface trajectories in online handwriting", *Cognitive Computation*, vol. 4, no. 2, pp. 195–205, 2012.

[45] O. Hornykiewicz, "Biochemical aspects of Parkinson's disease", *Neurology*, vol. 51, no. 2 Suppl 2, S2–S9, 1998.

[46] D. W. Dickson, "Parkinson's disease and parkinsonism: Neuropathology", *Cold Spring Harbor perspectives in medicine*, vol. 2, no. 8, a009258, 2012.

[47] L. S. Forno, "Neuropathology of Parkinson's disease", *Journal of Neuropathology & Experimental Neurology*, vol. 55, no. 3, pp. 259–272, 1996.

[48] S. Heinzel, D. Berg, T. Gasser, H. Chen, C. Yao, R. B. Postuma, and M. T. F. on the Definition of Parkinson's Disease, "Update of the MDS research criteria for prodromal Parkinson's disease", *Movement Disorders*, vol. 34, no. 10, pp. 1464–1470, 2019.

[49] P. Brodal, *The Central Nervous System: Structure and Function*, 3rd ed. Oxford University Press, 2003.

[50] A. Hughes, S. Daniel, and A. Lees, "The clinical features of Parkinson's disease in 100 histologically proven cases", *Advances in neurology*, vol. 60, p. 595, 1993.

[51] A. Berardelli, J. Rothwell, P. Thompson, and M. Hallett, "Pathophysiology of bradykinesia in Parkinson's disease", *Brain*, vol. 124, no. 11, pp. 2131–2146, 2001.

[52] F. B. Horak, D. Dimitrova, and J. G. Nutt, "Direction-specific postural instability in subjects with Parkinson's disease", *Experimental neurology*, vol. 193, no. 2, pp. 504–521, 2005.

[53] L. Brabenec, J. Mekyska, Z. Galaz, and I. Rektorova, “Speech disorders in Parkinson’s disease: Early diagnostics and effects of medication and brain stimulation”, *Journal of neural transmission*, vol. 124, no. 3, pp. 303–334, 2017.

[54] S. Pinto and J.-L. Velay, “Handwriting as a marker for PD progression: A shift in paradigm”, *Neurodegenerative disease management*, vol. 5, no. 5, pp. 367–369, 2015.

[55] M. Thomas, A. Lenka, and P. Kumar Pal, “Handwriting analysis in Parkinson’s disease: Current status and future directions”, *Movement disorders clinical practice*, vol. 4, no. 6, pp. 806–818, 2017.

[56] R. Saunders-Pullman, C. Derby, K. Stanley, A. Floyd, S. Bressman, R. B. Lipton, A. Deligtisch, L. Severt, Q. Yu, M. Kurtis, *et al.*, “Validity of spiral analysis in early Parkinson’s disease”, *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 4, pp. 531–537, 2008.

[57] J. McLennan, K. Nakano, H. Tyler, and R. Schwab, “Micrographia in Parkinson’s disease”, *Journal of the Neurological Sciences*, vol. 15, no. 2, pp. 141–152, 1972, ISSN: 0022-510X.

[58] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, and I. Schlesinger, “Handwriting as an objective tool for Parkinson’s disease diagnosis”, *Journal of neurology*, vol. 260, no. 9, pp. 2357–2361, 2013.

[59] A. A. Alhusaini, G. R. Melam, and S. Buragadda, “Short-term sensorimotor-based intervention for handwriting performance in elementary school children”, *Pediatrics International*, vol. 58, no. 11, pp. 111–1123, 2016.

[60] K. P. Feder and A. Majnemer, “Handwriting development, competency, and intervention”, *Developmental Medicine & Child Neurology*, vol. 49, no. 4, pp. 312–317, 2007.

[61] V. Matijević-Mikelić, T. Košiček, M. Crnković, Z. Trifunović-Maček, and S. Grazio, “Development of early graphomotor skills in children with neurodevelopmental risks”, *Acta clinica Croatica*, vol. 50, no. 3, pp. 317–321, 2011.

[62] A. P. Accardo, M. Genna, and M. Borean, “Development, maturation and learning influence on handwriting kinematics”, *Human movement science*, vol. 32, no. 1, pp. 136–146, 2013.

[63] A. Kushki, H. Schwellnus, F. Ilyas, and T. Chau, “Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia”, *Research in Developmental Disabilities*, vol. 32, no. 3, pp. 1058–1064, 2011.

[64] A. W. V. Gemmert and H.-L. Teulings, “Advances in graphonomics: Studies on fine motor control, its development and disorders”, *Human Movement Science*, vol. 25, no. 4-5, pp. 447–453, 2006.

[65] K. McHale and S. A. Cermak, “Fine motor activities in elementary school: Preliminary findings and provisional implications for children with fine motor problems”, *American Journal of Occupational Therapy*, vol. 46, no. 10, pp. 898–903, 1992.

[66] B. Engel-Yeger, L. Nagauker-Yanuv, and S. Rosenblum, “Handwriting performance, self-reports, and perceived self-efficacy among children with dysgraphia”, *American Journal of Occupational Therapy*, vol. 63, no. 2, pp. 182–192, 2009.

[67] A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association, 2013.

[68] W. H. Organization, *International statistical classification of diseases and related health problems (10th ed.)* 2016. [Online]. Available: `https://icd.who.int/browse10/2016/en`.

[69] B. C. Smits-Engelsman and G. P. Van Galen, “Dysgraphia in children: Lasting psychomotor deficiency or transient developmental delay?”, *Journal of experimental child psychology*, vol. 67, no. 2, pp. 164–184, 1997.

[70] S. T. Peverly, J. K. Garner, and P. C. Vekaria, “Both handwriting speed and selective attention are important to lecture note-taking”, *Reading and Writing*, vol. 27, no. 1, pp. 1–30, 2014.

[71] J. Medwell and D. Wray, “Handwriting automaticity: The search for performance thresholds”, *Language and Education*, vol. 28, no. 1, pp. 34–51, 2014.

[72] L. Hamstra-Bletz, J. DeBie, B. Den Brinker, *et al.*, “Concise evaluation scale for children’s handwriting”, *Lisse: Swets*, vol. 1, 1987.

[73] S. Rosenblum, “Development, reliability, and validity of the handwriting proficiency screening questionnaire (HPSQ)”, *The American Journal of Occupational Therapy*, vol. 62, no. 3, pp. 298–307, 2008.

[74] S. Rosenblum and L. Gafni-Lachter, “Handwriting proficiency screening questionnaire for children (HPSQ–C): Development, reliability, and validity”, *The American Journal of Occupational Therapy*, vol. 69, no. 3, 6903220030p1–6903220030p9, 2015.

[75] D. V. Cicchetti, “Assessing inter-rater reliability for rating scales: Resolving some basic issues”, *The British Journal of Psychiatry*, vol. 129, no. 5, pp. 452–456, 1976.

[76] S. Rosenblum, P. L. Weiss, and S. Parush, “Product and process evaluation of handwriting difficulties”, *Educational Psychology Review*, vol. 15, no. 1, pp. 41–81, 2003.

[77] V. Paz-Villagrán, J. Danna, and J.-L. Velay, “Lifts and stops in proficient and dysgraphic handwriting”, *Human movement science*, vol. 33, pp. 381–394, 2014.

[78] E. Sesa-Nogueras and M. Faundez-Zanuy, “Biometric recognition using online uppercase handwritten text”, *Pattern Recognition*, vol. 45, no. 1, pp. 128–144, 2012.

[79] M. D. S. T. F. on Rating Scales for Parkinson’s Disease, “The unified Parkinson’s disease rating scale (UPDRS): Status and recommendations”, *Movement Disorders*, vol. 18, no. 7, pp. 738–750, 2003.

[80] J.-Y. Lee, J. W. Kim, W. Y. Lee, J.-M. Kim, T.-B. Ahn, H.-J. Kim, J. Cho, and B. S. Jeon, “Daily dose of dopaminergic medications in Parkinson disease: Clinical correlates and a posteriori equation.”, *Neurology Asia*, vol. 15, 2010.

[81] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, “The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment”, *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, 2005.

[82] I. G. McKeith, T. J. Ferman, A. J. Thomas, F. Blanc, B. F. Boeve, H. Fujishiro, K. Kantarci, C. Muscio, J. T. O’Brien, R. B. Postuma, *et al.*, “Research criteria for the diagnosis of prodromal dementia with Lewy bodies”, *Neurology*, vol. 94, no. 17, pp. 743–755, 2020.

[83] H. Benedict, “Brief Visual Memory Test–revised: Professional manual”, *Odessa, FL: Psychological Assessment Resources*, 1997.

[84] O. Bezdicek, D. J. Libon, H. Stepankova, E. Panenkova, J. Lukavsky, K. D. Garrett, M. Lamar, C. C. Price, and M. Kopecek, “Development, validity, and normative data study for the 12-word Philadelphia Verbal Learning Test [czP (r) VLT-12] among older and very old Czech adults”, *The Clinical Neuropsychologist*, vol. 28, no. 7, pp. 1162–1181, 2014.

[85] D. Wechsler *et al.*, *WAIS-III WMS-III Technical Manual.* Psychological Corporation San Antonio, TX, 1997.

[86] M. Preiss, Z. Kalivodová, I. Kundrátová, L. Mrlinová, T. Ježková, M. Kubů, and P. Houbová, "Test verbální fluence–vodítka pro všeobecnou dospělou populaci", *Psychiatrie*, vol. 6, no. 2, pp. 74–77, 2002.

[87] E. K. Warrington and M. James, "The visual object and space perception battery", 1991.

[88] M. San Luciano, C. Wang, R. A. Ortega, Q. Yu, S. Boschung, J. Soto-Valencia, S. B. Bressman, R. B. Lipton, S. Pullman, and R. Saunders-Pullman, "Digitized spiral drawing: A possible biomarker for early Parkinson's disease", *PloS one*, vol. 11, no. 10, e0162799, 2016.

[89] G. D. Cascarano, C. Loconsole, A. Brunetti, A. Lattarulo, D. Buongiorno, G. Losavio, E. Di Sciascio, and V. Bevilacqua, "Biometric handwriting analysis to support Parkinson's disease assessment and grading", *BMC medical informatics and decision making*, vol. 19, no. 9, pp. 1–11, 2019.

[90] Z. Galaz, J. Mucha, V. Zvoncak, and J. Mekyska, *Handwriting features*, version 1.0.2, May 2022. [Online]. Available: `https://github.com/BDALab/handwriting-features/`.

[91] J. Mucha, Z. Galaz, V. Zvoncak, and J. Mekyska, *Handwriting sample*, version 1.0.0, Nov. 2021. [Online]. Available: `https://github.com/BDALab/handwriting-sample/`.

[92] ——, *HandAQUS – handwriting acquisition software*, version 1.0.0, 2021. [Online]. Available: `https://github.com/BDALab/HandAQUS`.

[93] ——, *HandAQUS for Cintiq*, version 1.0.0, 2021. [Online]. Available: `https://github.com/BDALab/HandAQUS_for_Cintiq`.

[94] T. Asselborn, M. Chapatte, and P. Dillenbourg, "Extending the spectrum of dysgraphia: A data driven strategy to estimate handwriting quality", *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020.

[95] J. Rusz, T. Tykalova, M. Novotny, D. Zogala, K. Sonka, E. Ruzicka, and P. Dusek, "Defining speech subtypes in De Novo Parkinson disease: Response to long-term levodopa therapy", *Neurology*, vol. 97, no. 21, e2124–e2135, 2021.

[96] M. Diaz, M. A. Ferrer, D. Impedovo, G. Pirlo, and G. Vessio, "Dynamically enhanced static handwriting representation for Parkinson's disease detection", *Pattern Recognition Letters*, vol. 128, pp. 204–210, 2019.

[97] M. Diaz, M. Moetesum, I. Siddiqi, and G. Vessio, "Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and bigrus", *Expert Systems with Applications*, vol. 168, p. 114 405, 2021.

[98] B. M. Morberg, A. S. Malling, B. R. Jensen, O. Gredal, L. Wermuth, and P. Bech, "The Hawthorne effect as a pre-placebo expectation in Parkinsons disease patients participating in a randomized placebo-controlled clinical study", *Nordic journal of psychiatry*, vol. 72, no. 6, pp. 442–446, 2018.

List of author's publications

This list contains only publications related to the topic of the thesis. The full list of the author's publications could be found in his ORCID profile[3]. Reported metrics were retrieved from the Web of Science (WoS) database on 9th June 2022. The numbers of citations are considered **without self-citations**.

5.3 Journal articles

[J1] Z. Galaz, P. Drotar, J. Mekyska, M. Gazda, J. Mucha, V. Zvoncak, Z. Smekal, M. Faundez-Zanuy, R. Castrillon, J. R. Orozco-Arroyave, S. Rapcsak, T. Kincses, L. Brabenec, and I. Rektorova, "Comparison of CNN-learned vs. handcrafted features for detection of Parkinson's disease dysgraphia in a multilingual dataset", *Frontiers in Neuroinformatics*, vol. 1, no. 1, pp. 1–26, 2022

Quartile:	Q1 (Mathematical & Computational Biology)
	Q2 (Neurosciences)
Impact factor:	4.081
Number of citations:	0
Author's contribution:	30 % (supervision, identification of knowledge gaps, data curation, data processing, investigation, methodology, visualisation, discussion)
Reprint:	Appendix A.1

[J2] L. Brabenec, P. Klobusiakova, J. Mekyska, and I. Rektorova, "Shannon entropy: A novel parameter for quantifying pentagon copying performance in non-demented Parkinson's disease patients", *Parkinsonism & Related Disorders*, vol. 94, pp. 45–48, 2022

Quartile:	Q1 (Clinical Neurology)
Impact factor:	4.891
Number of citations:	0
Author's contribution:	30 % (data curation, data processing, investigation, methodology, results, visualisation)
Reprint:	Appendix A.4

[J3] J. A. Nolazco-Flores, M. Faundez-Zanuy, V. De La Cueva, and J. Mekyska, "Exploiting spectral and cepstral handwriting features on diagnosing Parkinson's disease", *IEEE Access*, vol. 9, pp. 141 599–141 610, 2021

Quartile:	Q2 (Computer Science, Information Systems)
	Q2 (Engineering, Electrical & Electronic)
	Q2 (Telecommunications)
Impact factor:	3.367
Number of citations:	2
Author's contribution:	30 % (data curation, discussion)
Reprint:	Appendix A.5

[3] `https://orcid.org/0000-0002-6195-193X`

[J4] M. Faundez-Zanuy, J. Mekyska, and D. Impedovo, “Online handwriting, signature and touch dynamics: Tasks and potential applications in the field of security and health”, *Cognitive Computation*, vol. 13, no. 5, pp. 1406–1421, 2021

Quartile:	Q1 (Computer Science, Artificial Intelligence) Q1 (Neurosciences)
Impact factor:	5.418
Number of citations:	1
Author’s contribution:	40 % (data curation, data processing, investigation, methodology, results, visualisation, discussion)
Reprint:	Appendix A.6

[J5] K. Safarova, J. Mekyska, and V. Zvoncak, “Developmental dysgraphia: A new approach to diagnosis.”, *The International Journal of Assessment and Evaluation*, no. 1, 2021

Quartile:	not indexed by WoS
Impact factor:	0.000
Number of citations:	0
Author’s contribution:	20 % (identification of knowledge gaps, investigation, discussion)
Reprint:	Appendix A.7

[J6] K. Safarova, J. Mekyska, V. Zvoncak, Z. Galaz, P. Francova, B. Cechova, B. Losenicka, Z. Smekal, T. Urbanek, J. M. Havigerova, and S. Rosenblum, “Psychometric properties of screening questionnaires for children with handwriting issues”, *Frontiers in Psychology*, p. 2937, 2020

Quartile:	Q2 (Psychology, Multidisciplinary)
Impact factor:	2.988
Number of citations:	1
Author’s contribution:	20 % (data curation, investigation, discussion)
Reprint:	Appendix A.8

[J7] Z. Galaz, J. Mucha, V. Zvoncak, J. Mekyska, Z. Smekal, K. Safarova, A. Ondrackova, T. Urbanek, J. M. Havigerova, J. Bednarova, and M. Faundez-Zanuy, “Advanced parametrization of graphomotor difficulties in school-aged children”, *IEEE Access*, vol. 8, pp. 112 883–112 897, 2020

Quartile:	Q2 (Computer Science, Information Systems) Q2 (Engineering, Electrical & Electronic) Q2 (Telecommunications)
Impact factor:	3.367
Number of citations:	3
Author’s contribution:	30 % (supervision, identification of knowledge gaps, data curation, investigation, methodology, discussion)
Reprint:	Appendix A.9

[J8] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, V. Zvoncak, K. Safarova, T. Urbanek, J. M. Havigerova, J. Bednarova, and Z. Smekal, “Analysis of various fractional order derivatives approaches in assessment of graphomotor difficulties”, *IEEE Access*, vol. 8, pp. 218 234–

218 244, 2020

Quartile:	Q2 (Computer Science, Information Systems)
	Q2 (Engineering, Electrical & Electronic)
	Q2 (Telecommunications)
Impact factor:	3.367
Number of citations:	0
Author's contribution:	20 % (data curation, investigation, discussion)
Reprint:	Appendix A.10

[J9] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, K. Lopez-de-Ipina, V. Zvoncak, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Identification and monitoring of Parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting", *Applied Sciences*, vol. 8, no. 12, p. 2566, 2018

Quartile:	Q2 (Engineering, Multidisciplinary)
	Q2 (Physics, Applied)
	Q3 (Chemistry, Multidisciplinary)
	Q3 (Materials Science, Multidisciplinary)
Impact factor:	2.679
Number of citations:	11
Author's contribution:	20 % (data curation, investigation, discussion)
Reprint:	Appendix A.11

[J10] C. Alonso-Martinez, M. Faundez-Zanuy, and J. Mekyska, "A comparative study of in-air trajectories at short and long distances in online handwriting", *Cognitive Computation*, vol. 9, no. 5, pp. 712–720, 2017

Quartile:	Q1 (Computer Science, Artificial Intelligence)
	Q1 (Neurosciences)
Impact factor:	5.418
Number of citations:	6
Author's contribution:	30 % (data curation, data processing, investigation, discussion)
Reprint:	Appendix A.20

[J11] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and rating of developmental dysgraphia by handwriting analysis", *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, 2016

Quartile:	Q2 (Computer Science, Artificial Intelligence)
	Q2 (Computer Science, Cybernetics)
Impact factor:	2.968
Number of citations:	14
Author's contribution:	80 % (supervision, identification of knowledge gaps, data processing, investigation, methodology, results, visualisation, discussion)
Reprint:	Appendix A.22

[J12] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease", *Artificial intelligence in Medicine*, vol. 67, pp. 39–46, 2016

Quartile:	Q1 (Computer Science, Artificial Intelligence)
	Q1 (Engineering, Biomedical)
	Q1 (Medical Informatics)
Impact factor:	5.326
Number of citations:	107
Author's contribution:	40 % (supervision, data curation, investigation, methodology, discussion)
Reprint:	Appendix A.23

[J13] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Decision support framework for Parkinson's disease based on novel handwriting markers", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 508–516, 2014

Quartile:	Q1 (Rehabilitation)
	Q2 (Engineering, Biomedical)
Impact factor:	3.802
Number of citations:	65
Author's contribution:	40 % (supervision, data curation, investigation, methodology, discussion)
Reprint:	Appendix A.25

[J14] L. Masarova, P. Drotar, J. Mekyska, Z. Smekal, and I. Rektorova, "Hodnocení písma u pacientů s Parkinsonovou nemocí", *Cesk Slov Neurol N*, vol. 77, no. 110.4, pp. 456–462, 2014

Quartile:	Q4 (Neurosciences)
	Q4 (Surgery)
Impact factor:	0.350
Number of citations:	0
Author's contribution:	40 % (supervision, data curation, investigation, methodology, discussion)
Reprint:	Appendix A.26

[J15] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease", *Computer Methods and Programs in Biomedicine*, vol. 117, no. 3, pp. 405–411, 2014

Quartile:	Q1 (Computer Science, Interdisciplinary Applications)
	Q1 (Computer Science, Theory & Methods)
	Q1 (Engineering, Biomedical)
	Q1 (Medical Informatics)
Impact factor:	5.428
Number of citations:	49
Author's contribution:	40 % (supervision, data curation, investigation, methodology, discussion)
Reprint:	Appendix A.27

[J16] M. Faundez-Zanuy, A. Hussain, J. Mekyska, E. Sesa-Nogueras, E. Monte-Moreno, A. Esposito, M. Chetouani, J. Garre-Olmo, A. Abel, Z. Smekal, *et al.*, "Biometric applications related to human beings: There is life beyond security", *Cognitive Computation*, vol. 5, no. 1, pp. 136–151, 2013

Quartile:	Q1 (Computer Science, Artificial Intelligence) Q1 (Neurosciences)
Impact factor:	5.418
Number of citations:	33
Author's contribution:	30 % (identification of knowledge gaps, data curation, data processing, investigation, methodology, results, visualisation, discussion)
Reprint:	Appendix A.29

[J17] E. Sesa-Nogueras, M. Faundez-Zanuy, and J. Mekyska, "An information analysis of in-air and on-surface trajectories in online handwriting", *Cognitive Computation*, vol. 4, no. 2, pp. 195–205, 2012

Quartile:	Q1 (Computer Science, Artificial Intelligence) Q1 (Neurosciences)
Impact factor:	5.418
Number of citations:	28
Author's contribution:	30 % (data processing, investigation, methodology, results)
Reprint:	Appendix A.34

5.4 Papers in conference proceedings

[D1] Z. Galaz, J. Mekyska, J. Mucha, V. Zvoncak, Z. Smekal, M. Faundez-Zanuy, L. Brabenec, I. Moravkova, and I. Rektorova, "Prodromal diagnosis of Lewy body diseases based on the assessment of graphomotor and handwriting difficulties", in *20th Conference of the International Graphonomics Society (IGS)*, (in press), Springer, 2022, pp. 1–16

Number of citations:	0
Author's contribution:	60 % (supervision, identification of knowledge gaps, data curation, data processing, investigation, methodology, results, discussion)
Reprint:	Appendix A.2

[D2] J. Mucha, Z. Galaz, J. Mekyska, M. Faundez-Zanuy, V. Zvoncak, Z. Smekal, L. Brabenec, and I. Rektorova, "Exploration of various fractional order derivatives in Parkinson's disease dysgraphi analysis", in *20th Conference of the International Graphonomics Society (IGS)*, (in press), Springer, 2022, pp. 1–15

Number of citations:	0
Author's contribution:	20 % (data curation, investigation, discussion)
Reprint:	Appendix A.3

[D3] J. Mucha, J. Mekyska, M. Faundez-Zanuy, P. S.-C. Z. Galaz, V. Zvoncak, T. Kiska, Z. Smekal, K. Lopez-de-Ipina, and I. Rektorova, "Advanced analysis of online handwriting in a multilingual cohort of patients with Parkinson's disease", in *Advances in Signal Processing and Artificial Intelligence: Proceedings of the 1st International Conference on Advances in Signal Processing and Artificial Intelligence*, IFSA, 2019, pp. 144–147

Number of citations:	0
Author's contribution:	30 % (data curation, investigation, discussion)
Reprint:	Appendix A.12

[D4] J. Mekyska, Z. Galaz, K. Safarova, V. Zvoncak, J. Mucha, Z. Smekal, A. Ondrackova, T. Urbanek, J. M. Havigerova, J. Bednarova, and M. Faundez-Zanuy, "Computerised assessment of graphomotor difficulties in a cohort of school-aged children", in *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2019, pp. 1–6

Number of citations:	0
Author's contribution:	80 % (supervision, identification of knowledge gaps, data curation, data processing, investigation, methodology, results, visualisation, discussion)
Reprint:	Appendix A.13

[D5] V. Zvoncak, J. Mekyska, K. Safarova, Z. Smekal, and P. Brezany, "New approach of dysgraphic handwriting analysis based on the tunable Q-factor wavelet transform", in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, 2019, pp. 289–294

Number of citations:	1
Author's contribution:	30 % (supervision, identification of knowledge gaps, data curation, investigation, discussion)
Reprint:	Appendix A.14

[D6] J. Mucha, M. Faundez-Zanuy, J. Mekyska, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de-Ipina, "Analysis of Parkinson's disease dysgraphia based on optimized fractional order derivative features", in *2019 27th European Signal Processing Conference (EUSIPCO)*, IEEE, 2019, pp. 1–5

Number of citations:	0
Author's contribution:	20 % (data curation, investigation, discussion)
Reprint:	Appendix A.15

[D7] V. Zvoncak, J. Mucha, Z. Galaz, J. Mekyska, K. Safarova, M. Faundez-Zanuy, and Z. Smekal, "Fractional order derivatives evaluation in computerized assessment of handwriting difficulties in school-aged children", in *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2019, pp. 1–6

Number of citations:	0
Author's contribution:	30 % (data curation, investigation, discussion)
Reprint:	Appendix A.16

[D8] J. Mucha, J. Mekyska, M. Faundez-Zanuy, K. Lopez-De-Ipina, V. Zvoncak, Z. Galaz, T.

Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, “Advanced Parkinson’s disease dysgraphia analysis based on fractional derivatives of online handwriting”, in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2018, pp. 1–6

Number of citations:	1
Author’s contribution:	20 % (data curation, investigation, discussion)
Reprint:	Appendix A.17

[D9] V. Zvoncak, J. Mekyska, K. Safarova, Z. Galaz, J. Mucha, T. Kiska, Z. Smekal, B. Losenicka, B. Cechova, P. Francova, M. Faundez-Zanuy, and S. Rosenblum, “Effect of stroke-level intra-writer normalization on computerized assessment of developmental dysgraphia”, in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2018, pp. 1–5

Number of citations:	0
Author’s contribution:	30 % (supervision, identification of knowledge gaps, data curation, investigation, methodology, discussion)
Reprint:	Appendix A.18

[D10] J. Mucha, V. Zvoncak, Z. Galaz, M. Faundez-Zanuy, J. Mekyska, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de-Ipina, “Fractional derivatives of online handwriting: A new approach of parkinsonic dysgraphia analysis”, in *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 2018, pp. 1–4

Number of citations:	1
Author’s contribution:	20 % (data curation, investigation, discussion)
Reprint:	Appendix A.19

[D11] P. Drotar, J. Mekyska, Z. Smekal, I. Rektorova, L. Masarova, and M. Faundez-Zanuy, “Contribution of different handwriting modalities to differential diagnosis of Parkinson’s disease”, in *2015 IEEE international symposium on medical measurements and applications (MeMeA) proceedings*, IEEE, 2015, pp. 344–348

Number of citations:	19
Author’s contribution:	40 % (supervision, data curation, investigation, methodology, discussion)
Reprint:	Appendix A.24

[D12] M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure-Alcobe, A. Esposito, J. Mekyska, and K. Lopez-de-Ipina, “A preliminary study on aging examining online handwriting”, in *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)*, IEEE, 2014, pp. 221–224

Number of citations:	2
Author’s contribution:	30 % (data processing, investigation, methodology, results)
Reprint:	Appendix A.28

[D13] Z. Smekal, J. Mekyska, I. Rektorova, and M. Faundez-Zanuy, “Analysis of neurological disorders based on digital processing of speech and handwritten text”, in *International symposium* *on signals, circuits and systems ISSCS2013*, IEEE, 2013, pp. 1–6

Number of citations:	1
Author's contribution:	80 % (identification of knowledge gaps, data curation, data processing, investigation, methodology, results, visualisation, discussion)
Reprint:	Appendix A.30

[D14] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "A new modality for quantitative evaluation of Parkinson's disease: In-air movement", in *13th IEEE international conference on bioinformatics and bioengineering*, IEEE, 2013, pp. 1–4

Number of citations:	17
Author's contribution:	40 % (supervision, data curation, investigation, methodology, discussion)
Reprint:	Appendix A.31

[D15] M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure-Alcobe, J. Garre-Olmo, J. Mekyska, K. Lopez-de-Ipina, and A. Esposito, "A preliminary study of online drawings and dementia diagnose", in *Neural Nets and Surroundings*, Springer, 2013, pp. 367–374

Number of citations:	1 (Scopus)
Author's contribution:	20 % (data processing, investigation, methodology, results, visualisation)
Reprint:	Appendix A.32

[D16] P. Drotar, J. Mekyska, Z. Smekal, I. Rektorova, L. Masarova, and M. Faundez-Zanuy, "Prediction potential of different handwriting tasks for diagnosis of Parkinson's", in *2013 E-Health and Bioengineering Conference (EHB)*, IEEE, 2013, pp. 1–4

Number of citations:	24
Author's contribution:	40 % (supervision, data curation, investigation, methodology, discussion)
Reprint:	Appendix A.33

5.5 Chapters in books

[C1] M. Faundez-Zanuy and J. Mekyska, "Privacy of online handwriting biometrics related to biomedical analysis", in *User-Centric Privacy and Security in Biometrics*, Institution of Engineering and Technology, 2017, pp. 17–39

Number of citations:	3 (Scopus)
Author's contribution:	50 % (identification of knowledge gaps, data processing, investigation, methodology, results, visualisation, discussion)
Reprint:	Appendix A.21

Jiří Mekyska

Head of the Brain Diseases Analysis Laboratory
Brno University of Technology

Personal Information

Birth:	May 11, 1985
Address:	Na Pomezí 1534/13a, 747 06 Opava, Czech Republic
Cell phone:	+420 607 700 329
E-mail:	mekyska@vut.cz

Education

2010–2014	Brno University of Technology, Faculty of Electrical Engineering and Communication, Technická 3058/10, 616 00 Brno, Czech Republic, degree: Ph.D.
2008–2010	Brno University of Technology, Faculty of Electrical Engineering and Communication, Technická 3058/10, 616 00 Brno, Czech Republic, degree: M.Sc.
2005–2008	Brno University of Technology, Faculty of Electrical Engineering and Communication, Technická 3058/10, 616 00 Brno, Czech Republic, degree: B.Sc.
1997–2005	Grammar School Kojetín, Svatopluka Čecha 68, 752 01 Kojetín, Czech Republic

Additional Information on Education

2009–2010	Escola Universitària Politècnica de Mataró, Avda. Puig i Cadafalch 101-111, 083 03 Mataró (Barcelona), Spain
1999	Brighton International Summer School, PO Box 2831, East Sussex, Brighton, United Kingdom

Internships

2014	Centre de recherches INRIA Bordeaux Sud-Ouest, 200 rue de la Vieille Tour, 334 05 Talence Cedex, France
2011	Pompeu Fabra University, Escola Universitària Politècnica de Mataró, Avda. Puig i Cadafalch 101-111, 083 03 Mataró (Barcelona), Spain

Prizes and Awards

2021	Best Paper Award (paper: Automatic Segmentation of Actigraphy Data Utilising Gradient Boosting Algorithm) at TSP 2021
2019	Best Paper Award (paper: Comparing Parkinson's Disease Dysarthria and Aging Speech using Articulation Kinematics) at BIOSIGNALS 2019
2013	Joseph Fourier's Award for the scientific work in the field of non-invasive neurodegenerative disorders analysis
2010	Brno University of Technology rector's prize for the master thesis "Identification of persons via voice imprint"

2010	Brno University of Technology dean's prize for the master thesis "Identification of persons via voice imprint"
2010	Master's degree with honours
2008	Bachelor's degree with honours

Employment History

2010–now Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Telecommunications

- leadership: head of the Brain Diseases Analysis Laboratory, head of the Human-machine Interaction Group at the Signal Processing Laboratory, project management
- researcher: non-invasive analysis of neurodegenerative and neurodevelopmental disorders utilising acoustic and online handwriting analysis, actigraphy, machine learning, statistical analysis, programming in Matlab and Python

2007–2009 Honeywell, spol. s. r. o. – HTS CZ

- software developer: Honeywell Aerospace, activities in Brno Datalink team – implementation of GE Fanuc Condor 520 PCI ARINC 429 to Honeywell Airsim environment, Small C scripting, system testing in TATS
- software developer: Honeywell Aerospace, activities in Brno FADEC (Full Authority Digital Engine Control) team – Simulink models analysis, The Honeywell Integrated Lifecycle Tools and Environment (HiLiTE) analysis, development of tool which automatically generate input signal range of Simulink models

Memberships and Synergistic Activities

2022–now	Innovative Health Initiative, member of the Science and Innovation Panel
2021–now	International Parkinson and Movement Disorder Society, member of the Speech and Movement Disorders Study Group
2021–now	Czech Science Foundation, Chairman of the Expert Panel P103 (Cybernetics, Artificial Intelligence and Information Processing)
2018–now	Research, Development and Innovation Council of the Czech government, member of the Expert Advisory Body of Evaluators
2021	Technology Agency of the Czech Republic, consultations regarding a preparation of Horizon Europe project proposals
2018–2019, 2021	Innovative Medicines Initiative, Member of Expert Panel
2019–2021	Czech Science Foundation, Member of Expert Panel P103 (Cybernetics and Information Processing)

Fields of interest

- *Acoustic analysis* – design of acoustic features quantifying specific disorders in all dimensions of speech (respiration, phonation, articulation, prosody); monitoring progress of hypokinetic dysarthria (HD); monitoring effect of treatment (e.g. repetitive transcranial magnetic stimulation) on HD; revealing pathophysiological mechanisms common for HD and other motor/non-motor features of Parkinson's disease; predicting changes in time; differential analysis; identification of Lewy body diseases in prodromal state; development of epilepsy treatment methods based on musicotherapy

- *Online handwriting processing* – design of parameters quantifying Parkinson's disease and developmental dysgraphia; establishing complex dysgraphia profile; design of systems diagnosing and assessing developmental dysgraphia
- *Sleep analysis* – diagnosis of rapid eye movement (REM) sleep behavior disorder in subjects in the risk of having Lewy body diseases (employing actigraphy)
- *Signal processing and machine learning* – 1D signal parameterization and analysis, machine learning, statistical analysis
- *Health 4.0* – integration of proposed methods into monitoring and decision support systems

Participation in Projects
Principal investigator/co-investigator

2022–2025 Czech Ministry of Education, Youth and Sports project no. LX22NPO5107, *National Institute for Neurological Research*

2020–2023 Czech Ministry of Health, project no. NU20-04-00294, *Diagnostics of Lewy body diseases in prodromal stage based on multimodal data analysis*

2020–2023 Technology Agency of the Czech Republic, project no. TL03000287, *Software for advanced diagnosis of graphomotor disabilities*

2016–2019 Czech Ministry of Health, project no. 16-30805A, *Effects of non-invasive brain stimulation on hypokinetic dysarthria, micrographia, and brain plasticity in patients with Parkinson's disease*

2018–2020 Czech Science Foundation, project no. 18-16835S, *Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing*

Researcher

2022–2025 Czech Ministry of Health, project no. NU22J-04-00074, *Home-based non-invasive brain stimulation in combination with Lee Silverman Voice Treatment on hypokinetic dysarthria in Parkinson's disease*

2021–2024 Spanish Ministry of Science and Innovation, project no. PID2020-113242RB-I00, *Advanced Online Writing Preprocessing Techniques for Device Interoperability*

2019–2021 Interreg CENTRAL EUROPE, project niCE-life, *Development of an integrated concept for the deployment of innovative technologies and services allowing independent living of frail elderly*

2017–2021 Horizon 2020, Marie Skłodowska-Curie project CoBeN, *Novel Network-Based Approaches for Studying Cognitive Dysfunction in Behavioral Neurology*

2017–2020 Technology Agency of the Czech Republic, project no. VI20172020078, *System for centralized supervision of complex and large area objects of critical state infrastructure*

2015–2019 Czech Ministry of Education, Youth and Sports, project no. LO1401, *Interdisciplinary Research of Wireless Technologies*

2014–2017 Technology Agency of the Czech Republic, project no. TA04031666, *Intelligent Telematics Information System of Public Transportation II*

2013–2016	COST, project no. IC1206, *De-Identification for Privacy Protection in Multimedia Content*
2012–2015	Czech Ministry of Health, project no. NT13499, *Speech, its Impairment and Cognitive Performance in Parkinson's Disease*
2012–2015	Czech Ministry of Industry and Trade, project no. FR-TI4/696, *Localization and classification of vibrations by scattered fibre optic sensor along long distances*
2012–2014	Czech Ministry of Health, project no. GAP102/12/1104, *Study of Metabolism and Localization of High Grade Glioma using MR Imaging Techniques*
2011–2014	Czech Ministry of Education, Youth and Sports, project no. CZ.1.07/2.3.00/20.0094, *Support for Incorporating R&D Teams in International Cooperation in the Area of Image and Audio Signal Processing*
2010–2012	Czech Ministry of Education, Youth and Sports, project no. ME10123, *The Research of Algorithms for Processing of Digital Images and Image Sequences*
2010–2013	Czech Ministry of Education, Youth and Sports, project no. ED2.1.00/03.0072, *Centre of Sensor, Information and Communication Systems*
2010–2014	Czech Ministry of Interior, project no. VG20102014033, *Improvement of Risk Area Security Using Combined Methods for Biometrical Identification of Subjects*
2008–2010	COST, project no. OC08057, *Analysis and Enhancement of Speech and Image Signals form Noise for Cross-Modal Analysis of Verbal and Non-verbal Communication*

Invited Lectures

2021	*Data Science in Small Medical Data Sets: From Logistic Regression Towards Logistic Regression*, Johns Hopkins University, Whiting School of Engineering, Center for Language and Speech Processing, Hackerman 226, 3400 North Charles Street, Baltimore, MD 21218-2680, USA
2020	*Machine Learning in Limited Medical Data Sets: Doing Precision Guesswork on Unreliable Data Provided by Those with High Expectations*, The 12th International Congress on Ultra Modern Telecommunications and Control Systems (online)
2020	*Objective Diagnosis of Graphomotor Difficulties in Children with Developmental Dysgraphia*, Wacom Connected Ink 2020 (online)
2019	*Acoustic Analysis of Hypokinetic Dysarthria as a Monitoring and Predictive Tool in Parkinson's Disease*, Massachusetts Institute of Technology, 77 Massachusetts Avenue Cambridge, MA 02139, USA
2018	*Acoustic Analysis of Speech and Voice Disorders in Parkinson's Disease: a Decision Support and Monitoring Tool*, Acoustics Research Institute, Wohllebengasse 12-14, A-1040, Vienna, Austria
2017	*Acoustic Analysis of Hypokinetic Dysarthria in Patients with Parkinson's Disease: From Basics to Integration in mHealth Systems*, 12th Conference on Neurogenic Adult Communication Disorders, Brno, Czech Republic
2016	*Paraclinical Analysis of Parkinson's Disease: Moving from Diagnosis to Complex Assessment*, Institute for Technological Development and Innovation in Communications (IDeTIC), University of Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain
2014	*Neurological Disorders Analysis Using the Speech Signal Processing*, Department of Information Security and Communication Technology, The Norwegian University of Science and Technology, 2802 Gjøvik, Norway

2013	*Advanced Digital Handwriting Processing*, Faculty of Social Welfare and Health Sciences, University of Haifa, Mt. Carmel Haifa 31905, Israel
2012	*Neurological Disorders Analysis Using the Speech Signals*, Signal Analysis and Interpretation Laboratory (SAIL), Electrical Engineering Department, Viterbi School of Engineering, University of Southern California, California, USA
2011	*Selection of Optimal Parameters for the Parkinsonian Speech Analysis*, Department of Automation, USTB, No. 30 Xuyuan Road, Beijing 100083, P. R. China

Scientific Activity

- Web of Science: 99 indexed works, 890 citations without self-citations, h-index: 19
- Scopus: 98 indexed works, 1188 citations without self-citations, h-index: 23

Editorial Activity

2020	guest editor (special section: Multimodal Tracking of Functional Data in Parkinson's Disease and Related Disorders – Speech and Language Neuromotor and Cognitive Assessment), *Frontiers in Human Neuroscience*
2020	guest editor (special section: Behavioral Biometrics for eHealth and Well-Being), *IEEE Access*

Committee

2015, 2017–2020	steering committee, program committee, *IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*
2017–2021	program committee, *International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS)*
2021–2022	program committee, *International Conference on Telecommunications and Signal Processing (TSP)*
2015–2018	program committee, *International Congress on Neurotechnology, Electronics and Informatics (NEUROTECHNIX)*
2018–2019	program committee, *Iberoamerican Congress on Pattern Recognition (CIARP)*
2019	program committee, *International Conference on Neurotechnology and Physiological Computing Systems (NEUROPhyCS)*
2015	program committee, *International Conference on Non Linear Speech Processing (NOLISP)*
2014–2018	organizing committee, *Signal Processing Laboratory Workshop*

Reviews

- Advances in Cognitive Psychology
- Artificial Intelligence in Medicine
- Biomedical Signal Processing and Control
- Brain and Behavior
- Computer Methods and Programs in Biomedicine
- Computers in Biology and Medicine
- Current Alzheimer Research
- IEEE Access

- IEEE Journal of Selected Topics in Signal Processing
- IEEE Sensors Journal
- IEEE Transactions on Human-Machine Systems
- Information Fusion
- Journal of Electrical Engineering
- Logopaedica
- Movement Disorders
- Movement Disorders Clinical Practice
- Neurocomputing
- Neuropsychological Rehabilitation
- npj Parkinson's disease
- Parkinsonism & Related Disorders
- Patterns
- Pattern Recognition Letters
- Proceedings of The Royal Society A
- Radioengineering
- Sensors
- Science China
- Scientific Reports

Guaranteed Courses (Academic Year 2021/2022)

- BPA–ASI: Analysis of Signals (Electrical Engineering Study Programme; Brno University of Technology)

Lectured Courses (Academic Year 2021/2022)

- BPA–ASI: Analysis of Signals (Electrical Engineering Study Programme; Brno University of Technology)
- BPC–ASI: Analysis of Signals and Systems (Audio Engineering Study Programme; Microelectronics and Technology Study Programme; Telecommunication and Information Systems Study Programme; Brno University of Technology)
- MPA–CSI: Digital Signals and Systems (Communications and Networking Study Programme; Communications and Networking (Double-Degree) Study Programme; Brno University of Technology)
- MPC–ZRE: Speech Processing (Communications and Informatics Study Programme; Audio Engineering Study Programme; Brno University of Technology)

Tutorship

- Graduated PhD students: 4
- Active PhD students: 4
- Graduated MSc students: 17
- Graduated BSc students: 14

Five Selected Publications

- Brabenec, L., Klobusiakova, P., Mekyska, J., & Rektorova, I. (2022). Shannon entropy: A novel parameter for quantifying pentagon copying performance in non-demented Parkinson's disease patients. *Parkinsonism & Related Disorders*, 94, 45–48. (Q1, no. of citations without self-citations according to WoS: 0)
- Stillova, K., Kiska, T., Koritakova, E., Strycek, O., Mekyska, J., Chrastina, J., & Rektor, I. (2021). Mozart effect in epilepsy: Why is Mozart better than Haydn? Acoustic qualities-based analysis of stereoelectroencephalography. *European Journal of Neurology,* 28(5), 1463–1469. (Q1, no. of citations without self-citations according to WoS: 3)
- Drotar, P., Mekyska, J., Rektorova, I., Masarova, L., Smekal, Z., & Faundez-Zanuy, M. (2016). Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artificial Intelligence in Medicine*, 67, 39–46. (Q1, no. of citations without self-citations according to WoS: 107)
- Mekyska, J., Janousova, E., Gomez-Vilda, P., Smekal, Z., Rektorova, I., Eliasova, I., . . . & Lopez-de-Ipina, K. (2015). Robust and complex approach of pathological speech signal analysis. *Neurocomputing*, 167, 94–111. (Q1, no. of citations without self-citations according to WoS: 46)
- Drotar, P., Mekyska, J., Rektorova, I., Masarova, L., Smekal, Z., & Faundez-Zanuy, M. (2014). Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease. *Computer Methods and Programs in Biomedicine*, 117(3), 405–411. (Q1, no. of citations without self-citations according to WoS: 49)

List of appendices

A.1 Comparison of CNN-Learned vs. Handcrafted Features for Detection of Parkinson's Disease Dysgraphia in a Multilingual Dataset

frontiers | Frontiers in Neuroinformatics

ORIGINAL RESEARCH
published: 30 May 2022
doi: 10.3389/fninf.2022.877139

Comparison of CNN-Learned vs. Handcrafted Features for Detection of Parkinson's Disease Dysgraphia in a Multilingual Dataset

*Zoltan Galaz[1], Peter Drotar[2], Jiri Mekyska[1], Matej Gazda[2], Jan Mucha[1], Vojtech Zvoncak[1], Zdenek Smekal[1], Marcos Faundez-Zanuy[3], Reinel Castrillon[4,5], Juan Rafael Orozco-Arroyave[4,6], Steven Rapcsak[7], Tamas Kincses[8], Lubos Brabenec[9] and Irena Rektorova[9,10]**

[1] Department of Telecommunications, Brno University of Technology, Brno, Czechia, [2] Intelligent Information Systems Laboratory, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Košice, Slovakia, [3] Escola Superior Politecnica, Tecnocampus, Mataró, Spain, [4] Faculty of Engineering, Universidad de Antioquia—UdeA, Medellín, Colombia, [5] Faculty of Engineering, Universidad Católica de Oriente, Rionegro, Colombia, [6] Pattern Recognition Lab, Friedrich-Alexander-Universität, Erlangen, Germany, [7] Department of Neurology, College of Medicine, University of Arizona, Tucson, AZ, United States, [8] Department of Neurology, University of Szeged, Szeged, Hungary, [9] Applied Neuroscience Research Group, Central European Institute of Technology—CEITEC, Masaryk University, Brno, Czechia, [10] First Department of Neurology, Faculty of Medicine and St. Anne's University Hospital, Masaryk University, Brno, Czechia

OPEN ACCESS

Edited by:
Ludovico Minati, Tokyo Institute of Technology, Japan

Reviewed by:
Gennaro Vessio, University of Bari Aldo Moro, Italy
Rubén San-Segundo, Polytechnic University of Madrid, Spain

***Correspondence:**
Irena Rektorova
irena.rektorova@fnusa.cz

Received: 16 February 2022
Accepted: 19 April 2022
Published: 30 May 2022

Citation:
Galaz Z, Drotar P, Mekyska J, Gazda M, Mucha J, Zvoncak V, Smekal Z, Faundez-Zanuy M, Castrillon R, Orozco-Arroyave JR, Rapcsak S, Kincses T, Brabenec L and Rektorova I (2022) Comparison of CNN-Learned vs. Handcrafted Features for Detection of Parkinson's Disease Dysgraphia in a Multilingual Dataset. Front. Neuroinform. 16:877139. doi: 10.3389/fninf.2022.877139

Parkinson's disease dysgraphia (PDYS), one of the earliest signs of Parkinson's disease (PD), has been researched as a promising biomarker of PD and as the target of a noninvasive and inexpensive approach to monitoring the progress of the disease. However, although several approaches to supportive PDYS diagnosis have been proposed (mainly based on handcrafted features (HF) extracted from online handwriting or the utilization of deep neural networks), it remains unclear which approach provides the highest discrimination power and how these approaches can be transferred between different datasets and languages. This study aims to compare classification performance based on two types of features: features automatically extracted by a pretrained convolutional neural network (CNN) and HF designed by human experts. Both approaches are evaluated on a multilingual dataset collected from 143 PD patients and 151 healthy controls in the Czech Republic, United States, Colombia, and Hungary. The subjects performed the spiral drawing task (SDT; a language-independent task) and the sentence writing task (SWT; a language-dependent task). Models based on logistic regression and gradient boosting were trained in several scenarios, specifically single language (SL), leave one language out (LOLO), and all languages combined (ALC). We found that the HF slightly outperformed the CNN-extracted features in all considered evaluation scenarios for the SWT. In detail, the following balanced accuracy (BACC) scores were achieved: SL—0.65 (HF), 0.58 (CNN); LOLO—0.65 (HF), 0.57 (CNN); and ALC—0.69 (HF), 0.66 (CNN). However, in the case of the SDT, features extracted by a CNN provided competitive results: SL—0.66 (HF), 0.62 (CNN); LOLO—0.56 (HF), 0.54 (CNN); and ALC—0.60 (HF), 0.60 (CNN). In summary, regarding the SWT, the HF outperformed the CNN-extracted features over 6% (mean BACC of 0.66 for HF, and 0.60 for CNN). In the case of the SDT, both feature sets provided almost identical classification performance (mean BACC of 0.60 for HF, and 0.58 for CNN).

Keywords: machine learning, deep learning, feature extraction, Parkinson's disease dysgraphia, handwriting analysis

1. INTRODUCTION

Parkinson's disease (PD) is a chronic idiopathic disorder characterized by the progressive loss/degeneration of dopaminergic neurons in the *substancia nigra pars compacta* (Hornykiewicz, 1998; Dickson, 2012) with the development of α-synuclein-containing Lewy bodies within the dopaminergic neurons (Forno, 1996). PD is the second most frequent neurodegenerative disorder, with the prevalence rate estimated to be ∼2.0% for people aged over 65 years (Heinzel et al., 2019). To date, the gradual deficiency of dopaminergic neurons in the basal ganglia has been recognized as a major cause of parkinsonian symptoms (Brodal, 2003). In addition to a large variety of other motor symptoms, such as tremor at rest (Hughes et al., 1993), progressive bradykinesia (Berardelli et al., 2001), muscular rigidity (Hughes et al., 1993), postural instability (Horak et al., 2005), and hypokinetic dysarthria (Brabenec et al., 2017), one of the prominent motor symptoms of PD is so-called Parkinson's disease dysgraphia (PDYS) (Letanneux et al., 2014; Pinto and Velay, 2015; Thomas et al., 2017).

PDYS is a term describing a spectrum of neuromuscular difficulties, including motor-memory dysfunction (problems combining memory input with motor output), graphomotor production deficits (poor muscle coordination), motor feedback difficulties (over-activation of certain muscles and joints during handwriting as well as problems tracking the location of the pen's tip) and others. These cause a variety of handwriting difficulties (HD) manifesting as dysfluent, shaky, slow, and less readable handwriting; a progressive decrease in letter amplitude or width, namely, micrographia (McLennan et al., 1972; Rosenblum et al., 2013; Letanneux et al., 2014); etc. Hence, PDYS has serious consequences that significantly affect a person's everyday life, starting with slow and less legible handwriting and often progressing to lower self-esteem, poor emotional well-being, problematic communication, and social interaction, and many others. To introduce a timely and effective treatment to improve a patient's quality of life as much as possible, neurologists, and other experts could benefit from a remote, objective, fast, and low-cost decision support system. Such a system could employ artificial intelligence and provide information that might lie beyond human perception. It could enable specialists to combine their expertise with a large volume of data that are not available when utilizing a conventional in-clinic examination to identify and assess parkinsonian symptoms. Finally, such an approach could be implemented in decentralized clinical trials and could significantly suppress the Hawthorne effect (Morberg et al., 2018).

In general, the handwriting tasks that are traditionally employed in PDYS analysis can be classified into drawing, writing, and more complex tasks (Vessio, 2019). Usually, simple drawing or writing elements are performed repetitively and continuously as a single exercise. In the drawing task category, spirals, circles, meanders, and simple figures are frequently used for motor performance evaluation. These types of drawing tasks are effortless and well-tolerated and hence are suitable for studying motor control deficits in PD patients, especially for assessing tremor (San Luciano et al., 2016; Vessio, 2019). As PD patients commonly exhibit constructional apraxia (Garre-Olmo et al., 2017), their drawings may contain simplifications, lack of perspective, fewer angles, or spatial alterations. Letters, words, and sentences are commonly acquired during the examination process in the writing task category. As PD patients may produce slower and more irregular movements, mainly due to rigidity and bradykinesia, the results of repetitive writing tasks usually emerge in a more segmented fashion (Pullman, 1998; Drotar et al., 2016). Sentence writing requires a high degree of simultaneous processing, including motor planning; therefore, it is suitable for detecting micrographia (Bidet-Ildei et al., 2011), which is the most commonly observed handwriting abnormality in PD patients. Finally, more complicated handwriting tasks, such as the Clock Drawing Test (Agrell and Dehlin, 1998), may be used as well as part of a more complex examination involving cognitive and functional issues.

Currently, the most promising approach for the robust, objective, and computerized assessment of PDYS utilizes various signals describing the process/product of handwriting acquired by a digitizing tablet (Drotar et al., 2014, 2015). Such signals represent the movement of a digitizing stylus (pen) along both the horizontal and vertical axes, the pressure exerted on the surface of a digitizer, and the tilt and azimuth angles, acquired with respect to a specific series of timestamps to form a collection of time series describing the process of handwriting from beginning to end (referred to as online handwriting). In addition, modern digitizers have the ability to record not only the movement of a pen on the surface of the digitizer but also the movement above the surface (in-air movement; Alonso-Martinez et al., 2017). As shown in a variety of research studies focusing on the identification and assessment of HD in patients suffering from PD, Alzheimer's disease (AD), essential tremor (Drotar et al., 2014, 2016; Alonso-Martinez et al., 2017; Impedovo et al., 2018), etc., online handwriting capture provides the ability to characterize the process of handwriting in terms of its kinematic, dynamic, and temporal features, which are not accessible from the final handwritten product when using the conventional pen and paper methodology (referred to as offline handwriting).

At present, the following handcrafted features are conventionally used to describe the product/process of handwriting/drawing (Rosenblum et al., 2013; Thomas et al., 2017; De Stefano et al., 2019): (a) spatial features—width, height, and length; (b) temporal features—duration; (c) kinematic features—velocity, acceleration, and jerk; (d) dynamic features—pressure, tilt, and azimuth; and (e) other features—number of interruptions (pen elevations), etc. These features are computed either for an entire product or on a per-stroke basis utilizing on-surface and in-air movements. In the case of per-stroke computation, the investigated signals are broken down into the separate strokes forming the final handwritten product. A crucial characteristic of these conventional features is their clinical interpretability, allowing them to be linked with the real physiological phenomena behind the studied pathologies, which is extremely important for the mass adoption of this methodology in real clinical use cases.

Despite the broad use and indisputable success of these conventional handcrafted features, our recent studies (Mucha et al., 2018a,b; Mucha et al., 2019) concerning the computerized identification and assessment of PD and developmental dysgraphia (DD) have illustrated the necessity of additional research into novel and more advanced parametrization techniques for handwriting that could enable more robust and complex characterization of HD. For this reason, various nonlinear handwriting features based on modulation spectra, fractional-order derivatives (FD) and the tunable-Q wavelet transform have been developed and evaluated (Galaz et al., 2020; Mucha et al., 2020).

Conventional and nonlinear handcrafted features have shown promising potential for the quantification of hidden patterns in deficient handwriting. However, the necessity of manual design and development is still a severe limitation. Recent advancements in artificial neural networks offer new possibilities for automated feature extraction. By utilizing transfer learning, pre-trained convolutional neural networks (CNNs) can be advantageously used to extract features and, as such, provide an alternative solution in place of tedious and time-consuming manual feature design. This approach has already been used not only for handwriting processing (Gil-Martin et al., 2019; Moetesum et al., 2019; Gazda et al., 2021) but also in several other domains (Hagerty et al., 2019; Minaee et al., 2020). Nevertheless, in the area of handwriting processing, one apparent limitation of CNN feature extraction is that it utilizes only image data, and as such, it is limited only to offline handwriting processing. However, there have recently been some promising attempts to employ recurrent neural networks for the classification of handwriting signals (Diaz et al., 2021).

As seen from the above discussion, various parametrization techniques for offline and online handwriting have been developed. However, a major limitation of the current state of affairs is that these techniques are treated separately most of the time. Studies comparing the robustness of conventional handcrafted features with that of features extracted automatically using a pre-trained CNN for the identification and assessment of PDYS are lacking. Moreover, multilingual studies analyzing datasets acquired from subjects of different nationalities are very rare.

The primary goal of this work is to compare two different approaches for the identification of PDYS from drawing and handwriting. The first approach is based on online handwriting utilizing a set of conventional handcrafted features (baseline), whereas the second approach relies on automated feature extraction from offline handwriting utilizing a pre-trained CNN. The primary aim of this comparison is to reveal whether a set of features that are automatically extracted with no prior domain knowledge could compete with a set of handcrafted features designed by domain experts. The secondary goal of this work is to explore the power of both feature sets for the identification of PDYS in a multilingual dataset. In this study, we consider two different handwriting tasks, namely, the Archimedean spiral drawing task and the sentence writing task. The reason behind this selection is to examine a drawing task, which is independent of language, and a writing task, which is dependent on language. We note that except for our own previous work (Mucha et al., 2019), in which the Spanish and Czech sentence tasks were investigated together, this is the only study to date to consider a large multilingual cohort of PD patients, who were enrolled in the Czech Republic, the United States, Colombia, and Hungary. Such cross-language and cross-cultural clinical studies are essential to generalize the methodology used for PDYS diagnosis and assessment; therefore, the findings of this study could lay a foundation for future research in this area.

2. RELATED WORKS ON PD CLASSIFICATION FROM HANDWRITING

2.1. Online Handwriting

The most frequently used handcrafted features extracted from online handwriting can be divided into (a) conventional features (temporal, spatial, kinematic, and dynamic) and (b) advanced features (Vessio, 2019). Among conventional features, the following features have been utilized the most: (a) temporal—duration of writing, duration of strokes; (b) spatial—width, height, and length of a written product or of individual strokes; (c) kinematic—velocity, acceleration, jerk; and (d) dynamic—pressure, tilt, azimuth, etc. With respect to advanced features, various studies have explored designs based on entropy, the signal-to-noise ratio (SNR), empirical mode decomposition (EMD), cepstrum (Nolazco-Flores et al., 2021), sigma–lognormal models (O'Reilly and Plamondon, 2009), FD (Mucha et al., 2018b), etc.

To obtain a complete picture of the utilization of handcrafted features in PDYS diagnosis and assessment, we refer to comprehensive reviews published up through 2019 (Letanneux et al., 2014; Impedovo and Pirlo, 2018; De Stefano et al., 2019; Vessio, 2019). In the following discussion, we review a number of recent articles. Although the present work investigates conventional features only, the review below includes studies that have employed conventional features, advanced features, or both; the primary focus is the summarization of the latest works addressing the computerized assessment of HD in patients suffering from PD.

Impedovo et al. (2018) investigated whether a diagnosis of PD based on the quantitative analysis of online handwriting could be successful in early to mid stages of the disease. For this purpose, the PaHaW database was reduced to a subset of 65 subjects [36 healthy controls (HCs) and 29 PD patients] who fit the Hoehn and Yahr scale at scores from 1 to 2.5 (Goetz et al., 2004, 2008). Almost all of the extracted features were kinematic, whereas some of them utilized entropy. Significant discriminative power was achieved in the sentence task [accuracy (ACC) of 71.95% with a Gaussian naïve Bayes classifier], thus confirming the previously reported findings of Drotar et al. (2016) that the writing of a long sentence presents a higher cognitive demand such that the effect of PD can manifest itself in the aggravation of HD.

Intending to improve the computerized assessment of PD severity, Mucha et al. (2018b) deeply analyzed various advanced kinematic features based on FD. The newly designed features were compared to conventional ones for only those PaHaW subjects who completed all of the 9 tasks (Drotar et al., 2016) (69 subjects in total). The authors reported that the conventional in-air features outperformed the advanced ones in the differential analysis (ACC of 97.1% with an XGBoost classifier) as well as in the estimation of PD duration [estimation error rate (EER) of 23.6%], but in this specific case, the in-air parameters were combined with features extracted from the on-surface movement. On the other hand, the severity of PD in terms of the score on the Unified Parkinson's Disease Rating Scale, part V: Hoehn and Yahr scale (UPDRS V) was better estimated by the new FD-based metrics (EER of 12.5%), suggesting that fractional calculus can play a significant role in the assessment of PD.

In 2019, Rios-Urrego et al. (2019) analyzed the ability to use kinematic, geometric, spectral and nonlinear dynamic features to model HD and to discriminate between HCs and patients with PD. In that study, they enrolled 130 subjects from Colombia, who were asked to draw an Archimedean spiral and to write a short sentence. The results indicated an ACC of 83.3% [K-nearest neighbors (KNN) classifier] for the Archimedean spiral and ACC of 75% [support vector machine (SVM) classifier] in the case of the sentence writing task. The absence of nonlinear features in the trained models indicated that such features did not contribute to the classification accuracy as much as kinematic or geometric features.

Jerkovic et al. (2019) experimented with in-air handwriting features and multiclass linear discriminant analysis (cLDA) to differentiate between HCs, patients with PD and patients with atypical parkinsonism. Altogether, 43 subjects from Serbia were enrolled in the study. The task was to write a sentence in various scenarios, such as with or without looking at the monitor of the laptop during writing. Various kinematic features related to the in-air and on-surface trajectories were extracted. The combination of the on-surface and in-air features led to ACC of 86%, whereas a model trained only with in-air features had a slightly lower ACC of ∼79%. The results led to the conclusion that kinematic features based on both the in-air and on-surface trajectories are equally important in the quantitative analysis of the handwriting of PD patients with various types of motor impairments.

Impedovo (2019) investigated the use of new velocity-based signal processing techniques for the advance diagnosis of PD based on the discrete Fourier transform (DFT; for assessing rapidity and fluency), sigma–lognormal modeling (SLM; for quantifying the constant tremor pattern of PD utilizing cepstrum properties) and the Maxwell–Boltzmann distribution (MBD; for modeling handwriting velocity profiles). In his work, he utilized online handwriting records from the PaHaW database. The newly proposed features were extracted together with conventional features (baseline; Impedovo et al., 2018) for all tasks in the database. When classification was performed using all features and all tasks, the newly proposed features were selected among the 10 best-performing features (ACC of 94%, SVM classifier) and outperformed the baseline features (ACC of 88% SVM classifier). The author was able to increase the HC/PD classification accuracy to 98% when using only the most suitable tasks (the Archimedean spiral, "lll" and the word "lektorka").

In 2020, a study published by Aouraghe et al. (2020) introduced new kinematic features utilizing the discrete time wavelet transform (DTWT), the fast Fourier transform (FFT) and a Butter/adaptive filter in the diagnosis of PD. Altogether, 80 native Arabic speakers were enrolled. All of them wrote a particular segment of text on several lines. Additionally, to better predict the continuous degradation of PD handwriting, the output of the text task was segmented line by line using unsupervised K-means clustering (observing the variation in the x and y trajectories). All of the extracted features (new and conventional) were calculated for the whole text and for each segmented line separately (at least 4 lines). The best performance on the entire task corresponded to ACC of 85.7% (KNN classifier). The first line showed a slightly lower ACC of 78.6% when a decision tree (DT) classifier was used. The last line proved to be the most effective and discriminative segment in the study when utilizing the DWT (ACC of 92.9%). Segmentation proved to be a valid method, as the results confirmed the hypothesis that PD handwriting degradation, deterioration, and fatigue increase over time.

While the previous approaches relied on carefully designed handcrafted features, Vásquez-Correa et al. (2019) proposed directly feeding the raw captured signals and their derivatives into a 1D CNN. These authors utilized a rather small CNN with two convolutional and pooling layers. This procedure allowed ACC of 67% to be achieved in the classification of PD patients and HC subjects. The authors performed several experiments using only onset or offset data, constituting the 200 ms after the transition from on-surface to in-air movement or the transition from in-air to on-surface movement. However, this approach did not seem to improve the prediction accuracy.

There are also some other studies that confirm feasibility of the digitized spiral drawing for PD detection (Kamble et al., 2021) and PD stage classification (Zham et al., 2017).

2.2. Offline Handwriting

In contrast to approaches based on online handwriting, in which multiple modalities are available, offline handwriting approaches must rely on visual data only. This significantly limits the information that is available for classification. Moetesum et al. (2019) utilized a pretrained AlexNet CNN to extract features from images capturing handwriting samples. To further enhance the extraction of features and boost the performance, the authors combined three different types of image preprocessing techniques. With this approach, they obtained ACC of 76% on a single task from the PaHaW dataset and ACC of 83% when merging all tasks used for prediction.

Recently, Gazda et al. (2021) proposed the idea of multiple-fine-tuned CNNs for the classification of PD handwriting. Similar to the work of Moetesum et al. (2019), this approach relies on a pretrained CNN. However, Gazda et al. utilized datasets of handwriting samples to bridge the gap between the semantically different ImageNet dataset, which was used for network pretraining, and parkinsonian handwriting datasets. This approach enabled more efficient transfer learning, leading to ACC of 92.7% on the spiral drawing task from the NewHandPD dataset and ACC of 85.8% on the spiral drawing task from the PaHaW dataset.

Similarly, six pretrained CNNs (AlexNet, GoogLeNet, VGG16, VGG19, ResNet50, and ResNet101) were evaluated in Kamran et al. (2021) in terms of their performance on four different handwriting datasets. The obtained results strongly depended on the dataset, with the most challenging dataset being PaHaW. In this case, the classification accuracy was only 62.5%, compared to accuracies of over 90% for the HandPD, NewHandPD (Pereira et al., 2016) and Parkinson's Drawing (Zham et al., 2017) datasets.

Finally, the authors of Diaz et al. (2019) were able to merge the online and offline handwriting approaches by incorporating dynamic information into static images. This approach seemed to improve classification in cases where the task can be performed continuously without lifting the pen. The highest ACC of 75% was achieved using VGG as the feature extractor and a linear SVM as the classifier for a single drawing task (spiral). Further improvements were obtained by building an ensemble classifier based on the results from different tasks, yielding ACC of 86%.

For a better illustration, a summary of the related works is provided in **Table 1**. The overview of the related works based on online handwriting is in the upper part, and studies based on offline handwriting are in the bottom part of the table.

3. MATERIALS AND METHODS

3.1. Dataset

In total, 143 patients with PD (71 female and 72 male; mean age 66.32 ± 10.79 years) and 151 HCs (86 female and 65 male; mean age 64.79 ± 9.90 years) were enrolled in several geographical locations: the Czech Republic (CZ), Hungary (HU), the United States of America (US), and Colombia (CO). A corresponding multilingual dataset was created by fusing the following databases: PaHaW (Drotar et al., 2016), CoBeN (acquired under the Marie Skłodowska-Curie grant agreement no. 734718), and HWUDEA (Castrillon et al., 2019; Rios-Urrego et al., 2019). In the case of the PaHaW database, the participants performed 9 tasks (e.g., Archimedean spiral, letters, syllables, words, sentence) on A4 paper that was laid down and fixed to a digitizing tablet (Wacom Intuos 4M, with a sampling frequency of $f_s = 133$ Hz). A special Wacom inking pen was used to provide immediate visual feedback, i.e., simulating classical pen-and-paper writing/drawing. The participants enrolled for the acquisition of CoBeN underwent a protocol consisting of 8 tasks (e.g., connecting dots, overlapping pentagons, Archimedean spiral, sentences) using a similar paper–tablet setup; however, in this case, the data were recorded by a Wacom Intuos Pro L ($f_s = 133$ Hz). Finally, the HWUDEA database was acquired by employing a Wacom Cintiq 13HD Touch display tablet ($f_s = 180$ Hz). In total, 17 tasks were recorded for each participant (e.g., spring, alphabet, sentence, Archimedean spiral, house drawing). Although the databases were collected following different protocols, all of them share two tasks: the Archimedean spiral drawing task and a sentence writing task. Selected samples can be seen in **Figure 1**.

Demographic data with respect to each of the two tasks shared among all databases are reported in **Table 2**. Unfortunately, the databases are not annotated with the same clinical information (e.g., the CoBeN–HU dataset contains only information about sex and age); nevertheless, to provide at least limited insight into the characteristics of the PD patients, we summarize the available metadata in **Table 3**. None of the participants had a history or the presence of any psychiatric symptoms, cognitive impairment, or any disease affecting the central nervous system (other than PD in the PD cohort). All PD patients were diagnosed based on the diagnostic criteria for PD (Postuma et al., 2016). They were well-compensated on their stable dopaminergic medication and without major motor fluctuations or dyskinesias [they were examined while on their regular dopaminergic medication (ON state) ∼1 h after the L-dopa dose]. All subjects signed informed consent forms. The study was approved by the relevant local ethics committees.

3.2. Scenarios

We define three main scenarios to analyze the effect of linguality on the classification of PDYS:

1. Single language—In this scenario, we consider datasets for every language separately. As such, there are four different models: HU, US, CO, and CZ (the Czech dataset is created by merging the PaHaW and CoBeN datasets). In this scenario, each classification model is trained and tested on a dataset consisting of data samples that all come from the same language. This scenario is considered to correspond to internal model validation because the linguality of the datasets is not considered; rather, the robustness of the features is evaluated at the per-dataset level.
2. Leave one language out—In this scenario, the influence of different languages on the classification performance is evaluated by training each model on three out of four datasets and testing it on the remaining dataset. With this approach, we aim to investigate the effect of transferring knowledge between datasets coming from different language sources. We refer to this scenario as the leave-one-language-out scenario. This scenario is considered to correspond to

TABLE 1 | Overview of the related works.

References	Participants	Task	Features	Analysis	Results
Online handwriting					
Impedovo et al. (2018)	29 PD, 36 HC	PaHaW–all	Kinematic, enthropy	GNB	ACC = 72.0%
Mucha et al. (2018b)	33 PD, 36 HC		FD-based kinematic	XGBoost	ACC = 97.1%
		PaHaW–all			EER = 23.6% (PD dur)
					EER = 12.5% (UPDRS V)
Rios-Urrego et al. (2019)	39 PD, 70 HC	Archimedean spiral	Kinematic, geometric	KNN	ACC = 83.3% (spiral)
		Short sentence	Spectral, non-linear	SVM	ACC = 75.0% (sentence)
Jerkovic et al. (2019)	33 PD, 10 HC	Various sentences	Kinematic	cLDA	ACC = 86.0%
Impedovo (2019)	37 PD, 38 HC	PaHaW–all	DFT, SLM, MBD	SVM	ACC = 94.0%
Aouraghe et al. (2020)	40 PD, 40 HC	Segment of text	DTWT, FFT	KNN	ACC = 85.7% (full text)
			Butter/adaptive filter	decision tree	ACC = 78.6% (first line)
Vásquez-Correa et al. (2019)	44 PD, 40 HC	14 drawings/writings	Original signal	1D CNN	ACC = 67.0%
Offline handwriting					
Moetesum et al. (2019)	37 PD, 38 HC	PaHaW–all	AlexNet CNN	SVM	ACC = 83.0%
Gazda et al. (2021)	64 PD, 71 HC	Archimedean spiral	Pre-trained CNN and transfer learning (ImageNet→PD dataset)		ACC = 92.7% (NewHandPD)
	2 dataset				ACC = 85.8% (PaHaW)
Kamran et al. (2021)	PaHaW	Several drawings	AlexNet, GoogLeNet, VGG16 VGG16, ResNet50, ResNet101		ACC = 62.5% (PaHaW)
	HandPD				ACC = 91.4% (HandPD)
	NewHandPD				ACC = 98.4% (NewHandPD)
	PD Drawings				ACC = 90.0% (PD Drawings)
Diaz et al. (2019)	37 PD, 38 HC	PaHaW–all	VGG	SVM	ACC = 86.0%

PD, Parkinson's disease; HC, healthy control; PaHaW, Parkinson's disease handwriting database (Drotar et al., 2016); FD, fractional order derivative; ACC, accuracy; EER, estimation error rate; PD dur, PD duration; GNB, gaussian naïve bayes classifier; xGBoost, extreme gradient boosting tree; KNN, K-nearest neighbors; SVM, support vector machine; cLDA, multi-class linear discriminant analysis; CNN, convolution neural network; ResNet, residual neural network; VGG, very deep CNN; DFT, discrete fourier transformation; SLM, sigma-lognormal model; MBD, maxwell-boltzmann distribution; DTWT, discrete time wavelet transform; FFT, fast Fourier transform; UPDRS V, UPDRS, part V: Hoehn and Yahr scale (Fahn and Elton, 1987).

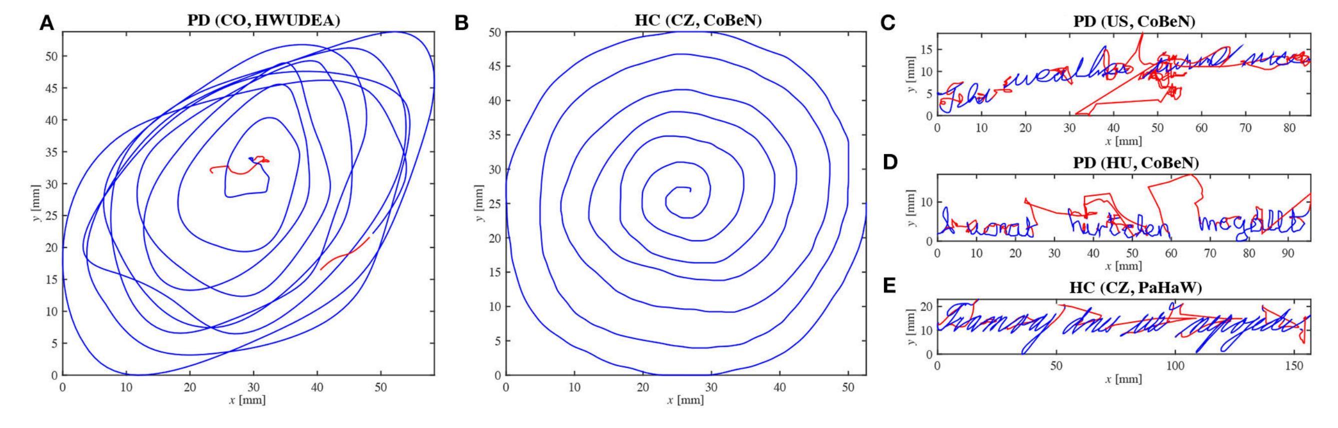


FIGURE 1 | Selected samples from the multilingual dataset (blue line – on-surface movement; red line – in-air movement). **(A)** Spiral drawing (PD patient); **(B)** Spiral drawing (HC); **(C)** English sentence (PD patient) "The weather turned nice"; **(D)** Hungarian sentence (PD patient) "A vonat hirtelen megállt"; **(E)** Czech sentence (HC) "Tramvaj dnes už nepojede".

external model validation because the multilinguality of the data is taken into account, i.e., the validation samples come from a different geographical location, as recommended in the TRIPOD guidelines (Collins and Moons, 2019).

3. All languages combined—In the last scenario, we combine all datasets of different languages into one complete dataset to evaluate the performance of the features on the mixed dataset.

3.3. Feature Extraction

Although the individual databases were acquired using different devices, all of them recorded the following information (time series): the x and y positions ($x[n]$ and $y[n]$), the timestamp ($t[n]$), a binary variable ($b[n]$) taking values of 0 for in-air movement (i.e., movement of the pen tip up to 1.5 cm above the tablet's surface) and 1 for on-surface movement (i.e., movement of the

TABLE 2 | Demographic characteristics.

Dataset	Language	PD (*N*; female)	PD (*N*; male)	PD (age)	HC (*N*; female)	HC (*N*; male)	HC (age)
Archimedean spiral							
PaHaW	CZ	18	15	69.21 ± 11.10	17	19	62.50 ± 11.70
CoBeN	CZ	6	13	66.48 ± 7.77	30	10	67.04 ± 6.07
CoBeN	US	3	6	68.56 ± 4.07	9	3	72.50 ± 8.37
CoBeN	HU	2	7	66.00 ± 9.96	7	5	64.92 ± 5.30
HWUDEA	CO	41	28	64.42 ± 11.85	22	27	62.69 ± 11.34
Sentence							
PaHaW	CZ	19	18	69.32 ± 10.97	18	20	62.42 ± 11.39
CoBeN	CZ	6	13	66.48 ± 7.77	30	9	67.21 ± 6.05
CoBeN	US	3	6	68.56 ± 4.07	9	3	72.50 ± 8.37
CoBeN	HU	2	6	65.88 ± 10.64	7	5	64.92 ± 5.30
HWUDEA	CO	13	4	63.88 ± 7.61	5	5	70.20 ± 10.67

TABLE 3 | Clinical characteristics of the PD patients.

Dataset	Language	Duration of PD [years]	LED [mg/day]	UPDRS III	UPDRS V
PaHaW	CZ	8.38 ± 4.80	1,432.19 ± 704.78	–	2.27 ± 0.85
CoBeN	CZ	4.00 ± 4.15	568.33 ± 508.03	7.00 ± 1.41	–
CoBeN	US	–	333.12 ± 240.40	–	–
CoBeN	HU	–	–	–	–
HWUDEA	CO	10.56 ± 11.16	–	36.78 ± 19.63	2.38 ± 0.61

LED, L-dopa equivalent daily dose (Lee et al., 2010); UPDRS III, Unified Parkinson's Disease Rating Scale, part III: motor examination (Fahn and Elton, 1987); UPDRS V, UPDRS, part V: Hoehn and Yahr scale (Fahn and Elton, 1987).

pen tip on the paper), the pressure exerted on the tablet's surface during writing ($p[n]$), the pen tilt ($a[n]$), and the pen azimuth ($az[n]$). First, we preprocessed the recordings for unit unification (e.g., we expressed the x and y positions in millimeters, time in seconds, etc.) and resampling [we resampled all signals to $f_s = 133$ Hz employing a finite impulse response (FIR) antialiasing low-pass filter]. Subsequently, we parameterized the signals employing the previously mentioned baseline and CNN-based features.

3.3.1. Baseline Features

To establish a good baseline for the evaluation of the CNN-based features, we consulted several recent articles and reviews (Impedovo and Pirlo, 2018; De Stefano et al., 2019; Vessio, 2019) and extracted the handcrafted features that are most commonly used for the quantitative assessment of PD dysgraphia. These features can be divided into six groups:

1. Temporal—duration of writing (DUR), ratio of the on-surface/in-air durations (DURR), duration of strokes (SDUR), and ratio of the on-surface/in-air stroke durations (SDURR)
2. Spatial—width (WIDTH), height (HEIGHT), and length (LEN) of the whole product as well as those of its individual strokes, i.e., stroke width (SWIDTH), height (SHEIGHT), and length (SLEN)
3. Kinematic—velocity (VEL), angular velocity (AVEL), and acceleration (ACC)
4. Dynamic—pressure (PRESS), tilt (TILT), and azimuth (AZIM)
5. Spiral-specific (San Luciano et al., 2016; Cascarano et al., 2019)—first-order smoothness of spiral (1stSm), second-order smoothness of spiral (2ndSm), spiral tightness (TGHTNS), first-order zero-crossing rate of spiral (1stZC), second-order zero-crossing rate of spiral (2ndZC), degree of spiral drawing severity (DoS), mean drawing speed of spiral (MDS), variability of spiral width (SWVI), and spiral precision index (SPI)
6. Other—number of interruptions or pen elevations (NINT), relative number of interruptions (RNINT), number of on-surface interstroke intersections (NIEI), relative number of on-surface interstroke intersections (RNIEI), number of on-surface intrastroke intersections (NIAI), relative number of on-surface intrastroke intersections (RNIAI), total number of on-surface intrastroke intersections (TNIAI), relative total number of on-surface intrastroke intersections (RTNIAI), relative number of changes in velocity profile (RNCV), relative number of changes in pressure profile (RNCP), relative number of changes in tilt profile (RNCT), and relative number of changes in azimuth profile (RNCA)

The spatial, temporal, and kinematic features were extracted from both the on-surface and in-air movements. In addition, the kinematic features were also analyzed for the horizontal and vertical projections of the movements. Features that are represented by time series were transformed into scalar values using the median, interquartile range (iqr), nonparametric coefficient of variation (ncv; defined as iqr/median), and slope by applying the Theil–Sen estimator (slope). In the case of the kinematic time series, we also calculated the 95th percentile (95p).

For each feature, we use the following notation: *INF: DIR-FN (HL)*, where *INF* denotes the processed information (ON for on-surface, AIR for in-air, PRESS for pressure, TILT for tilt, and AZIM for azimuth), *DIR* denotes the direction (H for horizontal and V for vertical), *FN* is the feature name, and *HL* is the statistic used for the transformation.

3.3.2. CNN-Based Features

Over the past decade, CNNs have demonstrated outstanding capabilities on various tasks, such as image recognition, medical image analysis, and handwriting recognition. Multiple state-of-the-art models exist, with a typical structure consisting of an input layer, a mix of convolutional and pooling layers, and one output layer. Deeper networks often produce better results than shallower ones; on the other hand, they have multiple times more parameters and require more data for training, especially when compared to traditional machine learning models. To overcome this problem, transfer learning techniques have been proposed.

The idea behind transfer learning is to take advantage of the features of a CNN trained on one task and use them for another task. Given a source domain D_s, a corresponding task T_s, a target domain D_t, and the corresponding task T_t, where $D_s \neq D_t$ and $T_s \neq T_t$, the goal of transfer learning is to reduce the error of the target predictive function $f_t(.)$ in D_t. For transfer learning, two main paradigms exist. The first is called fine tuning, in which a neural network or at least part of the neural network is retrained, thus changing the weights of the layers. In the second approach, a CNN is used to extract features. In the feature extraction model, the weights trained on the source task are frozen, and the corresponding representations are applied in the target task.

In case of CNN-based features we render images from data captured by the digitizing tablet. Specifically, we use only the x and y positions ($x[n]$ and $y[n]$). To extract CNN-based features, we employed the state-of-the-art CNN known as VGG16 (Simonyan and Zisserman, 2014), pretrained on the ImageNet dataset (Russakovsky et al., 2015). The VGG16 is well-known architecture that is still being frequently used thanks to its relative simplicity. The input images were resized to 224×224 by nearest-neighbor interpolation. We extracted features from the last convolutional layer in the VGG16 network. The extracted features capture abstract representations of the processed input image. Features were classified by CNN head consisting of fully connected layer and output layer.

3.4. Machine Learning

For the handcrafted features, we built binary classification models using an ensemble extreme gradient boosting algorithm known as XGBoost (Chen and Guestrin, 2016). The reason behind using such an advanced nonlinear classifier is to search for complex nonlinear patterns in a feature set composed of rather simple feature representations. To build models with the optimal hyperparameters, we applied a randomized search strategy to optimize the following set of hyperparameters: the learning rate [0.001, 0.01, 0.1, 0.2, 0.3], γ [0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.5], the maximum tree depth [6, 8, 10, 12, 15], the fraction of observations to be randomly sampled for each tree (subsample ratio) [0.5, 0.6, 0.7, 0.8, 0.9, 1.0], the subsample ratio for the columns at each level [0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], the subsample ratio for the columns when constructing each tree [0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], the minimum sum of the weights of all observations required in a child node [0.5, 1.0, 3.0, 5.0, 7.0, 10.0], and the balance between positive and negative weights [1, 2, 3, 4].

In contrast, the binary classification models for the CNN-based features were built using L2-regularized logistic regression (LR), also known as ridge regression. The reason behind using this much simpler linear classifier is the assumption that the underlying nonlinear representations are already captured by the CNN-extracted features. In addition, features extracted from convolutional layers tend to have very high dimensionality, and thus, using a simpler classifier minimizes the chance of overfitting and maximizes the computational efficiency. To find the optimal parameters of the LR classifier, we searched through the various settings for the regularization parameter C given by the following set: [0.001, 0.01, 0.1, 1, 10, 100, 1000].

The randomized search was conducted 500 times. In both cases, the objective of the hyperparameter search was to optimize the balanced accuracy score (BACC; described in more detail along with other evaluation scores below) *via* stratified five-fold cross-validation with five repetitions (the five-fold cross-validation scheme was chosen as a reasonable compromise between the numbers of samples in the training and validation folds, i.e., to provide the classifier with sufficient training samples while also testing its performance on a representative subset of the overall sample size).

Finally, the trained classification models were evaluated on a per-scenario basis: (a) single language—in this scenario, we conducted stratified five-fold cross-validation with five repetitions; (b) leave one language out—in this scenario, we tested the performance of each trained classifier on the remaining dataset that was not present in the training data; and (c) all languages combined—in this scenario, we again employed stratified five-fold cross-validation with five repetitions. Only one sample of Archimedean spiral or sentence was available from each subject. Therefore, all decisions are based on a per subject basis. The classification test performance was established using the following well-known and widely used classification metrics: BACC, sensitivity (SEN), specificity (SPE), and F1 score.

4. RESULTS

4.1. Single-Language Scenario

The classification performance of the models trained in this scenario is summarized in **Table 4**. First, we trained and tested the classification models using the spiral drawing task. The highest BACC values of 82% (handcrafted features) and 77% (CNN-based features) were achieved for the US dataset. These accuracies are notably higher than those achieved for the other

TABLE 4 | Classification performance in the single-language scenario.

Language	Features	BACC	F1	SEN	SPE
Spiral drawing					
CZ	Handcrafted	0.59 ± 0.08	0.590.07	0.82 ± 0.12	0.36 ± 0.14
	CNN	0.64 ± 0.03	0.65 ± 0.05	0.65 ± 0.09	0.65 ± 0.06
CO	Handcrafted	0.59 ± 0.12	0.72 ± 0.07	0.81 ± 0.09	0.37 ± 0.23
	CNN	0.61 ± 0.02	0.62 ± 0.02	0.62 ± 0.03	0.62 ± 0.02
HU	Handcrafted	0.64 ± 0.17	0.61 ± 0.20	0.72 ± 0.29	0.57 ± 0.34
	CNN	0.48 ± 0.03	0.52 ± 0.13	0.52 ± 0.16	0.52 ± 0.12
US	Handcrafted	0.82 ± 0.18	0.77 ± 0.28	0.84 ± 0.31	0.81 ± 0.23
	CNN	0.77 ± 0.02	0.77 ± 0.07	0.77 ± 0.11	0.77 ± 0.08
Sentence writing					
CZ	Handcrafted	0.66 ± 0.08	0.62 ± 0.08	0.64 ± 0.10	0.69 ± 0.12
	CNN	0.65 ± 0.04	0.66 ± 0.04	0.66 ± 0.04	0.66 ± 0.05
CO	Handcrafted	0.56 ± 0.18	0.72 ± 0.19	0.83 ± 0.22	0.28 ± 0.29
	CNN	0.50 ± 0.08	0.54 ± 0.07	0.54 ± 0.08	0.54 ± 0.09
HU	Handcrafted	0.75 ± 0.18	0.65 ± 0.30	0.82 ± 0.34	0.59 ± 0.34
	CNN	0.50 ± 0.06	0.48 ± 0.08	0.48 ± 0.10	0.48 ± 0.08
US	Handcrafted	0.65 ± 0.20	0.54 ± 0.28	0.58 ± 0.34	0.73 ± 0.32
	CNN	0.70 ± 0.04	0.70 ± 0.06	0.70 ± 0.08	0.70 ± 0.05

BACC, balanced accuracy; F1, F1 score; SEN, sensitivity; SPE, specificity.

datasets, which indicates that the US samples most likely carry certain recognizable patterns of PD related to the graphomotor difficulties manifested during spiral drawing. With respect to the comparison between the handcrafted and CNN-based features, the results show similar trends, with both types of features yielding the highest accuracy on the US dataset and quite similar results on the other datasets. More specifically, the CNN-based features outperformed the handcrafted features on the CZ dataset (BACCs of 64 vs. 59%) as well as on the CO dataset (BACCs of 61 vs. 59%) but yielded less accurate predictions on the US dataset. This shows that CNNs, even when provided with visual information only, can be competitive with handcrafted features on the spiral drawing task. However, there is one exception. From the performance of the CNN-based features on the HU dataset, it is evident that this model failed to provide reasonable predictions (BACC of 48% with the CNN-extracted features as opposed to BACC of 64% with the handcrafted features).

To interpret the machine learning models, we investigated the top ten most important features (see **Figure 2**). In the CZ dataset, most of these features are derived from the on-surface angular velocity. Other kinematic features are based on the on-surface velocity and the mean drawing speed of the spiral. Finally, the zero-crossing rate of the spiral, the pressure and the spiral smoothness all show some importance. The most important feature in the CO dataset is the ratio between the on-surface and in-air durations. It is followed by the relative number of interruptions and by the tilt-based and azimuth-based parameters. The important feature set also contains the in-air duration and spiral tightness. The rest of the features are based on the angular velocity and horizontal/vertical velocity. The most important set of features for the HU model contains two spatial parameters, width and height. The variation in azimuth plays an important role as well. Finally, the majority of the important features are kinematic (angular velocity, velocity, and acceleration). These features are also important in the US database. In addition, some spatial parameters (length and height), the pressure and the intraspiral intersections are identified as important.

Second, we evaluated the models for the sentence writing task in the same scenario. There are a few interesting points to note. First, prediction fails on the CO dataset for both types of features (BACC of 56% with the handcrafted features and BACC of 50% with the CNN-based features). The reason is most likely the small sample size; in the CO data, there are only 27 sentences, compared to the 118 spirals used in the previous experiment. Next, the model utilizing the handcrafted features clearly outperformed the model based on the CNN features on the HU dataset (BACC of 75% with the handcrafted features and BACC of 50% with the CNN-based features) and yielded slightly more accurate predictions on the CZ dataset (BACC of 66% vs. BACC of 65%). This is to be expected since for CNN-based features, a larger sample size is probably needed to learn the underlying patterns from a given sentence; compared with spiral drawing, sentence writing is much less restricted in terms of what the final handwritten product should look like. Finally, even though the US dataset contains spirals and sentences from the same patient group, the classification accuracy is significantly lower for the sentence writing task than for the spiral drawing task. Quite surprisingly, the CNN-extracted features outperformed the handcrafted features for the US cohort (BACCs of 70 vs. 65%).

Regarding the interpretation of the models shown in **Figure 3**, the most important features in the CZ dataset are based on the on-surface velocity, more specifically on its median and

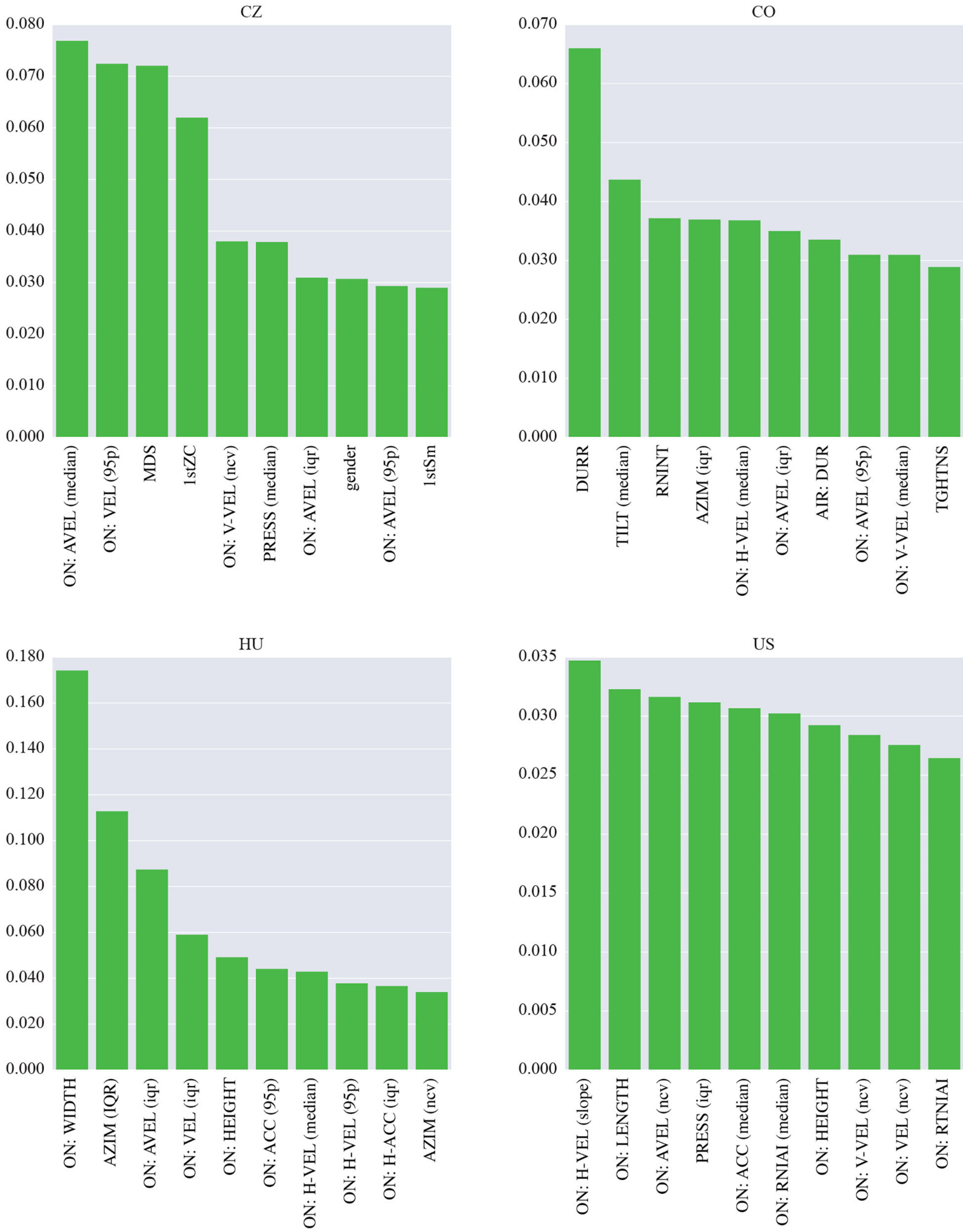


FIGURE 2 | Importance of the features used in the models in the single-language scenario (spiral drawing task).

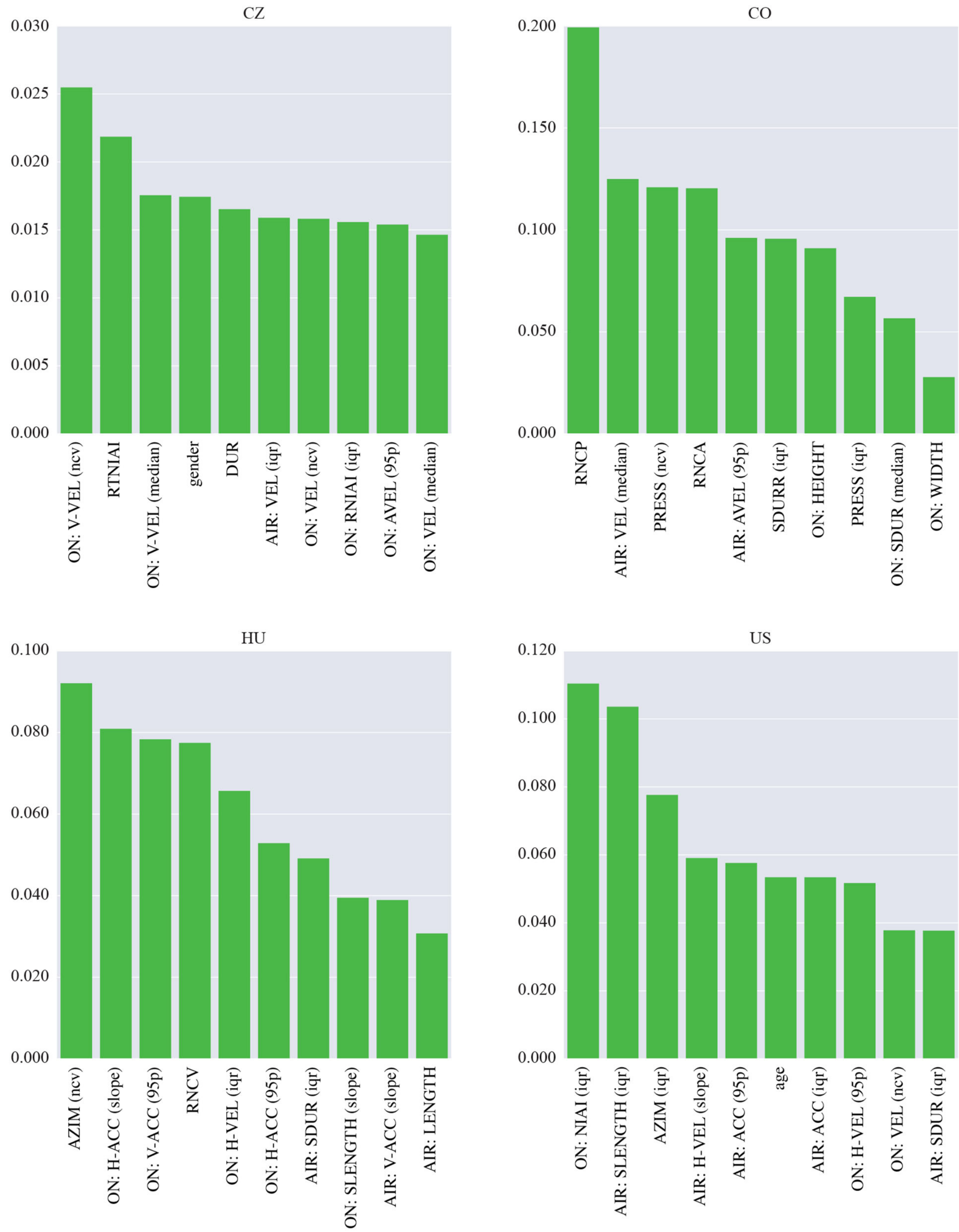


FIGURE 3 | Importance of the features used in the models in the single-language scenario (sentence writing task).

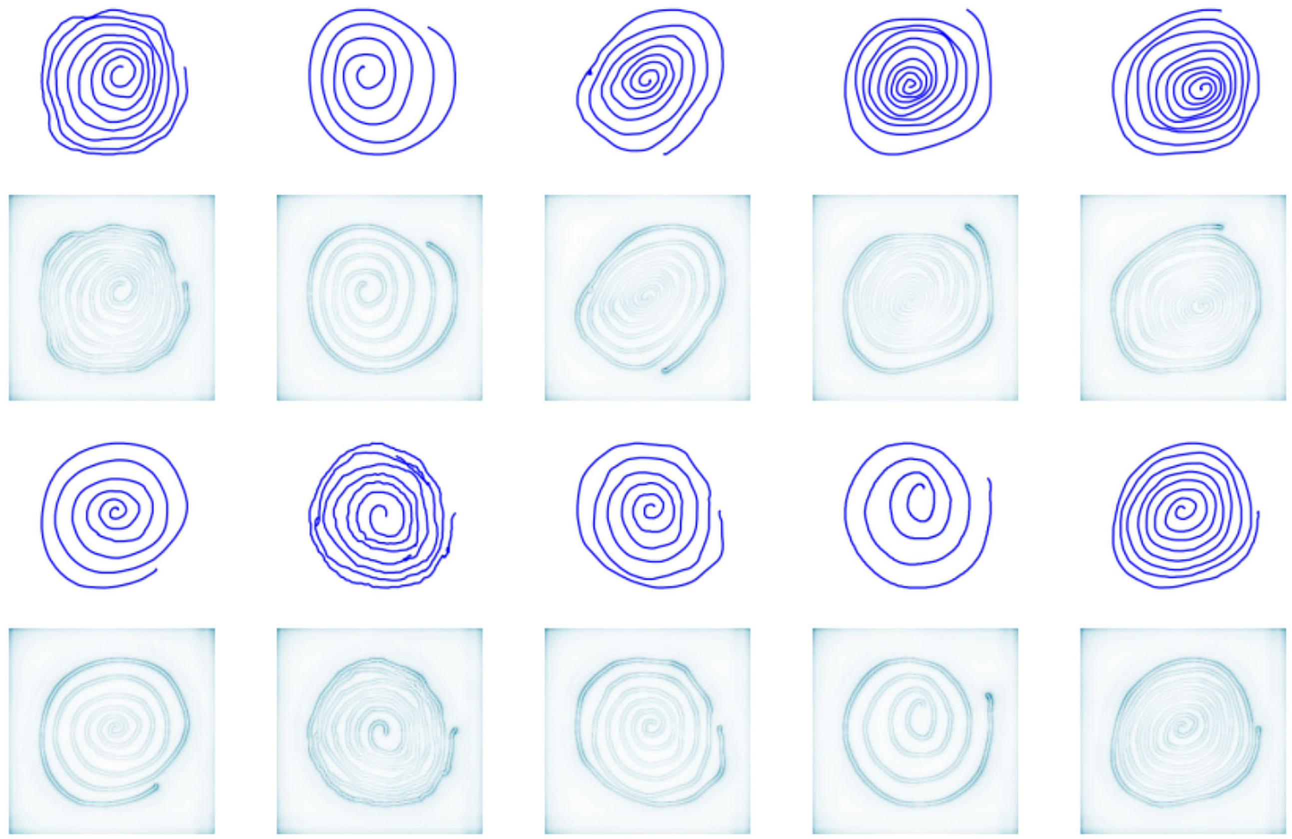


FIGURE 4 | Relevance maps for ten Archimedean spirals (two random samples from each dataset are depicted).

variation. In addition, the two highest-ranked velocity-based parameters are derived from the vertical projection. The most important feature set also contains the duration and number of intrastroke intersections. The most important feature in the CO dataset is the relative number of changes in the pressure profile, and two other pressure-based parameters (range and variation) were also selected. The last dynamic parameter is the number of changes in the azimuth profile. Regarding kinematic features, the set contains the in-air velocity and angular velocity. The stroke duration and spatial features such as width and height also play important roles. In the HU dataset, the most important feature is the variation in azimuth. Other significant features include the on-surface and in-air acceleration, and the on-surface horizontal velocity and the relative number of changes in the velocity profile are also important. Temporal features are represented by the in-air stroke duration. Finally, two important spatial parameters are identified: the on-surface stroke length and the overall length of the in-air movement. The three most important features in the US dataset are the number of on-surface intrastroke intersections, the in-air stroke length and the range of the azimuth. These are followed by mainly kinematic parameters, i.e., the in-air horizontal velocity, in-air acceleration, and on-surface velocity (including its horizontal projection). In terms of temporal features, the set also contains the in-air stroke duration.

The interpretation of CNN decisions is not straightforward since CNN models work in a black-box manner. We employ deep Taylor decomposition (Montavon et al., 2017) to gain a better understanding of the decisions made. Deep Taylor decomposition generates relevance maps illustrating the importance of single pixels in images. **Figures 4**, **5** show the relevance maps for ten spirals and four sentence writing samples, illustrating the pixels that were considered the most relevant for CNN-based feature extraction. Note that all figures that were used as CNN input were rendered at a resolution of 244×244 pixels. This resolution is optimal for the pretrained VGG network, but it created some deformation of the handwriting in the sentence writing task. This might have produced suboptimal results; however, using different resolutions would have required training the whole network from scratch, which would have been incompatible with the intention of this study.

4.2. Leave-One-Language-Out Scenario

The classification performance of the models trained in this scenario is summarized in **Table 5**. Naturally, the native language of a participant exerts no influence on the spiral drawing task; however, we can still investigate how the models performed on external validation datasets. When the CZ dataset was used as the test set, BACC degraded from 59 to 54% and

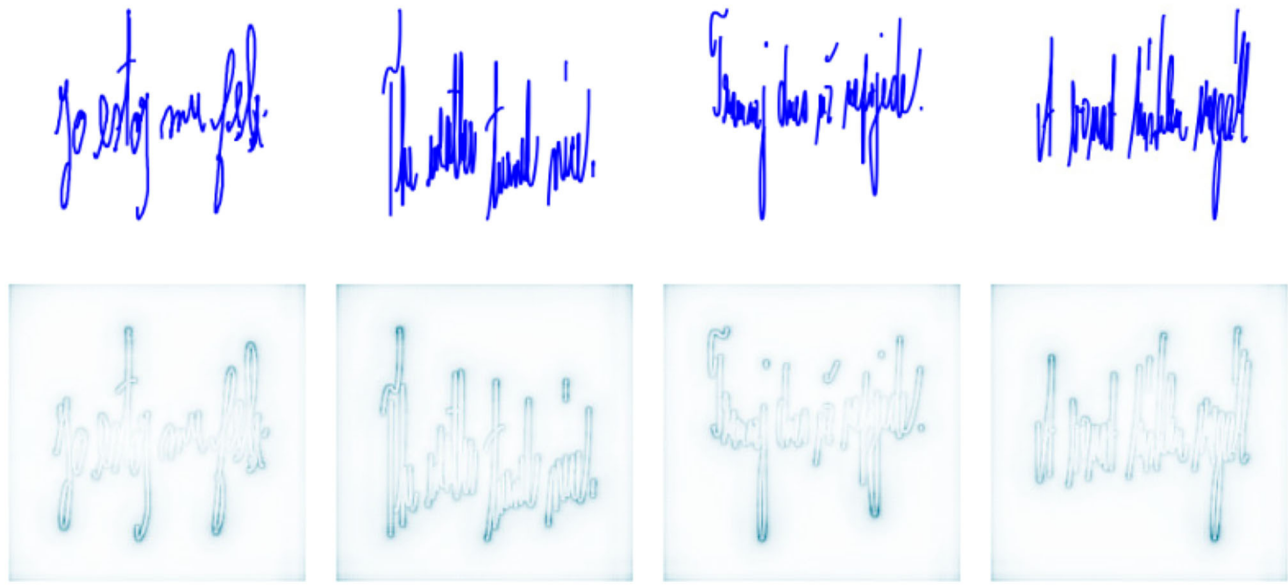


FIGURE 5 | Relevance maps for four sentences (one random sample is depicted for each language).

from 64 to 45% for the handcrafted and CNN-based features, respectively. In contrast, in the case of the CO test set, BACC with the handcrafted features decreased from 59 to 50%, while the performance of the CNN-based features slightly improved, specifically from 61 to 63%. In the case of the HU test set, BACC with the handcrafted features similarly degraded from 64 to 56%, but interestingly, when the CNN-extracted features were used, the classification performance improved from 48 to 71%, even higher than in the internal model validation in the previous experiment. This can be explained by the fact that the HU dataset is quite small, so the model was not able to learn well from data coming from the HU dataset only. Finally, the prediction performance on the US test set, which yielded optimistic results in the single-language scenario, decreased dramatically. For the handcrafted features, BACC decreased from 82 to 65%, and for the CNN-based features, the model completely failed to generalize, as BACC decreased from 77 to only 38%. This shows that the pattern responsible for the high classification accuracy in the internal model validation is most likely not present (or is less prominent) in the other datasets.

Regarding the sentence writing task, the language does exert an influence, and it is therefore important to look at the differences in the classification performance achieved in the internal and external validations. When the CZ dataset was used as the test set, BACC decreased from 66 to 63% and from 65 to 54% for the handcrafted and CNN-based features, respectively. In the case of the CO test set, BACC decreased from 56 to 50% for the handcrafted features and from 59 to 51% for the CNN-extracted features. With respect to the HU test set, BACC degraded from 75 to 67% for the handcrafted features but improved from 50 to 60% for the CNN-based features. This is consistent with the results of the spiral drawing task, for which the classifier based on the CNN-extracted features needed more data for training. In the case of the US test set, BACC improved for the handcrafted features, from 65 to 70%, but decreased for the CNN-extracted features, from 71 to 63%.

Interestingly, the classifiers utilizing the CNN-based features extracted from the spiral drawing task either outperformed those trained on the handcrafted features or failed to generalize, whereas the classifiers based on the handcrafted features

TABLE 5 | Classification performance in the leave-one-language-out scenario.

TRAIN	TEST	Features	BACC	F1	SEN	SPE
Spiral drawing						
CO+HU+US	CZ	Handcrafted	0.54	0.51	0.62	0.46
		CNN	0.45	0.41	0.48	0.42
CZ+HU+US	CO	Handcrafted	0.50	0.74	1.00	0.00
		CNN	0.63	0.62	0.54	0.71
CZ+CO+US	HU	Handcrafted	0.56	0.47	0.44	0.67
		CNN	0.71	0.67	0.67	0.75
CZ+CO+HU	US	Handcrafted	0.65	0.67	0.88	0.41
		CNN	0.38	0.32	0.33	0.42
Sentence writing						
CO+HU+US	CZ	Handcrafted	0.63	0.68	0.78	0.48
		CNN	0.54	0.58	0.80	0.29
CZ+HU+US	CO	Handcrafted	0.59	0.30	0.18	1.00
		CNN	0.51	0.72	0.82	0.20
CZ+CO+US	HU	Handcrafted	0.67	0.64	0.59	0.75
		CNN	0.60	0.46	0.38	0.83
CZ+CO+HU	US	Handcrafted	0.71	0.67	0.59	0.83
		CNN	0.63	0.46	0.33	0.92

TRAIN, training dataset; TEST, test dataset; BACC, balanced accuracy; F1, F1 score; SEN, sensitivity; SPE, specificity.

TABLE 6 | Classification performance in the scenario with all languages combined.

Task	Features	BACC	F1	SEN	SPE
Spiral	Handcrafted	0.60 ± 0.06	0.63 ± 0.06	0.73 ± 0.10	0.48 ± 0.07
	CNN	0.60 ± 0.01	0.61 ± 0.02	0.61 ± 0.04	0.61 ± 0.04
Sentence	Handcrafted	0.69 ± 0.05	0.65 ± 0.07	0.61 ± 0.09	0.78 ± 0.07
	CNN	0.66 ± 0.01	0.67 ± 0.01	0.67 ± 0.03	0.67 ± 0.03

BACC, balanced accuracy; F1, F1 score; SEN, sensitivity; SPE, specificity.

extracted from the sentence writing task yielded higher classification accuracy in all four experiments (with different combinations of training and test datasets). This was to be expected since in the latter case, the models were trained on sentences with orthography different from that in the test set. These findings confirm the hypothesis that the handcrafted features designed by domain experts are more robust than automatically extracted CNN-based features in cases in which different visual patterns are to be evaluated.

4.3. Scenario With All Languages Combined

In the last scenario, we combined the samples from all languages together to create a single heterogeneous dataset. The classification performance of the models trained in this scenario is summarized in **Table 6**. In the case of the spiral drawing task, the handcrafted features and CNN-based features show very similar performance, achieving 60% accuracy. The hypothesis that CNN-based features are more sensitive to the visual orthography of the sentence writing task is also confirmed by this last scenario, as the classifier based on handcrafted features outperformed the one trained on CNN-extracted features, achieving almost 70% accuracy (although in this case, the difference was much less prominent).

5. DISCUSSION

We compared the results of two different approaches to feature extraction: handcrafted features and features extracted by a CNN. In the case of the handcrafted features, we utilized a set of baseline features that are frequently used for handwriting analysis. We focused mainly on temporal, spatial, kinematic, and dynamic features, and we did not employ any advanced nonconventional features. Similarly, in the case of the CNN-extracted features, we used a pretrained VGG network to extract the features, although propositions have already emerged for improving the methodologies applied to diagnose PD from offline handwriting (Moetesum et al., 2019; Gazda et al., 2021). The motivations for this are two-fold. First, our aim was to establish baseline results that can be used as a reference in the future. Second, by using these baseline approaches, we could provide a fair comparison between the classification performance of handcrafted features and CNN-extracted features.

Regarding clinical interpretability, the models based on the Archimedean spiral drawing task mainly utilized kinematic features. This finding is reasonable because the cardinal symptoms of PD, such as rigidity, akinesia, and bradykinesia, have a significant impact on fine motor skills, including handwriting/drawing (Letanneux et al., 2014). Generally, PDYS is associated with reduced velocity (Ponsen et al., 2008; Rosenblum et al., 2013; Impedovo and Pirlo, 2018; De Stefano et al., 2019), which could occur more frequently than the most pronounced symptom, micrographia (Letanneux et al., 2014). Since the Archimedean spiral drawing task is a task in which subjects perform coordinated rotation, among the kinematic parameters, the angular velocity seems to play the most important role in the differentiation of PD/HC subjects.

Interestingly, features specifically designed for the assessment of Archimedean spiral drawing in PD patients (San Luciano et al., 2016; Cascarano et al., 2019; such as the smoothness of the spiral, the spiral tightness, the variability of the spiral width, and the spiral precision index) were not as important as we initially assumed. Similar to the dynamic features (e.g., pressure, tilt, azimuth), spatial features (width, height, length), and temporal features (duration), they were important only in some specific datasets.

Concerning the clinical interpretability of the models based on the sentence writing task, except for the CO database, all models were again based mainly on kinematic features, mostly extracted from the on-surface movement. In terms of projection, kinematic deficits were observed in both the horizontal and vertical movements. Nevertheless, in the largest database (CZ), deficits mainly dominated in the vertical projection. Kushki et al. (2011) reported that the finger system (which is mainly involved in vertical movement) is more affected by muscular fatigue than the wrist system (which controls horizontal movement). From an anatomical point of view, vertical movement requires coordinated movement and finer flexions/extensions of more joints (interphalangeal and metacarpophalangeal), i.e., it is more complex than ulnar abductions of the wrist (Van Galen, 1991; Dounskaia et al., 2000), and we assume this to be the reason why kinematic deficits were more strongly observed in this direction. This finding could also be somehow linked with progressive/consistent vertical micrographia, i.e., progressive/consistent reduction in letter amplitude (Thomas et al., 2017). However, this hypothesis requires further research because some studies suggest that the horizontal version of micrographia is even more common than the vertical version (Thomas et al., 2017).

Interestingly, except for the CZ database, the azimuth also played a significant role, more specifically its variation and range. We have identified one publication in which the authors advantageously utilized azimuth-based features in the semisupervised modeling of PDYS (Ammour et al., 2020). We assume that tremor could lead to improper coordination of the upper extremities, which could manifest as unstable azimuth features during the process of handwriting.

Temporal features (the duration of the whole process or of individual strokes) additionally played an important role in all models. In some studies, duration has not been found to be useful for discriminating between PD patients and HCs because although patients with PD write slowly, they also write smaller letters and thus ultimately spend the same time on, e.g., copying a sentence (Letanneux et al., 2014; Vessio, 2019). Nevertheless, in our case, with a few exceptions, spatial parameters were not found to be important in PDYS modeling.

Although it has been reported that PD patients generally apply less pressure (Rosenblum et al., 2013), we observed an important role of pressure-based features only in the CO model. Since only the CO database was recorded using the Wacom Cintiq tablet, the question arises of whether the corresponding discriminative power is associated solely with the disease or whether it is somehow enhanced by writing on a display.

In contrast to conventional shallow machine learning models, deep CNN models are quite challenging to interpret because of the dimensionality and complexity involved. However, as mentioned in the previous sections, we employed deep Taylor decomposition (Montavon et al., 2017) to create relevance maps illustrating the pixels that were considered most relevant for CNN-based feature extraction.

Regarding the spiral drawing task, as seen from **Figure 4**, the pixels that were assigned the highest weight for decisions lay along the outline of the drawn image. This indicates that the outer curve may convey information that can be explored to differentiate PD patients and HCs. We can hypothesize that this location in the spiral is strongly related to the shape and size of the spiral itself, which requires more focus and fine control over the kinematic and dynamic aspects of drawing. In the case of the sentence writing task, **Figure 5** shows that the most important pixels tend to be clustered around bends with high curvature. Again, this likely indicates that areas with higher differentiation potential are related to increased demand in terms of the kinematic and dynamic aspects of handwriting. This is an interesting observation showing that a CNN without any knowledge about the evolution of drawing/handwriting over time (as it is given only the final handwritten product) is able to identify the areas in handwritten images that require increased muscular control and focus. This observation could be consistent with the findings presented in Vásquez-Correa et al. (2019), where the transitions from non-moving to moving and from moving to non-moving states were shown to be highly informative. Additionally, this observation supports the importance of handcrafted features and poses an interesting research question of whether deep neural networks, when trained with adequately large and heterogeneous datasets, could provide more insights for the development of new features or whether the present knowledge about baseline handwriting features could be used for the development of novel deep neural networks specialized for automated feature extraction from handwriting/drawing.

5.1. Study Limitations

This work has several limitations. First, we need to be aware of the restricted statistical strength of any inferences regarding the population of patients with PD given the relatively limited sample size. In addition, although the clinical information is not complete for all of the datasets, it is evident that the PD cohort contains patients with different levels of PD progression; for example, based on the UPDRS III, the CO subjects are at a more severe stage than the CZ subjects. On the other hand, by fusing them together, we were able to train models that could support the diagnosis of PD in both severe and early stages.

Another limitation is associated with the effect of medication. Since we did not have information about LED for all PD subjects, we could not control for this effect in the statistical modeling. According to Zham et al. (2019), levodopa has a positive effect, especially on the performance in simple graphomotor tasks, such as the Archimedean spiral drawing task in our case. Nevertheless, the authors reported that no such benefit was observed in the sentence writing task, which imposes higher memory and cognitive loads. Therefore, we assume that controlling for the effect of medication in our analyses could further improve the performance of the models based on the spiral drawing task.

Next, although we performed unit unification and resampling on the signals so that they all had the same sampling frequency, the different recording conditions (e.g., paper vs. the display version of the tablets) could still have had some impact on the results.

In addition, various machine learning models should be trained and compared in future studies to obtain more information about the classification performance of the proposed features and to obtain the most robust models for PDYS identification. Finally, the relationship between the classification performance of the trained models and the feature space complexity as well as the cross-validation setup should be investigated to evaluate and confirm the robustness of the proposed methodology.

In summary, considering its limitations, this study should be viewed as a pilot study that is exploratory in nature, and its results should be confirmed by subsequent research studies.

6. CONCLUSION

We investigated several aspects of handwriting evaluation for the detection of PDYS. First, we compared the utilization of handcrafted features with the utilization of features extracted by a CNN. We found that the two approaches are competitive, especially for the spiral drawing task, which is independent of language. Handcrafted features (especially kinematic features) proved to be the better choice for the sentence writing task in multilingual scenarios. This is expected since CNN-based features are extracted only from offline handwriting samples, from which temporal information is not available. In addition, the orthography of a sentence is strongly affected by the language of the writer. Second, we analyzed the effect of multilinguality on the training and performance of classification models. Here, in contrast to our initial hypothesis, model validation performed on sentences written in a different language than the ones used for training did not result in performance degradation. In fact, the prediction accuracy improved in the case of the US and HU datasets. Finally, we compared the sentence writing task and the spiral drawing task. Here, the sentence writing task showed higher discrimination potential, even in multilingual scenarios.

Although there are several limitations, to the best of our knowledge, this is the first study to compare the classification performance of conventional handcrafted features designed by domain experts and features extracted automatically by a pretrained CNN from a multilingual dataset collected from patients suffering from PD. It also provides an objective evaluation of PDYS detection using two different and very promising approaches and analyzes several aspects of handwriting that are frequently neglected in the literature. Based on the results, we can conclude that both types of features have great potential to be used to describe various aspects of drawing/handwriting in both language-independent and language-dependent scenarios. In summary, our work can be perceived as establishing some initial baseline results for further research toward the introduction of new prediction models utilizing handcrafted features as well as CNN-based features that could provide more robustness and confidence in the identification of HD in patients with PD.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee of Masaryk University, Zerotinovo Namesti, 617/9, 601 77 Brno, Czech Republic. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

IR, TK, SR, JO-A, MF-Z, ZS, JMe, PD, and ZG: conceptualization. ZG, PD, JMe, MG, JMu, and VZ: research about the current state of knowledge. IR, LB, TK, SR, JO-A, RC, MF-Z, JMe, and ZG: database acquisition, development, and processing. ZG, PD, JMe, MG, JMu, and VZ: feature extraction, machine learning, and experiments. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 734718 (CoBeN), by the Czech Ministry of Health under grant no. NU20-04-00294, by the Slovak Research and Development Agency under contract no. APVV-16-0211, by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under contract VEGA 1/0327/20, and by the Spanish grant Ministerio de ciencia e innovación PID2020-113242RB-I00.

REFERENCES

Agrell, B., and Dehlin, O. (1998). The clock-drawing test. *Age Ageing* 27, 399–403. doi: 10.1093/ageing/27.3.399

Alonso-Martinez, C., Faundez-Zanuy, M., and Mekyska, J. (2017). A comparative study of in-air trajectories at short and long distances in online handwriting. *Cogn. Comput*. 9, 712–720. doi: 10.1007/s12559-017-9501-5

Ammour, A., Aouraghe, I., Khaissidi, G., Mrabti, M., Aboulem, G., and Belahsen, F. (2020). A new semi-supervised approach for characterizing the Arabic on-line handwriting of Parkinson's disease patients. *Comput. Methods Prog. Biomed*. 183:104979. doi: 10.1016/j.cmpb.2019.07.007

Aouraghe, I., Alae, A., Ghizlane, K., Mrabti, M., Aboulem, G., and Faouzi, B. (2020). A novel approach combining temporal and spectral features of Arabic online handwriting for Parkinson's disease prediction. *J. Neurosci. Methods* 339:108727. doi: 10.1016/j.jneumeth.2020.108727

Berardelli, A., Rothwell, J., Thompson, P., and Hallett, M. (2001). Pathophysiology of Bradykinesia in Parkinson's disease. *Brain* 124, 2131–2146. doi: 10.1093/brain/124.11.2131

Bidet-Ildei, C., Pollak, P., Kandel, S., Fraix, V., and Orliaguet, J.-P. (2011). Handwriting in patients with Parkinson disease: effect of l-DOPA and stimulation of the sub-thalamic nucleus on motor anticipation. *Hum. Mov. Sci*. 30, 783–791. doi: 10.1016/j.humov.2010.08.008

Brabenec, L., Mekyska, J., Galaz, Z., and Rektorova, I. (2017). Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. *J. Neural Transm*. 124, 303–334. doi: 10.1007/s00702-0171676-0

Brodal, P. (2003). *The Central Nervous System: Structure and Function, 3 Edn*. Oxford: Oxford University Press.

Cascarano, G. D., Loconsole, C., Brunetti, A., Lattarulo, A., Buongiorno, D., Losavio, G., et al. (2019). Biometric handwriting analysis to support Parkinson's disease assessment and grading. *BMC Med. Inform. Decis. Making* 19, 1–11. doi: 10.1186/s12911-019-0989-3

Castrillon, R., Acien, A., Orozco-Arroyave, J. R., Morales, A., Vargas, J., Vera-Rodríguez, R., et al. (2019). "Characterization of the handwriting skills as a biomarker for Parkinson's disease," in *2019 14th IEEE International Conference on Automatic Face* & *Gesture Recognition (FG 2019)* (Lille), 1–5. doi: 10.1109/FG.2019.8756508

Chen, T., and Guestrin, C. (2016). "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. doi: 10.1145/2939672.2939785

Collins, G. S., and Moons, K. G. M. (2019). Reporting of artificial intelligence prediction models. *Lancet* 393, 1577–1579. doi: 10.1016/S0140-6736(19)30037-6

De Stefano, C., Fontanella, F., Impedovo, D., Pirlo, G., and di Freca, A. S. (2019). Handwriting analysis to support neurodegenerative diseases diagnosis: a review. *Pattern Recogn. Lett*. 121, 37–45. doi: 10.1016/j.patrec.2018.05.013

Diaz, M., Ferrer, M. A., Impedovo, D., Pirlo, G., and Vessio, G. (2019). Dynamically enhanced static handwriting representation for Parkinson's disease detection. *Pattern Recogn. Lett*. 128, 204–210. doi: 10.1016/j.patrec.2019.08.018

Diaz, M., Moetesum, M., Siddiqi, I., and Vessio, G. (2021). Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and BiGRUs. *Expert Syst. Appl*. 168:114405. doi: 10.1016/j.eswa.2020.114405

Dickson, D. W. (2012). Parkinson's disease and Parkinsonism: neuropathology. *Cold Spring Harb. Perspect. Med*. 2:a009258. doi: 10.1101/cshperspect.a009258

Dounskaia, N., Van Gemmert, A., and Stelmach, G. (2000). Interjoint coordination during handwriting-like movements. *Exp. Brain Res*. 135, 127–140. doi: 10.1007/s002210000495

Drotar, P., Mekyska, J., Rektorova, I., Masarova, L., Smekal, Z., and Faundez-Zanuy, M. (2014). Analysis of in-air movement in handwriting: a novel marker for Parkinson's disease. *Comput. Methods Prog. Biomed*. 117, 405–411. doi: 10.1016/j.cmpb.2014.08.007

Drotar, P., Mekyska, J., Rektorova, I., Masarova, L., Smekal, Z., and Faundez-Zanuy, M. (2016). Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artif. Intell. Med*. 67, 39–46. doi: 10.1016/j.artmed.2016.01.004

Drotar, P., Mekyska, J., Rektorova, I., Masarova, L., Smekal, Z., and Zanuy, M. F. (2015). Decision support framework for Parkinson's disease based on novel handwriting markers. *IEEE Trans. Neural Syst. Rehabil. Eng*. 23, 508–516. doi: 10.1109/TNSRE.2014.2359997

Fahn, S., and Elton, R. L. (1987). *UPDRS Development Committee (1987) Unified Parkinson's Disease Rating Scale. Recent Developments in Parkinson's Disease*. Florham Park, NJ: Macmillan.

Forno, L. S. (1996). Neuropathology of Parkinson's disease. *J. Neuropathol. Exp. Neurol*. 55, 259–272. doi: 10.1097/00005072-199603000-00001

Galaz, Z., Mucha, J., Zvoncak, V., Mekyska, J., Smekal, Z., Safarova, K., et al. (2020). Advanced parametrization of graphomotor difficulties in school-aged children. *IEEE Access* 8, 112883–112897. doi: 10.1109/ACCESS.2020.3003214

Garre-Olmo, J., Faúndez-Zanuy, M., López-de Ipi na, K., Calvó-Perxas, L., and Turró-Garriga, O. (2017). Kinematic and pressure features of handwriting and drawing: preliminary results between patients with mild cognitive impairment, Alzheimer disease and healthy controls. *Curr. Alzheimer Res*. 14, 960–968. doi: 10.2174/1567205014666170309120708

Gazda, M., Hireš, M., and Drotár, P. (2021). Multiple-fine-tuned convolutional neural networks for Parkinson's disease diagnosis from offline handwriting. *IEEE Trans. Syst. Man Cybernet*. 1–12. doi: 10.1109/TSMC.2020.3048892

Gil-Martin, M., Montero, J. M., and San-Segundo, R. (2019). Parkinsons disease detection from drawing movements using convolutional neural networks. *Electronics* 8:907. doi: 10.3390/electronics8080907

Goetz, C. G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G. T., Counsell, C., et al. (2004). Movement disorder society task force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov. Disord*. 19, 1020–1028. doi: 10.1002/mds.20213

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., et al. (2008). Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov. Disord*. 23, 2129–2170. doi: 10.1002/mds.22340

Hagerty, J. R., Stanley, R. J., Almubarak, H. A., Lama, N., Kasmi, R., Guo, P., et al. (2019). Deep learning and handcrafted method fusion: Higher diagnostic accuracy for melanoma dermoscopy images. *IEEE J. Biomed. Health Informatics* 23, 1385–1391. doi: 10.1109/JBHI.2019.2891049

Heinzel, S., Berg, D., Gasser, T., Chen, H., Yao, C., Postuma, R. B., et al. (2019). Update of the MDS research criteria for prodromal Parkinson's disease. *Mov. Disord*. 34, 1464–1470. doi: 10.1002/mds.27802

Horak, F. B., Dimitrova, D., and Nutt, J. G. (2005). Direction-specific postural instability in subjects with Parkinson's disease. *Exp. Neurol*. 193, 504–521. doi: 10.1016/j.expneurol.2004.12.008

Hornykiewicz, O. (1998). Biochemical aspects of Parkinson's disease. *Neurology* 51(2 Suppl. 2), S2–S9. doi: 10.1212/WNL.51.2_Suppl_2.S2

Hughes, A., Daniel, S., and Lees, A. (1993). The clinical features of Parkinson's disease in 100 histologically proven cases. *Adv. Neurol*. 60:595.

Impedovo, D. (2019). Velocity-based signal features for the assessment of Parkinsonian handwriting. *IEEE Signal Process. Lett*. 26, 632–636. doi: 10.1109/LSP.2019.2902936

Impedovo, D., and Pirlo, G. (2018). Dynamic handwriting analysis for the assessment of neurodegenerative diseases: a pattern recognition perspective. *IEEE Rev. Biomed. Eng*. 12, 209–220. doi: 10.1109/RBME.2018.2840679

Impedovo, D., Pirlo, G., and Vessio, G. (2018). Dynamic handwriting analysis for supporting earlier Parkinson's disease diagnosis. *Information* 9:247. doi: 10.3390/info9100247

Jerkovic, V. M., Kojic, V., Miskovic, N. D., Djukic, T., Kostic, V. S., and Popovic, M. B. (2019). Analysis of on-surface and in-air movement in handwriting of subjects with Parkinson's disease and atypical parkinsonism. *Biomed. Eng*. 64, 187–194. doi: 10.1515/bmt-2017-0148

Kamble, M., Shrivastava, P., and Jain, M. (2021). Digitized spiral drawing classification for Parkinson's disease diagnosis. *Measurement* 16:100047. doi: 10.1016/j.measen.2021.100047

Kamran, I., Naz, S., Razzak, I., and Imran, M. (2021). Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease. *Fut. Gener. Comput. Syst*. 117, 234–244. doi: 10.1016/j.future.2020.11.020

Kushki, A., Schwellnus, H., Ilyas, F., and Chau, T. (2011). Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia. *Res. Dev. Disabil*. 32, 1058–1064. doi: 10.1016/j.ridd.2011.01.026

Lee, J. Y., Kim, J. W., Lee, W. Y., Kim, J. M., Ahn, T. B., Kim, H. J., et al. (2010). Daily dose of dopaminergic medications in Parkinson's disease: clinical correlates and a posteriori equation. *Neurol. Asia* 15, 137–143.

Letanneux, A., Danna, J., Velay, J.-L., Viallet, F., and Pinto, S. (2014). From micrographia to Parkinson's disease dysgraphia. *Mov. Disord*. 29, 1467–1475. doi: 10.1002/mds.25990

McLennan, J., Nakano, K., Tyler, H., and Schwab, R. (1972). Micrographia in Parkinson's disease. *J. Neurol. Sci*. 15, 141–152. doi: 10.1016/0022-510X(72)90002-0

Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., and Jamalipour Soufi, G. (2020). Deep-COVID: predicting covid-19 from chest x-ray images using deep transfer learning. *Med. Image Anal*. 65:101794. doi: 10.1016/j.media.2020.101794

Moetesum, M., Siddiqi, I., Vincent, N., and Cloppet, F. (2019). Assessing visual attributes of handwriting for prediction of neurological disorders-a case study on Parkinson's disease. *Pattern Recogn. Lett*. 121, 19–27. doi: 10.1016/j.patrec.2018.04.008

Montavon, G., Lapuschkin, S., Binder, A., Samek, W., and Müller, K.-R. (2017). Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recogn*. 65, 211–222. doi: 10.1016/j.patcog.2016.11.008

Morberg, B. M., Malling, A. S., Jensen, B. R., Gredal, O., Wermuth, L., and Bech, P. (2018). The Hawthorne effect as a pre-placebo expectation in Parkinson's disease patients participating in a randomized placebo-controlled clinical study. *Nordic J. Psychiatry* 72, 442–446. doi: 10.1080/08039488.2018.1468480

Mucha, J., Mekyska, J., Faundez-Zanuy, M., Lopez-de Ipina, K., Zvoncak, V., Galaz, Z., et al. (2018a). "Advanced Parkinson's disease dysgraphia analysis based on fractional derivatives of online handwriting," in *10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. doi: 10.1109/ICUMT.2018.8631265

Mucha, J., Mekyska, J., Galaz, Z., Faundez-Zanuy, M., Lopez-de Ipina, K., Zvoncak, V., et al. (2018b). Identification and monitoring of Parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting. *Appl. Sci*. 8:2566. doi: 10.3390/app8122566

Mucha, J., Mekyska, J., Galaz, Z., Faundez-Zanuy, M., Zvoncak, V., Safarova, K., et al. (2020). Analysis of various fractional order derivatives approaches in assessment of graphomotor difficulties. *IEEE Access* 8, 218234–218244. doi: 10.1109/ACCESS.2020.3042591

Mucha, J., Mekyska, J., Galáž, Z., Zvonč'ak, V., Kiska, T., and Smékal, Z. (2019). "Advanced analysis of online handwriting in a multilingual cohort of patients with Parkinson's disease," in *Advances in Signal Processing and Artificial Intelligence: Proceedings of the 1st International Conference on Advances in Signal Processing and Artificial Intelligence* (International Frequency Sensor Association (IFSA) Publishing, S. L.), 144–147.

Nolazco-Flores, J. A., Faundez-Zanuy, M., De La Cueva, V., and Mekyska, J. (2021). Exploiting spectral and cepstral handwriting features on diagnosing Parkinson's disease. *IEEE Access* 9, 141599–141610. doi: 10.1109/ACCESS.2021.3119035

O'Reilly, C. and Plamondon, R. (2009). Development of a sigma-lognormal representation for on-line signatures. *Pattern Recogn*. 42, 3324–3337. doi: 10.1016/j.patcog.2008.10.017

Pereira, C. R., Weber, S. A. T., Hook, C., Rosa, G. H., and Papa, J. P. (2016). "Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 340–346. doi: 10.1109/SIBGRAPI.2016.054

Pinto, S., and Velay, J.-L. (2015). Handwriting as a marker for PD progression: a shift in paradigm. *Neurodegener. Dis. Manage*. 5, 367–369. doi: 10.2217/nmt.15.29

Ponsen, M. M., Daffertshofer, A., Wolters, E. C., Beek, P. J., and Berendse, H. W. (2008). Impairment of complex upper limb motor function in *de novo* Parkinson's disease. *Parkinsonism Relat. Disord*. 14, 199–204. doi: 10.1016/j.parkreldis.2007.07.019

Postuma, R. B., Berg, D., Adler, C. H., Bloem, B. R., Chan, P., Deuschl, G., et al. (2016). The new definition and diagnostic criteria of Parkinson's disease. *Lancet Neurol*. 15, 546–548. doi: 10.1016/S1474-4422(16)00116-2

Pullman, S. (1998). Spiral analysis: a new technique for measuring tremor with a digitizing tablet. *Mov. Disord*. 13, 85–89. doi: 10.1002/mds.870131315

Rios-Urrego, C., Vásquez-Correa, J., Vargas-Bonilla, J., Nöth, E., Lopera, F., and Orozco-Arroyave, J. (2019). Analysis and evaluation of handwriting in patients with Parkinson's disease using kinematic, geometrical, and non-linear features. *Comput. Methods Prog. Biomed.* 173, 43–52. doi: 10.1016/j.cmpb.2019.03.005

Rosenblum, S., Samuel, M., Zlotnik, S., Erikh, I., and Schlesinger, I. (2013). Handwriting as an objective tool for Parkinson's disease diagnosis. *J. Neurol.* 260, 2357–2361. doi: 10.1007/s00415-013-6996-x

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y

San Luciano, M., Wang, C., Ortega, R. A., Yu, Q., Boschung, S., Soto-Valencia, J., et al. (2016). Digitized spiral drawing: a possible biomarker for early Parkinson's disease. *PLoS ONE* 11:e0162799. doi: 10.1371/journal.pone.0162799

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv[Preprint].arXiv:1409.1556.*

Thomas, M., Lenka, A., and Kumar Pal, P. (2017). Handwriting analysis in Parkinson's disease: current status and future directions. *Mov. Disord. Clin. Pract.* 4, 806–818. doi: 10.1002/mdc3.12552

Van Galen, G. P. (1991). Handwriting: Issues for a psychomotor theory. *Hum. Mov. Sci.* 10, 165–191. doi: 10.1016/0167-9457(91)90003-G

Vásquez-Correa, J. C., Arias-Vergara, T., Orozco-Arroyave, J. R., Eskofier, B., Klucken, J., and Nöth, E. (2019). Multimodal assessment of Parkinson's disease: a deep learning approach. *IEEE J. Biomed. Health Inform.* 23, 1618–1630. doi: 10.1109/JBHI.2018.2866873

Vessio, G. (2019). Dynamic handwriting analysis for neurodegenerative disease assessment: a literary review. *Appl. Sci.* 9:4666. doi: 10.3390/app9214666

Zham, P., Kumar, D., Viswanthan, R., Wong, K., Nagao, K. J., Arjunan, S. P., et al. (2019). Effect of levodopa on handwriting tasks of different complexity in Parkinson's disease: a kinematic study. *J. Neurol.* 266, 1376–1382. doi: 10.1007/s00415-019-09268-2

Zham, P., Kumar, D. K., Dabnichki, P., Poosapadi Arjunan, S., and Raghav, S. (2017). Distinguishing different stages of Parkinson's disease using composite index of speed and pen-pressure of sketching a spiral. *Front. Neurol.* 8:435. doi: 10.3389/fneur.2017.00435

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

A.2 Prodromal Diagnosis of Lewy Body Diseases Based on the Assessment of Graphomotor and Handwriting Difficulties

Prodromal Diagnosis of Lewy Body Diseases Based on the Assessment of Graphomotor and Handwriting Difficulties ⋆

Zoltan Galaz[1][0000-0002-8978-351X], Jiri Mekyska[1][0000-0002-6195-193X], Jan Mucha[1][0000-0001-5126-440X], Vojtech Zvoncak[1][0000-0002-1948-4653], Zdenek Smekal[1][0000-0002-8483-5448], Marcos Faundez-Zanuy[2][0000-0003-0605-1282], Lubos Brabenec[3][0000-0002-8348-5757], Ivona Moravkova[3,4,5][0000-0003-1050-685X], and Irena Rektorova[3,4][0000-0002-5455-4573]

[1] Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic
[2] Escola Superior Politecnica, Tecnocampus, Mataro, Barcelona, Spain
[3] Applied Neuroscience Research Group, Central European Institute of Technology – CEITEC, Masaryk University, Brno, Czech Republic
[4] First Department of Neurology, Faculty of Medicine and St. Anne's University Hospital, Masaryk University, Brno, Czech Republic
[5] Faculty of Medicine, Masaryk University, Brno, Czech Republic

Abstract. To this date, studies focusing on the prodromal diagnosis of Lewy body diseases (LBDs) based on quantitative analysis of graphomotor and handwriting difficulties are missing. In this work, we enrolled 18 subjects diagnosed with possible or probable mild cognitive impairment with Lewy bodies (MCI-LB), 7 subjects having more than 50 % probability of developing Parkinson's disease (PD), 21 subjects with both possible/probable MCI-LB and probability of PD > 50 %, and 37 age- and gender-matched healthy controls (HC). Each participant performed three tasks: Archimedean spiral drawing (to quantify graphomotor difficulties), sentence writing task (to quantify handwriting difficulties), and pentagon copying test (to quantify cognitive decline). Next, we parameterized the acquired data by various temporal, kinematic, dynamic, spatial, and task-specific features. And finally, we trained classification models for each task separately as well as a model for their combination to estimate the predictive power of the features for the identification of LBDs. Using this approach we were able to identify prodromal LBDs with 74 % accuracy and showed the promising potential of computerized objective and non-invasive diagnosis of LBDs based on the assessment of graphomotor and handwriting difficulties.

⋆ This work was supported by grant no. NU20-04-00294 (Diagnostics of Lewy body diseases in prodromal stage based on multimodal data analysis) of the Czech Ministry of Health and by Spanish grant of the Ministerio de Ciencia e Innovación no. PID2020-113242RB-I00.

Keywords: Lewy body diseases · online handwriting · graphomotor difficulties · handwriting difficulties · machine learning · prodromal diagnosis.

1 Introduction

Lewy body diseases (LBDs) is a term describing a group of neurodegenerative disorders characterized by a pathophysiological process of α-synuclein accumulation in specific brain regions leading to the formation of Lewy bodies and Lewy neurites resulting in cell death. LBDs consists of two major clinical entities: Parkinson's disease (PD) and dementia with Lewy bodies (DLB) [30, 39]. Although the phenotypes and temporal evolution of motor and cognitive symptoms of these two diseases vary, they share many clinical and pathophysiological features and are therefore referred to as LBDs spectrum. Together with Alzheimer's disease (AD), LBDs comprise the major part of all cases of neurodegenerative disorders.

It is known that LBDs do not start suddenly. At the time the clinical symptoms occur, the neurodegenerative process has reached a severe degree in which most of the targeted neurons have already been damaged. Before the clinical diagnosis based on the presence of typical clinical symptoms becomes possible, there is a long period of the underlying neurodegenerative process with subtle or nonspecific symptoms [18, 30] such as sleep disturbances, mood changes, smell loss, constipation, etc. This period of LBDs is called the prodromal stage.

One of the early markers of PD is PD dysgraphia (micrographia and other alterations in handwriting, e. g. kinematic and dynamic) [21, 33, 34]. Similarly, some manifestations of dysgraphia have been observed in the prodromal DLB as well [24]. Although modern approaches to the analysis of graphomotor and handwriting difficulties (utilising digitising tablets) were proved to work well during e. g. diagnosis of the clinical stage of PD [9, 11, 36], assessment of cognition in PD patients [4], or discrimination of AD and mild cognitive impairment (MCI) [15], to the best of our knowledge, no studies employed this technology (with high potential) in the prodromal diagnosis of LBDs in a larger scale.

Identification of the early stages of LBDs is crucial for the development of disease-modifying treatment since the neurodegeneration may be possibly stopped or treated before the pathological cascades start. Therefore, the goal of this study is to explore whether the computerised assessment of graphomotor and handwriting difficulties could support the prodromal diagnosis of LBDs, more specifically, we aim to:

1. identify which task significantly discriminates LBD patients and age- and gender-matched healthy controls (HC),
2. identify what conventional online handwriting features have good discrimination power.

2 Materials and methods

2.1 Dataset

We enrolled 39 subjects (19 females, 20 males, age = 69.53 ± 6.61) diagnosed with possible or probable MCI (based on the scores of the MoCA – Montreal Cognitive Assessment [26] and based on the CCB – Complex Cognitive Battery, see the explanation below) who were simultaneously diagnosed with possible or probable MCI-LB (i.e. mild cognitive impairment with Lewy bodies) based on the criteria published by McKeith et al. [23]. In this group, 21 subjects also had more than 50 % probability of developing PD (calculated following the MDS criteria published in [18]). In addition, we enrolled 7 subjects (2 females, 5 males, age = 66.41 ± 4.32) without possible/probable MCI-LB, but still with more than 50 % probability of developing PD. Finally, we enrolled 37 HC (26 females, 11 males, age = 67.60 ± 5.61). In the experiments, we stratified the subjects into two groups, HC vs. LBD (i. e. people with a high risk of developing PD or DLB).

CCB was used to evaluate four cognitive domains: 1) memory (The Brief Visuospatial memory test–revised [2], Philadelphia Verbal Learning Test [3]); 2) attention (Wechsler Adult Intelligence Scale-III: Letter-Number Sequencing, Digit Symbol Substitution [38]); 3) executive functions (Semantic and phonemic verbal fluency [31], Picture arrangement test [38]); and 4) visuospatial functions (Judgment of Line Orientation [37]). The cognitive domain z-scores were computed as the average z-scores of the tests included in the particular domain.

The participants were asked to perform a set of three tasks:

1. Archimedean spiral (spiral) – we consider this task as a graphomotor one, i. e. it is a building block of some letter shapes; in addition, it is a golden standard in PD dysgraphia diagnosis [36]
2. sentence “Tramvaj dnes už nepojede” (translation: “A tram will not go today.”) writing (sentence) – this handwriting task was used e. g. in the PaHaW database [11]
3. pentagon copying test (pentagons) – it is a task frequently used for quantification of cognitive decline [4]

All participants were right-handed and had Czech as their native language. They all signed an informed consent form that was approved by the local ethics committee.

2.2 Feature extraction

The participants were asked to perform the tasks (using the Wacom Ink pen) on an A4 paper that was laid down and fixed to a digitizing tablet Wacom Intuos 4 M (sampling frequency $f_s = 130\,\text{Hz}$). Before the acquisition, they had some time to get familiar with the hardware. The recorded time series (x and y position; timestamp; a binary variable, being 0 for in-air movement and 1 for on-surface movement, respectively; pressure exert on the tablet's surface during writing; pen tilt; azimuth) were consequently parameterised utilising the following set of features (we selected the set based on available reviews and based on our experience [9, 11, 36]):

1. temporal – duration of writing, ratio of the on-surface/in-air duration, duration of strokes, and ratio of the on-surface/in-air stroke duration
2. kinematic – velocity, and acceleration
3. dynamic – pressure, tilt, and azimuth
4. spatial – width, height, and length of the whole product, as well as its particular strokes, i. e. stroke width, height, and length
5. spiral-specific – degree of spiral drawing severity [32], mean drawing speed of spiral [32], second-order smoothness of spiral [32], spiral precision index [5], spiral tightness [32], variability of spiral width [32], and first-order zero-crossing rate of spiral [32]
6. other – number of interruptions (pen elevations), number of pen stops [28], tempo (number of strokes normalised by duration), number of on-surface intra-stroke intersections, relative number of on-surface intra-stroke intersections, number of on-surface inter-stroke intersections, and relative number of on-surface inter-stroke intersections, Shannon entropy [4], number of changes in the velocity profile, relative number of changes in the velocity profile

Most of the features were extracted using the recently released Python library handwriting-features (v 1.0.1) [14], the rest of them were coded in Matlab. Some features (mainly spatial, temporal and kinematic) were extracted from both on-surface and in-air movements. In addition, kinematic features were also analysed in horizontal and vertical projection. Features represented by vectors were consequently transformed to a scalar value using median, non-parametric coefficient of variation (nCV; interquartile range of feature divided by its median), slope and 95th percentile (95p).

2.3 Statistical analysis and machine learning

To compare the distribution of features between the HC and LBD subjects, we conducted Mann-Whitney U-test with the significance level of 0.05. Moreover, to assess the strength of a relationship between the features and the subject's clinical status (HC/LBD), we computed Spearman's correlation coefficient (ρ) with the significance level of 0.05. Finally, during this exploratory step, we calculated Spearman's correlation with the domains of CCB and the overall score of MDS–Unified Parkinson's Disease Rating Scale (MDS–UPDRS), part III (motor part) [16].

To identify the presence of graphomotor or handwriting difficulties, we built binary classification models using an ensemble extreme gradient boosting algorithm known as XGBoost [6] (with 100 estimators). This algorithm was chosen due to its robustness to outliers, ability to find complex interactions among features as well as the possibility of ranking their importance. To build models with an optimal set of hyperparameters, we conducted 1000 iteration of randomized search strategy via stratified 5-fold cross-validation with 10 repetitions aiming to optimize balanced accuracy score (BACC; described in more detail along with other evaluation scores below). The following set of hyperparameters were optimized: the learning rate [0.001, 0.01, 0.1, 0.2, 0.3], γ [0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.5], the maximum tree depth [6, 8, 10, 12, 15], the fraction of observations to be randomly sampled for each tree (subsample ratio) [0.5, 0.6, 0.7, 0.8, 0.9, 1.0], the subsample ratio for the columns at each level [0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], the subsample ratio for the columns when constructing each tree [0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0], the minimum sum of the weights of all observations required in a child node [0.5, 1.0, 3.0, 5.0, 7.0, 10.0], and the balance between positive and negative weights [1, 2, 3, 4].

The classification test performance was determined using the following classification metrics: Matthew's correlation coefficient [22] (MCC), balanced accuracy (BACC), sensitivity (SEN) also known as recall (REC), specificity (SPE), precision (PRE) and F1 score (F1). These metrics are defined as follows:

$$\mathrm{MCC} = \frac{TP \times TN + FP \times FN}{\sqrt{N}}, \tag{1}$$

$$\mathrm{BACC} = \frac{1}{2}\left(\frac{TP}{TP+FN}\frac{TN}{TN+FP}\right), \tag{2}$$

$$\mathrm{SPE} = \frac{TN}{TN+FP}, \tag{3}$$

$$\mathrm{PRE} = \frac{TP}{TP+FP}, \tag{4}$$

$$\mathrm{REC} = \frac{TP}{TP+FN}, \tag{5}$$

$$\mathrm{F1} = 2\frac{PRE \times REC}{PRE + REC} \tag{6}$$

where $N = (TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)$, TP (true positive) and FP (false positive) represent the number of correctly identified LBD subjects and the number of subjects incorrectly identified as having LBDs, respectively. Similarly, TN (true negative) and FN (false negative) represent the number of correctly identified HC and the number of subjects with LBDs incorrectly identified as being healthy.

To further optimize the trained classification models, we fine-tuned the models' decision thresholds via the receiver operating characteristics (ROC) curve. Using the fine-tuned decision thresholds, we evaluated the classification performance of the models using the leave-one-out cross-validation. The ROC curves were plotted using the probabilities of the predicted labels obtained via the cross-validation procedure that was employed during the final evaluation of the fine-tuned models.

And finally, to evaluate the statistical significance of the prediction performance obtained by the built classification models, a non-parametric statistical method named permutation test was employed [7, 29]. For this purpose, we applied 1 000 permutations with the significance level of 0.05. To estimate the performance of the models on the permuted data, we used the same classification setup as employed during the training phase [27].

3 Results

The results of the exploratory data analysis are summarized in Table 1 (sorted based on the p-value for the Mann-Whitney U-test). The following features were found as the most distinguishing ones in terms of the differentiation between HC and subjects with LBD (the top 4 features are listed; *, **, and *** denote the p-values for both the Mann-Whitney U-test and Spearman's correlation coefficient being bellow the significance level of 0.05, 0.01, and 0.001, respectively; if both p-values are bellow a different significance level, the weaker statistical significance is selected): a) spiral – nCV of acceleration (on-surface) $\rho = -0.2438^{*}$, variability of spiral width $\rho = 0.2439^{*}$, median of azimuth $\rho = 0.2378^{*}$, and spiral precision index $\rho = 0.2367^{*}$; b) sentence – number of pen stops $\rho = 0.3460^{**}$, slope of duration of stroke (in-air) $\rho = 0.2823^{**}$, median of vertical velocity (on-surface) $\rho = -0.2438^{*}$, and median of vertical acceleration (on-surface) $\rho = 0.2317^{*}$; and c) pentagons – width of writing (on-surface) $\rho = -0.3045^{**}$, median of length of stroke (on-surface) $\rho = -0.2894^{**}$, nCV of length of stroke (on-surface) $\rho = 0.2489^{*}$, and median of duration of stroke (on-surface) $\rho = -0.2327^{*}$.

Table 1. Results of the exploratory analysis.

feature	p(U)	ρ	p(ρ)
spiral			
nCV of acceleration (s)	0.0138	-0.2438	0.0263
variability of spiral width	0.0138	0.2439	0.0263
median of azimuth	0.0158	0.2378	0.0304
spiral precision index	0.0162	0.2367	0.0312
nCV of duration of stroke (s)	0.0438	-0.1892	0.0867
sentence			
number of pen stops	0.0009	0.3460	0.0014
slope of duration of stroke (a)	0.0054	0.2823	0.0097
median of vertical velocity (s)	0.0138	-0.2438	0.0263
median of vertical acceleration (s)	0.0182	0.2317	0.0351
rel. total number of intra-stroke intersections	0.0232	-0.2206	0.0451
pentagons			
width of writing (s)	0.0030	-0.3045	0.0051
median of length of stroke (s)	0.0045	-0.2894	0.0080
nCV of length of stroke (s)	0.0123	0.2489	0.0233
median of duration of stroke (s)	0.0178	-0.2327	0.0343
median of horizontal acceleration (s)	0.0182	0.2317	0.0351

[1] p(U) – p-value of Mann-Whitney U-test; ρ – Spearman's correlation coefficient; p(ρ) – p-value of ρ; (s) – on-surface movement; (a) – in-air movement.

Next, Table 2 presents the results of the correlation analysis (*, and ** denote the p-values for Spearman's correlation coefficient being below the significance level of 0.05 and 0.01, respectively) between the features summarized in Table 1 and the following clinical information: a) MDS–UPDRS, and b) CCB domains.

Table 2. Results of the correlation analysis.

feature	ρ (UPDRS)	ρ (V)	ρ (A)	ρ (E)
spiral				
nCV of acceleration (s)	-0.3411*	-0.0013	0.1130	0.1899
variability of spiral width	0.1653	-0.3973**	-0.2981*	-0.1666
median of azimuth	0.0442	-0.3656*	-0.1029	-0.0490
spiral precision index	0.0606	-0.0942	-0.3987**	-0.2126
nCV of duration of stroke (s)	-0.1089	-0.1344	-0.1618	-0.0469
sentence				
num. of pen stops	-0.1018	-0.1181	0.1012	-0.1956
slope of duration of stroke (a)	0.2620	-0.1928	-0.0513	-0.1025
median of vertical velocity (s)	0.0314	0.1106	0.0025	0.1794
median of vertical acceleration (s)	-0.2641	-0.0301	0.3246*	0.0193
rel. total num. of intra-stroke intersections	0.0477	0.1647	0.1143	0.0962
pentagons				
width of writing (s)	-0.3448*	0.2947*	0.1351	0.1362
median of length of stroke (s)	-0.1545	0.1607	0.0501	0.1511
nCV of length of stroke (s)	0.3065*	-0.2435	-0.1126	-0.1155
median of duration of stroke (s)	-0.0348	0.0080	-0.0085	-0.0269
median of horizontal acceleration (s)	0.3215*	-0.0226	-0.1632	-0.2060

[1] ρ – Spearman's correlation coefficient ($*$ denotes p-value < 0.05 and $**$ denotes p-value < 0.01); UPDRS – MDS–Unified Parkinson's Disease Rating Scale, part III (motor part) [16]; V – visuospatial domain of CCB; A – attention domain of CCB; E – executive functions domain of CCB; (s) – on-surface movement; (a) – in-air movement.

To visualize the difference in the distribution of the top 4 features summarized above for HC and subjects with LBD, the box-violin plots are presented in Figures 1–3. The Figure 1 shows the distribution of the features for the spiral drawing, the Figure 2 shows the distribution of the features for the sentence writing, and the Figure 3 is dedicated to the distribution of the features for the pentagon copying test.

The results of the classification analysis are summarized in Table 3. We trained 4 models in total: 3 models dedicated to each task separately and a model combining all of the tasks. The following results were achieved (where $*$ and $**$ denote p-value of the permutation test bellow < 0.05 and < 0.01, respectively): a) spiral – BACC = 0.6848**, SEN = 0.8696, SPE = 0.5000; b) sentence – BACC = 0.7283**, SEN = 0.9783, SPE = 0.4783 c) pentagons – BACC = 0.6848**, SEN = 0.9348, SPE = 0.4348; and d) all tasks combined – BACC = 0.7391**, SEN = 0.8043, SPE = 0.6739. The ROC curves of the trained models are shown in Figure 4.

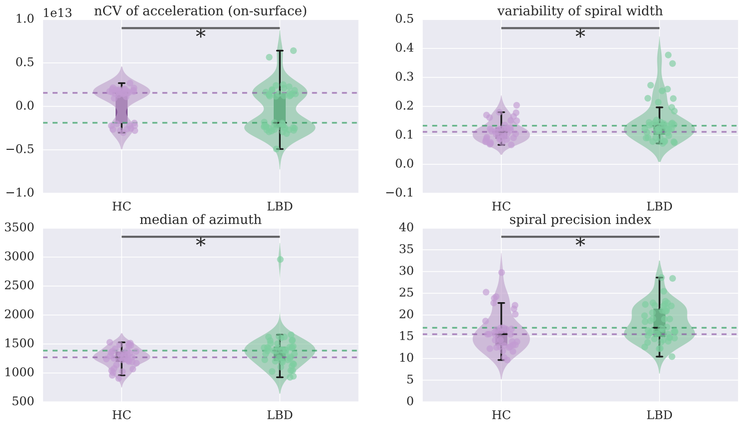


Fig. 1. Distribution of the top 4 most discriminating features (spiral drawing).

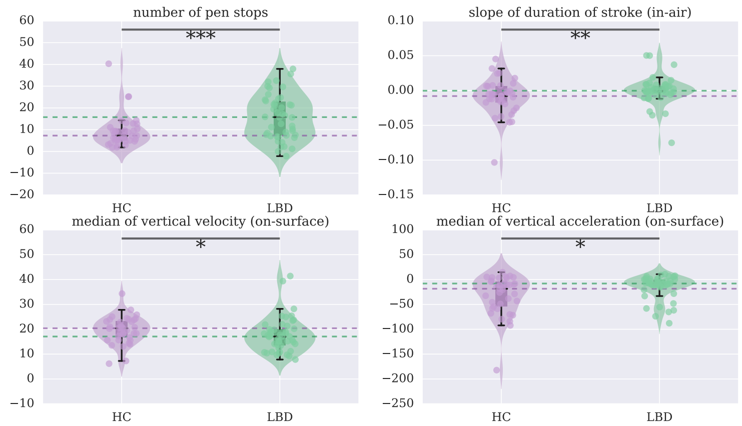


Fig. 2. Distribution of the top 4 most discriminating features (sentence writing).

4 Discussion

As mentioned in the methodology, the Archimedean spiral is considered as a gold standard, especially in the assessment of graphomotor difficulties in PD patients [5, 8, 32], nevertheless, it has been utilised during the quantitative analysis of Huntington's disease, essential tremor, or brachial dystonia as well [13]. Concerning the spiral features with the highest discrimination power (as identified by the Mann-Whitney U-test), we observed that the LBD group was associated

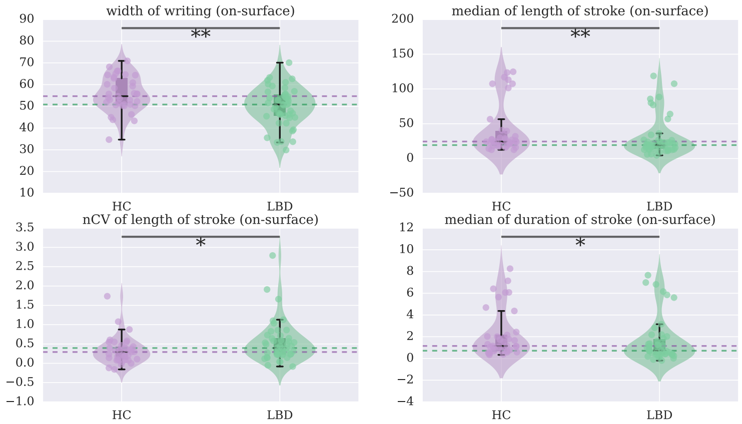


Fig. 3. Distribution of the top 4 most discriminating features (pentagons copying test).

Table 3. Results of the classification analysis.

task	MCC	BACC	SEN	SPE	PRE	F1	threshold	p
spiral	0.3977	0.6848	0.8696	0.5000	0.6349	0.7339	0.26	**
sentence	0.5271	0.7283	0.9783	0.4783	0.6522	0.7826	0.36	**
pentagons	0.4267	0.6848	0.9348	0.4348	0.6232	0.7478	0.13	**
all tasks combined	0.4824	0.7391	0.8043	0.6739	0.7115	0.7551	0.48	**

[1] MCC – Matthew's correlation coefficient; BACC – balanced accuracy; SEN – sensitivity; SPE – specificity; PRE – precision; F1 – F1 score; p – p-values computed by the permutation test (1 000 permutations, $*$ denotes p-value < 0.05 and $**$ denotes p-value < 0.01); threshold – fine-tuned decision threshold.

with a lower range in on-surface acceleration, which we suppose is caused by rigidity. This assumption is supported by the fact that the measure significantly correlates ($\rho = -0.3$, $p < 0.05$) with the overall score of MDS–UPDRS III. Next, the LBD group was not able to keep small variability of loop-to-loop spiral width index, which is in line with findings reported in [32]. We also observed a significant correlation between this feature and the visuospatial ($\rho = -0.4$, $p < 0.01$) and the attention ($\rho = -0.3$, $p < 0.05$) domain of CCB. On the other hand, the LBD group had generally higher values of the spiral precision index than the HC one, which is against our initial assumptions (also the correlation with the attention domain of CCB is surprisingly negative; $\rho = -0.4$, $p < 0.01$). Finally, the last significant correlation with the clinical status was identified in the median of azimuth, which was higher in the LBD group (in addition we observed a negative correlation with the visuospatial domain of CCB; $\rho = -0.4$, $p < 0.05$).

Regarding the classification analysis, based on the spiral features, we were able to discriminate the LBD and HC groups with 68 % balanced accuracy (area

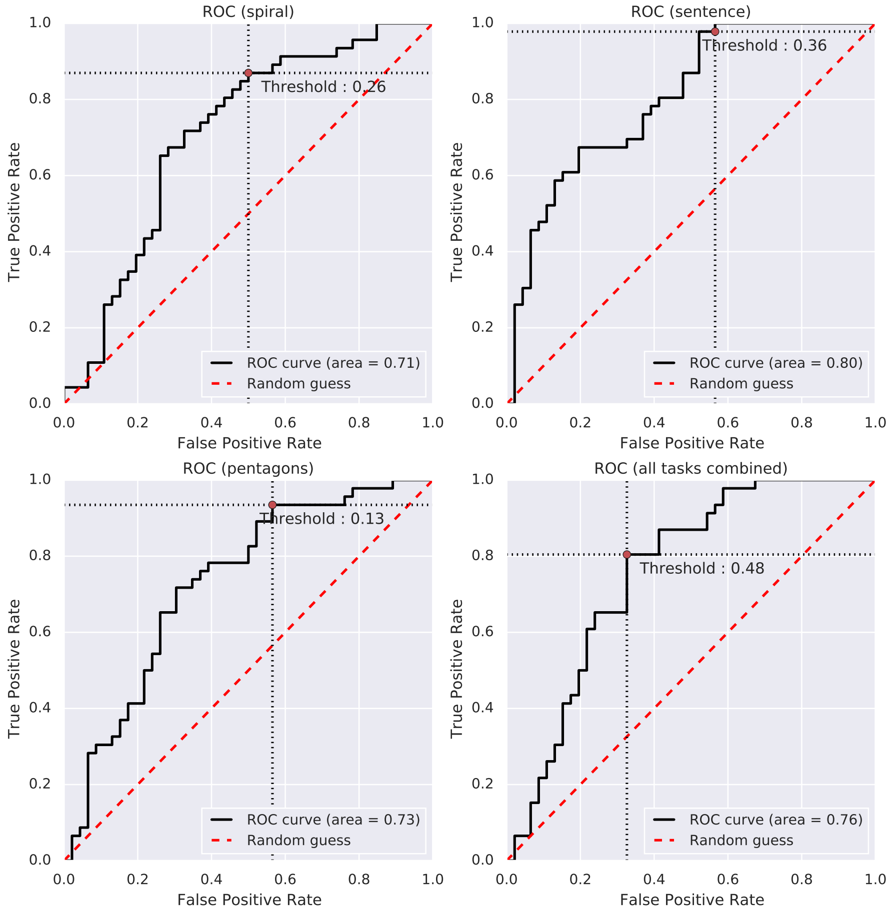


Fig. 4. Receiver operating characteristic curves for the trained models.

under the curve (AUC) = 71 %), which is the worst result when compared to other tasks and which supports our previous findings that even though the spiral is considered as a gold standard the sentence copy task accents the manifestations of dysgraphia much better [11].

Regarding the sentence, the most discriminative feature extracted from this task is the number of pen stops (i.e. a pen is in contact with the paper and does not vary its position for at least 30 ms [8]), which was higher in the LBD group. This parameter has been mainly employed in the diagnosis of developmental dysgraphia in children population [28], however, in one study, Danna et al. observed that this measure (but extracted from the spiral) was significantly different between PD patients in the OFF state and HC [8]. Initially, we assumed that the feature could be theoretically linked with cognitive deficits, but we did not observe any significant correlation with the visuospatial, attention, or executive functions domain of CCB. The second most significant feature was the slope of the duration of in-air strokes. The positive correlation coefficient suggests that the LBD subjects were associated with progressing fatigue [1, 12, 17]. Next, in the LBD group, we observed lower on-surface vertical velocity (this is in line with e.g. [21, 36]), but increased on-surface vertical acceleration. This could be probably explained by the slow and less smooth handwriting. In terms of projection, the reason why these deficits dominate in the vertical movement could be explained by the fact that the finger system (which is mainly involved in the vertical movement) is more affected by muscular fatigue than the wrist system (which controls horizontal movement) [20]. The vertical movement requires coordinated movement and finer flexions/extensions of more joints (interphalangeal and metacarpophalangeal), thus it is more complex than ulnar abductions of the wrist [10, 35] and could more accent the rigidity and bradykinesia. In addition, this manifestation could be associated with the progressive/consistent vertical micrographia, i. e., progressive/consistent reduction in letter amplitude [34].

In terms of classification, by modelling features extracted from the sentence, we were able to differentiate both groups with 73 % balanced accuracy (AUC = 80 %). In comparison with the state of the art in supportive LBD or PD diagnosis [9, 19, 36], it is not a competitive result, but on the other hand, we would like to highlight that we deal with results evaluating diagnosis of LBDs in the prodromal state that has not been targeted by other research teams yet.

Concerning the last (cognitive) task, all the top 5 discriminative features were extracted from the on-surface movement. In our recent article [4] we proved that in-air entropy-based parameters could be used to identify early cognitive deficits in PD without major cognitive impairment and that they correlate with the level of attention. In the current study, these in-air measures were not significant, but on the other hand, their on-surface variants (i.e. median of Shannon entropy calculated from the global/vertical movement) had the p-values of the Mann-Whitney U-test < 0.05, moreover, they significantly correlated with the visuospatial domain of CCB (e.g. $\rho = -0.3$, $p < 0.05$). The top 5 parameters consist of the width of the product, which was smaller in the LBD group. It slightly correlates with the lower median of the length of strokes ($\rho = 0.3$) and lower median of the duration of strokes ($\rho = 0.2$) and probably means that the subjects in the LBD group made the overlapped pentagons smaller. In addition, since the non-parametric coefficient of variation of the length of strokes was higher, we assume that the LBD subjects were not able to keep a stable length of strokes (nevertheless, based on the scoring published in [25], this is assumed as a very small deviation). Regarding the width, we also observed a negative correlation ($\rho = -0.3$, $p < 0.05$) with the overall score of MDS–UPDRS III.

The classification based on the pentagon copying test provided 68 % balanced accuracy (AUC = 0.73 %), which is slightly better than in the case of the spiral, but not as high as in the case of the sentence.

And finally, a machine learning model based on the whole set of features (tasks) enabled us to improve the accuracy to 74 % (AUC = 76 %). This shows that the combination of the graphomotor, handwriting and cognitive deficits can be used to achieve reasonable performance in the prodromal diagnosis of LBDs.

5 Conclusion

This study has several limitations. Our dataset has a small sample size and the HC and LBD groups are imbalanced, therefore to get better results in terms of their generalisation, a bigger database must be analysed. Next, due to the small sample size, we fused subjects with a high risk of developing PD or MCI-LB into one LBD group. Nevertheless, subjects with MCI-LB in its prodromal stage are associated mainly with cognitive (executive or visuospatial) decline, while subjects with prodromal PD experience mainly motor deficits. In other words, we suppose that further stratification of these participants into two groups could increase the classification accuracy (we hypothesise that MCI-LB would be more pronounced in the pentagon copying task and PD in the handwriting one). Finally, although we tried a correction of multiple comparisons during the statistical analysis, almost no significant features appeared after this adjustment. To sum up, concerning the limitations mentioned above, the study should be considered as a pilot one

In conclusion, despite the limitations, to the best of our knowledge, it is the first work exploring the impact of computerised analysis of a graphomotor, cognitive, and handwriting task on the prodromal diagnosis of these neurodegenerative disorders. It bridges the knowledge gap in the field of LBDs, and provides baseline results for future studies focusing on the prodromal diagnosis of LBDs via a computerized and objective analysis of graphomotor and handwriting difficulties.

Bibliography

[1] Aouraghe, I., Alae, A., Ghizlane, K., Mrabti, M., Aboulem, G., Faouzi, B.: A novel approach combining temporal and spectral features of Arabic online handwriting for Parkinson's disease prediction. Journal of Neuroscience Methods **339**, 108727 (2020)

[2] Benedict, H.: Brief Visual Memory Test–revised: Professional manual. Odessa, FL: Psychological Assessment Resources (1997)

[3] Bezdicek, O., Libon, D.J., Stepankova, H., Panenkova, E., Lukavsky, J., Garrett, K.D., Lamar, M., Price, C.C., Kopecek, M.: Development, validity, and normative data study for the 12-word Philadelphia Verbal Learning Test [czP (r) VLT-12] among older and very old Czech adults. The Clinical Neuropsychologist **28**(7), 1162–1181 (2014)

[4] Brabenec, L., Klobusiakova, P., Mekyska, J., Rektorova, I.: Shannon entropy: A novel parameter for quantifying pentagon copying performance in non-demented Parkinson's disease patients. Parkinsonism & related disorders **94**, 45–48 (2022)

[5] Cascarano, G.D., Loconsole, C., Brunetti, A., Lattarulo, A., Buongiorno, D., Losavio, G., Di Sciascio, E., Bevilacqua, V.: Biometric handwriting analysis to support Parkinson's disease assessment and grading. BMC medical informatics and decision making **19**(9), 1–11 (2019)

[6] Chen, T., Guestrin, C.: XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 16. ACM Press (2016). https://doi.org/10.1145/2939672.2939785, https://doi.org/10.1145/2939672.2939785

[7] Combrisson, E., Jerbi, K.: Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J. Neurosci. Methods **250**, 126–136 (jul 2015). https://doi.org/10.1016/j.jneumeth.2015.01.010

[8] Danna, J., Velay, J.L., Eusebio, A., Véron-Delor, L., Witjas, T., Azulay, J.P., Pinto, S.: Digitalized spiral drawing in Parkinson's disease: A tool for evaluating beyond the written trace. Human movement science **65**, 80–88 (2019)

[9] De Stefano, C., Fontanella, F., Impedovo, D., Pirlo, G., di Freca, A.S.: Handwriting analysis to support neurodegenerative diseases diagnosis: A review. Pattern Recognition Letters **121**, 37–45 (2019)

[10] Dounskaia, N., Van Gemmert, A., Stelmach, G.: Interjoint coordination during handwriting-like movements. Experimental Brain Research **135**(1), 127–140 (2000)

[11] Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., Faundez-Zanuy, M.: Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. Artificial intelligence in Medicine **67**, 39–46 (2016)

[12] Drotár, P., Mekyska, J., Smékal, Z., Rektorová, I., Masarová, L., Faundez-Zanuy, M.: Prediction potential of different handwriting tasks for diagnosis of Parkinson's. In: 2013 E-Health and Bioengineering Conference (EHB). pp. 1–4. IEEE (2013)

[13] Faundez-Zanuy, M., Mekyska, J., Impedovo, D.: Online handwriting, signature and touch dynamics: Tasks and potential applications in the field of security and health. Cognitive Computation **13**(5), 1406–1421 (2021)

[14] Galaz, Z., Mucha, J., Zvoncak, V., Mekyska, J.: Handwriting features. https://github.com/BDALab/handwriting-features (2022)

[15] Garre-Olmo, J., Faundez-Zanuy, M., López-de Ipiña, K., Calvó-Perxas, L., Turró-Garriga, O.: Kinematic and pressure features of handwriting and drawing: preliminary results between patients with mild cognitive impairment, alzheimer disease and healthy controls. Current Alzheimer research **14**(9), 960–968 (2017)

[16] Goetz, C.G., Tilley, B.C., Shaftman, S.R., Stebbins, G.T., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M.B., Dodel, R., et al.: Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. Movement disorders: official journal of the Movement Disorder Society **23**(15), 2129–2170 (2008)

[17] Harralson, H.H., Teulings, H.L., Farley, B.G.: Handwriting variability in movement disorder patients and effects of fatigue. In: of the fourteenth biennial conference of the International Graphonomics Society. pp. 103–107 (2009)

[18] Heinzel, S., Berg, D., Gasser, T., Chen, H., Yao, C., Postuma, R.B., on the Definition of Parkinson's Disease, M.T.F.: Update of the MDS research criteria for prodromal Parkinson's disease. Movement Disorders **34**(10), 1464–1470 (2019)

[19] Impedovo, D., Pirlo, G.: Dynamic handwriting analysis for the assessment of neurodegenerative diseases: a pattern recognition perspective. IEEE reviews in biomedical engineering **12**, 209–220 (2018)

[20] Kushki, A., Schwellnus, H., Ilyas, F., Chau, T.: Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia. Research in Developmental Disabilities **32**(3), 1058–1064 (may 2011)

[21] Letanneux, A., Danna, J., Velay, J.L., Viallet, F., Pinto, S.: From micrographia to Parkinson's disease dysgraphia. Movement Disorders **29**(12), 1467–1475 (2014)

[22] Matthews, B.W.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta (BBA) **405**(2), 442–51 (1975)

[23] McKeith, I.G., Ferman, T.J., Thomas, A.J., Blanc, F., Boeve, B.F., Fujishiro, H., Kantarci, K., Muscio, C., O'Brien, J.T., Postuma, R.B., et al.: Research criteria for the diagnosis of prodromal dementia with Lewy bodies. Neurology **94**(17), 743–755 (2020)

[24] Monvoisin-Joly, T., Furcieri, E., Chabran, E., Blanc, F.: Writing in prodromal and mild dementia with Lewy bodies: an exploratory and preliminary study. Geriatrie et Psychologie Neuropsychiatrie du Vieillissement **19**(3), 341–351 (2021)

[25] Nagaratnam, N., Nagaratnam, K., O'Mara, D.: Intersecting pentagon copying and clock drawing test in mild and moderate Alzheimer's disease. Journal of Clinical Gerontology and Geriatrics **5**(2), 47–52 (2014)

[26] Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., Chertkow, H.: The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society **53**(4), 695–699 (2005)

[27] Ojala, M., Garriga, G.: Permutation tests for studying classifier performance. In: Data Mining, 2009. ICDM '09. Ninth IEEE International Conference on. pp. 908–913 (Dec 2009). https://doi.org/10.1109/ICDM.2009.108

[28] Paz-Villagrán, V., Danna, J., Velay, J.L.: Lifts and stops in proficient and dysgraphic handwriting. Human movement science **33**, 381–394 (2014)

[29] Phipson, B., Smyth, G.K.: Permutation p-values should never be zero: Calculating exact p-values when permutations are randomly drawn. Stat. Appl. Genet. Mol. Biol. **9**(1) (jan 2010). https://doi.org/10.2202/1544-6115.1585

[30] Postuma, R.B., Berg, D., Stern, M., Poewe, W., Olanow, C.W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A.E., et al.: MDS clinical diagnostic criteria for Parkinson's disease. Movement disorders **30**(12), 1591–1601 (2015)

[31] Preiss, M., Kalivodová, Z., Kundrátová, I., Mrlinová, L., Ježková, T., Kubů, M., Houbová, P.: Test verbální fluence–vodítka pro všeobecnou dospělou populaci. Psychiatrie **6**(2), 74–77 (2002)

[32] San Luciano, M., Wang, C., Ortega, R.A., Yu, Q., Boschung, S., Soto-Valencia, J., Bressman, S.B., Lipton, R.B., Pullman, S., Saunders-Pullman, R.: Digitized spiral drawing: a possible biomarker for early Parkinson's disease. PloS one **11**(10), e0162799 (2016)

[33] Saunders-Pullman, R., Derby, C., Stanley, K., Floyd, A., Bressman, S., Lipton, R.B., Deligtisch, A., Severt, L., Yu, Q., Kurtis, M., et al.: Validity of spiral analysis in early Parkinson's disease. Movement disorders: official journal of the Movement Disorder Society **23**(4), 531–537 (2008)

[34] Thomas, M., Lenka, A., Kumar Pal, P.: Handwriting analysis in Parkinson's disease: current status and future directions. Movement disorders clinical practice **4**(6), 806–818 (2017)

[35] Van Galen, G.P.: Handwriting: Issues for a psychomotor theory. Human movement science **10**(2-3), 165–191 (1991)

[36] Vessio, G.: Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review. Applied Sciences **9**(21), 4666 (2019)

[37] Warrington, E.K., James, M.: The visual object and space perception battery (1991)

[38] Wechsler, D., et al.: WAIS-III WMS-III Technical Manual. Psychological Corporation San Antonio, TX (1997)

[39] Yamada, M., Komatsu, J., Nakamura, K., Sakai, K., Samuraki-Yokohama, M., Nakajima, K., Yoshita, M.: Diagnostic criteria for dementia with Lewy bodies: updates and future directions. Journal of Movement Disorders **13**(1), 1 (2020)

A.3 Exploration of Various Fractional Order Derivatives in Parkinson's Disease Dysgraphi Analysis

Exploration of Various Fractional Order Derivatives in Parkinson's Disease Dysgraphia Analysis *

Jan Mucha[1] [0000-0001-5126-440X], Zoltan Galaz[1] [0000-0002-8978-351X], Jiri Mekyska[1] [0000-0002-6195-193X], Marcos Faundez-Zanuy[2] [0000-0003-0605-1282], Vojtech Zvoncak[1] [0000-0002-1948-4653], Zdenek Smekal[1] [0000-0002-8483-5448], Lubos Brabenec[3] [0000-0002-8348-5757], and Irena Rektorova[3,4] [0000-0002-5455-4573]

[1] Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic
[2] Escola Superior Politecnica, Tecnocampus, Mataro, Barcelona, Spain
[3] Applied Neuroscience Research Group, Central European Institute of Technology – CEITEC, Masaryk University, Brno, Czech Republic
[4] First Department of Neurology, Faculty of Medicine and St. Anne's University Hospital, Masaryk University, Brno, Czech Republic

Abstract. Parkinson's disease (PD) is a common neurodegenerative disorder with a prevalence rate estimated to 2.0 % for people aged over 65 years. Cardinal motor symptoms of PD such as rigidity and bradykinesia affect the muscles involved in the handwriting process resulting in handwriting abnormalities called PD dysgraphia. Nowadays, online handwritten signal (signal with temporal information) acquired by the digitizing tablets is the most advanced approach of graphomotor difficulties analysis. Although the basic kinematic features were proved to effectively quantify the symptoms of PD dysgraphia, a recent research identified that the theory of fractional calculus can be used to improve the graphomotor difficulties analysis. Therefore, in this study, we follow up on our previous research, and we aim to explore the utilization of various approaches of fractional order derivative (FD) in the analysis of PD dysgraphia. For this purpose, we used the repetitive loops task from the Parkinson's disease handwriting database (PaHaW). Handwritten signals were parametrized by the kinematic features employing three FD approximations: Grünwald-Letnikov's, Riemann–Liouville's, and Caputo's. Results of the correlation analysis revealed a significant relationship between the clinical state and the handwriting features based on the velocity. The extracted features by Caputo's FD approximation outperformed the rest of the analyzed FD approaches. This was also confirmed by the results of the classification analysis, where the best model trained

* This work was supported by grant no. NU20-04-00294 (Diagnostics of Lewy body diseases in prodromal stage based on multimodal data analysis) of the Czech Ministry of Health and by Spanish grant of the Ministerio de Ciencia e Innovación no. PID2020-113242RB-I00.

by Caputo's handwriting features resulted in a balanced accuracy of 79.73 % with a sensitivity of 83.78 % and a specificity of 75.68 %.

Keywords: Fractional order derivatives · fractional calculus · Parkinson's disease · online handwriting · handwriting difficulties.

1 Introduction

Fractional calculus (FC) is a name of the theory of integrals and derivatives of an arbitrary order [28]. It has been developed simultaneously with the well-known differential calculus [16] and its principles have been successfully used in modern engineering and science in general [18, 32, 37]. The advances of FC have been employed in the modeling of different diseases as well, like the human immunodeficiency virus (HIV) [2] or malaria [27]. In addition, the FC has been widely utilized in several computer vision disciplines such as the super-resolution, motion estimation, image restoration or image segmentation [34]. Furthermore, in our recent research we developed new handwriting features extraction techniques based on the application of the fractional order derivatives (FD) [11, 21–25].

Parkinson's disease (PD) is a chronic idiopathic disorder, with the prevalence rate estimated to be approximately 2.0 % for people aged over 65 years [12]. It is characterized by the progressive loss of dopaminergic neurons in the *substancia nigra pars compacta* [6, 13], which is a major cause of the symptoms linked with the PD. Primary PD motor symptoms are tremor at rest, muscular rigidity, progressive bradykinesia, and postural instability [3, 14]. One of the essential motor symptoms of PD is PD dysgraphia [17, 36]. Additionally, a variety of non-motor symptoms such as cognitive impairment, sleep disturbances, depression, etc. may arise.

PD dysgraphia includes a spectrum of neuromuscular difficulties like motor-memory dysfunction, motor feedback difficulties, graphomotor production deficits and others [17, 31]. These disabilities leads to a variety of handwriting difficulties manifesting as dysfluent, shaky, slow, and less readable handwriting. The most commonly observed handwriting abnormality in PD patients is micrographia. Micrographia represents the progressive decrease of letter's amplitude or width [20]. Some PD patients never develop micrographia, but they still exhibit other handwriting difficulties. Accordingly, the consequences of PD dysgraphia significantly affect a person's quality of life. Starting with slow and less legible handwriting and often progressing to lower self-esteem, poor emotional well-being, problematic communication and social interaction, and many others. Nowadays, the most advanced approaches of the PD manifestations quantification contained in the handwriting are based on digitizing tablets [9, 21, 35]. These devices can acquire x and y trajectories along with temporal information, therefore the temporal, kinematic, or dynamic characteristics can be processed together with the spatial features. Handwritten signal acquired by the digitizing tablet is called online handwriting.

In the past decades, researchers have been exploring the effect of several handwriting/drawing tasks in PD dysgraphia analysis, including the simplest ones (loops, circles, lines, Archimedean spiral) together with more complex ones (words, sentences, drawings, etc.) [7–9, 21–23, 26]. Drotar et al. [7–9] reported classification accuracy up to 89 % using a combination of kinematic, pressure, energy or empirical mode decomposition features. The diagnosis of PD with accuracy of 71.95 % based on the kinematic and entropy features extracted from the sentence task was reported by Impedovo et al. [15]. Taleb et al. [35] reported up to 94 % accuracy of PD severity prediction using kinematic and pressure features in combination with adaptive synthetic sampling approach (ADASYN) for model training. Rios-Urrego et al. [30] achieved classification accuracy of 83.3 % using the kinematic, geometric, spectral and nonlinear dynamic features. New kinematic features utilizing the discrete time wavelet transform, the fast Fourier transform and a Butter/adaptive filter introduced by Aouraghe et al. [1] resulted in classification accuracy of 92.2 %.

Finally, in our recent works [21–23, 25] we introduced and evaluated a new advanced approach of PD dysgraphia analysis employing the FD as a substitution of the conventional differential derivative during the basic kinematic feature extraction. Newly designed handwriting features achieved classification accuracy up to 90 %, using the Grünwald-Letnikov approach only. In addition to PD dysgraphia analysis, we explored the FD-based handwriting features in analysis of graphomotor difficulties in school-aged children, where we examined three different FD approaches [24]. The results suggests that the employment of various FD approximations brings major differences in kinematic handwriting features. Therefore, as a next logical step, this study aims to:

1. extend our previous research in PD dysgraphia analysis by the utilization of various FD approaches,
2. explore the differences of various FD approaches in the analysis of PD dysgraphia,
3. compare the power of the FD-based handwriting features extracted by several FD approximations to distinguish between the PD patients and healthy controls (HC).

2 Materials and methods

2.1 Dataset

For the purpose of this study, we used the Parkinson's disease handwriting database (PaHaW) [7]. The database consists of several handwriting or drawing tasks acquired in 37 PD patients and 38 healthy controls (HC). The participants were enrolled at the First Department of Neurology, St. Anne's University Hospital in Brno, Czech Republic. All participants reported Czech language as their native language and they were right-handed. The patients completed their tasks approximately 1 hour after their regular dopaminergic medication (L-dopa). All participants signed an informed consent form approved by the local ethics committee. Demographic and clinical data of the participants involved in this study can be found in Table 1. For the purpose of this study, we selected the repetitive loop handwriting task. This task is missing for several participants of the PaHaW dataset, therefore, we processed 31 PD patients and 37 HC only.

Table 1. Demographic and clinical data of the participants.

Gender	N	Age [y]	PD dur [y]	UPDRS V	LED [mg/day]
Parkinson's disease patients					
Females	15	70.2 ± 8.4	7.9 ± 3.9	1.9 ± 0.4	1129.7 ± 572.9
Males	16	65.9 ± 13.1	7.0 ± 3.9	2.4 ± 0.9	1805.7 ± 743.3
All	31	68.0 ± 11.1	7.4 ± 3.9	2.2 ± 0.8	1478.6 ± 739.8
Healthy controls					
Females	17	61.6 ± 10.2	-	-	-
Males	20	63.3 ± 12.5	-	-	-
All	37	62.9 ± 11.5	-	-	-

[1] N – number of subjects; y – years; PD dur – PD duration; UPDRS V – Unified Parkinson's disease rating scale, part V: Modified Hoehn & Yahr staging score [10]; LED – L-dopa equivalent daily dose.

2.2 Data Acquisition

The PaHaW database [7] consists of nine handwriting tasks. For the purpose of this study we selected the repetitive loop task only. An example of the repetitive loop task for a PD patient and a HC can be seen in Figure 1. During the acquisition of the handwriting tasks, the participants were rested and seated in a comfortable position with a possibility to look at a pre-filled template. In case of some mistakes, they were allowed to repeat the task. A digitizing tablet (Wacom Intuos 4M) was overlaid with an empty paper and the participants wrote on that using the Wacom Inking pen. Online handwriting signals were recorded with $f_{\mathrm{s}} = 150$ Hz sampling rate, and the following time sequences were acquired: x and y coordinates ($x[t]$, $y[t]$); time-stamp (t); on-surface and in-air movement status ($b[t]$); pressure ($p[t]$); azimuth ($az[t]$); and tilt (also called altitude; $al[t]$).

2.3 Fractional Order Derivative

The main subject of this study is the exploration of the various FD approximations as a substitution of the conventional differential derivatives in the handwriting feature extraction process. We utilized three different FD approximations, namely: Grünwald-Letnikov (GL), Riemann–Liouville (RL), and Caputo (C), implemented by Valério Duarte in Matlab [38–40].

First approach employed in this study was developed by Grünwald and Letnikov. A direct definition of the derivation of the function $y(t)$ by the order α – $D^{\alpha}y(t)$ [28] is based on the finite differences of an equidistant grid in $[0, \tau]$,

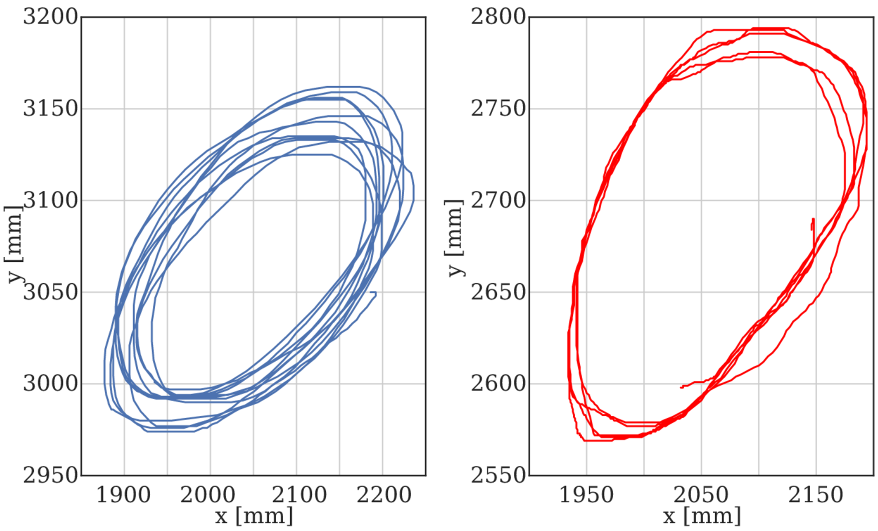


Fig. 1. Example of the repetitive loop task for a HC (left) and a PD patient (right).

assuming that the function $y(t)$ satisfies certain smoothness conditions in every finite interval $(0, t), t \leq T$, where T denotes the period. Choosing the grid

$$0 = \tau_0 < \tau_1 < ... < \tau_{n+1} = t = (n+1)h, \tag{1}$$

with

$$\tau_{k+1} - \tau_k = h, \tag{2}$$

and using the notation of finite differences

$$\frac{1}{h^\alpha}\Delta_h^\alpha y(t) = \frac{1}{h^\alpha}\left(y(\tau_{n+1}) - \sum_{v=1}^{n+1} c_v^\alpha y(\tau_{n+1-v})\right), \tag{3}$$

where

$$c_v^\alpha = (-1)^{v-1}\binom{\alpha}{v}. \tag{4}$$

The Grünwald–Letnikov definition from 1867 is defined as

$$^{GL}\mathrm{D}^\alpha y(t) = \lim_{h\to 0}\frac{1}{h^\alpha}\Delta_h^\alpha y(t), \tag{5}$$

where $^{GL}\mathrm{D}^\alpha y(t)$ denotes the Grünwald-Letnikov derivatives of order α of the function $y(t)$, and h represents the sampling lattice.

Second approach used in this study has been given by Riemann-Liouville. The left-inverse interpretation of $D^\alpha y(t)$ by Riemann-Liouville [28, 18] from 1869 is defined as

$$^{RL}\mathrm{D}^\alpha y(t) = \frac{1}{\Gamma(n-\alpha)}\left(\frac{d}{dt}\right)^n \int_0^t (t-\tau)^{n-\alpha-1} y(t)\ \mathrm{d}t, \tag{6}$$

where ${}^{RL}\mathrm{D}^{\alpha}y(t)$ denotes the Riemann–Liouville derivatives of order α of the function $y(t)$, Γ is the gamma function and $n-1 < \alpha \leq n, n \in \mathbf{N}, t > 0$.

Third and last FD approach involved in this study was developed by M. Caputo [4]. In contrast to the previous ones, the improvement hereabouts lies in the unnecessity to define the initial FD condition [18, 28]. The Caputo's definition from 1967 is

$$ {}^{C}\mathrm{D}^{\alpha}y(t) = \frac{1}{\Gamma(n-\alpha)} \int\limits_{0}^{t} (t-\tau)^{n-\alpha-1} y^{n}(t)\ \mathrm{d}t, \tag{7} $$

where ${}^{C}\mathrm{D}^{\alpha}y(t)$ denotes the Caputo derivatives of order α of the function $y(t)$, Γ is the gamma function and $n-1 < \alpha \leq n, n \in \mathbf{N}, t > 0$.

2.4 Feature extraction

Considering the nature of the selected task, on-surface handwriting features were extracted only. Since we did employ three FD approaches in the feature extraction process, three sets of the handwriting features were created. Digitizing tablet rarely omits 3–4 samples during the acquisition, therefore the in-signal outliers removal was performed (outliers were considered as elements more than three scaled median absolute deviations from the median). If not pre-processed, the differentiation of this gap would leave significant peaks in the output handwriting feature. All handwriting features were computed for α in the range of 0.1–1.0 (with the step of 0.1), where $\alpha = 1.0$ is equal to the full derivation. Furthermore, the statistical properties of all extracted handwriting features were described by the mean and the relative standard deviation (relstd). To sum up, each feature set consists of 180 computed kinematic features.

2.5 Statistical analysis and machine learning

Firstly, the normality test of the handwriting features using the Shapiro-Wilk test was performed [33]. Since most of the features were found to come from normal distribution, we did not apply any normalization on a feature basis. To control for the effect of confounding factors (also known as covariates), we controlled for the effect of age and gender of the subjects.

Next, Spearman's (ρ) and Pearson's (r) correlation coefficient with the significance level of 0.05 were computed to assess the strength of the monotonous and linear relationship between the handwriting features and the subject's clinical status (PD/HC). Finally, to control for the issue of multiple comparisons, p-values were adjusted using the False Discovery Rate (FDR) method.

Consequently, binary classification models were built in order to distinguish between the PD patients and HC utilizing the extracted handwriting features. An ensemble extreme gradient boosting algorithm known as XGBoost [5] (with 100 estimators) was used for this purpose. The XGBoost algorithm was selected due to its ability to find complex interactions among features as well as the possibility of ranking their importance and its robustness to outliers. Hyper-parameter space optimization (1000 iteration) by the randomized search strategy (stratified 5-fold cross-validation with 10 repetitions) was performed to optimize balanced accuracy. The set of hyper-parameters that were optimized can be found in the following table (Table 2).

Table 2. Hyper-parameters set.

hyper-parameter	values
learning rate	[0.001, 0.01, 0.1, 0.2, 0.3]
gamma	[0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.5]
maximum tree depth	[6, 8, 10, 12, 15]
subsample ratio	[0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
columns subsample ratio at each level	[0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
columns subsample ratio for each tree	[0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
balance between positive and negative weights	[1, 2, 3, 4]
minimum weights required in a child node	[0.5, 1.0, 3.0, 5.0, 7.0, 10.0]

The classification performance was evaluated by the following classification metrics: Matthew's correlation coefficient [19] (MCC), balanced accuracy (BACC), sensitivity (SEN) also known as recall (REC), specificity (SPE), precision (PRE) and F1 score (F1). These metrics are defined as follows:

$$\mathrm{MCC} = \frac{TP \times TN + FP \times FN}{\sqrt{N}}, \tag{8}$$

$$\mathrm{BACC} = \frac{1}{2}\left(\frac{TP}{TP + FN}\frac{TN}{TN + FP}\right), \tag{9}$$

$$\mathrm{SPE} = \frac{TN}{TN + FP}, \tag{10}$$

$$\mathrm{PRE} = \frac{TP}{TP + FP}, \tag{11}$$

$$\mathrm{REC} = \frac{TP}{TP + FN}, \tag{12}$$

$$\mathrm{F1} = 2\frac{PRE \times REC}{PRE + REC} \tag{13}$$

where $N = (TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)$, TP (true positive) and FP (false positive) represent the number of correctly identified PD patient and the number of subjects incorrectly identified as PD patient, respectively. Similarly, TN (true negative) and FN (false negative) represent the number of correctly identified HC and the number of subjects with PD incorrectly identified as being healthy.

For a better illustration, the overview of the performed analysis from the handwriting task selection to the evaluation of the results can be found in Figure 2.

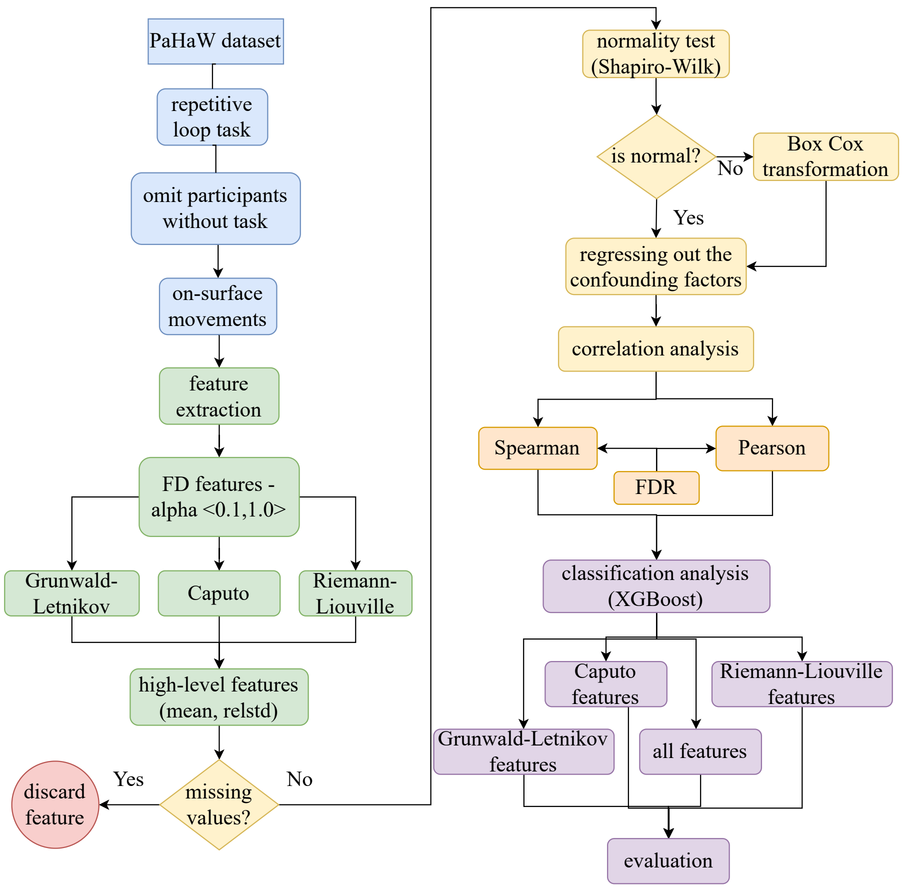


Fig. 2. Flow overview of the performed experiments.

3 Results

The results of the correlation analysis can be seen in Table 3, where the top 5 features per FD approximation according to the p-values of Spearman's correlation are shown. The most significant correlation (after the FDR adjustment) with the clinical state (PD/HC) of the participants was identified in the features extracted by the Caputo's FD approach. Nevertheless, all FD approaches provided the handwriting features that pass the selected significance level ($p < 0.05$), while features extracted by Caputo's and Riemann-Liouville's achieved the p-values very close to 0. Most of the top selected handwriting features are based on horizontal velocity, and all of them have α different from 1, which confirms the positive impact of the FD in PD dysgraphia analysis.

Table 3. Results of the correlation analysis between the subjects' clinical status (PD/HC) and the computed handwriting features ranked by the adjusted p-value (and the correlation coefficient) of Spearman's correlation.

feature name	ρ	p_s	p_s^*	r	p_p	p_p^*
Caputo						
relstd horizontal velocity-α=0.6	-0.5408	0.0001	0.0001	-0.5456	0.0001	0.0001
relstd horizontal velocity-α=0.5	-0.5122	0.0001	0.0001	-0.5204	0.0001	0.0001
relstd horizontal velocity-α=0.4	-0.4912	0.0001	0.0001	-0.5024	0.0001	0.0001
mean horizontal velocity-α=0.3	0.4791	0.0001	0.0001	0.4049	0.0006	0.0051
mean horizontal velocity-α=0.4	0.4716	0.0001	0.0001	0.4240	0.0003	0.0036
Grünwald-Letnikov						
relstd horizontal velocity-α=0.8	-0.4475	0.0001	0.0180	-0.4332	0.0002	0.0240
relstd horizontal velocity-α=0.9	-0.4310	0.0002	0.0180	-0.4184	0.0004	0.0240
relstd horizontal velocity-α=0.7	-0.4220	0.0003	0.0180	-0.4162	0.0004	0.0240
relstd horizontal velocity-α=0.6	-0.3964	0.0008	0.0324	-0.3682	0.0020	0.0720
relstd vertical velocity-α=0.9	-0.3949	0.0009	0.0324	-0.3801	0.0014	0.0630
Riemann-Liouville						
mean horizontal velocity-α=0.2	0.4882	0.0001	0.0001	0.3869	0.0011	0.0060
relstd horizontal velocity-α=0.2	-0.4716	0.0001	0.0001	-0.4643	0.0001	0.0013
mean horizontal velocity-α=0.3	0.4716	0.0001	0.0001	0.4240	0.0003	0.0022
relstd vertical velocity-α=0.2	-0.4686	0.0001	0.0008	-0.4654	0.0001	0.0013
relstd vertical velocity-α=0.3	-0.4475	0.0001	0.0008	-0.4483	0.0001	0.0013

[1] ρ – Spearman's correlation coefficient; p_s – p-value of Spearman's correlation; p_s^* – adjusted p-value of Spearman's correlation; r – Pearson's correlation coefficient; p_p – p-value of Pearson's correlation; p_p^* – adjusted p-value of Pearson's correlation; relstd – relative standard deviation; h. – horizontal; v. – vertical.

The results of the classification analysis are summarized in Table 4. In total, 4 models were trained: one model per each FD approach and one model combining all the features. The best classification performance was achieved by the Caputo's FD approach with BACC = 0.7973, SEN = 0.8378, SPE = 0.7568, PRE = 0.7750 and F1 = 0.8052. However, the highest SEN and SPE were achieved by the Riemann-Liouville approach (SPE = 0.8378, PRE = 0.8065).

Next, in Figure 3 the comparison of the horizontal velocity function for α = 0.6 across all of the utilized FD approximations is visualized. The handwriting features were extracted from the performance of the PD patient with high PD severity. And finally, an example of the dependency of the mean of horizontal velocity on the FD order α for all three FD approaches is shown in Figure 4.

4 Discussion

The main goal of this study is to explore various FD approximations and their differences in the analysis of the PD dysgraphia by online handwriting. For better illustration and more understanding of the differences as well as the common characteristics, the comparison of the identical handwriting feature extracted

Table 4. Results of the classification analysis.

FD approach	MCC	BACC	SEN	SPE	PRE	F1
C	**0.5966**	**0.7973**	**0.8378**	0.7568	0.7750	**0.8052**
RL	0.5204	0.7568	0.6757	**0.8378**	**0.8065**	0.7353
GL	0.4867	0.7432	0.7297	0.7568	0.7500	0.7397
ALL	0.5135	0.7568	0.7568	0.7568	0.7568	0.7568

[1] MCC – Matthew's correlation coefficient; BACC – balanced accuracy; SEN – sensitivity; SPE – specificity; PRE – precision; F1 – F1 score; GL – Grünwald-Letnikov; C – Caputo; RL – Riemann–Liouville; ALL (combination of all feature-types, i. e. 540 features).

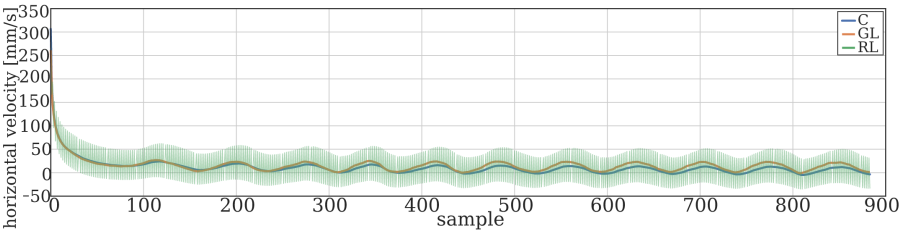


Fig. 3. Comparison of the horizontal velocity function (α=0.6) across all of the FD approximations (PD patient; C – Caputo; GL – Grünwald-Letnikov; RL – Riemann–Liouville).

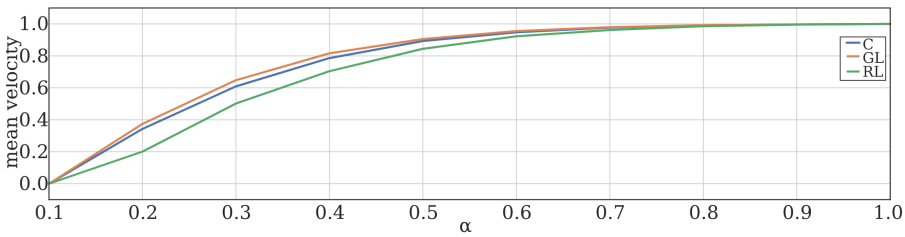


Fig. 4. Mean of horizontal velocity depending on FD order α (PD patient; C – Caputo; GL – Grünwald-Letnikov; RL – Riemann–Liouville).

for all three FD approaches can be found in Figure 3. The feature is extracted from the handwritten product of a PD patient and the feature represents the horizontal velocity for $\alpha = 0.6$. The velocity function extracted by the Riemann-Liouville's approximation dominates by its oscillatory nature in comparison to the other two approaches. Nevertheless the envelope of Riemann-Liouville's approach follows the local maximums and minimums of the functions computed by the Caputo's and Grünwald-Letnikov's approximation. A minor shift of the velocity function can be noticed between the Caputo's and Grünwald-Letnikov's approaches. This is due to the nature of the Caputo's FD approach, which differentiates input data before the convolution operation, so the temporal memory is applied to the velocity afterwards. Regarding the visualization in Figure 3, we can confirm the differences in the same handwriting feature extracted by various FD approximations. Additionally, the dependency comparison of the mean of horizontal velocity on the order α is provided in Figure 4. The oscillatory behaviour of the Riemann-Liouville's function results in the wider gap from the Caputo's and Grünwald-Letnikov's functions. Nevertheless, all three FD approaches converge to the same point as the order α is closer to 1.0. This behaviour is expected, because the full derivation has to be the same for all approaches.

Regarding the results of the correlation analysis, the most significantly correlated handwriting features (after the FDR adjustment) were extracted by the Caputo's FD. This observation is in line with our previous results [24], where we analysed the same three FD approaches in assessment of the graphomotor difficulties in school-aged children. The performance of the handwriting features extracted by the Riemann–Liouville's approach is almost as good as the Caputo's features. The Grünwald-Letnikov's handwriting features achieved weaker relationship, however the features are still below selected level of significance ($p < 0.05$). Most significantly correlated handwriting features are related to the horizontal velocity. In general, PD dysgraphia is linked with the reduced velocity, which could occur even more often than micrographia [15, 29, 31]. This strong relationship is reasonable due to the cardinal symptoms of PD, such as bradykinesia or rigidity, which have a significant impact on fine motor skills, including handwriting/drawing. Moreover, some studies suggest that the horizontal version of micrographia is even more common than the vertical version [36]. The values of the correlation coefficients for handwriting features described by the mean are positive, which means that the performance of the participant is worse with the higher values of the horizontal velocity. This can be confusing because just the opposite effect may be expected. However, this may be specific for the repetitive loop task, where the velocity for the healthy writer is more constant. On the other hand, the writer with PD dysgraphia performs the loop more jerkily, which leads to higher velocity with more variability. This is confirmed by the fact that the features described by the relative standard deviation are negative, which means that the handwriting performance is better with the lower variability of the horizontal velocity.

Based on the results of the classification analysis, the best classification performance was obtained by the handwriting features computed by Caputo's FD. The resulting balanced accuracy was 79.73 % with SEN = 83.78 % and SPE = 75.68 %. In our similar study [21] we achieved classification accuracy of 80.60 % with SEN = 79.4 % and SPE = 80.56 % using all of the handwriting tasks from the PaHaW database, but only the Grünwald-Letnikov FD was employed. In comparison to this study, we can conclude that the exploration of the various FD approaches improved the classification analysis, considering that we achieved almost the same performance only by one handwriting task and using the on-surface kinematic features only. The balanced accuracy of the Riemann-Liouville and Grünwald-Letnikov FD is approximately 5 % lower while the sensitivity is lower up to 15 % in comparison to the Caputo's FD. Considering the reported results, we can conclude that the Caputo's approach is the most suitable FD approximation of the kinematic analysis of the PD dysgraphia by online handwriting.

5 Conclusion

To the best of our knowledge, this is one of the first studies performing an investigation of the various FD approaches in the computerized analysis of the PD dysgraphia by online handwriting. For that reason, the outcomes should be considered as being rather exploratory and pilot in nature. Based on the reported results, Caputo's FD approximation outperformed the rest of the analysed FD approaches in all experiments. The correlation analysis resulted in the significant relationship between the clinical state and the handwriting features based on the velocity, which is in line with our previous findings. Additionally, the best classification model achieved the balanced accuracy of 79.73 % with SEN = 83.78 % and SPE = 75.68 %, which is a comparable result to our previous studies.

This study has several limitations and possible parts, that could be further improved. The processed dataset is relatively small in terms of the statistical validity of the achieved results. Next, the α order should be explored more sensitively (e. g. with a step of 0.01 or even less) in order to identify the optimal range for PD dysgraphia analysis. Additionally, other feature types, such as temporal, spatial, and dynamic, should be included in future comparisons. Moreover, the comparison of the various FD-based features with the conventionally used handwriting features should be performed. Besides, all handwriting tasks included in the PaHaW database have to be investigated by the various FD approaches. And finally, various machine learning models should be trained and compared in future studies.

References

1. Aouraghe, I., Alae, A., Ghizlane, K., Mrabti, M., Aboulem, G., Faouzi, B.: A novel approach combining temporal and spectral features of arabic online handwriting for parkinson's disease prediction. Journal of Neuroscience Methods **339**, 108727 (Jun 2020). https://doi.org/10.1016/j.jneumeth.2020.108727, https://doi.org/10.1016/j.jneumeth.2020.108727
2. Arshad, S., Baleanu, D., Bu, W., Tang, Y.: Effects of hiv infection on cd4+ t-cell population based on a fractional-order model. Advances in Difference Equations **2017**(1), 92 (Mar 2017). https://doi.org/10.1186/s13662-017-1143-0, https://doi.org/10.1186/s13662-017-1143-0
3. Brabenec, L., Mekyska, J., Galaz, Z., Rektorova, I.: Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. Journal of Neural Transmission pp. 1–32 (2017). https://doi.org/10.1007/s00702-017-1676-0

4. Caputo, M.: Linear Models of Dissipation whose Q is almost Frequency Independent—II. Geophysical Journal International **13**(5), 529–539 (11 1967). https://doi.org/10.1111/j.1365-246X.1967.tb02303.x, https://doi.org/10.1111/j.1365-246X.1967.tb02303.x
5. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794. ACM (2016)
6. Dickson, D.W.: Parkinson's disease and parkinsonism: neuropathology. Cold Spring Harbor perspectives in medicine **2**(8), a009258 (2012)
7. Drotar, P., Mekyska, J., Rektorova, I., Masarova, L., Smekal, Z., Faundez-Zanuy, M.: Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. Artificial Intelligence in Medicine **67**, 39 – 46 (2016). https://doi.org/https://doi.org/10.1016/j.artmed.2016.01.004
8. Drotar, P., Mekyska, J., Rektorova, I., Masarova, L., Smekal, Z., Zanuy, M.F.: Decision support framework for Parkinson's disease based on novel handwriting markers. IEEE Transactions on Neural Systems and Rehabilitation Engineering pp. 1–1 (2015). https://doi.org/10.1109/tnsre.2014.2359997
9. Drotar, P., Mekyska, J., Smekal, Z., Rektorova, I., Masarova, L., Faundez-Zanuy, M.: Contribution of different handwriting modalities to differential diagnosis of Parkinson's disease. In: Medical Measurements and Applications (MeMeA), 2015 IEEE International Symposium. pp. 1–5 (2015)
10. Fahn, S., Elton, R.L.: UPDRS Development Committee (1987) Unified Parkinson's Disease Rating Scale. Recent developments in Parkinson's Disease. Macmillan, Florham Park (1987)
11. Galaz, Z., Mucha, J., Zvoncak, V., Mekyska, J., Smekal, Z., Safarova, K., Ondrackova, A., Urbanek, T., Havigerova, J.M., Bednarova, J., Faundez-Zanuy, M.: Advanced parametrization of graphomotor difficulties in school-aged children. IEEE Access **8**, 112883–112897 (2020). https://doi.org/10.1109/ACCESS.2020.3003214
12. Heinzel, S., Berg, D., Gasser, T., Chen, H., Yao, C., Postuma, R.B., on the Definition of Parkinson's Disease, M.T.F.: Update of the MDS research criteria for prodromal Parkinson's disease. Movement Disorders **34**(10), 1464–1470 (2019)
13. Hornykiewicz, O.: Biochemical aspects of Parkinson's disease. Neurology **51**(2 Suppl 2), S2–S9 (1998)
14. Hughes, A., Daniel, S., Lees, A.: The clinical features of parkinson's disease in 100 histologically proven cases. Advances in neurology **60**, 595 (1993)
15. Impedovo, D., Pirlo, G., Vessio, G.: Dynamic handwriting analysis for supporting earlier parkinson's disease diagnosis. Information **9**(10) (2018). https://doi.org/10.3390/info9100247, http://www.mdpi.com/2078-2489/9/10/247
16. Lazarević, M.: Further results on fractional order control of a mechatronic system. Scientific Technical Review, ISSN **206** (2013)
17. Letanneux, A., Danna, J., Velay, J.L., Viallet, F., Pinto, S.: From micrographia to Parkinson's disease dysgraphia. Movement Disorders **29**(12), 1467–1475 (2014)
18. Luchko, Y., Gorenflo, R.: An operational method for solving fractional differential equations with the caputo derivatives. Acta Mathematica Vietnamica **24**(2), 207–233 (1999)
19. Matthews, B.W.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta (BBA) **405**(2), 442–51 (1975)
20. McLennan, J., Nakano, K., Tyler, H., Schwab, R.: Micrographia in Parkinson's Disease. Journal of the Neurological Sciences **15**(2), 141–152 (1972)

21. Mucha, J., Faundez-Zanuy, M., Mekyska, J., Zvoncak, V., Galaz, Z., Kiska, T., Smekal, Z., Brabenec, L., Rektorova, I., Lopez-de-Ipina, K.: Analysis of parkinson's disease dysgraphia based on optimized fractional order derivative features. In: 2019 27th European Signal Processing Conference (EUSIPCO). pp. 1–5 (2019)
22. Mucha, J., Mekyska, J., Faundez-Zanuy, M., Lopez-de Ipina, K., Zvoncak, V., Galaz, Z., Kiska, T., Smekal, Z., Brabenec, L., Rektorova, I.: Advanced parkinson's disease dysgraphia analysis based on fractional derivatives of online handwriting. In: 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). (2018)
23. Mucha, J., Mekyska, J., Galaz, Z., Faundez-Zanuy, M., Lopez-de Ipina, K., Zvoncak, V., Kiska, T., Smekal, Z., Brabenec, L., Rektorova, I.: Identification and monitoring of parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting. Applied Sciences **8**(12), 2566 (2018)
24. Mucha, J., Mekyska, J., Galaz, Z., Faundez-Zanuy, M., Zvoncak, V., Safarova, K., Urbanek, T., Havigerova, J.M., Bednarova, J., Smekal, Z.: Analysis of various fractional order derivatives approaches in assessment of graphomotor difficulties. IEEE Access **8**, 218234–218244 (2020). https://doi.org/10.1109/ACCESS.2020.3042591
25. Mucha, J., Zvoncak, V., Galaz, Z., Faundez-Zanuy, M., Mekyska, J., Kiska, T., Smekal, Z., Brabenec, L., Rektorova, I., Lopez-de Ipina, K.: Fractional derivatives of online handwriting: A new approach of parkinsonic dysgraphia analysis. In: 2018 41st International Conference on Telecommunications and Signal Processing (TSP). pp. 214–217. IEEE (2018)
26. Nackaerts, E., Broeder, S., Pereira, M.P., Swinnen, S.P., Vandenberghe, W., Nieuwboer, A., Heremans, E.: Handwriting training in Parkinson's disease: A trade-off between size, speed and fluency. PLOS ONE **12**(12), e0190223 (dec 2017). https://doi.org/10.1371/journal.pone.0190223
27. Pinto, C.M., Machado, J.T.: Fractional model for malaria transmission under control strategies. Computers & Mathematics with Applications **66**(5), 908 – 916 (2013). https://doi.org/https://doi.org/10.1016/j.camwa.2012.11.017, http://www.sciencedirect.com/science/article/pii/S0898122112006785, fractional Differentiation and its Applications
28. Podlubny, I.: Fractional differential equations an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications. Academic Press, San Diego (1999)
29. Ponsen, M.M., Daffertshofer, A., Wolters, E.C., Beek, P.J., Berendse, H.W.: Impairment of complex upper limb motor function in de novo Parkinson's disease. Parkinsonism & Related Disorders **14**(3), 199–204 (2008)
30. Rios-Urrego, C., Vásquez-Correa, J., Vargas-Bonilla, J., Nöth, E., Lopera, F., Orozco-Arroyave, J.: Analysis and evaluation of handwriting in patients with parkinson's disease using kinematic, geometrical, and non-linear features. Computer Methods and Programs in Biomedicine **173**, 43–52 (May 2019). https://doi.org/10.1016/j.cmpb.2019.03.005, https://doi.org/10.1016/j.cmpb.2019.03.005
31. Rosenblum, S., Samuel, M., Zlotnik, S., Erikh, I., Schlesinger, I.: Handwriting as an objective tool for Parkinson's disease diagnosis. Journal of neurology **260**(9), 2357–2361 (2013)
32. Samko, S.G., Kilbas, A.A., Marichev, O.I., et al.: Fractional integrals and derivatives. Theory and Applications, Gordon and Breach, Yverdon **1993**, 44 (1993)
33. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). Biometrika **52**(3/4), 591–611 (1965)

34. Sun, H., Zhang, Y., Baleanu, D., Chen, W., Chen, Y.: A new collection of real world applications of fractional calculus in science and engineering. Communications in Nonlinear Science and Numerical Simulation **64**, 213 – 231 (2018). https://doi.org/https://doi.org/10.1016/j.cnsns.2018.04.019, http://www.sciencedirect.com/science/article/pii/S1007570418301308
35. Taleb, C., Khachab, M., Mokbel, C., Likforman-Sulem, L.: A reliable method to predict parkinson's disease stage and progression based on handwriting and resampling approaches. In: 2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR). pp. 7–12. IEEE (2018)
36. Thomas, M., Lenka, A., Kumar Pal, P.: Handwriting analysis in parkinson's disease: Current status and future directions. Movement Disorders Clinical Practice **4**(6), 806–818 (2017). https://doi.org/10.1002/mdc3.12552, https://onlinelibrary.wiley.com/doi/abs/10.1002/mdc3.12552
37. Uchaikin, V.V.: Fractional derivatives for physicists and engineers, vol. 2. Springer (2013)
38. Valério, D.: Variable order derivatives, https://www.mathworks.com/matlabcentral/fileexchange/24444-variable-order-derivatives
39. Valério, D., Sá da Costa, J.: Ninteger: a fractional control toolbox for Matlab. In: Fractional Differentiation and its Applications. Bordeaux (2004)
40. Valério, D., Sá da Costa, J. (eds.): An Introduction to Fractional Control. IET, Stevenage (2013)

A.4 Shannon entropy: A novel parameter for quantifying pentagon copying performance in non-demented Parkinson's disease patients

Parkinsonism and Related Disorders 94 (2022) 45–48

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis

Short communication

Shannon entropy: A novel parameter for quantifying pentagon copying performance in non-demented Parkinson's disease patients

Lubos Brabenec [a], Patricia Klobusiakova [a,b,c], Jiri Mekyska [d], Irena Rektorova [a,e,*]

[a] Applied Neuroscience Research Group, Central European Institute of Technology – CEITEC, Masaryk University, Kamenice 753/5, 625 00, Brno, Czech Republic
[b] Faculty of Medicine, Masaryk University, Kamenice 753/5, 625 00, Brno, Czech Republic
[c] Surgeon General Office of the Slovak Armed Forces, Ul. generala Milosa Vesela 21, 03401, Ruzomberok, Slovak Republic
[d] Department of Telecommunications, Brno University of Technology, Technicka 12, 616 00, Brno, Czech Republic
[e] First Department of Neurology, Faculty of Medicine, St. Anne's University Hospital, Masaryk University, Pekarska 664/53, 656 91, Brno, Czech Republic

ARTICLE INFO

Keywords:
Parkinson's disease
Kinematic analysis
Pentagon copying test
In-air movement
Shannon entropy

ABSTRACT

Introduction: Impaired copy of intersecting pentagons from the Mini-Mental State Examination (MMSE), has been used to assess dementia in Parkinson's disease (PD). We used a digitizing tablet during the pentagon copying test (PCT) as a potential tool for evaluating early cognitive deficits in PD without major cognitive impairment. We also aimed to uncover the neural correlates of the identified parameters using whole-brain magnetic resonance imaging (MRI).
Methods: We enrolled 27 patients with PD without major cognitive impairment and 25 age-matched healthy controls (HC). We focused on drawing parameters using a digitizing tablet. Parameters with between-group differences were correlated with cognitive outcomes and were used as covariates in the whole-brain voxel-wise analysis using voxel-based morphometry; familywise error (FWE) threshold $p < 0.001$.
Results: PD patients differed from HC in attention domain z-scores ($p < 0.0001$). In terms of tablet parameters, the groups differed in Shannon entropy (horizontal in-air, $p = 0.003$), which quantifies the movements between two strokes. In PD, a correlation was found between the median of Shannon entropy (horizontal in-air) and attention z-scores ($R = -0.55$, $p = 0.006$). The VBM revealed an association between our drawing parameter of interest and gray matter (GM) volume variability in the right superior parietal lobe (SPL).
Conclusion: Using a digitizing tablet during the PCT, we identified a novel entropy-based parameter that differed between the nondemented PD and HC groups. This in-air parameter correlated with the level of attention and was linked to GM volume variability of the region engaged in spatial attention.

1. Introduction

Subtle cognitive deficits are very common in Parkinson's disease (PD) and mostly include altered attention and executive functions that are particularly related to dopaminergic deficits and dysfunction of association basal ganglia circuitry [1], although other neurotransmitters seem to be involved as well [2]. In addition to the abovementioned profile of cognitive impairment, other cognitive domains may be affected [3]. For a quick and easy assessment of executive and visuospatial functions, a task involving copying two intersecting pentagons is often used. Performance in the pentagon copying test (PCT) has been shown to predict cognitive decline in PD [4]; however, results may vary depending on the scoring method [5]. We have shown that handwriting kinematic parameters, assessed with the help of a digitizing tablet, can precisely quantify both "on-surface" and "in-air" hand movements [6], and the in-air kinematic parameters distinguished PD from healthy controls (HC) with higher accuracy than the well-described on-surface handwriting parameters [7]. Hesitation and uncertainty between two handwriting/drawing strokes lead to excessive in-air movements that may reflect disturbed planning of movement execution and/or cognitive deficits. We have previously shown that Shannon entropy, i.e. a numerical measure of the randomness or uncertainty of a signal, is a good in-air parameter that reflects alterations of PD handwriting [8]. Therefore, in the frame of this study, using a digitizing tablet and exploiting the Shannon entropy, we aimed to identify a more precise parameter that would quantify PCT and distinguish nondemented PD subjects from

* Corresponding author. Applied Neuroscience Research Group, Central European Institute of Technology – CEITEC, Masaryk University, First Department of Neurology, School of Medicine, Masaryk University, St. Anne's Teaching Hospital, Pekarska 53, 656 91, Brno, Czech Republic.
E-mail address: irena.rektorova@fnusa.cz (I. Rektorova).

https://doi.org/10.1016/j.parkreldis.2021.11.037
Received 15 September 2021; Received in revised form 8 November 2021; Accepted 30 November 2021
Available online 4 December 2021

HC. We also aimed at identifying their cognitive and neural correlates.

2. Methods

2.1. Participants

We enrolled 27 patients with clinically established PD and 25 HC. All participants were right-handed. None of the subjects had a history or presence of psychosis, hallucinations, depression, or dementia [3]. All PD patients were on a stable dopaminergic medication and were tested in the ON medication state without dyskinesias. All participants signed an informed consent form that was approved by the local ethics committee.

2.2. PCT parameters and visual scoring

The participants were asked to perform a drawing on an A4 paper that was laid down and fixed to a digitizing tablet Wacom Intuos 4 M. Collected signals are described in Suppl. Material.

During the parameterization of the PCT drawings, we focused on six features included in evaluating the on-surface drawing: 1) spatial features – height and length; 2) temporal feature – duration of drawing; 3) kinematic feature – relative standard deviation of acceleration which is associated with the fluency of drawing; and 4) entropy-based features – median of Shannon entropy extracted from in-air hand movements (horizontal, vertical) between consecutive strokes. A stroke is a product of a drawing on paper performed between two pen elevations, e.g. see the blue lines in Fig. 1 and Suppl. Material for more details. Considering that the horizontal in-air movement is represented by time-series X with n unique samples x_i, then its Shannon entropy is calculated as $H(X) = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i)$, where $p(x_i)$ is the probability density function [8]. Analogically, the formula can be applied to the time series of the vertical in-air movement.

The PCT was scored by a psychologist (LB), using the qualitative scoring of pentagon test (QSPT) method [9]. The total score ranged from 0 to 13 points; for details see the Supplementary materials.

2.3. Neuropsychological assessment

Four cognitive domains (visuospatial, memory, attention, and executive domains) were examined using a complex neuropsychological assessment. The cognitive domain z-scores were computed as the average z-scores of the tests included in the particular domain and were correlated with PCT parameters. For details see the Supplementary materials.

2.4. MRI sequences

Subjects were scanned with a 3T Siemens Prisma MR scanner (Siemens, Erlangen, Germany). High-resolution anatomical T1-weighted images were acquired (TR = 2300 ms, TE = 2.33 ms, FA = 8°, FOV = 224 mm, slice thickness 1 mm, 240 sagittal slices, matrix size 224 × 224).

2.5. Association between PCT parameters and regional GM volumes

SPM12 software was used to pre-process anatomical T1-weighted images. MR images were segmented into gray and white matter segments and the DARTEL imported versions of GM and white matter were obtained for each subject. They were then spatially registered to the MNI coordinate system using the DARTEL toolbox [10]. GM probability maps were Jacobian-modulated in order to preserve the original GM volume and smoothed using a spatial filter with the Gaussian kernel (FWHM = 10 mm). Lastly, the values of images were divided by total intracranial volume (TIV) to correct for the effects of overall brain size. In the second-level whole-brain voxel-wise analysis, we investigated the presence of significant linear correlations between regional volumes and drawing features of interest (i.e. with significant differences between both groups) using the general linear model separately in the HC and PD groups. Age, gender, education, and levodopa equivalent dose (LED) were included as covariates of no interest. Results were considered significant if $p < 0.05$ after FWE correction was performed with the initial threshold being $p < 0.001$.

2.6. Statistical analysis

We used the Mann-Whitney U test to assess differences between HC and PD in PCT parameters, cognitive domain z-scores, and PCT visual scores. Spearman's correlations between PCT parameters of interest (i.e. with significant differences between both groups) and cognitive domains z-scores and PCT visual scores were calculated separately in the PD and HC groups. Additional partial correlation analyses were performed with LED and Unified Parkinson's Disease Rating Scale III (UPDRS III) scores as covariates in order to regress out the effects of motor impairment and/or dopaminergic medication. Bonferroni correction was used to control for multiple testing.

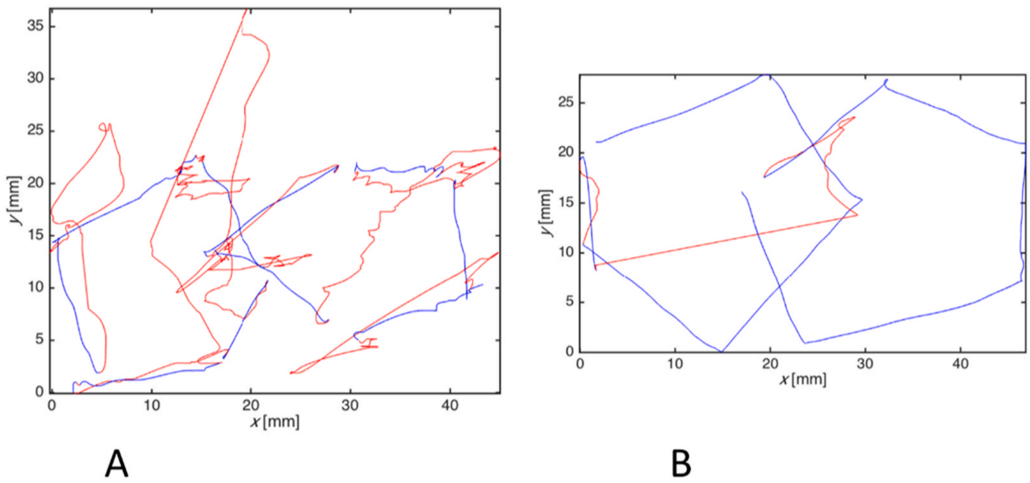


Fig. 1. PCT performed by a PD patient (A) and a HC subject (B). The blue line represents on-surface (on-paper) movement and the red line the in-air one. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3. Results

3.1. Clinical and cognitive outcomes

All 52 subjects completed the study. Altogether 13 patients had left-sided PD symptom predominance, 10 had right-sided symptom predominance, and 4 had bilateral PD. For the demographic and clinical/cognitive data for PD and HC groups, see Table 1. The PD and HC groups differed in all cognitive domains, but only the attention cognitive domain z-scores survived Bonferroni correction (for four measurements). Only 3 out of 27 PD patients had z-scores lower than −1.5 in at least two cognitive tests in one or more cognitive domains and were classified as PD-MCI [3]. The groups slightly differed in the QSPT ($p = 0.016$); however, as expected in non-demented subjects, the scores in both the HC and PD groups displayed a ceiling effect, therefore reducing variability in the data. These significant differences also remained after removal of these 3 subjects (see Suppl. Materials). There were no significant differences between right and left sided dominant patients, see Supplementary Table S4.

3.2. PCT parameters

We observed a significant difference between the HC and PD groups in the median of Shannon entropy (horizontal in-air) ($p = 0.003$), median of Shannon entropy (vertical in-air) ($p = 0.044$), and relative STD of acceleration (on-surface) ($p = 0.014$); see Table 1. Only the Shannon entropy (horizontal in-air) survived the Bonferroni correction for six measurements. For an illustration of in-air movements in PD and HC, see Fig. 1 A,B.

Table 1
Demographic and clinical/cognitive variables.

	PD, N = 27	HC, N = 25	Mann-Whitney
Gender (M/F)	17/10	7/18	p = 0.012
Age (years)	Med. = 67.0 IQR = 11.0	Med. = 67.1 IQR = 7.12	p = 0.806
Education (years)	Med. = 13 IQR = 5	Med. = 17 IQR = 5	p = 0.273
MOCA	Med. = 26 IQR = 6	Med. = 28 IQR = 4	p = 0.019
PD duration (years)	Med. = 4.0 IQR = 8.0	NA	NA
LED (mg)	Med. = 960.0 IQR = 910.0	NA	NA
UPDRS III	Med. = 11.0 IQR = 9.0	NA	NA
Memory domain z-scores	Med. = 0.475 IQR = 1.35	Med. = 1.058 IQR = 1.34	p = 0.033
Attention domain z-scores	Med. = −0.47 IQR = 0.74	Med. = 0.33 IQR = 0.57	p < 0.0001
Executive domain z-scores	Med. = −0.11 IQR = 0.90	Med. = 0.27 IQR = 0.76	p = 0.030
Language domain z-scores	Med. = −0.25 IQR = 1.5	Med. = 0.50 IQR = 1.13	p = 0.014
Visuospatial domain z-scores	Med. = 0.36 IQR = 1.16	Med. = 0.78 IQR = 0.47	p = 0.020
Height of drawing (mm)	Med. = 32.64 IQR = 6.83	Med. = 35.21 IQR = 11.06	p = 0.489
Length of drawing (mm)	Med. = 327.3 IQR = 136.2	Med. = 349.0 IQR = 96.2	p = 0.749
Duration of drawing (s)	Med. = 21.85 IQR = 9.13	Med. = 17.43 IQR = 15.62	p = 0.403
relative STD of acceleration (on-surface) higher values associated with more dysfluent movement	Med. = 20.01 IQR = 10.76	Med. = 14.91 IQR = 5.56	p = 0.014
median of Shannon entropy (horizontal in-air) higher values associated with excessive movements in-air	Med. = 5.4 IQR = 1.12	Med. = 4.54 IQR = 0.95	p = 0.003
median of Shannon entropy (vertical in-air) higher values associated with excessive movements in-air	Med. = 5.37 IQR = 1.23	Med. = 4.56 IQR = 0.92	p = 0.044
QSPT	Med. = 12.0 IQR = 1.0	Med. = 13.0 IQR = 1.5	p = 0.016

Med. – median, IQR – interquartile range.

3.3. Correlation analyses between PCT parameter of interest and cognitive outcomes

In the PD group, we found a significant correlation between the median of Shannon entropy (horizontal in-air) and attention domain z-scores ($R = -0.554$, $p = 0.006$). The result survived Bonferroni correction for four measurements (four cognitive domains z-scores), see Fig. S1 and Table S1 in the Supplementary materials. Similar results were found after regressing out the effects of LED and UPDRS III Motor Assessment ($R = -0.668$, $p = 0.005$).

In the HC group, we found significant correlation between median of Shannon entropy (horizontal in-air) and executive domain z-scores ($R = -0.447$, $p = 0.042$). The result lost significance after correction for multiple testing.

No significant correlations were found between visual PCT scores and cognitive domains in either PD or HC groups; see Table S2 in the Supplementary materials. No association was found between visual PCT scores and the median of Shannon entropy (horizontal in-air) either ($R = 0.096$; $p = 0.697$).

3.4. Correlation between MRI regional GM volumes and PCT parameters of interest

In the PD group, we found a significant negative correlation between our PCT parameter of interest (median of Shannon entropy, horizontal in-air) and GM volume in the right SPL (Brodmann area 7; Cluster size 585 voxels; MNI coordinates 25.5 –64.5 39.0; $p = 0.001$; see Fig. S2 in the Supplementary materials). In the HC group, there were no significant correlations between regional GM volumes and drawing features.

4. Discussion

Our study demonstrated that Shannon entropy extracted from in-air movement between two consecutive strokes, significantly differed between non-demented PD and HC while drawing intersecting pentagons.

Shannon entropy quantifies excessive in-air movements that could be associated with the following activities: movement preparation, in-air motor start hesitation, and movement uncertainty, as well as cognitive impairment or lapses of attention. Previous research focused on handwriting showed more alterations in the horizontal direction of handwriting than in the vertical direction, which may be due to wrist extension stiffness in PD [7,11]. We showed that the variability of horizontal Shannon entropy was closely linked to the level of attention, even after regressing out the effects of motor impairment as assessed by motor score and LED. Attention had been clearly affected in our non-demented PD subjects when compared to HC despite the fact that only 3 out of 27 PD subjects met the criteria for PD-MCI [3] and these significant differences also remained after removal of these 3 subjects. Notably, this correlation was not found with the visual PCT scores, probably due to the ceiling effect and low variability of visual PCT scores in non-demented subjects. We did not find any significant difference between right and left sided dominant patients.

Few studies have focused on assessing the neural correlates of pentagon drawing in PD, with variable results. Filoteo et al. [5] found that PCT accuracy, based on their modified visual scoring system (scores ranged from 3 to 0), significantly correlated with cortical volume variability in the left rostral middle frontal cortex, the right supplementary motor area, the pars triangularis, and the left cuneus in PD patients, i.e. regions involved in the frontoparietal, motor, language, and visual networks, respectively. Another study by Garcia-Diaz et al. [12] used cortical thickness measures and found that PD patients with abnormal pentagon drawings, as assessed visually, had significant cortical thinning of the right precentral and postcentral gyri, superior parietal region, and posterior cingulate cortex, i.e., in regions linked to higher order visual processing as well as attention and movement execution. Unlike in our study, the PCT scores were associated with widespread cortical region atrophy, meaning that visual PCT scores cannot identify and monitor early brain changes.

We demonstrated that changes in our in-air PCT parameter of interest were significantly related to GM volume variability solely of the right SPL. This region is a part of the dorsal attention network (DAN) [13] as well as the dorsal visual stream. Both pathways are involved in visual (spatial) attention control in PD [13] and atrophic changes of the SPL have been demonstrated early in the course of the disease [14]. Therefore, we assume that disrupted in-air movements during PCT, as assessed by the digitizing tablet, could be an early manifestation of attention-related posterior cortical volume changes that have been shown to predict cognitive decline during the PD course [4]. Future prospective longitudinal studies should assess whether this parameter may serve as an early marker of cognitive impairment and dementia in PD.

In conclusion, we identified a novel in-air parameter for quantitative PCT assessment. This parameter is closely linked to attention levels and to the GM volume variability of the posterior cortical region engaged in both visual attention and visual-spatial processing. Our results indicate that this in-air parameter could be used to evaluate early cognitive changes that precede disturbed pentagon drawing (as assessed visually) and this may be clinically relevant. Future longitudinal studies should assess whether the Shannon entropy of in-air movement will become a good marker for MCI and dementia conversion in PD.

Acknowledgement

This project received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 734718 (CoBeN), from a grant from the Czech Ministry of Health, NU20-04-00294 and Spanish grant Ministerio de ciencia e innovación PID2020-113242RB-I00. We also acknowledge the core facility MAFIL of CEITEC supported by the MEYS CR (LM2018129 Czech-BioImaging).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.parkreldis.2021.11.037.

References

[1] D. Ferrazzoli, P. Ortelli, G. Madeo, N. Giladi, G.M. Petzinger, G. Frazzitta, Basal ganglia and beyond: the interplay between motor and cognitive aspects in Parkinson's disease rehabilitation, Neurosci. Biobehav. Rev. 90 (2018) 294–308, https://doi.org/10.1016/j.neubiorev.2018.05.007.

[2] A. Yarnall, L. Rochester, D.J. Burn, The interplay of cholinergic function, attention, and falls in Parkinson's disease, Mov. Disord. 26 (2011) 2496–2503, https://doi.org/10.1002/MDS.23932.

[3] I. Litvan, J.G. Goldman, A.I. Tröster, B.A. Schmand, D. Weintraub, R.C. Petersen, B. Mollenhauer, C.H. Adler, K. Marder, C.H. Williams-Gray, D. Aarsland, J. Kulisevsky, M.C. Rodriguez-Oroz, D.J. Burn, R.A. Barker, M. Emre, Diagnostic criteria for mild cognitive impairment in Parkinson's disease: movement Disorder Society Task Force guidelines, Mov. Disord. 27 (2012) 349–356, https://doi.org/10.1002/mds.24893.

[4] C.H. Williams-Gray, S.L. Mason, J.R. Evans, T. Foltynie, C. Brayne, T.W. Robbins, R.A. Barker, The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort, J. Neurol. Neurosurg. Psychiatry 84 (2013) 1258–1264, https://doi.org/10.1136/jnnp-2013-305277.

[5] J.V. Filoteo, J.D. Reed, I. Litvan, D.L. Harrington, Volumetric correlates of cognitive functioning in nondemented patients with Parkinson's disease, Mov. Disord. 29 (2014) 360–367, https://doi.org/10.1002/mds.25633.

[6] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, M. Faundez-Zanuy, Analysis of in-air movement in handwriting: a novel marker for Parkinson's disease, Comput. Methods Progr. Biomed. 117 (2014) 405–411, https://doi.org/10.1016/j.cmpb.2014.08.007.

[7] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, K. Lopez-de-Ipina, V. Zvoncak, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, Identification and monitoring of Parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting, Appl. Sci. 8 (2018) 2566, https://doi.org/10.3390/app8122566.

[8] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, M. Faundez-Zanuy, Decision support framework for Parkinson's disease based on novel handwriting markers, IEEE Trans. Neural Syst. Rehabil. Eng. 23 (2015) 508–516, https://doi.org/10.1109/TNSRE.2014.2359997.

[9] P. Caffarra, S. Gardini, F. Dieci, S. Copelli, L. Maset, L. Concari, E. Farina, E. Grossi, The qualitative scoring MMSE pentagon test (QSPT): a new method for differentiating dementia with Lewy Body from Alzheimer's disease, Behav. Neurol. 27 (2013) 213–220, https://doi.org/10.3233/BEN-120319.

[10] J. Ashburner, A fast diffeomorphic image registration algorithm, Neuroimage 38 (2007) 95–113, https://doi.org/10.1016/j.neuroimage.2007.07.007.

[11] H.I. Ma, W.J. Hwang, S.H. Chang, T.Y. Wang, Progressive micrographia shown in horizontal, but not vertical, writing in Parkinson's disease, Behav. Neurol. 27 (2013) 169–174, https://doi.org/10.3233/BEN-120285.

[12] A.I. Garcia-Diaz, B. Segura, H.C. Baggio, M.J. Marti, F. Valldeoriola, Y. Compta, P. Vendrell, N. Bargallo, E. Tolosa, C. Junque, Structural MRI correlates of the MMSE and pentagon copying test in Parkinson's disease, Park. Relat. Disord. 20 (2014) 1405–1410, https://doi.org/10.1016/j.parkreldis.2014.10.014.

[13] N. Nemcova Elfmarkova, M. Gajdos, I. Rektorova, R. Marecek, S.Z. Rapcsak, Neural evidence for defective top-down control of visual processing in Parkinson's and Alzheimer's disease, Neuropsychologia 106 (2017) 236–244, https://doi.org/10.1016/j.neuropsychologia.2017.09.034.

[14] J. Kunst, R. Marecek, P. Klobusiakova, Z. Balazova, L. Anderkova, N. Nemcova-Elfmarkova, I. Rektorova, Patterns of grey matter atrophy at different stages of Parkinson's and Alzheimer's diseases and relation to cognition, Brain Topogr. 32 (2019) 142–160, https://doi.org/10.1007/s10548-018-0675-2.

A.5 Exploiting spectral and cepstral handwriting features on diagnosing Parkinson's disease

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 2, 2021, accepted October 3, 2021, date of publication October 8, 2021, date of current version October 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3119035

Exploiting Spectral and Cepstral Handwriting Features on Diagnosing Parkinson's Disease

JUAN A. NOLAZCO-FLORES [1], (Senior Member, IEEE), MARCOS FAUNDEZ-ZANUY [2], V. M. DE LA CUEVA [1], (Member, IEEE), AND JIRI MEKYSKA [3]

[1]School of Engineering and Science, Tecnológico de Monterrey, Monterrey, Nuevo León 64849, Mexico
[2]Escola Superior Politecnica, TecnoCampus Mataro-Maresme, 08302 Mataro, Spain
[3]Department of Telecommunications and SIX Research Center, Brno University of Technology, 61600 Brno, Czechia

Corresponding authors: Juan A. Nolazco-Flores (jnolazco@tec.mx) and V. M. de la Cueva (vcueva@tec.mx)

This work was supported in part by MINECO Spanish Grant PID2020-113242RB-I00.

ABSTRACT Parkinson's disease (PD) is the second most frequent neurodegenerative disease associated with several motor symptoms, including alterations in handwriting, also known as PD dysgraphia. Several computerized decision support systems for PD dysgraphia have been proposed, however, the associated challenges require new approaches for more accurate diagnosis. Therefore, this work adds spectral and cepstral handwriting features to the already-used temporal, kinematic and statistics handwriting features. First, we calculate temporal and kinematic features using displacement; statistic features (SF) using displacement, and horizontal and vertical displacement; spectral (SDF) and cepstral (CDF) using displacement, horizontal and vertical displacement and pressure. Since the employed dataset (PaHaW) contains only 37 PD patients and 38 healthy control subjects (HC), then as the second step, we augment the percentage of the smaller training set to equal the larger. Next, we augment both classes to increase the training patient's data and added random Gaussian noise in all augmentations. Third, the most relevant features were selected using the modified fast correlation-based filtering method (mFCBF). Finally, autoML is employed to train and test more than ten plain and ensembled classifiers. Experimental results show that adding spectral and cepstral features to temporal, kinematics and statistics features highly improved classification accuracy to 98.57%. Our proposed model, with lower computational complexities, outperforms conventional state-of-the-art models for all tasks, which is 97.62%.

INDEX TERMS Parkinson's disease, dysgraphia, online handwriting, feature extraction, data augmentation, autoML.

I. INTRODUCTION

Biometrics can be used for e-security and e-health [1] and can be grouped based on two traits. Morphological biometrics, such as fingerprints or eye pupils, use direct measurements of the physical traits of the human body [2], [3]. Behavioral biometrics, such as handwriting and drawing, use specific drawing and handwriting tasks performed by the subjects involved in data collection [4]. From a health perspective, online handwriting biometrics are more appealing and informative on states of diseases, such as dementia, than other popular biometrics traits, such as fingerprints or iris [3], [4] because they make part of routine functional activities from which evidence are drawn affected by the disease.

In the last two decades, online handwriting processing has been employed in the computerized assessment of neurodegenerative disorders (e.g., Parkinson's disease (PD)) [5], [6]. Patients with PD experience progressive loss of dopaminergic neurons in substantia nigra pars compacta (located in the midbrain), which is consequently associated with cardinal motor symptoms such as bradykinesia, rigidity, resting tremor, or postural instability [7]–[9]. Therefore, especially during the clinical phase of the disorder, we can observe freezing of gait [10], hypokinetic dysarthria [11], hypomimia [12], or alterations in handwriting [6]. The latter was initially linked with micrographia, i.e., a progressive decline in amplitude (vertical micrographia) or with

The associate editor coordinating the review of this manuscript and approving it for publication was Gang Mei.

(horizontal micrographia) of letters [13]. Nevertheless, micrographia is one manifestation of altered handwriting in patients with PD. Others include more pronounced changes in kinematics and dynamics too. Letanneux et. reported a connection to developmental dysgraphia and proposed a new and more general term, PD dysgraphia [14].

Recently, several designs of decision support systems for diagnosing different PDs based on speech/voice analysis [15]–[20] or gait monitoring [21]–[23], have been proposed. However, compared with online handwriting processing, both speech assessment and gait monitoring require more technical equipment and are vulnerable to low signal quality due to a noncontrolled environment. Speech assessment requires high-quality recording conditions without background noise and further postprocessing of recorded speech. This includes human-operated speech segmentation, making the whole process more difficult. Gait monitoring or tremor assessment techniques require specialized equipment, such as motion capture systems, accelerometers, and gyroscopes. However, the diagnosis of PD using handwriting processing can be easily administered at the clinic or a patient's home. Handwriting acquisition is simple and natural and requires no timing or exhaustive repetitions.

A comprehensive review of quantitative analysis of PD dysgraphia and its computerized diagnosis for published works until 2019 has been summarized [4]–[6], [14]. Furthermore, we review the state-of-the-art designs published in 2020 and 2021, focusing on articles using online handwriting.

The rest of the paper is organized as follows: Section II reviews related works and presents state-of-the-art results obtained in PD diagnosis based on the PaHaW database. Section III describes the H2O platform used in this work. Section IV describes the PAHAW database. Section V describes the feature extraction process used and describes the type of feature obtained. Section VI presents a brief explanation of the modified version of the fast correlation-based filtering feature selection methodology. Section VII describes the front-end hyperparameters. Section VIII describes the experiments conducted and their results. Finally, in Section IX, we present final remarks, comments, and conclusions.

II. RELATED WORKS

This section reviews related works and state-of-the-art results obtained for PD diagnosis. Table 1 shows a summary of the state-of-the-art results.

Ammour *et al.* [24] quantitatively analyzed online handwriting in 28 PD patients and 28 age-matched healthy controls (HC). They quantified the performance of subjects (when writing a text in Arabic letters) according to 1482 kinematics (velocity, acceleration, jerk, etc.), dynamic (pressure, tilt, azimuth, etc.), temporal (e.g., duration), and some additional features. From a semi-supervised approach (employing cluster analysis), they differentiated the PD group with 71.44% accuracy. Furthermore, they concluded that the complications of fine motor skills in PD patients were mainly manifested in the kinematic feature set.

Liaqat Ali, *et al.* [25] propose a method for dealing with the highly unbalanced handPD dataset. To improve the PD detection accuracy on this dataset, they developed a cascaded learning system that cascades a Chi2 model with an adaptive boosting (Adaboost) model. Experimental results confirmed that the proposed cascaded system outperforms six similar cascaded systems using six state-of-the-art machine learning models, respectively.

Taleb *et al.* [26] introduced a PD diagnosis concept that uses convolutional neural networks (CNN) fed by spectrograms (calculated from various online handwriting/drawing tasks) and CNN bidirectional long-short-term memory networks (CNN-BLSTM) fed by raw time series. In the publicly available dataset called HandPDMultiMC, containing 21 PD and HC, respectively, a classification accuracy of approximately 97.62% was achieved by combining CNN-BLSTM models trained with jittering and synthetic data augmentation. They trained 204,060 parameters model for one day using an NVIDIA GTX 1080 GPU of 8 GB.

Gupta *et al.* [27] explored the effect of age and gender on the performance of classification models. Thus, they used the PaHaW database [28] containing 37 PD patients and 38 HC. The subjects performed seven tasks including a sentence or isolated words. The data were parametrized using kinematic, entropic, and energetic features and fed into age- and gender-dependent support vector machine (SVM) models. A distinct set of discriminative features was observed in each category (age vs. gender). The results showed an improved classification accuracy of a general model from 75.76% to 83.75% and 79.55% in a female and male set, respectively.

Aouraghe *et al.* [29] focused on the effect of progressing fatigue in PD dysgraphia. They enrolled 40 PD patients and HC, respectively, copying a multiline paragraph in Arabic letters. First, the paragraph was segmented into individual lines and then, each line processed separately using a set of temporal, kinematic, dynamic, spectral, entropy-based, and wavelet-based features. The feature space was modeled by k-nearest neighbor classifier (KNN), SVM and decision trees. An accuracy of 92.86% was obtained when processing the last line of the paragraph, i.e., the line where the fatigue is mostly accented.

Deharab *et al.* [30] introduced a novel online handwriting parameterization using dynamic writing traces warping (DWTW).DWTW was applied to kinematic patterns of handwriting and returned parameters linked with the similarity between normative and pathological time series. The features were modeled using SVM and were evaluated on the PaHaW dataset (29 PD and 32HC; all eight tasks including handwriting and drawing of Archimedean spiral), and an accuracy of 88.33% was achieved.

TABLE 1. State -of -the -art in PD diagnosis based onUSING the PaHaW database. Legend: SVM – support vector machine; RF – random forests; ET – extremely randomized trees; ADA – AdaBoost, TKEO – Teager-Kaiser energy operator; EMD – empirical mode decomposition; DWTW – dynamic writing traces warping; CGP – cartesian genetic programming; 1DCL – 1-dimensional convolutional layer; BiGRUs – bidirectional gated recurrent units; ACC – accuracy; SEN – sensitivity; SPE – specificity; PRE – precision; REC – recall, AUC – area under the ROC curve.

Reference	Tasks	Features	Feature selection	ML algorithms	Performance	Conclusion
Drotar et al. 2016 [28]	all	temporal (e.g. duration), kinematic (e.g. velocity, acceleration, jerk), pressure-based.	Mann-Whitney U test	SVM	ACC = 81.3%, SEN = 87.4%, SPE = 80.9%	This is the study introducing the PaHaW dataset along with the baseline classification results.
Mucha et al. 2018 [35]	continuous and/or repetitive task, such as Archimedean spiral	Fractional-order-derivatives		SGBoost	ACC=97.14%	Features based on Fractional-Order Derivatives improve PD severity assesment.
Impedovo 2019 [34]	all	temporal (e.g. duration), kinematic (e.g. velocity, acceleration, jerk), dynamic (e.g. azimuth, tilt, pressure), entropic (e.g. Shannon entropy), TKEO, EMD-based, sigma-lognormal features, parameters based on the Maxwell-Boltzmann distribution and the discrete Fourier transform	Mann-Whitney U test, Relief	SVM	ACC=93.79% ACC = 98.44%	The accuracy for all task is 93.79. Accuracy of 98.44% was obtained when combining 3 tasks (Archimedean spiral, grapheme "l", and a word).
Angelillo et al. 2019 [42]	all	temporal (e.g. duration), spatial (e.g. length of stroke), kinematic (e.g. velocity, acceleration, jerk), dynamic (e.g. pressure), entropic (e.g. Shannon entropy), EMD-based, other (e.g. number of strokes)	SVM ranking	SVM	ACC = 91.67%	The best performance was obtained when combining 3 tasks (grapheme "l", bigram "le" and a word).
Diaz et al. 2019 [43]	all	static handwriting processed by CNN and enhanced by velocity and in-air movement	SVM ranking	ensemble (SVM, RF, ET, ADA)	ACC = 86.67%, SEN = 89.17%, SPE = 80.83%	The authors proved that a dynamic enhance of static handwriting could provide better results than considering the online/offline handwriting separately.
Mucha et al. 2019 [44]	all	kinematic features based on fractional calculus, dynamic (e.g. pressure, tilt)	-	XGBoost	ACC = 80.60%, SEN = 79.41%, SPE = 80.56%	Features based on fractional calculus enable complex assessment of kinematic aspects of graphomotor difficulties.
Gupta et al. 2020 [27]	all, excluding Archimedean spiral	kinematic (e.g. velocity, acceleration, jerk), entropic (e.g. Shannon entropy), and energetic (e.g. signal-to-noise ratio)	Mann-Whitney U test, SVM ranking	SVM	general: ACC = 75.76%, PRE = 97.72%, REC = 81.02%; gender-specific: ACC = 83.75%, SEN = 94.40%, SPE = 85.07%	Gender- and age-specific models reached better performance.
Deharab et al. 2020 [30]	all	features based on DWTW	-	SVM	ACC = 88.33%, SEN = 86.43%, SPE = 89.50%	The best performance was based on sentence, utilising only two features.
Taleb, et.al. 2020 [26]	all	horizontal and vertical displacment time series		CNN-BLSTM	ACC=97.62	High performance and high computer complexity for training.
Parziale et al. 2021 [31]	all	temporal (e.g. duration), kinematic (e.g. velocity, acceleration, jerk), dynamic (e.g. pressure), other (e.g. features base on rising/falling edge)	-	CGP	ACC = 71.18%, SEN = 70.33%, SPE = 73.83%	The classification model provides a good trade-off between discrimination power and interpretability.
Diaz et al. 2021[33]	all	spatial (e.g. x/y coordination, displacement), kinematic (e.g. velocity, acceleration, jerk), dynamic (e.g. pressure, tilt)	-	1DCL, BiGRUs	AUC = 96.88%, SEN = 92.50%, SPE = 100.00%	The study confirms the effectiveness of the sequence learning paradigm for processing sequential handwriting data.

Parziale *et al.* [31] addressed a recurrent issue in most published works, clinical interpretability. More specifically, authors frequently use handcrafted features poorly linked to physiological processes and employ less interpretable machine learning models (so-called black boxes).

Such systems are unacceptable for clinicians. Therefore, cartesian genetic programming (CGP) (which provides a tradeoff between performance and interpretability) was used in comparison with the more common classifiers. The proposed methodology was evaluated on two datasets, PaHaW (37 PD and 38 HC; all tasks were used) and NewHandPD (31 PD and 35 HC; subjects performed a spiral and a meander). Using conventional temporal, kinematic, and dynamic features, CGP produced more accurate results than white-box methods (reaching 71.18% in PaHaW and 80.39% in NewHandPD) and more interpretable than the black boxes.

Lamba *et al.* [32] analyzed basic temporal (e.g., duration) and kinematic (e.g., velocity, acceleration, jerk) measures in 62 PD patients and 15 HC (enrolled in the frame of the Irvine (UCI) Parkinson's disease spiral drawings dataset). Due to high imbalance, the synthetic minority-oversampling technique was employed to balance the cohort. Next, data were modeled by several machine learning models, e.g., SVM, AdaBoost, and XGBoost. Finally, a classification accuracy of 96.02% was reported for AdaBoost.

Diaz *et al.* [33] discussed processed time series of online handwriting (including velocity, acceleration, jerk, displacement, pressure, etc.) using one-dimensional convolutions and bidirectional gated recurrent units (BiGRUs). This end-to-end pipeline was applied to PaHaW (37 PD and 38 HC; all tasks were used) and NewHandPD (31 PD and 35 HC; all tasks were used). The method provided competitive results (96.25% accuracy in PaHaW and 94.44% in NewHandPD), thus confirming the effectiveness of the sequence learning paradigm for processing sequential handwriting data.

Impedovo *et al.* [34] investigate different velocity-based signal processing techniques to address PD assessment. He uses kinematic, energy, and cepstral features. The energy and cepstral features are similar to the ones used in this work, but they do **not** use filterbank, and they so not use the filterbank output to calculate the cepstral. An accuracy result of 93.7% for all tasks, and 98.44% when he selects the top three tasks was reported.

Mucha *et al.* [35] combine kinematic features with fractional-order derivatives and reported an accuracy of 97.14%, for the continuous and repetitive task, such as Archimedean spiral.

Finally, in [36] we proposed the use of spectral and cepstral features for emotion recognition. Here, we concatenated these features with very simple temporal features.

This study explores new approaches of online handwriting parameterization, augmentation, analysis, and modeling with a special focus on improved diagnostic accuracy. Furthermore, we explore the impact of newly proposed spectral and cepstral features on classification accuracy and improve the pipeline by adding data augmentation and modified fast correlation-based filtering feature selection method.

TABLE 2. List of machine learning models.

Classification Algorithms
Deep learning (Deep neural networks-DNN)
Distributed random forest (DRF)
Extremely randomized trees (XRT) models)
Generalized linear model (GLM)
Generalized additive model (GAM)
Gradient boosting machine (GBM)
Naïve bayes classifier (NBC)
Rulefit (RF)
Stacked ensembles (SE)
XGBoost (XGB)
Support vector machine (SVM)

III. DATA MODELING

For data modeling, we use autoML H2O [37], [38]. Automatic machine learning (AutoML) is the process of automating algorithm selection, feature generation, hyperparameter tuning, iterative modeling, and model assessment. It eases training and evaluation of machine learning models. AutoML includes many ML models, however, we limit the number of models to the ones shown in Table 1. Also, it ensembles the best models that outperform individual models. Furthermore, it uses the area under the ROC curve as the default ranking metric for binary classification. The configuration is such that individual models are tuned using a two-fold cross-validation set. AutoML automatically performs Bayesian hyperparameter optimization.

Since a default performance metric for each machine learning task is specified internally, the leaderboard is sorted by that metric.

In Table 2, ML models include stacked ensemble models. The stacked ensemble is an efficient ensemble method, such that the predictions, from machine learning algorithms, are used as inputs in a second layer learning algorithm. In the second layer, H20 ensembles all models, (StackedEnsemble_AllModels), and the best of family, (StackedEnsemble_BestOfFamily), including the best models of each kind in the final ensemble.

IV. PAHAW DATABASE

This study employed the Parkinson's disease handwriting database (PaHaW), containing 37 PD patients and 38 age- and gender-matched HC enrolled at the department of neurology, St. Anne's university hospital in Brno [28]. Besides age and gender, the PD group is described in terms of PD duration, unified Parkinson disease rating scale part V score, and levodopa equivalent daily dose.

All subjects have no history or presence of any psychiatric symptom or disease affecting the central nervous system, except for PD. The acquisition was performed when the patients were in their ON state, i.e., approximately one hour after taking levodopa.

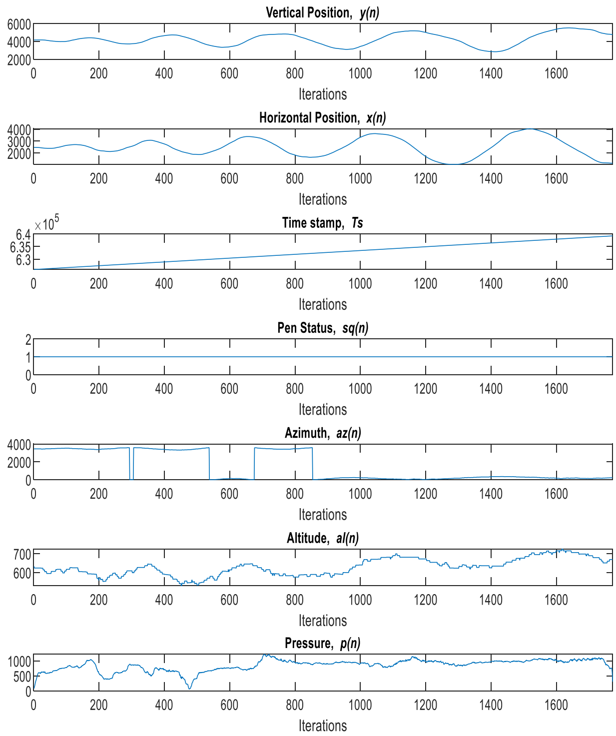


FIGURE 1. Online-drawing time series of the first task (Archimedean spiral drawings.

During the acquisition, the subjects were rested and seated in front of a table in a comfortable position. They completed a protocol on a printed template at a comfortable speed. The prefilled template was shown to the subjects; no restrictions on the number of repeated syllables/words in tasks or their heights were given. The signals were recorded at a 133 Hz sampling rate using the Intuos 4 M (Wacom technology) digitizing tablet and Wacom inking pen.

The protocol consists of the following eight tasks: Task 1 asks the user to draw, from inside out, an Archimedean spiral; tasks 2, 3, and 4 asks the user to repetitively write a cursive letter "l," syllable "le," and trigram "les," respectively; tasks 5, 6, and 7 asks the user to repetitively write a simple orthography and an easy syntax word, such that they are written in one continuous movement; finally, task 8 requires the user to write a longer sentence.

When the user was writing or drawing on the tablet (Fig. 1), the application captured the horizontal and vertical displacements of the pen tip in the x-axis, $x(n)$ and the y-axis, $y(n)$, respectively. Furthermore, the on-surface/in-air pen position information or status (touching/not-touching tablet's surface), $sq(n)$, the altitude of the pen with respect to the tablet's surface, $al(n)$, the pressure applied by the pen tip, $p(n)$, the azimuth angle of the pen with respect to the tablet's surface, $az(n)$, and the signal's timestamp, Ts, were captured.

A. DATA AUGMENTATION

Since the training database is small and unbalanced, we augment the smaller class such that both are equal in size. Then, we augment both classes to increase the training set.

Augmentation of the training data is performed as follows:

1. C_m = Identify the class with few samples.
2. N_s = Calculate the number of samples to compensate for the different number of samples.
3. Randomly select N_s of C_m.
4. For each selected sample, calculate the new feature vector by adding Gaussian random noise to the original features.

$$\mathrm{FV}_a = \mathrm{FV}_{u*} + \alpha * GV$$

where FV_{u*} is the feature vector of a random user, α is a value less than 0.2, and GV a vector with Gaussian random values. FV_a, FV_{u*}, and GV are vectors with equal dimensions.

V. FEATURE EXTRACTION

The front-end used in this study is shown in Fig. 2. This section describes the kinematic, statistics, spectral- and cepstral domain features used in the front-end. Definitions for calculating these features are provided in the next subsections and its graphical representation is shown in Fig. 3.

A. TEMPORAL AND KINEMATIC FEATURES

The row vector of temporal and kinematic features (KF) [34] for task τ and user u, applied to displacement, is defined as follows:

$$TKF^{\tau,u}_{d^{\tau,u}(n)} = [\mathcal{S}^{\tau,u}_1, F^{\tau,u}_1, F^{\tau,u}_2, \dot{F}^{\tau,u}_1, \dot{F}^{\tau,u}_2, NCV^{\tau,u}, NCA^{\tau,u}, NCV^{\tau,u}_r, NCA^{\tau,u}_r, \mathscr{r}^{\tau,u}],$$

where,

$d^{\tau,u}(n) = \sqrt{x^{\tau,u}(n)^2 + y^{\tau,u}(n)^2}$, is the displacement,

$\mathcal{S}^{\tau,u}_1$ is the trajectory during handwriting divided by the duration of writing,

$F^{\tau,u}_1 = \sum_{n=1}^{N-1} \mathscr{d}(2n)$, this is the on-air pen duration,

$F^{\tau,u}_2 = \sum_{n=0}^{N-1} \mathscr{d}(2n+1)$, this is the on-paper pen duration,

$\mathscr{d}(i)$ is the duration of the stroke i; when $i \bmod 2 = 0$, the pen is on- air, otherwise it is on the tablet surface,

$\dot{F}^{\tau,u}_1 = F^{\tau,u}_1/T$ represents the $F^{\tau,u}_1$ normalized to writing duration,

$\dot{F}^{\tau,u}_2 = F^{\tau,u}_2/T$ represents the $F^{\tau,u}_2$ normalized to writing duration,

$\mathscr{r}^{\tau,u}$ is the ratio of time the pen spent in-air/on the tablet's surface,

$NCV^{\tau,u} = 1/(n-1)\sum_{i=1}^{N-1} |v(i) - v(i+1)|$ represents the number of changes in velocity direction (The mean number of local extrema of velocity),

$NCA^{\tau,u} = 1/(n-2)\sum_{i=1}^{N-2} |a(i) - a(i+1)|$ represents the number of changes in acceleration direction (The mean number of local extrema of acceleration),

$NCV^{\tau,u}_r = NCV^{\tau,u}/(T-1)$ represents the $NCV^{\tau,u}$ relative to writing duration,

$NCA^{\tau,u}_r = NCA^{\tau,u}/(T - 2 * Ts)$ represents the $NCA^{\tau,u}$ relative to writing duration, Ts is the sampling time

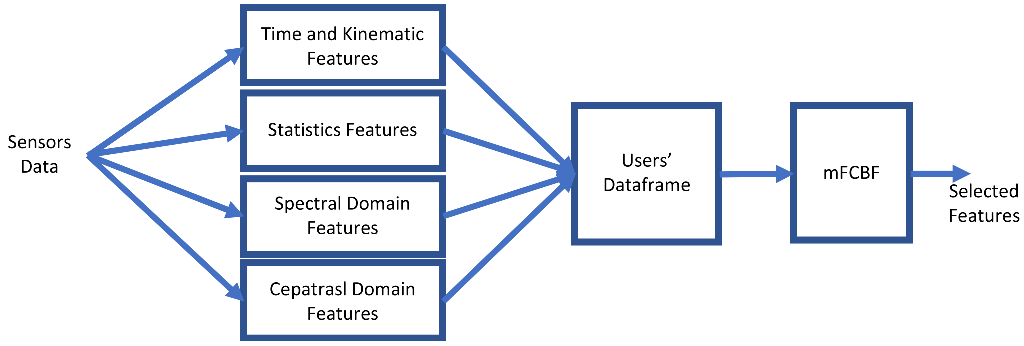


FIGURE 2. System Front-end mFCBF is the modified Fast Correlation-Based Filtering.

and $\mathcal{T} = spiral, \ letterl$, syllable le, $trigrammles$, $word1$, $word2$, $word3$, $sentence$ is the set of tasks to perform.

Then, the TKF row vector is the concatenation of TKF of each task τ; mathematically shown below, using relational algebra:

$$TKF^{U} = \left[\bigcup_{\tau \in \mathcal{T}} \left[TKF^{\tau,u}_{d(n)} \right]^{\mathrm{T}} \right]^{\mathrm{T}}.$$

We observe that to concatenate columns, we transpose the row vector, then we append the resulting column vectors using the union function. Finally, the row feature vector is obtained by transposing the column vector.

B. STATISTICS FEATURES

The statistics are obtained from the kinematic and stroke signals [39]. First, consider the set for task τ and user u:

$$\mathcal{g}^{\tau,u}(n) = \left\{ k^{\tau,u}_{w}(n), \mathfrak{s}^{\tau,u}(n) \right\},$$

where
$n = 1, \ldots, T$,
$\mathfrak{s}^{\tau,u}(n)$ is the stroke signal,
$k^{\tau,u}_{w}(n) = \{v^{\tau,u}_{w}(n), a^{\tau,u}_{w}(n), j^{\tau,u}_{w}(n)\}$ is the set of kinematic signals, applied to signal in set $w^{\tau,u}(n)$,
$w^{\tau,u}(n) = \{d^{\tau,u}(n), x^{\tau,u}(n), y^{\tau,u}(n)$ is the set containing discrete, horizontal, and vertical displacements,
$d^{\tau,u}(n) = \sqrt{x^{\tau,u}(n)^2 + y^{\tau,u}(n)^2}$, is the discrete displacement,
$x^{\tau,u}(n)$ is the horizontal displacement,
$y^{\tau,u}(n)$ is the vertical displacement,
$v^{\tau,u}_{w}(n) = \frac{w^{\tau,u}(n) - w^{\tau,u}(n-1)}{T_s}$, is the velocity applied to signals in $w^{\tau,u}(n)$,
$a^{\tau,u}_{w}(n) = \frac{v^{\tau,u}_{w}(n) - v^{\tau,u}_{w}(n-1)}{T_s}$, is the acceleration applied to signals in $w^{\tau,u}(n)$,
$j^{\tau,u}_{w}(n) = \frac{a^{\tau,u}_{w}(n) - a^{\tau,u}_{w}(n-1)}{T_s}$, is the jerk applied to signal in $w^{\tau,u}(n)$, and $\mathcal{T} = spiral, \ letterl$, syllable le, $trigrammles$, $word1$, $word2, word3$, $sentence$ is the set of tasks performed for each user.

Statistics features row vector [Drotar *et al.*, 2014; 2016] for task τ and user u, is defined as follows:

$$SF^{\tau,u}_{\mathcal{g}} = \left[\mathcal{B}^{\tau,u}_{\mathcal{g}}, \mathfrak{M}^{\tau,u}_{\mathcal{g}}, \mathcal{M}^{\tau,u}_{\mathcal{g}} \right],$$

where
$\mathcal{B}^{\tau,u}_{\mathcal{g}(n)}$ is the row vector of basic statistics features,
$\mathfrak{M}^{\tau,u}_{\mathcal{g}(n)}$ is the row vector of mean features and
$\mathcal{M}^{\tau,u}_{\mathcal{g}(n)}$ is the row vector of momentum features.
They are all applied to all signals in $\mathcal{g}^{\tau,u}(n)$.
The row vector of basic statistics features is defined as follows:

$$\mathcal{B}^{\tau,u}_{\mathcal{g}} = \left[\overleftrightarrow{\mathcal{g}}^{\tau,u}, \breve{\mathcal{g}}^{\tau,u}, \ddot{\mathcal{g}}^{\tau,u}, \dddot{\mathcal{g}}^{\tau,u}, \tilde{\dddot{\mathcal{g}}}^{\tau,u} \right],$$

where
$\overleftrightarrow{\mathcal{g}}^{\tau,u}$ is the range,
$\breve{\mathcal{g}}^{\tau,u}$ is the median,
$\ddot{\mathcal{g}}^{\tau,u}$ is the mode,
$\dddot{\mathcal{g}}^{\tau,u} = \left(1/n \sum_{n=1}^{n=T} (\mathrm{g}^{\tau,u}(n) - \bar{\mathrm{g}}^{\tau,u}) \right)^{1/2}$ is the standard deviation
$\overleftrightarrow{\dddot{\mathcal{g}}}^{\tau,u}$ is the outlier robust range (percentile 99th–percentile 1st); all above definitions applied to all signals in set $\mathcal{g}^{\tau,u}(n)$, and
$\mathcal{T}$ is the set of tasks to perform.
The row vector of mean features is defined as follows:

$$\mathfrak{M}^{\tau,u}_{g} = \left[\bar{\mathcal{g}}^{\tau,u}, \bar{\bar{\mathcal{g}}}^{\tau,u}, \overbrace{\mathcal{g}^{\tau,u}}^{tri} \right],$$

where
$\bar{\mathcal{g}}^{\tau,u} = 1/n \sum_{n=1}^{n=T} \mathcal{g}^{\tau,u}(n)$ is the arithmetic mean,
$\bar{\bar{\mathcal{g}}}^{\tau,u} = \left(\prod_{t=1}^{t=T} \mathcal{g}^{\tau,u}(n) \right)^{1/n}$ is the geometric mean,
$\overbrace{\mathcal{g}^{\tau,u}}^{tri} = \bigcup \overbrace{\mathcal{g}^{\tau,u}_{i}}^{tri} \forall i = 5, 10, 20, 30, 40, 50$, is the set of trimmed means for each of the values in i of $\mathcal{g}^{\tau,u}(n)$; the trimmed mean is the mean after removing the outliers. For example, suppose $\tilde{\mathcal{g}}^{\tau,u}_{i}$ has n values, the trimmed mean is the mean of $\tilde{\mathcal{g}}^{C}_{i}$ excluding the highest and lowest k data values, where $k = n * (i/100)/2.$; all above definitions applies to all signals in $\mathcal{g}^{\tau,u}(n)$, and
$\mathcal{T}(asdefined)$.

The row vector of momentum statistics features is defined as follows:

$$\mathcal{M}_g^{\tau,u} = \left[\overbrace{g^{\tau,u}}^{qua} \overbrace{g^{\tau,u}}^{per} \overbrace{g^{\tau,u}}^{mom} \mathcal{k}^{\tau,u} \right],$$

where

$\overbrace{g^{c}}$ *is* the row vector of quartiles ($Q_3 = 25(lower, Q_1 = 75/upper)$,

$\overbrace{g^{\tau,u}}^{per} = \bigcup \overbrace{g_i^{\tau,u}}^{per}, \forall i = 1, 5, 10, 20, 30, 90, 95, 100$, is the row vector of percentils,

$\overbrace{g^{\tau,u}}^{mom} = \bigcup \overbrace{g_i^{\tau,u}}^{mom}, \forall i = 3th, 4th, 5th, 6th$, is the row vector of i moments,

$\mathcal{k}^{\tau,u} = 1/\sigma^4 \left(\sum_{n=1}^{n=T} \left(g^{\tau,u}(n) - \bar{g}^{\tau,u} \right) \right)^{1/4}$, is the kurtosis; all above definitions applies to all signals in $g^{\tau,u}(n)$, and $\mathcal{T} = \{spiral, letterl,$ syllable$le, trigrammles, word1, word2, word3, sentence$

Then, the row vector of the statistics feature for user U, using relational algebra, is shown below:

$$SF^U = \left[\bigcup_{\tau \in \mathcal{T}} \bigcup_{\mathcal{G} \in g^C(n)} \left[SF_{\mathcal{G}}^{\tau,u} \right]^{\mathrm{T}} \right]^{\mathrm{T}}.$$

C. SPECTRAL-DOMAIN FEATURES

Spectral-domain feature row vectors for task τ and user u, applied to signals is $s^{\tau,u}(n)$, is defined as follows:

$$SDF_s^{\tau,u} = \left[FBCC_s^{\tau,u}(1), \ldots, FBCC_s^{\tau,u}(M) \right]$$

where

$LEFB_s^{\tau,u}(m) = \underset{\theta=1,2,\ldots,M}{filterbank}\{E_s^{\tau,u}(k)\}, \ for \ m = 1, 2, \ldots, M,$

$E_s^{\tau,u}(k) = log_2\left(|S^{\tau,u}(k)|^2\right)$, is the log energy spectrum,

$S^{\tau,u}(k) = \sum_{n=0}^{N-1} s^{\tau,u}(n) e^{-\frac{2\pi i}{N}kn}, \ for \ k = 0, 1, \ldots, K$, is the discrete Fourier transform of the signal and

$s^{\tau,u}(n) = x^{\tau,u}(n), y^{\tau,u}(n), p^{\tau,u}(n)$,

$x^{\tau,u}(n)$ is the horizontal displacement,

$y^{\tau,u}(n)$ is the vertical displacement, and

$p^{\tau,u}(n)$ is the pressure signal.

Then, the row vector of the spectrum-domain features is the concatenation of the SDF of each of the task τ for each signal in $s^{\tau,u}(n)$:

$$SDF^U = \left[\bigcup_{\tau \in \mathcal{T}} \bigcup_{\mathcal{S} \in s} \left[SDF_{\mathcal{S}}^{\tau,u} \right]^{\mathrm{T}} \right]^{\mathrm{T}}.$$

D. CEPSTRAL DOMAIN FEATURES

Cepstral domain feature row vectors for task τ and user u, applied to signals in $s^{\tau,u}(n)$, is defined as follows:

$$CDF_s^{\tau,u} = \left[LEFB_s^{\tau,u}(1), \ldots, LEFB_s^{\tau,u}(M) \right]$$

where

$$FBCC_s^{\tau,u}(q) = \sum_{m=0}^{M-1} LEFB_s^{\tau,u}(m) e^{-\frac{2\pi i}{N}qm},$$

$$For \ q = 1, 2, \ldots, Q.$$

M is the number of Filterbanks, $Q = M/2$ is the number of filterbanks, in the number of filterbanks divided by 2.

Then, the row vector of the cepstral domain features for user U, again using relational algebra, is shown below:

$$CDF^U = \left[\bigcup_{\tau \in \mathcal{T}} \bigcup_{\psi \in \Psi} \left[FBCC_s^{\tau,u}(q) \right]^{\mathrm{T}} \right]^{\mathrm{T}}.$$

E. USERS FEATURE

The row feature vector FV_u of each user, using relational algebra is shown as follows:

$$FV^{\mathbf{u}} = \left[\left[TKF^u \right]^T \cup \left[SF^u \right]^T \cup \left[SDF^u \right]^T \cup \left[CF^u \right]^T \right]^{\mathrm{T}}.$$

Alternatively, we define the disease state for each user as follows:

$$\mathcal{D}^u = \begin{cases} 0, & Normal \\ 1, & above\ Normal \end{cases} \ for \ all \ u = 1 \ldots U,$$

The row vector relating features to the emotional state is

$$FV\mathcal{D}^u = \left[\left[FV^{\mathbf{u}} \right]^{\mathrm{T}} \cup \left[\mathcal{D}^u \right]^{\mathrm{T}} \right]^{\mathrm{T}}$$

The data frame is defined as the union $FV\mathcal{D}^u$ of all users, and can be expressed using relational algebra notation as follows:

$$FV\mathcal{D} = \bigcup_{u=1}^{U} FV\mathcal{D}^u.$$

In this dataframe, the rows represent the number of users and the columns represent the features and its users' disease state.

VI. FEATURE SELECTION

Feature selection is a popular and common premodeling step in machine learning, especially in high-dimensional databases. Irrelevant features decrease the accuracy of data models because models also learn irrelevant information. Thus, selecting the right number of features increases the performance of the machine learning method.

Several methods for selecting features exist. All of them aim to obtain the best features and most do so by employing statistical tools with certain correlations to selection. The major difference between these tools is the selection criterion. Each patient has a considerable number of features, so we reduce the dimension of the number of features using a modified fast correlation-based filtering (FCBF) [40].

FCBF is based on two correlation factors: correlation between each feature and output and correlations among major difference between these tools is the selection criterion.

Each patient has a considerable number of features, so we reduce the dimension of the number of features using a modified fast correlation-based filtering (FCBF) [40].

FCBF selection is based on two steps. In the first step, the selected features are the ones whose correlation with the output are higher than a threshold. In the second step,

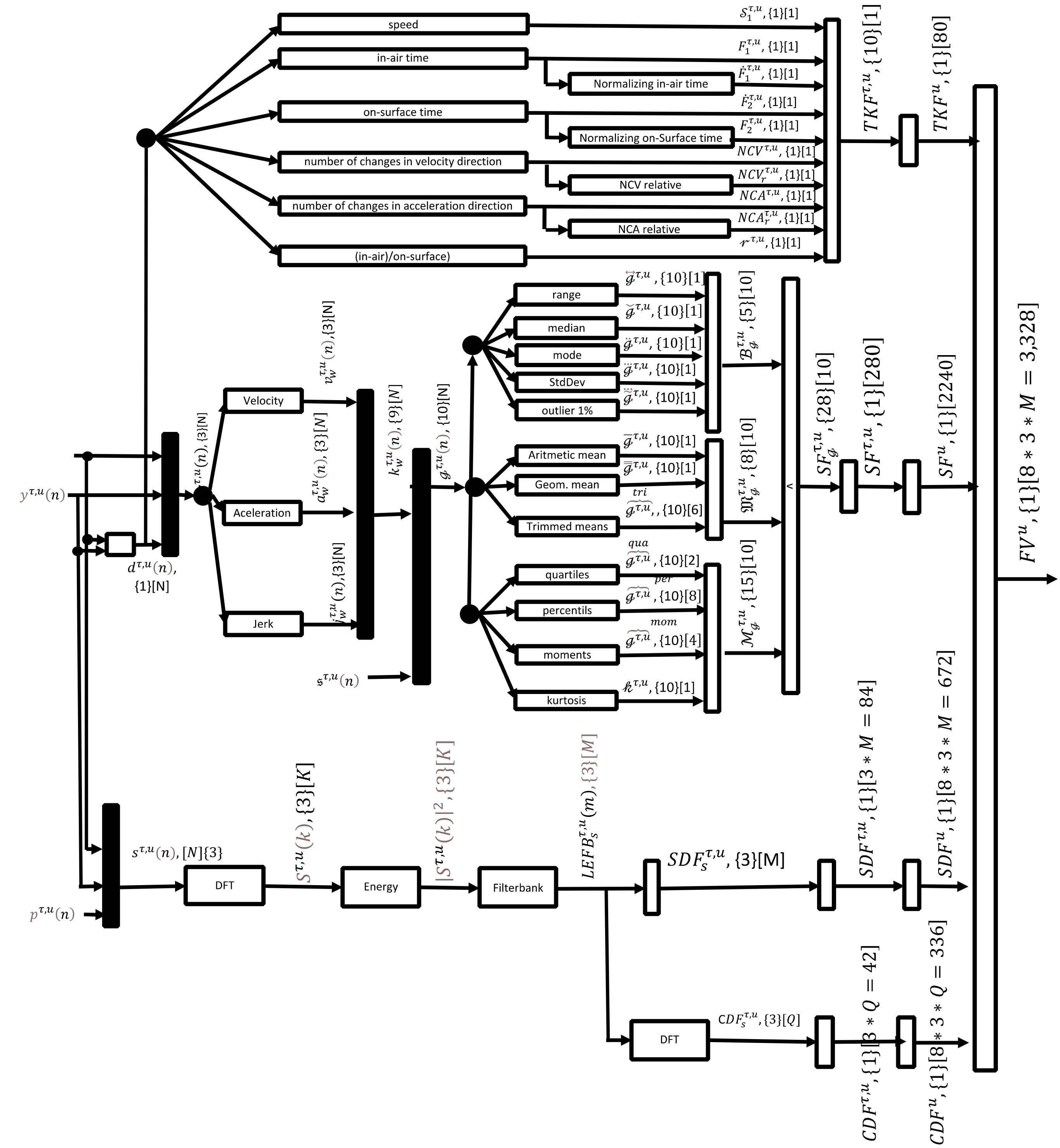


FIGURE 3. **Feature Processing. Notation: {} indicates the number of members in the set; [] indicates the vector dimension of each element of the set. Black vertical rectangles indicate that the output is a set of the inputs; no-filled vertical rectangles indicate the outputs is a column concatenation of the inputs vectors. The processing of the horizontal rectangles is made for each element of the input set, creating an output of the same dimension as the input. In this figure, we assume a filterbanks dimension of 28 (M = 28) and cepstral coefficients number of 14 (Q = 14).**

it takes the features of the first steps and selectes the ones with correlation lower than a threshold. In our modified version [36], mFCBF differs of the original FCBF at step 5, where the selected feature is the one having higher correlation with the output. Algorithm 1 shows a pseudo-code for the mFCBF process. mFCBF algorithm receives, as inputs, a dataframe and thresholds *oTh* and *iTh*. *oTh* is used to set the lower correlation value of each of the selected features and the output; *iTh* is used to set the higher correlation value between features. Using the values of *oTh* and *iTh*, we can

Algorithm 1 The mFCBF Algorithm Receives the Users Feature Matrix ($\mathbf{O^E}$), Minimum Correlation Threshold ($\mathbf{oTh}$), and the Maximum Correlation Threshold ($\mathbf{iTh}$) and Returns the Selected Set of Features

1: **function** mFCBF($\mathbf{O}$, $\mathbf{oTh}$, $\mathbf{iTh}$)
2: Calculate corr($\mathbf{O}$)
3: $\mathbf{O_{tmp}}$ $\leftarrow$ Select columns whose output correlation is > $\mathbf{oTh}$
4: Calculate corr($\mathbf{O_{tmp}}$)
5: $\hat{O}_{\mathbf{Th,iTh}}$ $\leftarrow$ Select columns whose correlation with the input is < $\mathbf{iTh}$ and with the highest correlation with the output.
6: **return** ($\hat{O}_{\mathbf{Th,iTh}}$)
7: **end function**

find the right features to maximize the performance of the machine learning method. This operation is expressed as follows:

$$\widehat{FV\mathcal{D}}_{oTh,iTh} = mFCBF_{oTh,iTh}\,(FV\mathcal{D})\,.$$

Note that in $\widehat{FV\mathcal{D}}_{oTh,iTh}$ is a 2-D array, where one dimension represents the number of users and the other, the number of selected features.

Feature selectivity is controlled with oTh and iTh values. For example, given 370 user feature vectors, then for $iTh = 0.15$ and $oTh = 0.7$, the number of selected features reduce to 26, 28, and 20 for the depression, anxiety, stress states, respectively.

One way to visualize the relevance of features in improving performance is to use RadViz [41]. In RadViz, each data frame sample is represented inside the circle using the value in each series according to a physical metaphor. Each point is attached to each characteristic with a force proportional to the value the sample takes in the corresponding series. This implies that the final position is the equilibrium position between all forces representing the characteristics. Figs. 2 and 3 show the RadViz of the 2658 features and the selected 47 features, respectively. RadViz shows the dominant proportional values (DPV) of the features. In the graph, the higher cloud dispersion means higher DPV, whereas, a higher DPV means that features are easily exploited to improve the classification. Furthermore, these 47 features have higher DPVs than the complete set of 2658 features.

VII. FRONT-END HYPERPARAMETERS

Spectral-domain features (SDF) is a function of the filterbank bandwidth ($fbbw$), the bandwidth of the filters on the filterbank (fbw), the filterbank's initial frequency (if), and the overlap between filters on the filterbank (ov).

Conversely, features selection (FS) depends on the feature-output-correlation threshold (oTh) and the intra-feature correlation threshold (iTh).

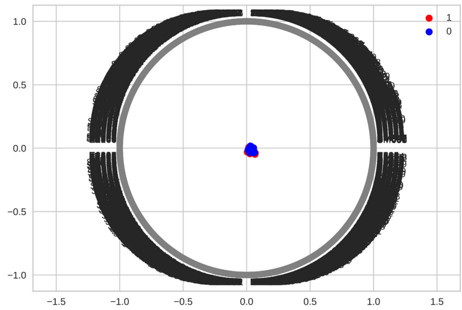


FIGURE 4. RadViz of the 2658 features. We can observe that there are no features with dominant proportional values (DPVs).

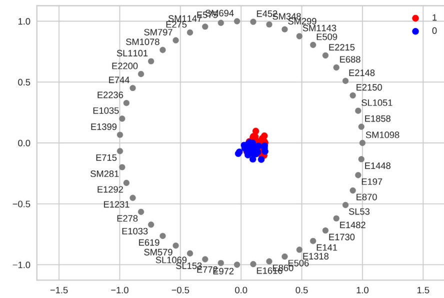


FIGURE 5. RadViz of 47 selected features. Note that selecting features increases the number of features with dominant proportional values (DPVs).

Therefore, the parameters for the final vector of features are

$$(fbbw, fbw, if, ov, oTh, iTh)$$

For practical, the range of values for each parameter is defined as follows:

$$\begin{aligned} iTh_{range} &= [0.2-1]\,,\\ oTh_{range} &= [0-0.20]\,,\\ fbbw_{range} &= \;[075],\;\text{ in Hz},\\ fbw_{range} &= [0.5-3]\,,\;\text{ in Hz};\\ if_{range} &= [0.5]\;\text{ and}\\ ov_{range} &= [0]\;\text{ in }\%. \end{aligned}$$

A different set of features is selected for each combination of values. More so, each set of features produces a corresponding performance accuracy. Since one of these combinations is optimal, we find the combination that maximizes the ML accuracy.

Since augmentation of the training data, user selection, and Gaussian noise is random, we are unaware of which users and random sequences generate a better model. Therefore, we train and test the model for different sets of users and different random sequences, and we select the maximum accuracy.

VIII. RESULTS

The Leave-Percentage-Out (LPO) was used for testing. Here, the data model is trained with all database registers but a percentage, and the test is performed on registers that were out.

TABLE 3. Accuracy (%) results for different sets of coefficients with(FS)/without(NO_FS) feature selection applying mFCBF as features selection.

Features	NO_FS	FS
KF	80	88.57
$KF \cup SF$	80	94.28
$KF \cup SF \cup SDF$	82.85	97.14
$\boldsymbol{KF \cup SF \cup SDF \cup CDF}$	88.57	**98.57**

This was repeated until all possibilities were checked, then, we averaged the accuracy of all tests. In our experiments, we leave the 15% out.

We sample different filterbank's hyperparameters as follows:

$$\begin{aligned} fbbw^s_{range} &= [15, 20, \ldots 50], \text{ in Hz}, \\ fbw^s_{range} &= [0.2, 0.3, , \ldots, 1.0], \\ fi_{range} &= [0.5] \text{ and} \\ ov_{range} &= [0]. \end{aligned}$$

and, the mFCBF hyperparameters (oTh and iTh) are given as follows:

$$\begin{aligned} oTh^s_{range} &= [0, 0.02, 0.04, 0.06, \ldots, 0.18, 0.20], \\ iTh^s_{range} &= [0.2, 0.30, \ldots, 1.0], \end{aligned}$$

Therefore, we find the combination of this sample space that maximizes ML accuracy.

Table 3 shows the different accuracies for different feature sets. The accuracy results for TKF, when using feature selection or not, are 88.87% and 80%, respectively. The accuracy results for concatenating SF and using either feature selection or not are 94.28% and 80%, respectively.

From these two experiments, we find that adding statistics feature when combined with TKF improves the result accuracy.

Table 3 shows that the accuracy of the results when concatenating SDF using either feature selection or not are 97.14% and 82.85%, respectively. The last accuracy result is for concatenating CDF using either feature selection or not, are 98.57% and 88.57%, respectively.

The training data for all experiments were augmented by 80%, and the amplitude of the random Gaussian (α) was set to 0.2.

IX. CONCLUSION AND FUTURE WORK

We applied spectral and cepstral features on Parkinson's disease detection. Although spectral and cepstral features have been successfully applied for emotion detection, here, we concatenate them with kinetic and statistical features.

Similar features were used in [52] without the filterbank, thereby providing the flexibility for changing filterbank's bandwidth, filterbank's filters bandwidth, filterbank's filters overlapping, and filterbank's initial frequency to improve performance.

As the first step, we calculated TKF using the displacement signals; SF using displacement, and horizontal and vertical displacement; the SDF and CDF from the displacement and the horizontal and vertical displacement, and pressure signals.

Since the employed dataset (PaHaW) contains 37 PD patients and 38 HC subjects, then as a second step, we augmented the smaller class of the training set such that both are equal in size; next, we augment both classes of the training data by randomly selecting 80% of the training patient's data and added random Gaussian noise in all augmentations. For the third step, we selected the most relevant features using mFCBF method. Finally, autoML was employed to train and test more than ten plain and ensembled classifiers.

Experimental results show that adding spectral and cepstral features to the kinematics and statistics features highly improves the classification accuracy, reaching a combined classification accuracy of 98.57%. This result shows that our proposed model outperforms the best state-of-the-art result, which sits at 97.62%. Moreover, the state-of-the-art model has higher computational complexity and is required to train 204,060 parameters model for one day using an NVIDIA GTX 1080 GPU of 8 GB.

REFERENCES

[1] M. Faundez-Zanuy, J. Fierrez, M. A. Ferrer, and R. Plamondon, "Handwriting biometrics: Applications and future trends in e-security and e-health," *Cogn. Comput.*, vol. 12, pp. 940–953, Aug. 2020.

[2] R. Plamondon and S. N. Srihari, "Online and off-line handwriting recognition: A comprehensive survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 63–84, Jan. 2000.

[3] S. Z. Li and A. K. Jain, *Encyclopedia of Biometrics*. New York, NY, USA: Springer, 2015.

[4] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. S. di Freca, "Handwriting analysis to support neurodegenerative diseases diagnosis: A review," *Pattern Recognit. Lett.*, vol. 121, pp. 37–45, Apr. 2018.

[5] D. Impedovo and G. Pirlo, "Dynamic handwriting analysis for the assessment of neurodegenerative diseases: A pattern recognition perspective," *IEEE Rev. Biomed. Eng.*, vol. 12, pp. 209–220, 2018.

[6] G. Vessio, "Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review," *Appl. Sci.*, vol. 9, no. 21, p. 4666, Nov. 2019.

[7] W. Poewe, K. Seppi, C. M. Tanner, G. M. Halliday, P. Brundin, J. Volkmann, A. E. Schrag, and A. E. Lang, "Parkinson disease," *Nat. Rev. Dis. Prim.*, vol. 3, no. 17013, pp. 1–21, 2017.

[8] R. B. Postuma, D. Berg, M. Stern, W. Poewe, C. W. Olanow, W. Oertel, J. Obeso, K. Marek, I. Litvan, A. E. Lang, G. Halliday, C. G. Goetz, T. Gasser, B. Dubois, P. Chan, B. R. Bloem, C. H. Adler, and G. Deuschl, "MDS clinical diagnostic criteria for Parkinson's disease," *Movement Disorders Off. J. Movement Disorder Soc.*, vol. 30, no. 12, pp. 1591–1601, Oct. 2015.

[9] J. Stochl, A. Boomsma, E. Ruzicka, H. Brozova, and P. Blahus, "On the structure of motor symptoms of Parkinson's disease," *Movement Disorders*, vol. 23, no. 9, pp. 1307–1312, 2008.

[10] J. Jankovic, "Parkinson's disease: Clinical features and diagnosis," *J. Neurol., Neurosurg. Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[11] J. Rusz, T. Tykalova, L. O. Ramig, and E. Tripoliti, "Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders," *Movement Disorders*, vol. 36, no. 4, pp. 803–814, Apr. 2021.

[12] S. Argaud, M. Vérin, P. Sauleau, and D. Grandjean, "Facial emotion recognition in Parkinson's disease: A review and new hypotheses," *Movement Disorders*, vol. 33, no. 4, pp. 554–567, Apr. 2018.

[13] H.-I. Ma, W.-J. Hwang, S.-H. Chang, and T.-Y. Wang, "Progressive micrographia shown in horizontal, but not vertical, writing in Parkinson's disease," *Behavioural Neurol.*, vol. 27, no. 2, pp. 169–174, 2013.

[14] A. Letanneux, J. Danna, J.-L. Velay, F. Viallet, and S. Pinto, "From micrographia to Parkinson's disease dysgraphia," *Movement Disorders*, vol. 29, no. 12, pp. 1467–1475, Oct. 2014.
[15] L. Brabenec, J. Mekyska, Z. Galaz, and I. Rektorova, "Speech disorders in Parkinson's disease: Early diagnostics and effects of medication and brain stimulation," *J. Neural Transmiss.*, vol. 124, no. 3, pp. 303–334, Mar. 2017.
[16] L. Moro-Velazquez and N. Dehak, "A review of the use of prosodic aspects of speech for the automatic detection and assessment of Parkinson's disease," in *Automatic Assessment of Parkinsonian Speech* (Communications in Computer and Information Science). Cham, Switzerland: Springer, Sep. 2020, pp. 42–59.
[17] L. Moro-Velazquez, J. A. Gomez-Garcia, J. D. Arias-Londoño, N. Dehak, and J. I. Godino-Llorente, "Advances in Parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102418.
[18] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Syst. Appl.*, vol. 137, pp. 22–28, Dec. 2019.
[19] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 4, pp. 828–834, Jul. 2013.
[20] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," *IEEE J. Transl. Eng. Health Med.*, vol. 7, pp. 1–10, 2019.
[21] E. Balaji, D. Brindha, and R. Balakrishnan, "Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease," *Appl. Soft Comput.*, vol. 94, Sep. 2020, Art. no. 106494.
[22] D. Joshi, A. Khajuria, and P. Joshi, "An automatic non-invasive method for Parkinson's disease classification," *Comput. Methods Programs Biomed.*, vol. 145, pp. 135–145, Jul. 2017.
[23] L. Brognara, P. Palumbo, B. Grimm, and L. Palmerini, "Assessing gait in Parkinson's disease using wearable motion sensors: A systematic review," *Diseases*, vol. 7, no. 1, p. 18, Feb. 2019.
[24] A. Ammour, I. Aouraghe, G. Khaissidi, M. Mrabti, G. Aboulem, and F. Belahsen, "A new semi-supervised approach for characterizing the Arabic on-line handwriting of Parkinson's disease patients," *Comput. Methods Programs Biomed.*, vol. 183, Jan. 2020, Art. no. 104979.
[25] L. Ali, C. Zhu, N. A. Golilarz, A. Javeed, M. Zhou, and Y. Liu, "Reliable Parkinson's disease detection by analyzing handwritten drawings: Construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model," *IEEE Access*, vol. 7, pp. 116480–116489, 2019.
[26] C. Taleb, L. Likforman-Sulem, C. Mokbel, and M. Khachab, "Detection of Parkinson's disease from handwriting using deep learning: A comparative study," *Evol. Intell.*, pp. 1–12, Sep. 2020.
[27] U. Gupta, H. Bansal, and D. Joshi, "An improved sex-specific and age-dependent classification model for Parkinson's diagnosis using handwriting measurement," *Comput. Methods Programs Biomed.*, vol. 189, Jun. 2020, Art. no. 105305.
[28] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," *Artif. Intell. Med.*, vol. 67, pp. 39–46, Feb. 2016.
[29] I. Aouraghe, A. Alae, K. Ghizlane, M. Mrabti, G. Aboulem, and B. Faouzi, "A novel approach combining temporal and spectral features of Arabic online handwriting for Parkinson's disease prediction," *J. Neurosci. Methods*, vol. 339, Jun. 2020, Art. no. 108727.
[30] E. D. Deharab and P. Ghaderyan, "Parkinson's disease detection using dynamic writing traces warping," in *Proc. 28th Iranian Conf. Electr. Eng. (ICEE)*, Aug. 2020, pp. 1–4.
[31] A. Parziale, R. Senatore, A. Della Cioppa, and A. Marcelli, "Cartesian genetic programming for diagnosis of Parkinson disease through handwriting analysis: Performance vs. interpretability issues," *Artif. Intell. Med.*, vol. 111, Jan. 2021, Art. no. 101984.
[32] R. Lamba, T. Gulati, K. A. Al-Dhlan, and A. Jain, "A systematic approach to diagnose Parkinson's disease through kinematic features extracted from handwritten drawings," *J. Reliable Intell. Environ.*, vol. 7, no. 3, pp. 253–262, Sep. 2021.
[33] M. Diaz, M. Moetesum, I. Siddiqi, and G. Vessio, "Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and BiGRUs," *Expert Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114405.
[34] D. Impedovo, "Velocity-based signal features for the assessment of Parkinsonian handwriting," *IEEE Signal Process. Lett.*, vol. 26, no. 4, pp. 632–636, Apr. 2019.
[35] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, K. Lopez-de-Ipina, V. Zvoncak, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Identification and monitoring of Parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting," *Appl. Sci.*, vol. 8, no. 12, p. 2566, Dec. 2018.
[36] J. A. Nolazco-Flores, M. Faundez-Zanuy, O. A. Velázquez-Flores, G. Cordasco, and A. Esposito, "Emotional state recognition performance improvement on a handwriting and drawing task," *IEEE Access*, vol. 9, pp. 28496–28504, 2021.
[37] *AutoML in H2O.Ai*. H2O.Ai. Accessed: Aug. 11, 2021. [Online]. Available: https://www.h2o.ai/products/h2o-automl/
[38] *The H2O Python Module*. The H2O Python Module—H2O Documentation. Accessed: Aug. 11, 2021. [Online]. Available: https://docs.h2o.ai/h2o/latest-stable/h2o-py/docs/intro.html
[39] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease," *Comput. Methods Programs Biomed.*, vol. 117, no. 3, pp. 405–411, 2014.
[40] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proc. 12th Int. Conf. Mach. Learn. (IGML)*, 2003, pp. 856–863.
[41] *RadViz Visualizer*. RadViz Visualizer—Yellowbrick v1.3.Post1 Documentation. Accessed: Aug. 11, 2021. [Online]. Available: https://www.scikit-yb.org/en/latest/api/features/radviz.html
[42] M. T. Angelillo, D. Impedovo, G. Pirlo, and G. Vessio, "Performance-driven handwriting task selection for Parkinson's disease classification," in *AI*IA 2019—Advances in Artificial Intelligence* (Lecture Notes in Computer Science), 2019, pp. 281–293.
[43] M. Diaz, M. A. Ferrer, D. Impedovo, G. Pirlo, and G. Vessio, "Dynamically enhanced static handwriting representation for Parkinson's disease detection," *Pattern Recognit. Lett.*, vol. 128, pp. 204–210, Dec. 2019.
[44] J. Mucha, M. Faundez-Zanuy, J. Mekyska, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de-Ipina, "Analysis of Parkinson's disease dysgraphia based on optimized fractional order derivative features," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2019, pp. 1–5.

JUAN A. NOLAZCO-FLORES (Senior Member, IEEE) was born in Gral. Terán, Nuevo León, Mexico. He received the B.Sc. and M.Sc. degrees from the Tecnológico de Monterrey, Monterrey, México, in 1986 and 1988, respectively, and the M.Phil. and Ph.D. degrees from the University de Cambridge, Cambridge, U.K., in 1992 and 1995, respectively.

From 2017 to 2020, he was the Dean of the School of Engineering and Science, South Region, Tecnológico de Monterrey. From 2011 to 2017, he was the Head of Information Technology, Tec of Monterrey. He was a Visiting Scholar with Marburg University and Mannheim University, Germany; Carnegie Mellon University, USA; Zaragoza University, Spain; and Cambridge University, U.K. He is currently the Head of the Data Science Hub, Tec of Monterrey. He is the author of more than 60 articles indexed in the Scopus citation report. His research interests include signal processing and data science applied to speech, security, and health.

Dr. Nolazco-Flores is a member of the Mexican Academy of Science, a member of the Mexican Research System, and an ACM Senior Member. He was awarded in 2005 and 2009 the ITESM Teaching and Research Award. He is a certified ABET Evaluator, and has participated as an Evaluator of the computer science programs at the University of West Florida; Grand Canyon University, Phoenix, NV, USA; and University of Iowa, Iowa City, IA, USA.

MARCOS FAUNDEZ-ZANUY was born in Barcelona, Spain. He received the Ph.D. degree from the Polytechnic University of Catalunya, in 1998.

From 2009 to 2018, he was the Dean of Escola Superior Politecnica Tecnocampus (the Polytechnic University of Catalunya, in 2015, and Pompeu Fabra University, after 2015. From 2010 to 2019, he was the Head of Research at Tecnocampus. He is currently a Full Professor at ESUP Tecnocampus Mataro, where he is the Head of the Signal Processing Group. He is the author of more than 100 articles indexed in ISI journal citation report, more than 100 conference papers, around ten books, and responsible for ten national and European research projects. His research interests include biometrics applied to security and health.

Dr. Faundez-Zanuy was the Initiator and the Chairperson of the European COST Action 277 Nonlinear Speech Processing, and the Secretary of COST Action 2102 Cross-Modal Analysis of Verbal and Nonverbal Communication. He is a Spanish Liaison of EURASIP.

V. M. DE LA CUEVA (Member, IEEE) was born in Xalapa, Veracruz, México. He received the B.Sc., M.Sc., and Ph.D. degrees from the Tecnológico de Monterrey, Monterrey, México, in 1986, 1993, and 1998 respectively.

From 1988 to 2021, he had been the Head of Department of Mechatronics on Campus Santa Fe; the Head of the School of Engineering; the Head of Postgraduate Studies of Engineering; and the Head of Research on Campus Ciudad de México, Tecnológico de Monterrey. He is currently the Head of the School of the Engineering and Science, Tecnológico de Monterrey, Toluca. He is the author of more than 30 articles published in journals and research congresses. He is the coauthor of three books in computer science and robotics. His research interests include artificial intelligence, robotics, and education.

JIRI MEKYSKA was born in Holesov, Czech Republic. He received the Ph.D. degree in telecommunications from the Brno University of Technology, Czech Republic, in 2014. He is currently the Head of the Brain Diseases Analysis Laboratory, Brno University of Technology, where he is also the Head of the Signal Processing Laboratory, Human–Machine Interaction Group, Department of Telecommunications. In cooperation with neurologists/psychologists from different countries, he develops diagnostic/monitoring systems focused on Lewy body diseases and developmental dysgraphia. He has coauthored more than 80 publications indexed by the Web of Science. He deals with the research of noninvasive neurodegenerative and neurodevelopmental disorders analysis based on speech, sleep, and handwriting processing.

A.6 Online Handwriting, Signature and Touch Dynamics: Tasks and Potential Applications in the Field of Security and Health

Cognitive Computation (2021) 13:1406–1421
https://doi.org/10.1007/s12559-021-09938-2

Online Handwriting, Signature and Touch Dynamics: Tasks and Potential Applications in the Field of Security and Health

Marcos Faundez-Zanuy[1] · Jiri Mekyska[2] · Donato Impedovo[3]

Received: 18 April 2021 / Accepted: 12 September 2021 / Published online: 1 October 2021

Abstract
Advantageous property of behavioural signals (e.g. handwriting), in contrast to morphological ones (e.g. iris, fingerprint, hand geometry), is the possibility to ask a user to perform many different tasks. This article summarises recent findings and applications of different handwriting/drawing tasks in the field of security and health. More specifically, it is focused on on-line handwriting and hand-based interaction, i.e. signals that utilise a digitizing device (specific devoted or general-purpose tablet/smartphone) during the realization of the tasks. Such devices permit the acquisition of on-surface dynamics as well as in-air movements in time, thus providing complex and richer information when compared to the conventional "pen and paper" method. Although the scientific literature reports a wide range of tasks and applications, in this paper, we summarize only those providing competitive results (e.g. in terms of discrimination power) and having a significant impact in the field.

Keywords Drawing · Online handwriting · Signature · Tasks · Touch dynamics

Introduction

Signature/handwriting recognition can be split into two categories: off-line and on-line [1–3]. In the former case, just the result of the signature/writing (i.e. static 2D image) is known because it is acquired after the realization (writing) process. On the other hand, online signature/writing consists of acquiring the signal during the realization process. This provides a large set of raw data:

- absolute spatial coordinates (x, y) of the tip of the pen,
- pressure exerted on the surface—of course, this value is zero when the pen is not touching the surface,
- angles of the pen: altitude and azimuth,
- time stamp of the moment where the previous values have been acquired.

When pressure is different from zero, the movement is considered to be on-surface and the whole set of information described before is acquired. When pressure is zero, the movement is considered to be in-air. If the distance from the tip of the pen to the paper surface is below one centimetre (depending on the specific acquisition tool) the whole set of information described before is acquired with the unique exception of pressure, which is always zero. A deeper discussion linked with the in-air movement can be found in our previous works [1, 3, 4].

From a pattern recognition perspective, off-line systems deal with image processing, while on-line ones with time-sequence signal processing. However, it must be argued that, so far, some solutions developed for off-line systems have been adopted to on-line ones and vice-versa [5]. An emerging and very interesting aspect discussed in this article deals with the possibility to sign and/or write and in general interact with the finger on a screen of a mobile device (e.g. smartphone or tablet) [6].

Four components are embedded in the signing/writing/drawing process [6, 7]:

- The physiological component is constituted by the writing system which includes muscles, arm, wrist, hand, fingers, etc.;

✉ Marcos Faundez-Zanuy
faundez@tecnocampus.cat

1 Escola Superior Politecnica, Tecnocampus, Avda. Ernest Lluch 32, Mataro 08302, Barcelona, Spain

2 Department of Telecommunications, Brno University of Technology, Technicka 10, Brno 61600, Czech Republic

3 Dipartimento di Informatica, Université degli Studi di Bari Aldo Moro, Via E. Orabona 4, Bari 70125, Italy

- The learned component deals with personalization, schooling, culture, habits, etc.;
- The cognitive one can be referred to mental abilities (learning, thinking, reasoning, remembering, problem-solving, decision-making, and attention);
- The contour component: given the above, some noise can be introduced due to the writing device, posture, spatial constraints, emotional state, etc.

Variations of these components are reflected into variations of the acquired signal and represent the intra-writer variability. The variability is then, typically, observed over short (day-to-day or trial-to-trial basis) or long periods (months, years, etc.). In the former case, the contour component has a major role on the overall variability [6], while in the latter one all the different components could have a significance and different impact [7]. It is quite intuitive that the handwriting signal can be used for multiple purposes: handwriting recognition [8], script recognition, drawing recognition [9], health evaluation, assessment of specific learning disabilities, gender recognition [10], fatigue detection [11], emotional state recognition [12], forensic studies, writer identification (based on signature or handwriting) and signature/writer verification.

However, handwriting does not only include the writing of cursive/capital letters or scripts, in fact drawings can be considered too. More specifically, the different handwriting tasks can be classified as [7]:

1. Simple drawing tasks: straight lines, circles, spirals, meanders, swipes, etc. These tasks are also referred to as graphomotor elements, because they represent the basic building blocks of letters;
2. Simple writing tasks: nonsense words, single characters, single tap, etc.;
3. Complex tasks: they simultaneously involve motor, cognitive, and functional issues (e.g. copying tasks, the clock-drawing task, etc.).

It has been demonstrated so far that, given a specific classification problem (e.g. writer identification, health status assessment, etc.), a specific task is more profitable than others. In fact, intuitively, given a specific writing task, one of the previous mentioned components could have a different impact on the acquired signal.

Handwriting is a cognitive task in which synchronized neuromotor orders are fired from the cortex to carry out the planned action [13]. The knowledge of these cognitive tasks performed by human brain is a milestone in the development of computerized models to simulate the human thought process in complex situations where the answers may be ambiguous and uncertain. In fact, cognitive systems include self-learning technologies that use data mining, pattern recognition and natural language processing (NLP) to mimic the way the human brain works.

Automatic handwriting-based analysis can be based on many different tasks performed by using a pen-based tool. These tasks are described and discussed in detail in section 2 along with a review of several relevant scientific works. Nevertheless, many finger-based interactions, related to handwriting, can occur on a wide set of touchscreen devices (e.g. smartphone, tablet, etc.): Section 3 reports a review of the most interesting results. Section 4 presents a re-organization and a discussion of all the different tasks (previously discussed) in terms of applications (security and/or health) also according to an effort-based taxonomy. Section 5 concludes the article.

Handwritten Tasks

Signature-Based Analysis

Figure 1 shows an example of a signature acquired with a Wacom Intuos digitizing tablet. The blue colour represents the on-surface movement, while the red colour the in-air one. The relevance of the in-air movement has been clearly described in [2, 3, 14]. Handwritten signature is the most widespread behavioural biometric trait: it has a socially accepted role as a proof of identity as well as a demonstration of the willing of the writer to accept the content of the document. For this reason, it has been extensively analysed [15]. Signature is adopted on a daily basis for commercial and banking payments/transactions and in many other sectors (e.g. express courier, education, healthcare, etc.). Several international competitions exist that facilitate a fair comparison between competing algorithms [8].

Although it is not massively used in health applications sometimes interactions appear between security and health, such as in documents signed by a user suffering from dementia or some other temporary/permanent health problem that can invalidate the signature. An example has been reported in [16]. The interesting aspect is that usually, security and health implications are present both together and cannot be considered as isolated application fields [16].

Micrographia (the abnormal progressive reduction in amplitude of letters) has been observed in the off-line and on-line signing tasks as well as in sentences of patients with Parkinson's disease [17, 18]. Signature position with respect to a dotted line (on or below) and other cognitive functions have been investigated, and it has been observed that it may be a marker of vulnerability of visuospatial abilities [19].

Recently, it has been demonstrated that, when dealing with on-line writing, velocity-related features play a very crucial role [20, 21]. A similar result has been observed on signatures when considering features related to the

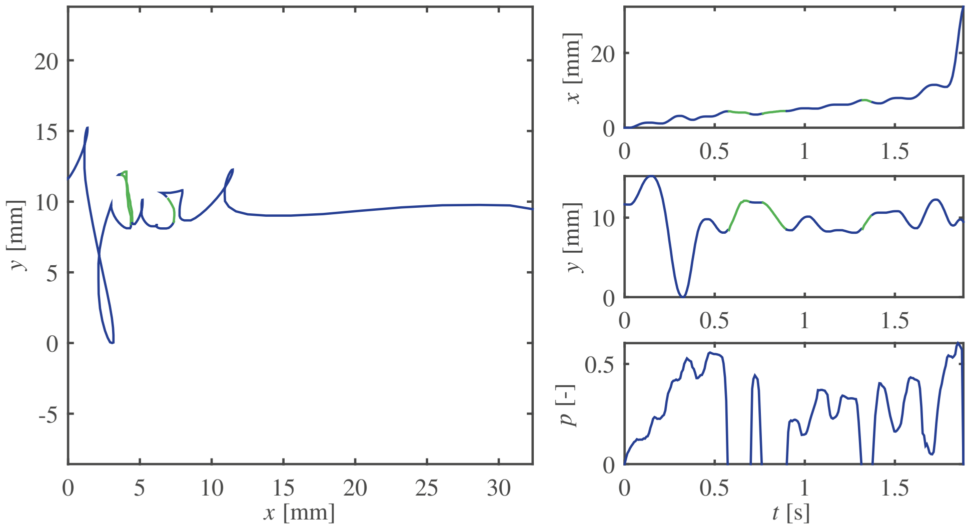


Fig. 1 Online signature: the product is depicted on the left side, the right side of the figure contains associated information about horizontal/vertical movement and pressure pattern (the blue colour represents the on-surface movement, while the green colour the in-air one)

Sigma-Lognormal model coupled with a Bagging CART classifier [22]. In this case, the approach has been able to discriminate dementia affected users from the healthy counterpart with 3% of Equal Error Rate (ERR), however the main limitation of this work is a reduced dataset. A more recent study has investigated the relation between signatures of persons with Alzheimer's disease (AD) and those written by age-matched healthy controls (HC) [23]. In this case, authors adopted a simple statistical evaluation on parameter features evaluated upon dynamic raw data (e.g. stroke duration, stroke amplitude, peak vertical velocity, average normalized jerk, etc.) and categorized signatures within three classes: text based, mixed, or stylized. No significative differences were observed apart from an association between increased feature variability and increased dementia severity for stylized and mixed signatures. Signatures were also acquired after one year during which a hard decline was observed in the cognitive status: signature features remained stable. Authors conclude the work by stating that dementia has a residual impact on signature formation. A similar finding is reported by Reiner et al. [24] who acquired two samples of signature and a spontaneous writing from 36 persons with Mild Cognitive Impairment (MCI) diagnosis and 38 HC. Cognitive functions in decision-making were also evaluated: while a significant correlation between spontaneous writing and neuropsychological test results was observed, signature deterioration did not appear to be correlated with the level of cognitive decline. However, it must be underlined that the style of the signature plays a role, in fact the speed for flourish signatures is higher than that of text-based ones, moreover muscles involved in the movement are more active in the generation of the flourish ones [25]. These results call for further and extended research.

The relation between handwritten signatures and personality traits has been evaluated considering static and dynamic features. It is interesting to report that aspects as gender and personality can be predicted effectively using signature velocity characteristics [26].

On-line signatures have been also used (coupled with speech) to distinguish among three psychophysiological states: normal, drowsiness and alcohol-intoxicated [27]. Dynamic and static features were adopted to test Bayesian hypothesis reporting an overall average error of 14.5%.

Unfortunately, very few works provide a comparison of performance when adopting writing, drawing and signing tasks. From an intuitive point of view, handwriting should be able to provide a wider range of information. More evidence is reported in the following paragraphs.

Text-Based Recognition Analysis

An example of a cursive handwriting could be seen in Fig. 2. Several security applications based on handwritten text have been proposed, such as [8] for capital letters or [28] for cursive drawing on a whiteboard, which is not a very usual writing scenario. However, they have not attracted too much attention of the scientific community. Especially when compared to signatures.

Drawing-Based Analysis

In security applications drawing analysis has attracted some attention especially in graffiti performed by gangs. Gangs use specific clothing, brands, symbols, tattoos, and graffiti to identify their group and interchange messages. Graffiti are any type of public markings that may appear in forms that range from simple written words to elaborate wall paintings [29]. However, due to its nature, they are off-line. Preliminary results in online recognition show a potential to identify people using some drawings [9].

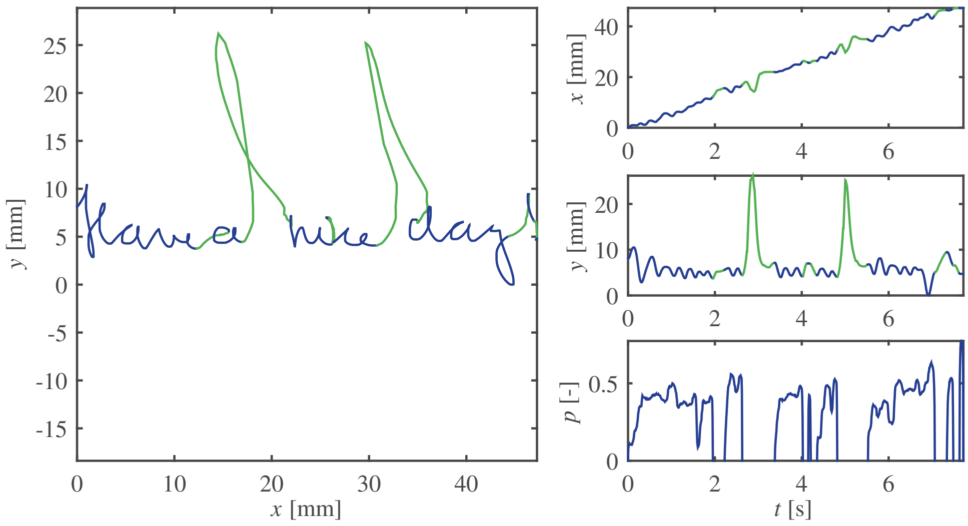


Fig. 2 Text "Have a nice day!" written in cursive letters: the product is depicted on the left side, the right side of the figure contains associated information about horizontal/vertical movement and pressure pattern (the blue colour represents the on-surface movement, while the green colour the in-air one)

Generally, in the area of diagnostics in medical context, drawings are widely used. Some common drawings and their potential usage in medical field are mentioned below.

Pentagon Test

The test is used, e.g. in the Mini-subject classification Mental State Examination (MMSE) to assess cognitive impairment [30]. It consists of copying a drawing, which includes two pentagons overlapping into a rhombus (see 3). It is of interest to report that it has been adopted to differentiate dementia associated with Lewy Body (DLB) from Alzheimer's disease (AD). To the aim, visual parameter features such as number of angles, distance/intersection, closure/opening, rotation and closing-in were considered with an artificial neural network classifier [31]. Park et al. [32] have recently adopted a mobile device to acquire timestamps, x-y coordinates and touch-events. In this case, raw data were processed by means of a U-Net (a convolutional network) to automatically segment angles, distance/intersection between two drawn figures, and closure/opening of the drawn figure contours. Moreover, tremor was also evaluated. It is worth noting that the evaluation of these parameters is associated with a specific disease scaling (interested readers can refer to [31]). Errors which occur in the copying/drawing tasks can be related to damages of the brain: it has been observed that the score connected to the pentagon copy task is associated with parietal grey matter volume and not with frontal, temporal, and occipital ones [33].

Clock-Drawing Test (CDT)

The test can be utilized as a precursory measure to indicate the likelihood of further/future cognitive deficits. It

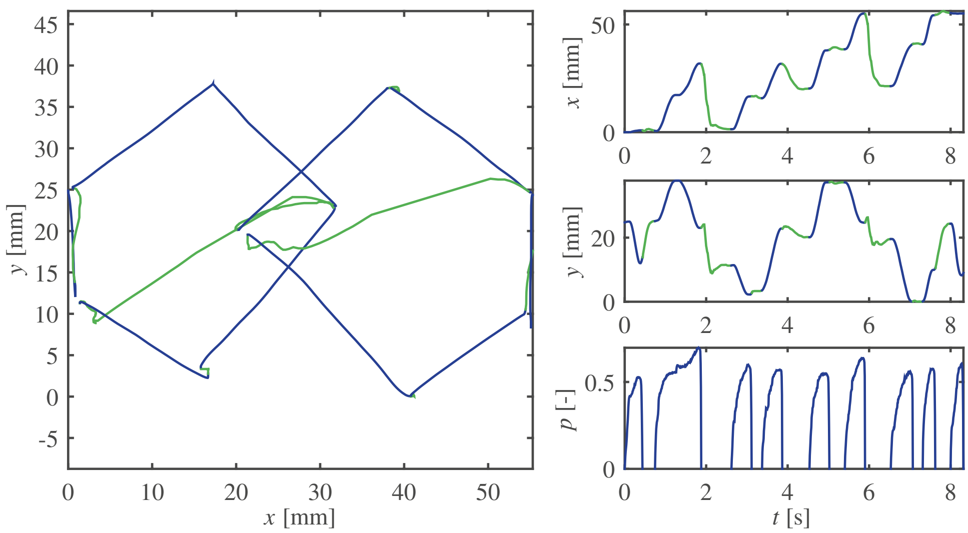


Fig. 3 Pentagon test: the product is depicted on the left side, the right side of the figure contains associated information about horizontal/vertical movement and pressure pattern (the blue colour represents the on-surface movement, while the green colour the in-air one)

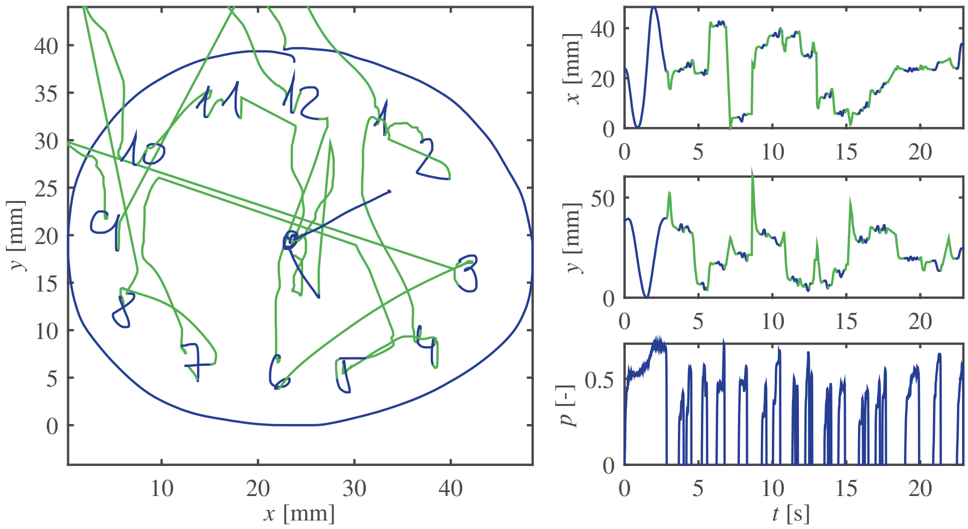


Fig. 4 Clock-drawing test: the product is depicted on the left side, and the right side of the figure contains associated information about horizontal/vertical movement and pressure pattern (the blue colour represents the on-surface movement, while the green colour the in-air one)

is used, e.g. in the Addenbrooke's Cognitive Examination – Revised (ACE-R) test [34] (see Fig. 4). The use of an on-line acquisition tool gives the possibility to evaluate the process of the clock construction and not only the final drawing. In the last decade digital on-line versions of the CDT have been considered [35]. Harbi et al. [36] used a set of features extracted at stroke level (evaluated upon the set of on-line raw data) and an SVM (Support vector machines algorithm) to identify connected components in normal and abnormal drawings. The same authors also proposed a multi-expert approach [37]. More specifically two systems were developed: the first one considered static images obtained by plotting the x-y coordinates and derived a set of static features evaluated by means of a CNN. The same CNN provided a final decision. The second system was based on the x-y coordinates sequences. It was showed that the combination of both systems was able to outperform individual classifiers in the dementia vs healthy subject classification task. Muller et al. [38] investigated the diagnostic value of a digital version of the CDT by comparing it to the standard pencil-paper version. To the aim, 20 patients with early dementia, 30 with MCI and 20 HC were considered. It was observed that in-air time provided by the digital version is able to provide a higher diagnostic accuracy (MCI vs HC) than the use of the traditional test.

House Drawing Copy

This test is used for identification of Alzheimer's disease [39, 40] (see Fig. 5).

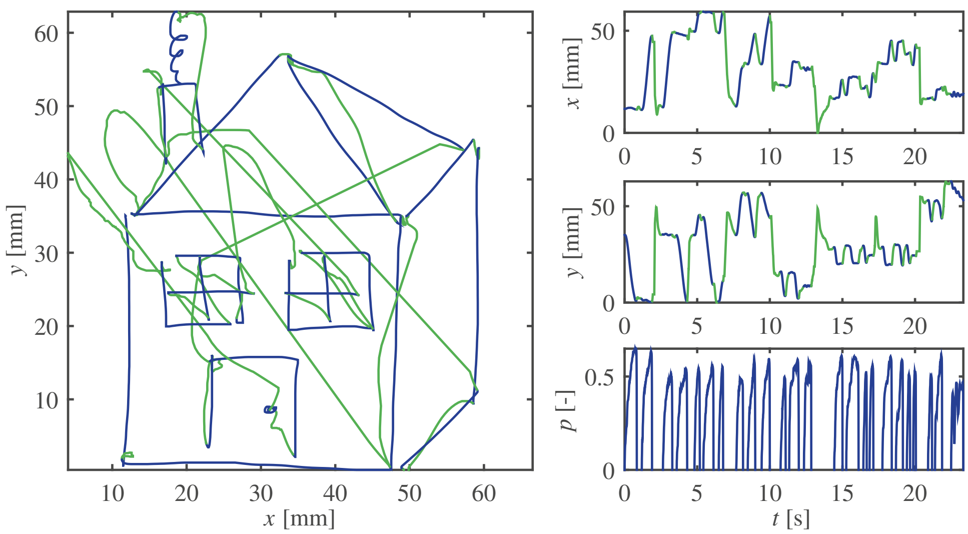


Fig. 5 House drawing test: the product is depicted on the left side, the right side of the figure contains associated information about horizontal/vertical movement and pressure pattern (the blue colour represents the on-surface movement, while the green colour the in-air one)

Archimedes Spiral and Straight Line (Drawing Between Points)

These tasks are useful to support diagnosis of, e.g. Parkinson's disease [41, 42], Huntington's disease [43], essential tremor [44–47], developmental dysgraphia [48], fatigue [11], or brachial dystonia [49]. See Figure 6. In the case of the Archimedes spiral acquisition and straight lines, the participants can have a printed spiral on a sheet of paper and a couple of dots to be connected and they are asked to trace it by a pen without touching the spiral neither the bars (see Fig. 7). Or, the spiral is shown them on a template and they are asked to replicate it on a blank sheet of paper.

Overlapped Circles (Ellipses)

It can be used for quantitative analysis of schizophrenia or Parkinson's disease [21, 50]. See Fig. 8, which represents some simple kinematic features that can be used for an effective diagnosis.

Spring Task (Connected l or Loops)

Several variants exist, such as the connected loops (see Fig. 9), inverted connected loops, connected f, etc. This task is interesting to check the skills to produce rhythmic movements, as well as sudden changes of direction (start-stop-start sequences), useful to evaluate problems to initiate movement [51].

Rey-Osterrieth Complex Figure Test (ROCF)

ROCF consists of copying a complex drawing [52]. It is frequently used to further explain any secondary effect of brain injury in neurological patients, to test for the presence of dementia, or to study the degree of cognitive development in children. In this task patients have to memorize an image and later they have to replicate it without looking at the example. Rey–Osterrieth complex text is depicted in Fig. 10.

Bank-Check Copying

It is, as for most cases of copying tasks, a functional writing task. To properly copy the bank check (Fig. 11), the user should be able to identify the source and corresponding target fields, to locate them and to write the correct content. The single movement and the corresponding stroke could be correct, but the task must be evaluated in its total execution. Patients affected by dementia could result in poor execution producing simplified figures, misplacement of the text, modifications in spatial relationships among strokes, etc. [53]. Considering the example reported in Fig. 11, in-air movements which reveal the action of locating the source and the corresponding target field performed by a mild stage dementia patient can be clearly observed.

Trail-Making Test (TMT)

The test is composed of two parts, in part A the user is required to connect a sequence of consecutive numeric targets (Fig. 12), in part B numbers and letters must be alternated in progressive order (i.e. 1-A, 2-B, etc.). The test involves attentional skills, motor planning, and working memory [54]. This test is adopted for a wide range of cases of brain dysfunction [55]; moreover, normative data are available for several countries according to relevant factors such as age, education level and gender [55]. Patients must complete the task as quickly as possible, if an error occurs then the examiner requests to correct it: this increases the total duration (time) thus reducing the final score assigned

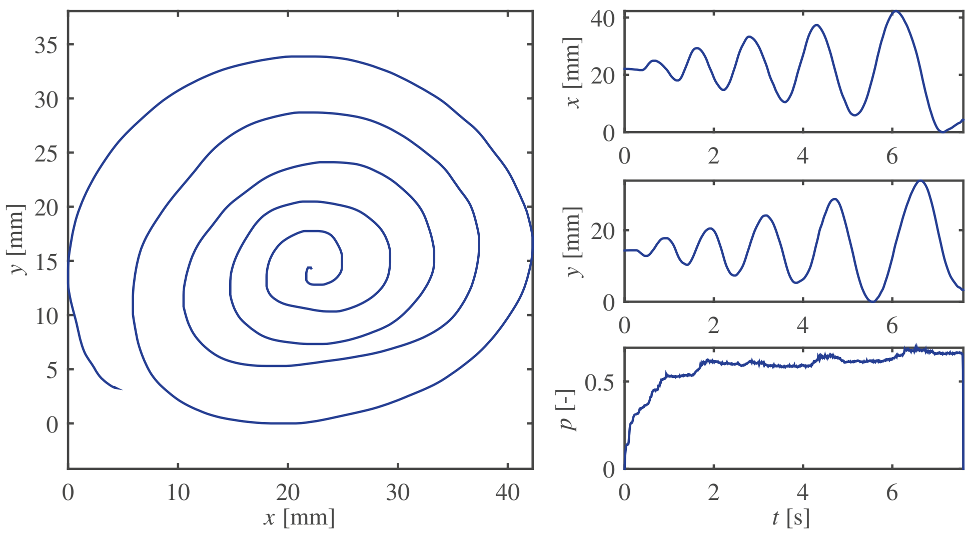


Fig. 6 Archimedes spiral: the product is depicted on the left side, the right side of the figure contains associated information about horizontal/vertical movement and pressure pattern

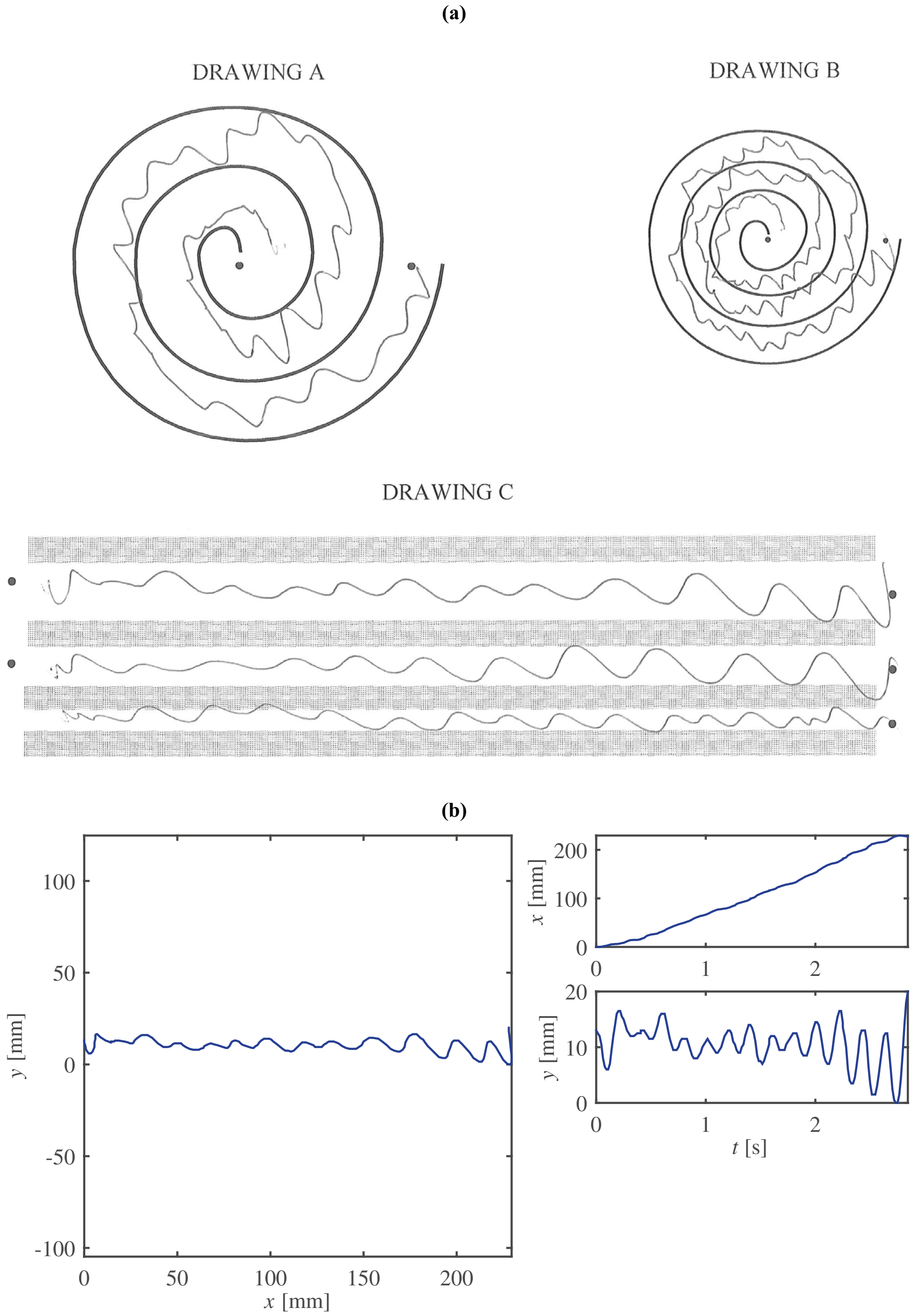


Fig. 7 **a**) Archimedes spirals and straight lines performed by a subject with essential tremor on a sheet of paper; **b**) reconstruction of the first straight line (information about the pressure is missing)

by the examiner (which is typically based only on the time spent). The test is asked to be performed without lifting the pen from the paper/tablet, however a wide set of in-air movements can be observed in the example reported in Figure 12 revealing the need for a mild dementia patient, to ideally retrace the path already written to be able to move forward from an error or hesitation point. The equivalence between the standard (pencil-paper) and the digital (pen-tablet) version of the TMT has been recently verified [56].

It is of interest to report that crossed pentagons, TMT and CDT tests have been recently adopted and compared within the context of handwriting processing to discriminate between HC, MCI and AD [57]. To the aim the following features were considered: pressure, numbers of segments, velocity, acceleration, jerk, in-air and on-the-pad total time. Accuracy of pentagons, TMT-partA, TMT-partB and CDT were, respectively, of 66.2%, 69.0%, 63.3% and 67.6%. The combination of all tasks was able to

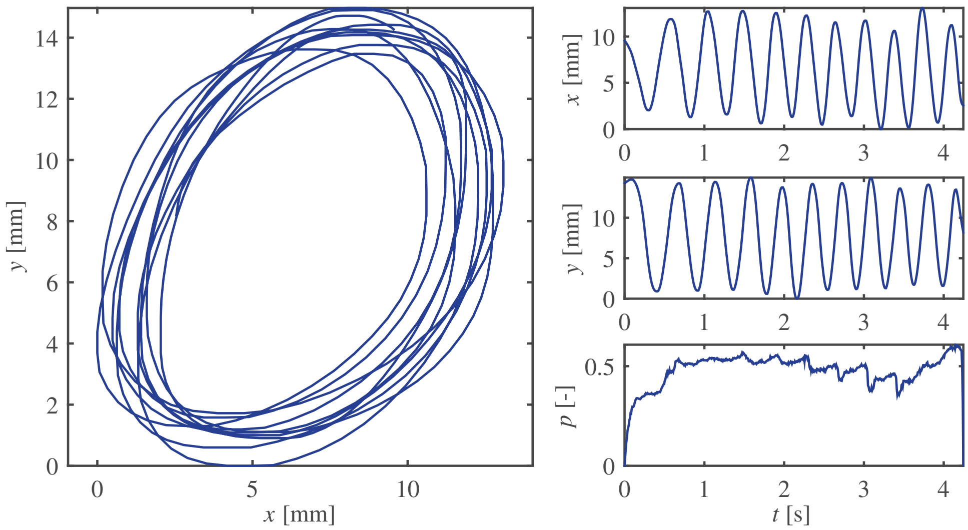


Fig. 8 Overlapped circles: the product is depicted on the left side, the right side of the figure contains associated information about horizontal/vertical movement and pressure pattern

provide increased performance thus revealing they have a certain degree of complementarity.

Cancellation Test of Digits

These tasks are selective attention tests based on a cancellation task. The patient is asked to find targets (e.g. the number 5 in the example in Fig. 13) within a short time constraint. So far, it has been reported that they are useful not only to discriminate AD and HC, but also to monitor the evolution of the cognitive decline [58]. Clinicians typically consider errors performed by patients; however, a digitized version of the test is also able to provide information related to the searching pattern (in air movement in Fig. 13).

Keystroke/Tactile/Touch Analysis

Keystroke dynamics have been extensively used for identification aims on physical keyboards [59–61] and recently on virtual keyboards when considering smartphones and tablets [62]. In this last situation, a wider range of interactions can be considered including finger-swiping patterns drawing, touch-dynamics and, of course, signatures [6, 63]. It is evident that aspects involved in handwriting/signing described in the previous sections are not far from those involved in more general hand-based interaction tasks because they involve the same hand motor area within the brain [64]. So far it has been underlined that typewriting includes a cognitive phase, an associative phase and an autonomous phase [65].

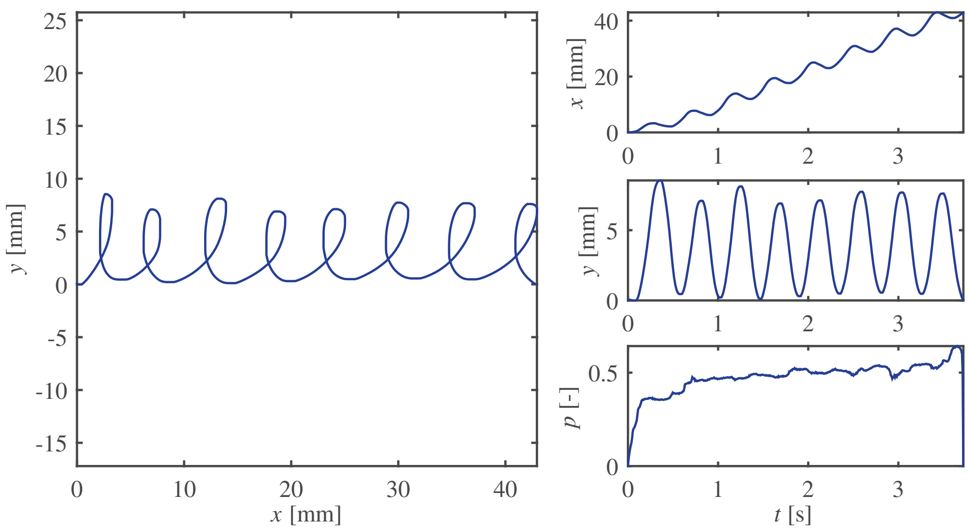


Fig. 9 Spring task: the product is depicted on the left side, the right side of the figure contains associated information about horizontal/vertical movement and pressure pattern

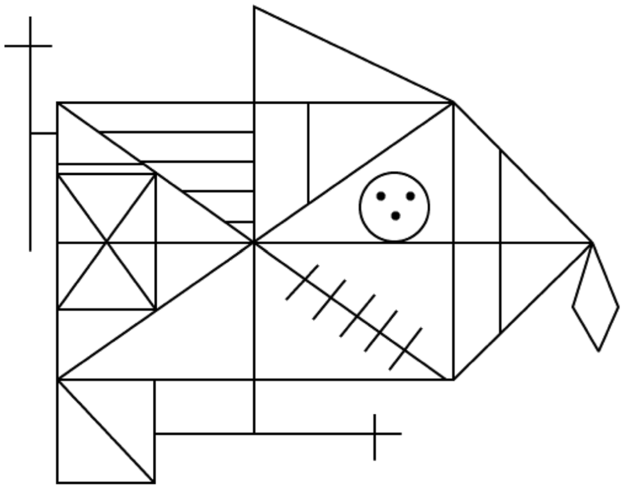


Fig. 10 Rey-Osterrieth complex figure test

Based on the previous observations, touch dynamics can be used also for health evaluation. However, it must be underlined that writing/drawing/signing and, more in general, interacting with a finger on a screen is different from using a pen/tablet system: dynamics as well as the final 2D drawing can be very different [6]. Main reasons are:

– habits in using the pen instead of the finger and/or vice versa;
– finger size compared to the pen's one (in terms of contact point);
– non-rigidity of the finger;
– friction between the finger and the screen.

It is worth noting that, so far, a useful tool for evaluating motor skills is finger-tapping, a test based on a special tool that allows to count the number of key taps within a given time interval (e.g. 30 seconds). This test is used for assessing the presence of bradykinesia, that is, an unnatural slowness in initiating and carrying out simple voluntary movements [51, 66]. Other interesting tasks can be considered. Iakovakis et al. [67] acquired fragmentary typing of short text on a touchscreen smartphone involving 18 PD patients and 15 HC. In this case features adopted were those of the typical key-stroke domain: hold times (time between the pressing and the releasing of a key), flight times (time between the releasing of a key and the pressing of the next one), etc. The adopted classification schema reported 0.82 and 0.81 of, respectively, sensitivity and specificity. Noyce et al. [68] adopted the following parameters for the PD vs HC classification: Kinesia Score (KS30) as, number of key taps in 30 seconds, Akinesia Time (AT30) as the mean dwell time on a key, Incoordination Score (IS30) as the variance of flight time between two consecutive keys and Dysmetria Scores (DS30) related to accuracy of key presses. It was observed that KS30, AT30 and IS30 were significantly able to discriminate PD patients from HC, moreover the same parameters were also correlated with UPDRS motor scores. Similar results have been obtained considering key hold time series and early PD patients [69]. Typing activity on smartphones, independently from the text, has been also considered [64]. In this last case, participants were requested to copy a randomly selected text for five minutes. The time sequence of flight times was used to compute parameter features to be fed to a set of different classifiers: a sensitivity/specificity of 0.81/0.81 has been reported in the binary PD/HC classification task. A very recent work has investigated and compared different touch gestures on the same device: flick, drag, handwriting, pinch, tap, and alternating finger tapping [70]. A wide set of spatial, velocity, time and pressure-based features was considered with the aim to distinguish between early PD patients and HC. The following results were observed: PD subjects resulted in less-efficient finger trajectories, less stable speed, less stable pressure and, higher tremor than HC. Touch gestures and typing

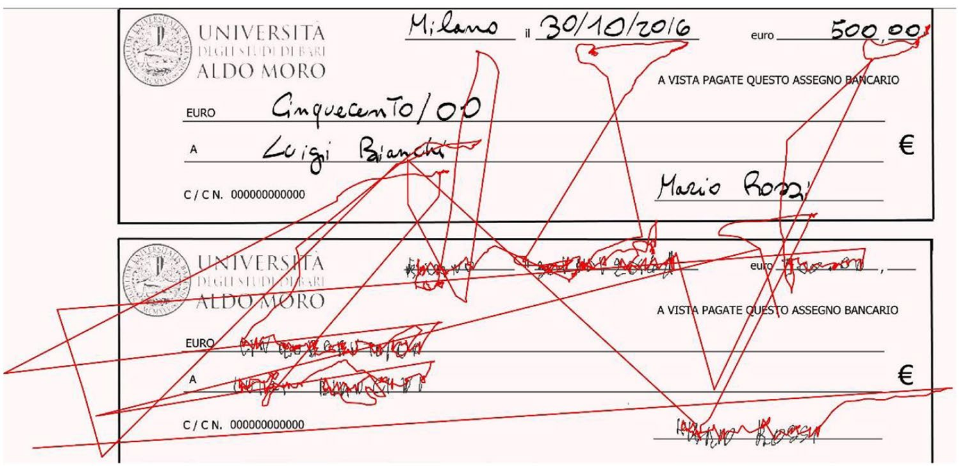


Fig. 11 Bank-check copying (the black colour represents the on-surface movement, while the red colour the in-air one)

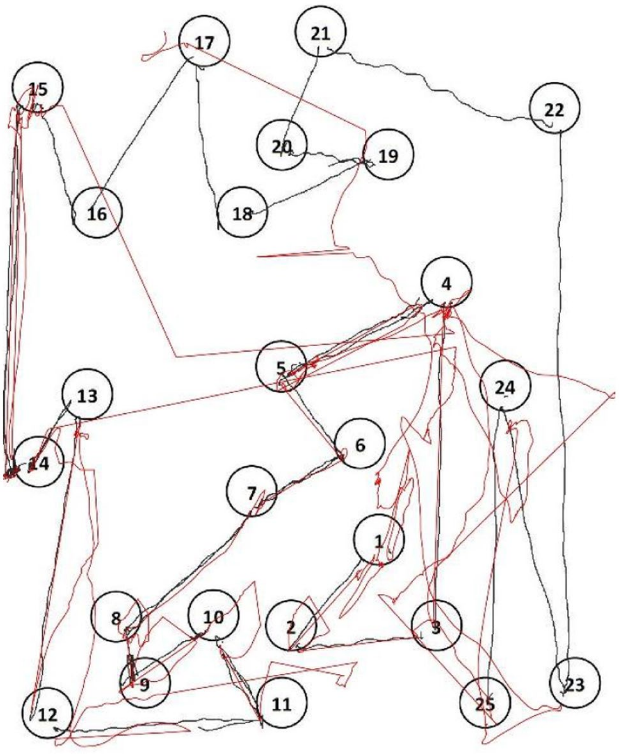


Fig. 12 Trail-Making Test (the black colour represents the on-surface movement, while the red colour the in-air one)

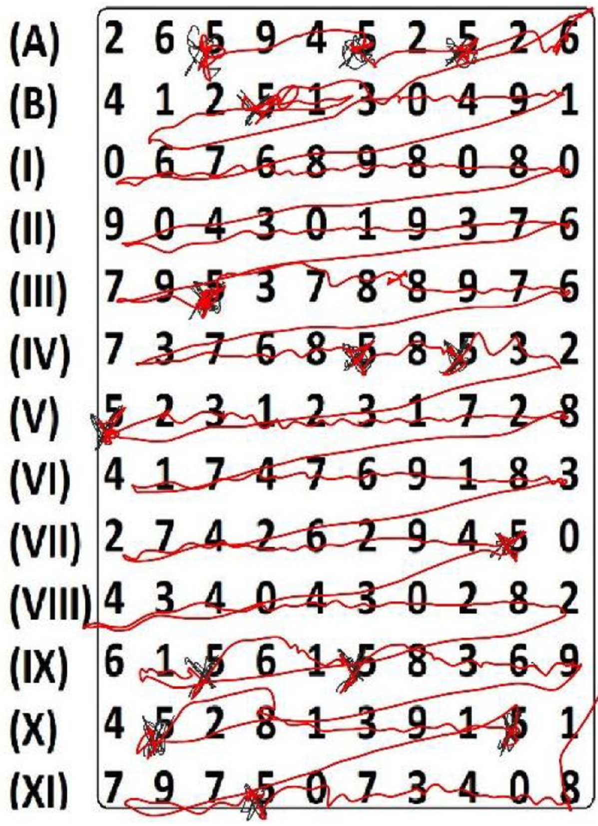


Fig. 13 Cancellation test of the digit '5' (the black colour represents the on-surface movement, while the red colour the in-air one)

appeared to be complementary tasks and an analysis of each task reported drag gestures most performing for classification aims. The best performance was achieved by using all categories of features. Lipsmeier et al. [71] considered also finger tapping (recording all touchscreen events) within a set of many tests related to the use of a smartphone (sustained phonation, rest tremor, postural tremor, balance and gait). The study involved 44 PD and 35 HC. The finger tapping appeared to be the less performing task, however it must be underlined that only intratap variability was considered as a feature.

Problems in hand movements are often the first symptoms of neurological disorders, which do not include only PD, but also Essential Tremor (ET) and Huntington's disease (HD) [72, 73]. On the other hand, dementia diseases, as for example Alzheimer Disease (AD), first result in cognitive rather than motor degradation. In fact, it is well-known that complex tasks including cognitive load (e.g. clock drawing and pentagons) are generally considered [7]. However also coping tasks can be considered. Van Waes et al. [65] requested to a set of 20 young HC, 20 cognitively healthy elderly and 12 age-matched elderly with mild cognitive impairment (MCI) or mild dementia due to AD to perform a typing copy task. Different performances were observed among the three groups.

More in general, a comprehensive user analysis must involve the monitoring of multiple behavior including typing, menu navigation, swipes, drawing and activity understanding [74], [75]. Unfortunately, no works are still available in this direction considering a mobile device, so that it can be considered an open field of research. Very few works are available demonstrating the possibility of using touch dynamics for emotion recognition by considering common unlock Android touch patterns [76] or typing on touchscreens [77].

Tasks Classification

The results presented in this section are based mainly on our own quantitative and qualitative assessment of earlier works. According to our previous work [11], handwritten tasks can be classified into three categories:

– Mechanical tasks—with no cognitive effort, this task can be performed without any heavy load because it is a repetitive movement that the user can do in an automatic way. The user is habituated to do it regularly in her/his life. This is the case of the handwritten signature, handwriting text in capital letters, and handwritten text in cursive letters. Usually, these kinds of tasks are quite straight forward and trivial. These kinds of tasks are quite frequent in the healthy population and find an important niche of applications in biometric recognition of people (user identification and verification [2]).

– Cognitive effort tasks—these tasks require some psychical effort to copy a complex drawing. In this case, the user requires a strategy to start the task. Some cognitive aspects are important because the user needs to know the parts of the drawing already done and the parts that are missing. This kind of task is especially challenging for those people affected by cognitive impairment, such as dementia, mild cognitive impairment, etc. This is the case of the house drawing task, pentagon drawing test, clock-drawing test, Osterrieth complex figure test, trail-making test, cancellation test of a specific digit and Bank-check copying. The clock-drawing test is a special case, because no model is presented prior or during the task. It has to be imagined during the performance.
– Fine motor control tasks—these tasks do not require a heavy cognitive load as the drawing itself is simple and straightforward to understand and memorize at a glance. However, good motor control is required to perform the task. This is the case of the Archimedes spiral drawing test, straight line test, spring drawing test, and concentric circle drawing test.

Table 1 summarizes the best tasks for each application purpose. This table is focused on on-line acquired signals. Although the table presents security and health applications separately, it is important to point out that the same signal can reveal identity and pathologies. Thus, privacy is another interesting research topic that involves both security and health [39].

Table 1 Summary of tasks classified by applications in the field of security and health (each task is classified into one of these three categories: (M)—mechanical task, (C)—cognitive task, (F)—fine motor control task)

Tasks	Security (user identification and verification)	Health
Signature (M)	classical application with increasing popularity in online cases (supermarkets, post offices, etc.) [1, 2, 5, 6, 78]	although pathologies can be detected (e.g. Alzheimer's disease [16, 22, 24, 26, 79]) this is not a popular task in health applications requiring more investigation due to controversial results [23]
	personality assessment [26]	
	international competitions exist to compare different algorithms [80–82]	
Handwriting (M)	capital letters [8]	Parkinson's disease [21, 51, 83–85]
	cursive letters [28]	Huntington's disease [43]
	letter level writer identification [86]	developmental dysgraphia [87, 88]
	gender recognition [10, 89–93]	attention deficit hyperactivity disorder [94]
	writer identification [95–98]	autism spectrum disorder [99]
	competitions in writer identification [90]	obsessive-compulsive disorder [100]
	competitions in gender identification [101]	fatigue [11]
		depression, stress, etc. [12], although better results are found using drawing tasks
		drug abuse, such as Marijuana [102], alcohol [103], caffeine [104]
Drawing (C or F)	graffiti's author identification (offline) [29, 105]	*Pentagon test (C)*
	preliminary results in on-line cases [9]	Alzheimer's disease [40]
		Clock drawing test (C)
		Alzheimer's disease [35, 38, 40, 106]
		mild cognitive impairment [40]
		mild major depressive disorder [107]
		House drawing (C)
		Alzheimer's disease [40, 108]
		mild cognitive impairment [40, 108]
		hypoxemic patient analysis [109]
		fatigue [11]

Table 1 (continued)

Tasks	Security (user identification and verification)	Health
		Archimedes spiral, meanders and straight lines (F)
		Parkinson's disease [41, 42]
		Huntington's disease [43]
		essential tremor [44–47]
		developmental dysgraphia [48]
		fatigue [11]
		brachial dystonia [49]
		Single or overlapped circles (F)
		Huntington's disease [43]
		schizophrenia [50]
		Spring task (F)
		fatigue [11]
		developmental dysgraphia [48, 110]
		schizophrenia [50, 111]
		bipolar disorder [111]
		Parkinson's disease [51]
		Huntington's disease [43]
		Rainbow task (F)
		developmental dysgraphia [48, 112]
		Saw task (F)
		developmental dysgraphia [48]
		Rey-Osterrieth complex figure test (C)
		mild cognitive impairment [113]
		Alzheimer's disease [114]
		Multiple geometrical figures copying (C)
		dementia [53]
		Trail making test (C)
		Alzheimer's disease [57]
		Cancellation test (C)
		Alzheimer's disease [58]
		Tree drawing (C)
		Alzheimer's disease [115]
		mild cognitive impairment [115]

Conclusions

Handwriting is probably one of the most complex tasks that human beings can perform. In addition to being considered a personal behavioral trait (suitable for biometric recognition in security applications), it can also reveal health aspects (when analyzing its quality).

A large amount of scientific literature exists in both application fields: security and health. However, there is no unified activity to be performed by hand writers. Depending on the specific application field, there are some tasks that can unveil richer information than others. Thus, we have tried to systematically review the existing tasks and applications with the goal to serve as a guide for presenting the main alternatives and the topics where they have succeeded.

On the first level, we can classify the tasks into three categories: signature, handwriting (cursive or capital letters), and drawings, being the latest one being richer in possibilities.

On the second level, we can classify the tasks into three different categories according to the specific skills required to perform the task: mechanical, cognitive effort and fine motor control. This second level of tasks classification is mainly relevant for health applications. However, contrary to the first classification, these are not disjoint sets, as each task requires some amount of effort from the other classes. Thus, it just depicts the predominant effort.

While large number of possible tasks exists, one research goal to be addressed is to find the best tasks for each application: those which require a short realization time and provide good discrimination capability (for instance, to differentiate essential tremor from Parkinson's disease).

We forecast new potential applications in the future based on online handwriting, especially in health. We encourage scientific community to test several handwriting tasks in order to find the optimal one. This paper summarizes the main successful ones and can serve as a potential task catalogue to explore when studying new or existing problems.

Acknowledgements This work was supported by Spanish grant PID2020-113242RB-I00, and by grants NU20-04-00294 (Diagnostics of Lewy body diseases in prodromal stage based on multimodal data analysis), and PRIN2017 – BullyBuster project – A framework for bullying and cyberbullying action detection by computer vision and artificial intelligence methods and algorithms. CUP: H94I19000230006.

Declarations

Conflicts of Interest The authors declare that they have no conflict of interest.

Ethical Standard All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this type of study formal consent is not required. This chapter does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

1. Alonso-Martinez C, Faundez-Zanuy M. On the use of time information at long distance in biometric online signature recognition. In: International Tyrrhenian Workshop on Digital Communication. 2017. pp. 3–7.
2. Faundez-Zanuy M. Signature recognition state-of-the-art. IEEE Aerosp Electron Syst Mag. 2005;20(7):28–32. https://doi.org/10.1109/MAES.2005.1499249.
3. Sesa-Nogueras E, Faundez-Zanuy M, Mekyska J. An information analysis of in-air and on-surface trajectories in online handwriting. Cogn Comput. 2012;4(2):195–205.
4. Alonso-Martinez C, Faundez-Zanuy M, Mekyska J. A comparative study of in-air trajectories at short and long distances in online handwriting. Cogn Comput. 2017;9(5):712–20.
5. Impedovo D, Pirlo G. Automatic signature verification: The state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2008;38(5):609–635.
6. Impedovo D, Pirlo G. Automatic signature verification in the mobile cloud scenario: survey and way ahead. IEEE Trans Emerg Topics Comp. 2018.
7. Impedovo D, Pirlo G. Dynamic handwriting analysis for the assessment of neurodegenerative diseases: a pattern recognition perspective. IEEE Rev Biomed Eng. 2018;12:209–20.
8. Sesa-Nogueras E, Faundez-Zanuy M. Biometric recognition using online uppercase handwritten text. Pattern Recogn. 2012;45(1):128–44.
9. Lopez-Xarbau J, Faundez-Zanuy M, Garnacho-Castaño M. Preliminary study on biometric recognition based on drawing tasks. In: Neural Approaches to Dynamics of Signal Exchanges. pp. 485–494. Springer. 2020.
10. Sesa-Nogueras E, Faundez-Zanuy M, Roure-Alcobé J. Gender classification by means of online uppercase handwriting: a text-dependent allographic approach. Cogn Comput. 2016;8(1):15–29.
11. Garnacho-Castaño MV, Faundez-Zanuy M, Lopez-Xarbau J. On the handwriting tasks' analysis to detect fatigue. Appl Sci. 2020;10(21):7630.
12. Likforman-Sulem L, Esposito A, Faundez-Zanuy M, Clémençon S, Cordasco G. EMOTHAW: A novel database for emotional state recognition from handwriting and drawing. IEEE Transactions on Human-Machine Systems. 2017;47(2):273–84.
13. Marcelli A, Parziale A, Senatore R. Some observations on handwriting from a motor learning perspective. In: AFHA. 2013. vol. 1022, pp. 6–10.
14. Impedovo D, Pirlo G, Vessio G, Angelillo MT. A handwriting-based protocol for assessing neurodegenerative dementia. Cogn Comput. 2019;11(4):576–86.
15. Vielhauer C. A behavioural biometrics. Public Service Review: European Union. 2005;9:113–5.
16. Faundez-Zanuy M, Hussain A, Mekyska J, Sesa-Nogueras E, Monte-Moreno E, Esposito A, Chetouani M, Garre-Olmo J, Abel A, Smekal Z, et al. Biometric applications related to human beings: there is life beyond security. Cogn Comput. 2013;5(1):136–51.
17. Van Gemmert AW, Teulings HL, Stelmach GE. Parkinsonian patients reduce their stroke size with increased processing demands. Brain Cogn. 2001;47(3):504–12.
18. Zhi N, Jaeger BK, Gouldstone A, Sipahi R, Frank S. Toward monitoring Parkinson's through analysis of static handwriting samples: A quantitative analytical framework. IEEE J Biomed Health Inform. 2016;21(2):488–95.
19. Wang D, Zhang Y, Yao C, Wu J, Jiao H, Liu M. Toward force-based signature verification: A pen-type sensor and preliminary validation. IEEE Trans Instrum Meas. 2009;59(4):752–62.
20. Impedovo D. Velocity-based signal features for the assessment of Parkinsonian handwriting. IEEE Signal Process Lett. 2019;26(4):632–6.
21. Mucha J, Mekyska J, Galaz Z, Faundez-Zanuy M, Lopez-de Ipina K, Zvoncak V, Kiska T, Smekal Z, Brabenec L, Rektorova I. Identification and monitoring of Parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting. Appl Sci. 2018;8(12):2566.
22. Pirlo G, Diaz M, Ferrer MA, Impedovo D, Occhionero F, Zurlo U. Early diagnosis of neurodegenerative diseases by handwritten signature analysis. In: International Conference on Image Analysis and Processing. 2015. pp. 290–297. Springer.
23. Caligiuri MP, Mohammed L. Signature dynamics in Alzheimer's disease. Forensic Sci Int. 2019;302.
24. Renier M, Gnoato F, Tessari A, Formilan M, Busonera F, Albanese P, Sartori G, Cester A. A correlational study between signature, writing abilities and decision-making capacity among people with initial cognitive impairment. Aging Clin Exp Res. 2016;28(3):505–11.
25. Carmona-Duarte C, de Torres-Peralta R, Diaz M, Ferrer MA, Martin-Rincon M. Myoelectronic signal-based methodology for the analysis of handwritten signatures. Hum Mov Sci. 2017;55:18–30.
26. Miguel-Hurtado O, Guest R, Stevenage SV, Neil GJ. The relationship between handwritten signature production and personality traits. In: IEEE International Joint Conference on Biometrics. 2014. pp. 1–8. IEEE.
27. Lozhnikov P, Sulavko A, Borisov R, Zhumazhanova S. Perspectives of subjects' psychophysiological state identification

using dynamic biometric features. In: Journal of Physics: Conference Series, vol. 1050, p. 012046. IOP Publishing. 2018.
28. Schlapbach A, Liwicki M, Bunke H. A writer identification system for on-line whiteboard data. Pattern Recogn. 2008;41(7):2381–97.
29. Jain AK, Lee JE, Jin R. Graffiti-ID: matching and retrieval of graffiti images. In: Proceedings of the First ACM workshop on Multimedia in forensics. 2009. pp. 1–6.
30. Folstein MF, Folstein SE, McHugh PR. mini-mental state: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3):189–98.
31. Caffarra P, Gardini S, Dieci F, Copelli S, Maset L, Concari L, Farina E, Grossi E. The qualitative scoring MMSE pentagon test (QSPT): a new method for differentiating dementia with Lewy body from Alzheimer's disease. Behav Neurol. 2013;27(2):213–20.
32. Park I, Kim YJ, Kim YJ, Lee U. Automatic, qualitative scoring of the interlocking pentagon drawing test (PDT) based on U-net and mobile sensor data. Sensors. 2020;20(5):1283.
33. Van der Stigchel S, De Bresser J, Heinen R, Koek HL, Reijmer YD, Biessels GJ, Van Den Berg E, et al. Parietal involvement in constructional apraxia as measured using the pentagon copying task. Dement Geriatr Cogn Disord. 2018;46(1–2):50–9.
34. Larner AJ. Addenbrooke's cognitive examination-revised (ACE-R) in day-to-day clinical practice. Age Ageing. 2007;36(6):68–686.
35. Kim H, Cho YS, Do EYL. Computational clock drawing analysis for cognitive impairment screening. In: Proceedings of the fifth international conference on tangible, embedded, and embodied interaction. 2010. pp. 297–300.
36. Harbi Z, Hicks Y, Setchi R, Bayer A. Segmentation of clock drawings based on spatial and temporal features. Procedia Computer Science. 2015;60:1640–8.
37. Harbi Z, Hicks Y, Setchi R. Clock drawing test digit recognition using static and dynamic features. Procedia Computer Science. 2016;96:1221–30.
38. Müller S, Preische O, Heymann P, Elbing U, Laske C. Increased diagnostic accuracy of digital vs. conventional clock drawing test for discrimination of patients in the early course of Alzheimer's disease from cognitively healthy individuals. Frontiers in aging neuroscience. 2017;9:101.
39. Faundez-Zanuy M, Mekyska J. Privacy of online handwriting biometrics related to biomedical analysis. 2017.
40. Garre-Olmo J, Faúndez-Zanuy M, López-de Ipiña K, Calvó-Perxas L, Turró-Garriga O. Kinematic and pressure features of handwriting and drawing: preliminary results between patients with mild cognitive impairment, Alzheimer disease and healthy controls. Curr Alzheimer Res. 2017;14(9):960–8.
41. San Luciano M, Wang C, Ortega RA, Yu Q, Boschung S, Soto-Valencia J, Bressman SB, Lipton RB, Pullman S, Saunders-Pullman R. Digitized spiral drawing: a possible biomarker for early Parkinson's disease. PLoS ONE. 2016;11(10).
42. Saunders-Pullman R, Derby C, Stanley K, Floyd A, Bressman S, Lipton RB, Deligtisch A, Severt L, Yu Q, Kurtis M, et al. Validity of spiral analysis in early Parkinson's disease. Mov Disord. 2008;23(4):531–7.
43. Caligiuri M, Snell C, Park S, Corey-Bloom J. Handwriting movement abnormalities in symptomatic and premanifest Huntington's disease. Movement disorders clinical practice. 2019;6(7):586–92.
44. Ferleger BI, Sonnet KS, Morriss TH, Ko AL, Chizeck HJ, Herron JA. A tablet-and mobile-based application for remote diagnosis and analysis of movement disorder symptoms. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020. pp. 5588–5591.
45. Louis ED, Yu Q, Floyd A, Moskowitz C, Pullman SL. Axis is a feature of handwritten spirals in essential tremor. Movement disorders: official journal of the Movement Disorder Society. 2006;21(8):1294–5.
46. Motin MA, Peters J, Perju-Dumbrava L, Ding C, Eller M, Raghav S, Ali SM, Kempster P, Radcliffe P, Kumar DK. Computerized screening of essential tremor and level of severity using consumer tablet. IEEE Access. 2021;9:15404–12.
47. Pullman SL. Spiral analysis: a new technique for measuring tremor with a digitizing tablet. Mov Disord. 1998;13(S3):85–9.
48. Galaz Z, Mucha J, Zvoncak V, Mekyska J, Smekal Z, Safarova K, Ondrackova A, Urbanek T, Havigerova JM, Bednarova J, et al. Advanced parametrization of graphomotor difficulties in school-aged children. IEEE Access. 2020;8:112883–97.
49. Ratliff J, Ortega RA, Ooi HY, Mirallave A, Glickman A, Yu Q, Raymond D, Bressman S, Pullman S, Saunders-Pullman R. Digitized spiral analysis may be a potential biomarker for brachial dystonia. Parkinsonism Relat Disord. 2018;57:16–21.
50. Caligiuri MP, Teulings HL, Dean CE, Lohr JB. The nature of bradykinesia in schizophrenia treated with antipsychotics. Psychiatry Res. 2019;273:537–43.
51. Smits EJ, Tolonen AJ, Cluitmans L, Van Gils M, Conway BA, Zietsma RC, Leenders KL, Maurits NM. Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinson's disease. PLoS ONE. 2014;9(5).
52. Corwin J, Bylsma F. Psychological examination of traumatic encephalopathy. Clinical Neuropsychologist. 1993;7:3–21.
53. De Lucia N, Peluso S, De Rosa A, Salvatore E, De Michele G, et al. Constructional apraxia is related to different cognitive defects across dementia. J Alzheimers Dis Parkinsonism. 2016;6(244):2161–0460.
54. Dahmen J, Cook D, Fellows R, Schmitter-Edgecombe M. An analysis of a digital variant of the trail making test using machine learning techniques. Technol Health Care. 2017;25(2):251–64.
55. Llinàs-Reglà J, Vilalta-Franch J, López-Pousa S, Calvó-Perxas L, Torrents Rodas D, Garre-Olmo J. The trail making test: association with other neuropsychological measures and normative values for adults aged 55 years and older from a Spanish-Speaking Population-based sample. Assessment. 2017;24(2):183–96.
56. Park SY, Schott N. The trail-making-test: Comparison between paper-and-pencil and computerized versions in young and healthy older adults. Appl Neuropsych: Adult. 2020. pp. 1–13.
57. Ishikawa T, Nemoto M, Nemoto K, Takeuchi T, Numata Y, Watanabe R, Tsukada E, Ota M, Higashi S, Arai T, et al. Handwriting features of multiple drawing tests for early detection of Alzheimer's disease: A preliminary result. In: MedInfo. 2019. pp. 168–172.
58. Della Sala S, Laiacona M, Spinnler H, Ubezio C. A cancellation test: its reliability in assessing attentional deficits in Alzheimer's disease. Psych Med London. 1992;22:885–885.
59. Bhatt S, Santhanam T. Keystroke dynamics for biometric authentication — a survey. In: 2013 international conference on pattern recognition, informatics and mobile engineering. 2013. pp. 17–23.
60. Monrose F, Rubin AD. Keystroke dynamics as a biometric for authentication. Futur Gener Comput Syst. 2000;16(4):351–9.
61. Teh PS, Teoh ABJ, Yue S. A survey of keystroke dynamics biometrics. Sci World J. 2013.
62. Teh PS, Zhang N, Teoh ABJ, Chen K. A survey on touch dynamics authentication in mobile devices. Comput Secur. 2016;59:210–35.
63. Ellavarason E, Guest R, Deravi F, Sanchez-Riello R, Corsetti B. Touch-dynamics based behavioural biometrics on mobile devices - a review from a usability and performance perspective. ACM Computing Surveys (CSUR). 2020;53(6):1–36.
64. García AM, Ibáñez A. A touch with words: dynamic synergies between manual actions and language. Neuroscience & Biobehavioral Reviews. 2016;68:59–95.

65. Van Waes L, Leijten M, Mariën P, Engelborghs S. Typing competencies in Alzheimer's disease: an exploration of copy tasks. Comput Hum Behav. 2017;73:311–9.
66. Teulings HL, Contreras-Vidal JL, Stelmach GE, Adler CH. Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control. Exp Neurol. 1997;146(1):159–70.
67. Iakovakis D, Hadjidimitriou S, Charisis V, Bostantzopoulou S, Katsarou Z, Hadjileontiadis LJ. Touchscreen typing-pattern analysis for detecting fine motor skills decline in early-stage parkinson's disease. Sci Rep. 2018;8(1):1–13.
68. Noyce AJ, Nagy A, Acharya S, Hadavi S, Bestwick JP, Fearnley J, Lees AJ, Giovannoni G. Bradykinesia-akinesia incoordination test: validating an online keyboard test of upper limb function. PLoS ONE. 2014;9(4).
69. Giancardo L, Sanchez-Ferro A, Arroyo-Gallego T, Butterworth I, Mendoza CS, Montero P, Matarazzo M, Obeso JA, Gray ML, Estépar RSJ. Computer keyboard interaction as an indicator of early Parkinson's disease. Sci Rep. 2016;6(1):1–10.
70. Tian F, Fan X, Fan J, Zhu Y, Gao J, Wang D, Bi X, Wang H. What can gestures tell? Detecting motor impairment in early Parkinson's from common touch gestural interactions. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019. pp. 1–14.
71. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, Cheng WY, Fernandez-Garcia I, Siebourg-Polster J, Jin L, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord. 2018;33(8):1287–97.
72. Deuschl G, Wenzelburger R, Löffler K, Raethjen J, Stolze H. Essential tremor and cerebellar dysfunction clinical and kinematic analysis of intention tremor. Brain. 2000;123(8):1568–80.
73. Smith MA, Brandt J, Shadmehr R. Motor disorder in Huntington's disease begins as a dysfunction in error feedback control. Nature. 2000;403(6769):544–9.
74. Stringer G, Couth S, Brown L, Montaldi D, Gledson A, Mellor J, Sutcliffe A, Sawyer P, Keane J, Bull C, et al. Can you detect early dementia from an email? A proof of principle study of daily computer use to detect cognitive and functional decline. Int J Geriatr Psychiatry. 2018;33(7):867–74.
75. Gledson A, Asfiandy D, Mellor J, Ba-Dhfari TOF, Stringer G, Couth S, Burns A, Leroi I, Zeng X, Keane J, et al. Combining mouse and keyboard events with higher level desktop actions to detect mild cognitive impairment. In: 2016 IEEE International Conference on Healthcare Informatics (ICHI). 2016. pp. 139–145.
76. Balducci F, Impedovo D, Macchiarulo N, Pirlo G. Affective states recognition through touch dynamics. Multimedia Tools and Applications. 2020;79(47):35909–26.
77. Mastoras RE, Iakovakis D, Hadjidimitriou S, Charisis V, Kassie S, Alsaadi T, Khandoker A, Hadjileontiadis LJ. Touchscreen typing pattern analysis for remote detection of the depressive tendency. Sci Rep. 2019;9(1):1–12.
78. Diaz M, Ferrer MA, Impedovo D, Malik MI, Pirlo G, Plamondon R. A perspective analysis of handwritten signature technology. Acm Computing Surveys (Csur). 2019;51(6):1–39.
79. Wang Z, Abazid M, Houmani N, Garcia-Salicetti S, Rigaud AS. Online signature analysis for characterizing early stage Alzheimer's disease: A feasibility study. Entropy. 2019;21(10):956.
80. Houmani N, Mayoue A, Garcia-Salicetti S, Dorizzi B, Khalil MI, Moustafa MN, Abbas H, Muramatsu D, Yanikoglu B, Kholmatov A, et al. BioSecure signature evaluation campaign (BSEC'2009): Evaluating online signature algorithms depending on the quality of signatures. Pattern Recogn. 2012;45(3):993–1003.
81. Malik MI, Ahmed S, Marcelli A, Pal U, Blumenstein M, Alewijns L, Liwicki M. ICDAR2015 competition on signature verification and writer identification for on-and off-line skilled forgeries (SigWIcomp2015). In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR). 2015. pp. 1186–1190.
82. Yeung DY, Chang H, Xiong Y, George S, Kashi R, Matsumoto T, Rigoll G. SVC2004: First international signature verification competition. In: International conference on biometric authentication. 2004. pp. 16–22. Springer.
83. Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M. Decision support framework for Parkinson's disease based on novel handwriting markers. IEEE Trans Neural Syst Rehabil Eng. 2014;23(3):508–16.
84. Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. Artif Intell Med. 2016;67:39–46.
85. Parziale A, Senatore R, DellaCioppa A, Marcelli A. Cartesian genetic programming for diagnosis of Parkinson disease through handwriting analysis: Performance vs. interpretability issues. Art Int Med. 2021;111:101984.
86. Chen Z, Yu HX, Wu A, Zheng WS. Level online writer identification. Int J Comp Vis. 2021. pp. 1–16.
87. Gargot T, Asselborn T, Pellerin H, Zammouri I, Anzalone MS, Casteran L, Johal W, Dillenbourg P, Cohen D, Jolly C. Acquisition of handwriting in children with and without dysgraphia: A computational approach. PloS one. 2020;15(9):e0237575
88. Zvoncak V, Mekyska J, Safarova K, Smekal Z, Brezany P. New approach of dysgraphic handwriting analysis based on the tunable Q-Factor wavelet transform. In: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). 2019. pp. 289–294. IEEE.
89. Ahmed M, Rasool AG, Afzal H, Siddiqi I. Improving handwriting based gender classification using ensemble classifiers. Expert Syst Appl. 2017;85:158–68.
90. Djeddi C, Al-Maadeed S, Gattal A, Siddiqi I, Souici-Meslati L, ElAbed H. ICDAR2015 competition on multi-script writer identification and gender classification using 'QUWI' database. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR). 2015. pp. 1191–1195.
91. Gattal A, Djeddi C, Siddiqi I, Chibani Y. Gender classification from offline multi-script handwriting images using oriented basic image features (oBIFs). Expert Syst Appl. 2018;99:155–67.
92. Liwicki M, Schlapbach A, Bunke H. Automatic gender detection using on-line and off-line information. Pattern Anal Appl. 2011;14(1):87–92.
93. Liwicki M, Schlapbach A, Loretan P, Bunke H. Automatic detection of gender and handedness from on-line handwriting. In: Proc. 13th Conf. of the Graphonomics Society. 2007. pp. 179–183.
94. Laniel P, Faci N, Plamondon R, Beauchamp MH, Gauthier B. Kinematic analysis of fast pen strokes in children with ADHD. Appl Neuropsychol Child. 2020;9(2):125–40.
95. Li B, Sun Z, Tan T. Online text-independent writer identification based on stroke's probability distribution function. In: International conference on biometrics. 2007. pp. 201–210.
96. Venugopal V, Sundaram S. An improved online writer identification framework using codebook descriptors. Pattern Recogn. 2018;78:318–30.
97. Venugopal V, Sundaram S. Modified sparse representation classification framework for online writer identification. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2018.
98. Yang W, Jin L, Liu M. Deepwriterid: An end-to-end online text-independent writer identification system. IEEE Intell Syst. 2016;31(2):45–53.
99. Rosenblum S, Ben-Simhon HA, Meyer S, Gal E. Predictors of handwriting performance among children with autism spectrum disorder. Research in Autism Spectrum Disorders. 2019;60:16–24.

100. Schabos O, Hoffmann K, Enzi B, Juckel G, Mavrogiorgou P. Kinematic analysis of handwriting movements in individuals with intellectual disabilities with and without obsessive compulsive symptoms. Psychopathology. 2019;52(6):346–57.
101. Hassaïne A, AlMaadeed S, Aljaam J, Jaoua A. ICDAR 2013 competition on gender prediction from handwriting. In: 2013 12th International Conference on Document Analysis and Recognition. 2013. pp. 1417–1421.
102. Foley RG, Miller AL. The effects of marijuana and alcohol usage on handwriting. Forensic Sci Int. 1979;14(3):159–64.
103. Phillips JG, Ogeil RP, Müller F. Alcohol consumption and handwriting: A kinematic analysis. Hum Mov Sci. 2009;28(5):619–32.
104. Tucha O, Walitza S, Mecklinger L, Stasik D, Sontag TA, Lange KW. The effect of caffeine on handwriting movements in skilled writers. Hum Mov Sci. 2006;25(4–5):523–35.
105. Tong W, Lee JE, Jin R, Jain AK. Gang and moniker identification by graffiti matching. In: Proceedings of the 3rd international ACM workshop on Multimedia in forensics and intelligence. 2011. pp. 1–6.
106. Sunderland T, Hill JL, Mellow AM, Lawlor BA, Gundersheimer J, Newhouse PA, Grafman JH. Clock drawing in Alzheimer's disease: a novel measure of dementia severity. J Am Geriatr Soc. 1989;37(8):725–9.
107. Heinik J, Werner P, Dekel T, Gurevitz I, Rosenblum S. Computerized kinematic analysis of the clock drawing task in elderly people with mild major depressive disorder: an exploratory study. Int Psychogeriatr. 2010;22(3):479.
108. Müller S, Preische O, Heymann P, Elbing U, Laske C. Diagnostic value of a tablet-based drawing task for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. Journal of Alzheimer's Disease. 2017;55(4):1463–9.
109. Fiz JA, Faundez-Zanuy M, Monte-Moreno E, Alcobé JR, Andreo F, Gomez R, Manzano JR. Short term oxygen therapy effects in hypoxemic patients measured by drawing analysis. Comput Methods Programs Biomed. 2015;118(3):330–6.
110. Vaivre-Douret L, Lopez C, Dutruel A, Vaivre S. Phenotyping features in the genesis of pre-scriptural gestures in children to assess handwriting developmental levels. Sci Rep. 2021;11(1):1–13.
111. Crespo Y, Ibañez A, Soriano MF, Iglesias S, Aznarte JI. Handwriting movements for assessment of motor symptoms in schizophrenia spectrum disorders and bipolar disorder. PLoS ONE. 2019;14(3).
112. Mekyska J, Faundez-Zanuy M, Mzourek Z, Galaz Z, Smekal Z, Rosenblum S. Identification and rating of developmental dysgraphia by handwriting analysis. IEEE Transactions on Human-Machine Systems. 2016;47(2):235–48.
113. Cheah WT, Chang WD, Hwang JJ, Hong SY, Fu LC, Chang YL. A screening system for mild cognitive impairment based on neuropsychological drawing test and neural network. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). 2019. pp. 3543–3548.
114. Kim KW, Lee SY, Choi J, Chin J, Lee BH, Na DL, Choi JH. A comprehensive evaluation of the process of copying a complex figure in early-and late-onset Alzheimer disease: A quantitative analysis of digital pen data. J Med Internet Res. 2020;22(8).
115. Robens S, Heymann P, Gienger R, Hett A, Müller S, Laske C, Loy R, Ostermann T, Elbing U. The digital tree drawing test for screening of early dementia: An explorative study comparing healthy controls, patients with mild cognitive impairment, and patients with early dementia of the Alzheimer type. Journal of Alzheimer's Disease. 2019;68(4):1561–74.

A.7 Developmental Dysgraphia: A New Approach to Diagnosis

VOLUME 28 ISSUE 1

The International Journal of

Assessment and Evaluation

Developmental Dysgraphia

A New Approach to Diagnosis

KATARÍNA ŠAFÁROVÁ, JIŘÍ MEKYSKA, AND VOJTĚCH ZVONČÁK

COMMON GROUND

THELEARNER.COM

THE INTERNATIONAL JOURNAL OF ASSESSMENT AND EVALUATION
https://thelearner.com
ISSN: 2327-7920 (Print)
ISSN: 2327-8692 (Online)
https://doi.org/10.18848/2327-7920/CGP (Journal)

First published by Common Ground Research Networks in 2021
University of Illinois Research Park
60 Hazelwood Drive
Champaign, IL 61820 USA
Ph: +1-217-328-0405
https://cgnetworks.org

The International Journal of Assessment and Evaluation is a peer-reviewed, scholarly journal.

Common Ground Research Networks, a member of Crossref

PEER REVIEW
Articles published in *The International Journal of Assessment and Evaluation* are peer reviewed using a two-way anonymous peer review model. Reviewers are active participants of The Learner Research Network or a thematically related Research Network. The publisher, editors, reviewers, and authors all agree upon the following standards of expected ethical behavior, which are based on the Committee on Publication Ethics (COPE) Core Practices. More information can be found at https://cgnetworks.org/journals/publication-ethics.

ARTICLE SUBMISSION
The International Journal of Assessment and Evaluation publishes biannually (June, December). To find out more about the submission process, please visit https://thelearner.com/journals/call-for-papers.

ABSTRACTING AND INDEXING
For a full list of databases in which this journal is indexed, please visit https://thelearner.com/journals/collection.

RESEARCH NETWORK MEMBERSHIP
Authors in *The International Journal of Assessment and Evaluation* are members of The Learner Research Network or a thematically related Research Network. Members receive access to journal content. To find out more, visit https://thelearner.com/about/become-a-member.

SUBSCRIPTIONS
The International Journal of Assessment and Evaluation is available in electronic and print formats. Subscribe to gain access to content from the current year and the entire backlist. Contact us at cgscholar.com/cg_support.

ORDERING
Single articles and issues are available from the journal bookstore at https://cgscholar.com/bookstore.

OPEN RESEARCH
The International Journal of Assessment and Evaluation is Hybrid Open Access, meaning authors can choose to make their articles open access. This allows their work to reach an even wider audience, broadening the dissemination of their research. To find out more, please visit https://cgnetworks.org/journals/open-research.

Developmental Dysgraphia: A New Approach to Diagnosis

Katarína Šafárová,[1] Masaryk University, Czech Republic
Jiří Mekyska, Brno University of Technology, Czech Republic
Vojtěch Zvončák, Brno University of Technology, Czech Republic

Abstract: Writing is a complex skill. Issues in this process, which are usually associated with developmental dysgraphia (DD), could consistently cause problems in everyday life, like for example, lower self-esteem and poorer academic achievement. That is why the correct diagnosis of DD is crucial for further child development. DD belongs to the category of specific learning disabilities and according to different studies, its prevalence ranges between 0.1 and 30 percent. Diagnosing a child with DD relies, in the first place, on teachers. After that, psychologists, or special educational specialists (in the Czech Republic) commonly use qualitative evaluation of the written process, where the child is observed when he or she is writing. Nevertheless, there are no objective tests or standardized examinations for the assessment of handwriting deficiency either in special educational or psychological practices. In the frame of current research, a new quantitative approach to handwriting proficiency assessment was developed. Digitizing tablets (Wacom Intuos Pro L) with a special inking pen (Wacom Ink Pen) are used to record the online handwriting process and graphomotor skills of children. Administration templates contain simple graphomotor elements and complex figures related to DD symptoms and cognitive (memory and visuospatial) abilities. This new approach to diagnose handwriting issues will be presented in this article.

Keywords: Developmental Dysgraphia, Diagnosis, Online Process, Machine Learning, Graphomotor Disabilities Rating Scale

Introduction

The aim of this article is to present the experimental research on graphomotor disabilities (GD) and developmental dysgraphia (DD) performed by an interdisciplinary team of psychologists, educationists, and engineers in Brno, Czech Republic. The main goal is to create an objective and accurate way to detect problems with handwriting and to help specialists in the practice of this method. In the first part of the article, GD and DD will be described and the problems with their existing definitions will be discussed. The second part will compare the methods of diagnosis of handwriting issues in the Czech Republic with those in the rest of the world, and the problems with the process of diagnosis in the Czech Republic will be clarified. In the third part, a new method will be presented—the graphomotor disabilities rating scale (GDRS)—which is currently being researched and developed.

There is a huge problem not only with diagnostic processes but with diagnostic methods themselves. Experts such as psychologists, occupational therapists, or teachers lack objective diagnostic methods to detect handwriting issues. One part of this problem is that there is a lack of research concerning this topic and that dysgraphia is marginalized. On the web, there are 12,120 results for the keyword "dyslexia" (reading disorder) and only 538 results for the keyword "dysgraphia" (writing disorder).

Theory and Current State

Graphomotor Skills and Handwriting

Graphomotor (GM) skills are psychomotor abilities which primarily comprise writing and drawing. The GM process is considered as an outcome of cognitive, motor, and perceptual skills

[1] Corresponding Author: Katarína Šafárová, Arne Nováka 1, 602 00, Brno, Department of Psychology, Faculty of Art, Masaryk University, Brno, Czech Republic, 602 00. email: katas@mail.muni.cz

The International Journal of Assessment and Evaluation
Volume 28, Issue 1, 2021, https://thelearner.com

ISSN: 2327-7920 (Print), ISSN: 2327-8692 (Online)
https://doi.org/10.18848/2327-7920/CGP/v28i01/143-160 (Article)

and their interactions. Hence, if a child has issues with GM skills, there is a disruption between ideas and the ability to express them through writing. It is for this reason that writing has been described as "language by hand" by Berninger and her colleagues (2002). The action of writing by hand begins with ideation and planning processes, followed by the application of language rules (grammar, syntax, spelling, punctuation, etc.), with the outcome of text production at the motor level. The lower arm, wrist, and fingers must cooperate to create a final handwriting product. Moreover, other processes like evaluation and self-monitoring, which serve as effective feedback, must be taken into consideration. All of these look like they are consecutive processes, but in reality, they all take place concurrently.

Nowadays, nearly all models of handwriting (Van Galen 1991; Kandel et al. 2011; McCloskey and Rapp 2017; Feder and Majnemer 2007; Cornhill and Case-Smith 1996; Berninger and Amtmann 2003; Flower and Hayes 1981) differentiate between several levels, which could be summarized as: (1) higher cognitive levels and (2) lower motoric-perception levels. For example, one of the oldest theories developed by Flower and Hayes (1981) distinguishes between three levels of mental representation. At the conceptual level, the ideas or preverbal messages, stored in the long-term memory, are created during the planning processes. At the linguistic level, translating processes are involved, and preverbal messages are translated into verbal messages. This is how the conceptual structure from the first stage gets its grammatical properties (syntax, morphology, and spelling, stored in the mental lexicon). Finally, at the motor level, the verbal message is converted into a sequence of motor plans and the written output is created. It comprises the regulation of handwriting parameters such as size, speed, spacing, force–form alignment, and slant (Hayes and Gradwohl-Nash 1996; Bock and Levelt 1994; Kellogg 1996; McCloskey and Rapp 2017).

Feder and Majnemer (2007) described two components of the writing process: motor and perceptual. Perceptual components comprise sensory modalities, visual perception, and sustained attention. Motor components pertain to fine-motor control (in-hand manipulation, bilateral integration, motor planning, and kinesthesia). Both components are linked to visual–motor integration as an important part of the whole handwriting process, which is the ability to coordinate visual information with motor processes. Other authors (Christensen 2005; Medwell, Strand and Wray 2009) have agreed that writing is not only a motor skill. According to them, the memory and orthographic processes work together to recall the letter shapes and translate its patterns automatically. A central part of the writing model proposed by Berninger and Amtmann (2003) involves the working memory. It is linked to long-term memory through the process of composing and to short-term memory through the process of reviewing. The executive functions like conscious attention, planning, revising, and strategies for self-evaluation are involved as well. This model argues that orthographic–motor integration (OMI) contributes more to handwriting skills than fine-motor skills (Graham and Weintraub 1996; Abbott and Berninger 1993). The OMI allows the child to recall the correct shape of the letters or whole words from his or her mind and write it down without focusing attention on it. It means that the process of writing is automated. Research shows that OMI accounts for more than 50 percent of the variance in written language performance in individuals from primary through secondary school and even into adulthood (Bourdin and Fayol 2002; Graham et al. 1997; Jones and Christensen 1999).

As Palmis and her colleagues (2017) state, the automation of handwriting means that writing does not require conscious effort. Therefore, the cognitive resources could be allocated elsewhere, for example, to planning, organizing, or creating the content of the story (ideation). Automaticity is based on experience and training and so the movements become more fluent. At this point, the Matthew effect should be mentioned (Cunningham and Chen 2014; Stanovich 1993), which describes the improvement in the reading and/or writing skills of children who have a better level of those abilities, and conversely, stagnation among children who have poor skills. From previous paragraphs, it is evident that writing is not merely a motor process, but other higher cognitive and executive processes are also involved. Richards et al. (2011, 512) wrote that writing is a "brain-based skill that facilitates meaning-making as writers externalize their cognitions through letter forms, the building blocks of written words and text." From this point of view, handwriting issues and dysgraphia should not be seen only as a distortion of a written product, but we should be able to see those higher cognitive processes as well and incorporate them into the whole picture of the normal writing process and also into writing disabilities. For example, when the child must focus on the execution of motor plans (lower motoric perception functions), the working memory capacity is overloaded and there is no space for higher cognitive processes. Likewise, the automatization of writing is not possible in that kind of setting.

Developmental Dysgraphia and Its Diagnosis

DD is defined as the disturbance of the process of written production, which is related to the mechanics of handwriting. In ICD-10 (WHO 1992), the problems of handwriting belong to Chapter 5: Mental, behavioral, and neurodevelopmental disorders; Section F80-F89: Pervasive and specific developmental disorders; Category F81: Specific developmental disorders of scholastic skills; and the final code F81.81: Disorder of written expression. To define handwriting issues, the ICD-10 was used instead of the DSM-IV (APA 2000), because it is the common diagnostic system in Europe. Although there is a definition of this disorder, there are missing diagnostic criteria both in the ICD-10 and the DSM-IV. Moreover, in the literature, different terms are used for describing dysgraphia. A child could be dysgraphic or could be named as having poor handwriting. A child could have handwriting issues or difficulties and those could be part of special learning disorders or disabilities. Also, the motoric part of dysgraphia could be linked to agraphia or developmental coordination disorder**.** Children with DD are of at least average intelligence and they have not been identified as having any neurological problems (Hamstra-Bletz and Blöte 1990). The prevalence ranges between 10 and 34 percent (Döhla and Heim 2016; Cermak and Bissell 2014) depending on the country and study in question. The data reported in Czech Republic concern only specific learning disorders, with estimates ranging from 3 to 5 percent (Kejřová and Krejčová 2015; Zelinková 2015).

Another problem with diagnosis is that of comorbidity. Writing issues are present in 30 to 47 percent of children with reading issues (Chung, Patel and Nizami 2020). Authors added that there are 90 to 98 percent children with neurodevelopmental issues such as attention-deficit hyperactivity disorder or autism, who also struggle with handwriting. It is also the case that developmental coordination disorder affects the handwriting process. In the Czech Republic, the numbers of comorbidities with other specific learning disorders are missing in the data.

Usually, the following symptoms are listed: (1) problems with size control—letters are not consistent in size; (2) slant—written letters are not even; (3) alignment—child is not able to follow the lines; (4) pressure—too high, usually linked to incorrect grip which causes fatigue or pain during the writing; (5) poor spacing between letters and words; (6) messy organization of the text on the page; (6) problems with letter differentiation and spelling—inversions of b and d, a and o, etc.; (7) added or missing strokes; (8) problems with beginning of writing—child does not know where or how to start; (9) grammar mistakes—child does not check the outcome or she/he erases or crosses out the text a lot. According to the literature (Rosenblum, Weiss, and Parush 2003) there are two main outcomes which are used for assessing and defining poor handwriting: (1) legibility, which is the combination of all the above-mentioned symptoms, and (2) performance time, which could be assessed as writing duration.

Diagnostic Process

The primary aim of researchers in this area is to develop standardized evaluation capable of producing quantitative scores for handwriting quality. The dilemma is how to define "readability" or "quality of handwriting" (Ayres 1912). Rosenblum, Weiss, and Parush (2003) distinguished two types of evaluation: product and process evaluation. To enhance the methodology, questionnaires were implemented as a third type.

First, there is product evaluation. As the name suggests, in this case, final and static outcomes of written products are evaluated by experts. This evaluation is based on two types of scales: (1) global and (2) analytical. Global scales consist solely of one factor, legibility, which means that the assessment is based on an overall judgment of this factor. Analytical scales are created with many criteria such as the shape of letters, spacing, or speed. In this approach, these specific features are linked to the general legibility of writing. In practice, there are different tests which serve this purpose. An overview of common tests used for diagnosing handwriting problems is shown in Table 1 (Feder and Majnemer 2003; Roston, Hinojosa, and Kaplan 2008). The major drawback of these tools is a lack of evaluation of psychometric qualities (see validity rows). Only the TOHL (Test of Legible Handwriting; Larsen and Hammill 1989) has documented evidence for criterion, content, and construct validity, but the test itself was designed in 1989.

In the last 20 years, there has been a new approach called process evaluation, where computerized technology, software, and digitizers, are used to record the process of handwriting itself (Rosenblum, Weiss, and Parush 2003; Longstaff and Heath 1997). This interdisciplinary field, called graphonomics, focuses on handwriting movement analysis (e.g., Van Gemmert and Teulings 2006). Most importantly, this approach addresses the limitations of previous measurement tools. In addition, as the child is writing on the paper with inking pen, it maintains the measure as ecologically valid. This online process records not only the process, but it also allows the researcher to measure the features which underlie writing, and which are not detectable by the naked eye. That is why this quantitative online examination is considered to be more precise and objective. Unfortunately, the diagnostic tool which should be established on the outcomes of these online and objective measurements does not exist yet. As a consequence, although there have been studies with promising outcomes (e.g., Asselborn, Chapatte and Dillenbourg 2020), actual diagnostic applications have not yet been pursued.

A third approach is evaluations using questionnaires, where the children judge themselves (self-evaluation) or are judged by others (teachers, parents, psychologists, etc.). These questionnaires are usually focused on different manifestations of handwriting issues or on experiences of disability and well-being. Handwriting problems are evaluated from a subjective point of view without assessing the handwriting itself, contrary to product evaluation. An example is the Handwriting Proficiency Screening Questionnaire for Children (HPSQ-C; Rosenblum and Gafni-Lachter 2015).

Having covered the different approaches to handwriting analysis, a description of the diagnostic process from an expert's point of view will ensue. The profession of occupational therapy is, among other things, responsible for correct GD and DD assessment and remediation in many jurisdictions. In the Czech Republic, the diagnosis process is distributed between teams of psychologists and a special educationist (SE) who takes part in the process. Working in collaboration, the psychologists are responsible for taking anamnesis and for testing intelligence and memory. Further diagnosis, especially of specific learning disorders (e.g., dyslexia, dysgraphia, dyscalculia, etc.), is carried out by a SE. When a diagnosis of dysgraphia is returned, the SE assesses the child in several ways:

1) writing assessment, which is created from a transcription task (cursive writing → cursive writing), a copying task (block letters → cursive writing), a dictation task, and/or drawing;
2) experienced but subjective observation during the writing process, which includes observation of (a) how the child is sitting during writing, (b) the grip, i.e., how the child is holding a pen or pencil, (c) the relaxation of hand during writing, (d) how the child is writing, i.e., the stroke smoothness, pressure, size of the letters, slant, etc., (e) speed of writing, (f) jerky movements, (g) increased focus on writing process, which could cause more grammar mistakes;
3) test of laterality or handedness;
4) visual perception;
5) right-to-left and spatial orientation;
6) analysis of homework and exercise books (Pokorná 2015; Zelinková 2015).

On looking at these criteria, we perceive several problems. Firstly, there is no screening tool in the Czech Republic for teachers or parents, who often first recognize handwriting issues. Neither do the practitioners have a specialized test for detecting or diagnosing handwriting issues, but as has been noted previously, nor are current, internationally used tests dependable, as they have questionable or unknown psychometric properties. The graphonomics approach is also yet to yield a diagnostic tool, while the questionnaire approach relies on subjective and nonuniform criteria for evaluating handwriting issues. For the latter, even when special educationists are experts in their field, there could be inconsistent outcomes from the diagnostic process.

Table 1: Summary of the Handwriting Evaluation Tests and its Psychometric Properties

	MHA	*ETCH-M*	*CHES-M*	*DRHP*	*TOLH*	*DHA*	*WOLD*	*THS-R*	*HHE*	*DASH*	*BHK*	*HST*
Age								6–18		9–16	6–12	3–12
Grade	1–2	2	2	3+	2–12	3–8	1–5		2–3			
Standards	✓				✓			✓		✓		✓
Alphabet writing	✓	✓				✓		✓	✓	✓		
Numeral writing		✓										
Near-point copying	✓	✓	✓			✓		✓	✓	✓	✓	✓
Far-point copying		✓				✓						
Dictation		✓				✓		✓	✓			
Composition		✓		✓	✓							
Speed	✓	✓	✓				✓		✓	✓	✓	✓
Global readability	✓	✓	✓		✓		✓	✓	✓		✓	
Inter-rater	0.87 – 0.98	0.75 – 0.92	0.85 – 0.93 ICC	0.61 – 0.65	0.95				0.75 – 0.79	0.85 – 0.99 ICC	0.71 – 0.89	0.99 ICC
Test-retest	0.58 – 0.94	0.63 – 0.77	×	×	0.90			0.82		0.50 – 0.92 ICC	0.51 – 0.55	0.99 ICC
Criterion	×	×	×	✓	✓	×	×	✓	×	×	×	×
Construct	×	×	×	×	✓	×	×	×	✓	×	×	×
Content	✓	✓	×	×	✓	×	×	×	×	✓	×	×

Notes: MHA: Minnesota Handwriting Assessment (Reisman 1993); ETCH-M: Evaluation Tool for Children's Handwriting-Manuscript (Amundson 1995); CHES-M: Children's Handwriting Evaluation Scale-Manuscript (Phelps and Stempel 1988); DRHP: Diagnosis and Remediation of Handwriting problems (Stott, Moyes and Henderson 1985); TOLH: Test of Legible Handwriting (Larsen and Hammill 1989); DHA: Denver Handwriting Analysis (Anderson 1983); WOLD: –Wold Sentence Copy Test (Maples 2003); THS-R: Test of Handwriting (Milone 2007); HHE (Erez and Parush 1999); DASH (Barnett et al. 2007); BHK (Hamstra-Bletz, DeBie, and Den Brinker 1987); HST (Wallen, Bonney and Lennox 1996).
Source: MHA-TOHL: Feder and Majnemer 2003; DHA-WOLD: Roston, Hinojosa, and Kaplan 2008

New Diagnostic Approach

Planned Sample

In the research project, children were enrolled via psychological counselling centers and via schools. The sample is divided into three groups: (1) children with diagnosed dysgraphia; (2) children who were not diagnosed, but the teachers or children themselves reported some problems with handwriting; and (3) typical development children. The sample consists of children from kindergarten to fourth grade, which makes five groups. There are 100 children planned for each group, that is, there should be approximately 500 children in the whole sample. Previous studies used only small samples, with approximately 20 children in each group on average (experimental/comparative) (e.g., Engel-Yeger, Nagauker-Yanuv, and Rosenblum 2009; Smits-Engelsman and Van Galen 1997).

With this type of sample distribution, differences between groups could be compared in three ways: (1) horizontally—between the condition of age/grade (five levels); (2) vertically—between the condition of diagnosis (three levels); (3) and a combination of the two conditions. The first axis of comparison is very important, because the internalization of the writing process is believed to occur between the third and fourth grades. It means that this process becomes more automatic, and attention becomes more focused on the process of ideation and the content of the text, which frees up capacity for the working memory (McCutchen 1996). Furthermore, changes or improvements in the writing could be tracked. Some studies suggest that handwriting characteristics of typical development in children should be smaller, of even size, and smoother (Meulenbroek and Van Galen 1989; Zesiger, Mounoud, and Hauert 1993). Older children also use less bounded cursive writing, but the letters are closer (Blöte and Hamstra-Bletz 1991).

Instruments: Digitizer and Software

For the research, the specially created software HandAQUS for handwriting acquisition is used together with digitizing tablets (Wacom Intuos Pro L) with a special inking pen (Wacom Ink Pen). This system allowed the researchers to capture information about the position of the pen (x and y axes) on the tablet's surface or up to 1.5 cm above it. It also allowed the measurement of pressure on the surface, the pen's tilt, and azimuth. Finally, the movement was sampled with 150 Hz sampling frequency, where each sample was associated with a time stamp, which enabled the reconstruction of kinematic characteristics. More specifically, it distinguishes between several categories of conventional handwriting features (Table 2).

Table 2: List of Clinical Features Used in the Study

Category/Features	*On-surface*			*In-air*		
Temporal	Global	Vertical	Horizontal	Global	Vertical	Horizontal
Duration of handwriting	✓			✓		
Spatial	Global	Vertical	Horizontal	Global	Vertical	Horizontal
Height of written product	✓			✓		
Width of written product	✓			✓		
Angle of written product	✓			✓		
Height of stroke	✓			✓		
Width of stroke	✓			✓		
Angle of stroke	✓			✓		
Kinematic	Global	Vertical	Horizontal	Global	Vertical	Horizontal
Velocity	✓	✓	✓	✓	✓	✓
Acceleration	✓	✓	✓	✓	✓	✓
Jerk	✓	✓	✓	✓	✓	✓
Dynamic	Global	Vertical	Horizontal	Global	Vertical	Horizontal
Pressure	✓					
Altitude	✓			✓		
Azimuth	✓			✓		
Other	Global	Vertical	Horizontal	Global	Vertical	Horizontal
Number of pen elevations	–	–	–	–	–	–
Number of changes in velocity	✓	✓	✓	✓	✓	✓

Source: Mekyska

The Protocol

The protocol for this study was created by a SE with extensive experience in diagnosing special learning disabilities, who is also the author of several remediation publications (e.g., Bednářová and Šmardová 2006; Bednářová 2017). All tasks were designed taking into consideration the different grades, as the complexity and difficulty were gradually increased. The assessment took approximately 50 minutes depending on the specific sample conditions. Overall, the protocol consisted of three types of tasks:

1) Seven graphomotor elements (see Figure 1): These represent the basic forms of Latin cursive letters, and for kindergarten children and first graders they are usually the very first attempt to create more complex graphomotor manifestations. For example, in this exercise the differences between the first and the second Archimedean spirals are noted. With the second one, it is more difficult to perform fine-motor movements because the loops are closer together. The last task is also very hard to accomplish for younger children, because contrary to the previous tasks the child has to use spatial abilities and combine the upper and lower loops (Mekyska et al. 2019).

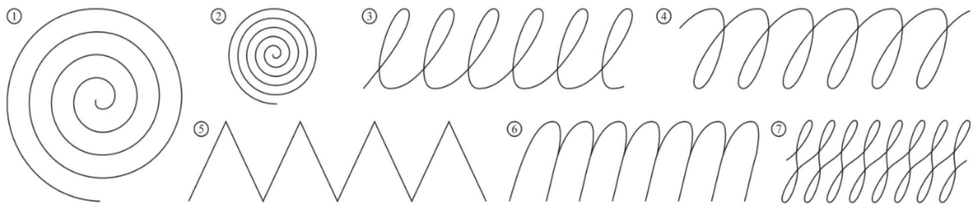


Figure 1: Examples of Graphomotor Elements
Source: Bednářová and Mekyska

2) Four sets of tasks for assessing cognitive processes (see Figure 2): These are intended to examine visuospatial abilities and working memory. In the first set of tasks, the child is asked to copy the figure into the empty box as precisely as she or he can (Task 8 in Figure 2). Immediately after that the assessor covers the upper part (the pattern and the copy) and the child is asked to draw the figure from memory (Task 9 in Figure 2). The same principle is applied to the remaining four tasks. The second set of four tasks is very similar to the previous one, where the child is asked to copy the complex figures to empty boxes (Task 18 in Figure 2). The principle of the third set is to draw five reverse figures (tasks 22 and 24 in Figure 2). The last task is based on the Rey-Osterreith complex figure principle (Rey 1959) where the child is asked to copy the complex figure (Task 27 in Figure 2), and after three minutes the child must recall the figure and draw it again (Task 28 in Figure 2).

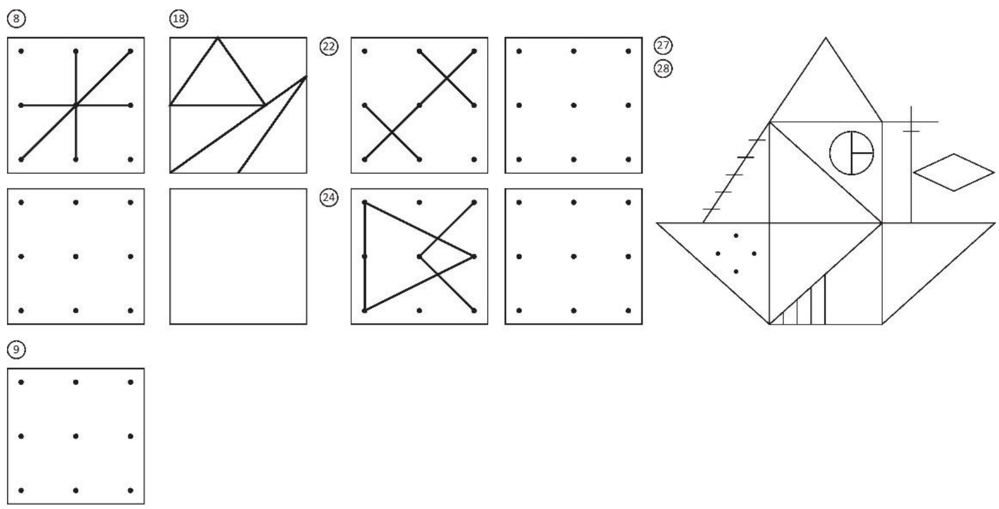


Figure 2: Examples of Cognitive Tasks
Source: Bednářová and Mekyska

3) Four sets of writing tasks: These were used to assess the handwriting process. Firstly, the child is asked for her or his signature. After this, the copying task follows, where the child copies one sentence from cursive script into cursive writing. The third task is a transcription one, where the child transcribes sentences from block letters to cursive writing, and the last one is a dictation task. According to some studies (Fryburg 1997; Margolin 1984; Parush et al. 2010), there should be differences between these types of tasks, because copying and transcriptions are based on processes where the inputs (stimuli) and outputs (final product) are visual. In the dictation task, the input is auditory, and the output is visual.

During the assessment, anamnestic data are taken to control demographic variables, such as age, sex, class, native language, parents' level of education, number of children in the family, and grades (Czech language, English language, Math); other disabilities, such as special learning disabilities diagnosis or other conditions (neurological, psychiatry, orthopedics, etc.). In addition, the Handwriting Proficiency Screening Questionnaire for Children (HPSQ-C; Rosenblum and Gafni-Lachter 2015) is used to check the child's self-evaluation. It contains ten items grouped in three factors: (1) legibility, (2) performance time, and (3) physical and emotional well-being. Every item is scored on a 5-point Likert scale (0 = never; 4 = always) and the total score is computed as the sum of points for each item. In the Czech sample, the mean is 12.86 with a standard deviation of ± 5.68. Based on this, two cut-off scores were created: the lower cut-off is 7 and the upper cut-off score is 19 (Šafárová et al. 2020).

Case Studies

Because the research is still ongoing, an example of a comparison between two children will be presented in the form of case studies. Children were chosen for this part of the study based on three criteria: (1) diagnosis, (2) HPSQ-C score, and (3) the classification provided by an expert. One child is a boy in the fourth grade from the experimental group with diagnosed dysgraphia (hereinafter DYS) and with a HPSQ-C score of 22. His scores for single HPSQ-C factors are: (1) legibility = 5; (2) performance time = 6; (3) physical and emotional well-being = 11. The other child is a girl in the fourth grade from the control group with typical handwriting development (hereinafter THD) with a HPSQ-C score of 5. Her scores for single HPSQ-C factors are: (1) legibility = 1; (2) performance time = 3; (3) physical and emotional well-being = 1. The DYS boy perceived his handwriting as less readable, it takes him more time to copy or write something, and he is not comfortable with the whole process of handwriting (e.g., feels pain, does not want to write, or feels tired).

The final output from the software is shown in Figure 3. The children are asked to transcribe the following text: Gusta, Lenka, Hana, and Stáňa are classmates. They will get a school report soon. After vacations they will attend the fourth grade. This text is a translation from the Czech version and Gusta, Lenka, Hana, and Stáňa are common Czech names. They were chosen for this task because the handwritten capital letters G, L, H, and S are difficult to remember and write for children with handwriting issues. In each image, the trajectory of the pen on the surface is depicted with blue lines and the trajectory of the pen in-air above the surface is depicted with red lines. In addition, in both figures, the pressure is represented by different shades of blue. Dark blue represents more pressure and light blue represents less pressure on the surface of the digitizer.

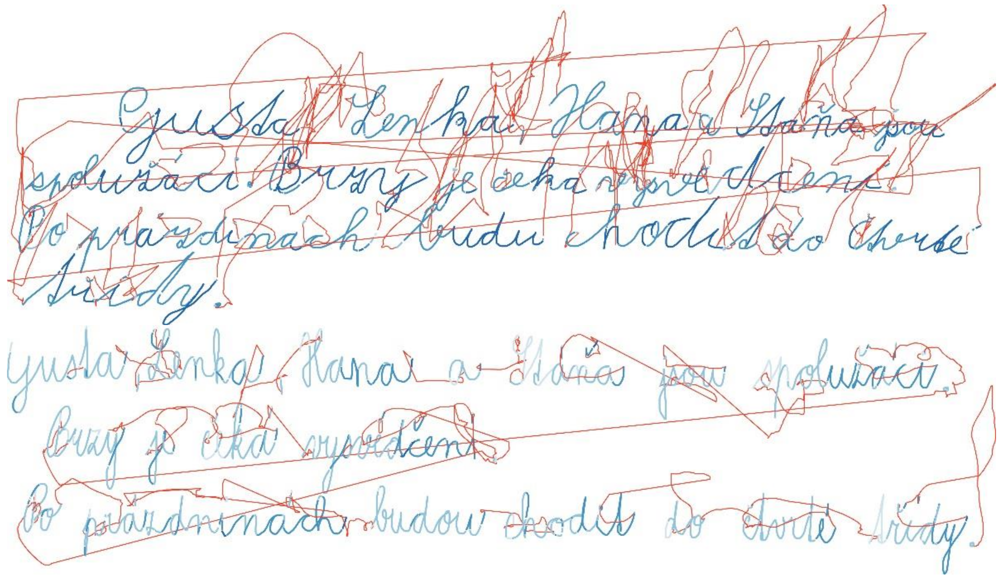


Figure 3: Comparison of Handwriting from the Transcription Task of the DYS Child (upper image) and THD Child (lower image)
Source: Zvončák

It is evident that the first writing sample is that of a child with handwriting issues. The letters are not even, they differ in size and slant, there are incorrect forms, the diacritics are missing, and there are missing letters. At first glance from a clinical point of view, it is obvious that the script is harder to read. However, with the new approach, it was found out that the DYS child spent significantly much more time in-air than the THD child (see red lines). The study showed that children with dysgraphia tend to have much longer in-air trajectories (Rosenblum, Parush, and Weiss 2003b) as they think longer about what the letter looks like and how to write it. This previously unrecognized phenomenon may indicate problems with the working memory and orthographic coding of graphemes. It also allows us to look at how the child thinks about the organization of the overall written text on the page.

Table 3 shows the values recorded for different handwriting features. In the first column, there is a differentiation between on-surface and in-air movements, corresponding with the blue lines (on-surface) and red lines (in-air) from both texts depicted in Figure 3. The features are divided into two categories, one quantifying the overall product (text) and the other quantifying particular strokes.

Table 3: Selected Handwriting Features of One DYS and One THD Child

		THD	DYS
Features of Whole Text			
On-surface	Width	49.29	63.52
	Height	192.23	250.42
	Length	2,636.84	4,107.46
	Duration	113.48	136.68
In-air	Width	46.88	79.34
	Height	194.66	257.25
	Length	1,838.15	4,857.65
	Duration	47.11	87.78
	Number of interruptions	98.00	124.00
	Mean of pressure	0.41	0.78
	Azimuth	1,220.47	1,142.45
	Altitude	568.80	672.04
Features of Strokes (Mean)			
On-surface	Width	7.92	11.65
	Height	5.91	5.90
	Length	37.69	40.78
	Duration	2.27	2.17

		THD	DYS
Features of Strokes (Mean)			
In-air	Width	16.93	27.76
	Height	7.49	16.20
	Length	29.33	67.67
	Duration	0.96	1.42
On-surface	Speed	14.58	17.49
	Velocity	16.68	18.88
	Acceleration	24.81	31.30
	Jerk	5,458.77	6,126.42
In-air	Speed	30.07	49.42
	Velocity	29.17	50.49
	Acceleration	−2.84	−111.53
	Jerk	19,289.18	−2,182.68

Source: Šafárová

Firstly, the focus will be on spatial features (given in millimeters) and the temporal features (given in seconds). The on-surface values demonstrate that the whole text of the DYS child is wider and higher, and that the overall trajectory of the written text is longer than that of the THD child. The same results concerning the mean width, height, and length of all strokes could be seen. By stroke is meant one continuous movement of the pen. One stroke can correspond to one letter, a group of letters, or even a whole word.

Also, the DYS child takes longer to complete the whole task, which indicates possible writing problems. Between the third and fourth grades of elementary school, children´s writing becomes smaller and faster. Dysgraphic children write significantly slower, and their font is large. At the same time, problems with fine-motor skills (vertical hand movements made up mainly of fingertips—height of strokes) and coarse-motor skills (horizontal hand movement made up of wrist—width of stroke) can be considered. These differences between the children described here are further accentuated by the spatial and temporal parameters above the paper surface.

The kinematic parameters are lower for the THD child. A higher value of velocity means that the task is written faster. Higher acceleration indicates higher fluctuation in the velocity of writing and the jerk feature refers to impulsiveness in writing and to the degree of sketching. From previous information, it may be concluded that the DYS child wrote the task much faster, with fewer fluctuations in speed (i.e., with less dynamics) and more sketching. However, some children without diagnosed problems have the opposite results. It can therefore be concluded that these children are slow writers because they are too focused on trying to make the script neater.

Dynamic features are represented by the pressure, altitude, and azimuth of the pen. It can be seen in the selected case that the DYS child had exerted more pressure on the tip of the pen on the surface of digitizer (see Figure 3, darker blue trajectory). Usually, clinicians or teachers report that children with dysgraphia have higher pressure but Rosenblum and Dror (2017) did not find any differences in pressure between children with and without developmental dysgraphia. The altitude parameter determines the angle between the surface and the pen. When the pen lies on the surface it is represented by 0°, and when the pen is perpendicular it is represented by 90°. The azimuth parameter specifies the position of the pen on the circle. At present, it is generally accepted that grip does not have a relationship with handwriting problems (Burton and Dancisak 2000; Sassoon, Nimmo-Smith and Wing 1986; Schwellnus et al. 2012).

Table 3 shows that the THD child interrupted the writing process several times during transcription. While it might seem that this is more typical for dysgraphic children (Chang and Yu 2013), Blöte and Hamstra-Bletz (1991) found out that in girls, especially during the automation of writing and during the creation of their own manuscript, cursive script is abandoned and there are gaps between individual letters—that is, the child makes more movements that are processed faster.

In addition, the coefficient of variation was used to compare the dynamic features between the children, which allowed us to examine differences in movement variability. Table 4 presents that the DYS child had higher variances of overall and in-air azimuth values. This means that the DYS child produced more circular movements. In addition, these differences are even more significant in air. This could be interpreted in the way that the DYS child had a problem with controlling the pen when the pen is above the surface. On the contrary, the THD child had higher variances of altitude values. That means that the variability of angles between the pen and surface is higher. The combination of both parameters could be understood as an indicator of fine-motor movements. If the position of the pen during writing is correct, the thumb and middle finger are used for creating the fine-motor movements and the index finger is used for stabilization. With the correct grip, during the writing the pen is oscillating back and forth, which causes altitude changes, but the position of the pen is more or less steady, which means no or minor azimuth changes. This combination could be seen in the handwriting of the THD child.

Table 4: Coefficients of Variation for Dynamic Parameters of One DYS and One THD Child

Dynamic Parameters	Coefficient of Variation	THD	DYS
Azimuth	Overall	0.24	0.43
	In-air	0.43	0.63
	On-surface	0.09	0.07
Altitude	Overall	0.17	0.15
	In-air	0.26	0.14
	On-surface	0.07	0.06
Pressure	Overall	0.65	0.71

Source: Šafárová

At this point it should be emphasized that the values used in the comparison of case studies are just absolute values. For some parameters (i.e., dynamic features), it will be suitable to use standard deviations or coefficient of variations because they indicate the variability in the writing process. Handwriting is a very dynamic process and in future research the focus should be on variances or changes in writing, instead of simple parameters. Lurija (1973) called the process of writing as a kinetic melody, which is apt. For those reasons it is necessary to search for and create new parameters or use the parameters from other fields of biosignal detection, or to explore the interactions between current parameters.

The Diagnostic Application: Graphomotor Disabilities Rating Scale

The general goal of the ongoing research project is to combine the process-oriented approach and the psychometric approach to create a new concept of objective GD and DD diagnosis and rating based on quantitative analysis of online handwriting. The outcome will be the graphomotor abilities rating scale (GDRS). Discrepancies between theoretical claims (e.g., dysgraphic children have problematic grip or are slower writers) and recorded clinical features will be further examined. Also, novel and nonconventional parameters, which will be able to better quantify motor skills and cognitive abilities, will be designed, as was presented in this article. Furthermore, mathematical models and machine learning that will calculate the GD score based on parameters and sociodemographic data are established. When completed, the project will produce a new DD scale, with a final score used as a rate of GD, rather than a binary classification (diagnosed/typical development).

Summary

This article referred to the current problematic state of diagnosing dysgraphia and explained that even when the disorder of written expression is defined by ICD-10 or DSM-IV, good diagnostic criteria are still missing (APA 2000; WHO 1992). Also, the relative lack of scientific interest in this topic in contrast with the related but more well-known condition of dyslexia was outlined. Present worldwide methods for measuring problems with handwriting are product-oriented and do not have sufficiently established psychometric characteristics or follow old norms and standardizations. Although there exists a possible process-oriented technique of measurement, this is currently not yet at the stage of being applied to diagnostic work. Moreover, in the Czech Republic there are no good methods specific for the Czech context for diagnosing handwriting problems. Collectively, these issues may result in the inconsistent evaluation of children with handwriting issues.

This discussion implies that a more objective assessment is needed. In the current work, process-oriented online assessment is applied. For this project, collaborating experts have created a new protocol which uses performance in tasks like graphomotor elements and complex figures to examine not only the handwriting process but also cognition (visuospatial abilities and working memory). The sampling strategy assumed distributions according to grades and diagnostic condition and included children who do not have any current diagnosis. The plan is to collect a larger sample size than has been used in similar previous research.

The main goal of the project is to create a graphomotor abilities rating scale (GASR). The GASR will be based on selected graphomotor, cognitive, and writing tasks which will distinguish among children with different degrees of handwriting difficulties. Using clinical, anamnestic, and digitally captured mechanical properties, a mathematical model will be built, which can then be used to compute scores for different categories of features (temporal, kinetics, etc.) and generate a final score so as to place a child along a scale. We think this would be an important step in generating a practical application from the research.

In this article, the new approach using data from two children who were sampled (DYS and THD) was illustrated. Comparison of these results showed that the DYS boy perceived himself as having less readable handwriting, taking more time to copy or write something, and being uncomfortable with the whole process of handwriting (e.g., feeling pain, not wanting to write, or feeling tired). From a mechanical and orthographic point of view, the DYS child's text was higher and wider on the page and it took more time to finish the writing task, with this excess duration reflecting in both on-surface and in-air measurements. The DYS child also made more interruptions during writing and exerted more pressure on the tip of the pen. Moreover, the DYS child wrote the task much faster, with fewer fluctuations in speed (i.e., with less dynamics), and more jerks. Some of these observations correspond with the classical picture of handwriting problems denoted in the literature, like for example, children with poor handwriting are slower writers (Rosenblum, Parush, and Weiss 2003a), spending more time in-air (Rosenblum, Parush, and Weiss 2003b; Rosenblum and Dror 2017), making more pauses or interruptions (Chang and Yu 2013); but some did not fit existing notions, like for example, less fluctuation in speed (Danna, Paz-Villagrán, and Velay 2013).

The field needs a better understanding of measured parameters and its relationships to create a meaningful picture of handwriting issues. This is because there are findings in the literature that seem to contradict claims made by other scholars, such as the influence of incorrect grip (Burton and Dancisak 2000; Sassoon, Nimmo-Smith, and Wing 1986; Schwellnus et al. 2012) or handedness (Ziviani and Elkins 1986). We hope that the ongoing research will enable others to empirically test some of the ideas presented in this article and contribute to the outstanding questions raised by other studies.

The case studies presented here, as well as the rest of the larger research project, have several limitations. First of all, given that it is the case that intact intelligence has been used as a rule-of-thumb to detect dysgraphia, we have chosen not to control for IQ. However, the sample was enrolled from: (1) elementary schools, which usually do not admit children with intellectual disabilities; and (2) special centers, where a diagnosis of mental retardation is an exclusion criterion for a diagnosis of dysgraphia. Given this, this limitation should not have a significant impact on the results. Another limitation is missing sociodemographic data of some of the children in the cohort. In some of the special centers, they are dealing with the European General Data Protection Regulation, known as the GDPR policy. This law enables counselling centers to share information about the children.

First, the process of handwriting itself should be mapped, as must its development and issues, and interindividual differences. Based on this, a new scale—the GDRS—will be created. At this point in the current research, the different settings of the feature groups (temporal, kinetics, etc.) are being investigated as there could be different combinations of dysgraphic issues. We hope that this scale will be helpful for professionals (special educationists and/or occupational therapists). Thus, the very first step should be to have a valid method for diagnosing, which will be also reliable. Subsequent work could then focus on remediation and therapy. With knowledge about the process and its development, researchers would be able to concentrate on the child's improvements or on contextual variables.

Problems with handwriting and graphomotor issues are long-standing and even today with computers and other smart devices, fine-motor movements are indispensable. Furthermore, writing continues to be a daily routine at schools as a significant part of the school day. Problems with writing are related not only to quantity and quality of written expression (Graham 1990), but with the child's academic achievement, which is usually assessed by her or his handwriting. Studies point to the relationship between worse grades and handwriting neatness (Brackett et al. 2013; Briggs 1980; Chase 1986; Graham, Harris, and Fink 2000; Hammerschmidt and Sudsawad 2004; Klein and Taub 2005). This approach could affect the child's self-esteem and well-being. In conclusion, if we want to offer these children better targeted care (remediation), we should understand their specific difficulties; this understanding is, of course, related to correct diagnosis.

Acknowledgement

This work was supported by a grant from the Czech Science Foundation 18-16835S (Research into advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing).

REFERENCES

Abbott, Robert D., and Virginia W. Berninger. 1993. "Structural Equation Modelling of Relationships among Developmental Skills and Writing Skills in Primary- And Intermediate-Grade Writers." *Journal of Educational Psychology* 85 (3): 478–508. https://psycnet.apa.org/doi/10.1037/0022-0663.85.3.478.

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR.* Washington, DC: American Psychiatric Association.

Amundson, Susan J. 1995. *Evaluation Tool of Children's Handwriting: ETCH Examiner's Manual*. Homer, AK: OT Kids.

Anderson, Peggy L. 1983. *Denver Handwriting Analysis: Manual*. Novato, CA: Academic Therapy Publications.

Asselborn, Thibault, Mateo Chapatte, and Pierre Dillenbourg. 2020. "Extending the Spectrum of Dysgraphia: A Data Driven Strategy to Estimate Handwriting Quality." *Scientific Reports* 10 (1): 1–11. https://doi.org/10.1038/s41598-020-60011-8.

Ayres, Leonard Porter. 1912. *A Scale for Measuring the Quality of Handwriting of School Children.* New York: Russell Sage Foundation.

Barnett, Anna, Sheila Henderson, Scheib Beverly, and Joerg Schulz. 2007. *The Detailed Assessment of Speed of Handwriting (DASH). Manual*. New York: Pearson Education.

Bednářová, Jiřina. 2017. *Mezi Námi Pastelkami* [Between Us Crayons]. Praha: Albatros Media as.

Bednářová, Jiřina, and Vlasta Šmardová. 2006. *Rozvoj grafomotoriky: Jak Rozvíjet Kreslení a Psaní* [Development of Graphomotorics: How to Develop Drawing and Writing]. Brno, Czech Republic: Edika.

Berninger, Virginia W., Robert D. Abbott, Sylvia P. Abbott, Steve Graham, and Todd Richards. 2002. "Writing and Reading: Connections Between Language by Hand and Language by Eye." *Journal of Learning Disabilities* 35 (1): 39–56. https://doi.org/10.1177/002221940203500104.

Berninger, Virginia W., and Dagmar Amtmann. 2003. "Preventing Written Expression Disabilities through Early and Continuing Assessment and Intervention for Handwriting and/or Spelling Problems: Research into Practice." In *Handbook of Learning Disabilities*, edited by H. Lee Swanson, Karen R. Harris, and Steve Graham, 345–363. New York: The Guilford Press.

Blöte, Anke W., and Lisa Hamstra-Bletz. 1991. "A Longitudinal Study on the Structure of Handwriting." *Perceptual and Motor Skills* 72:983–994. https://doi.org/10.2466%2Fpms.1991.72.3.983.

Bock, K., and W. Levelt. 1994. "Language Production: Grammatical Encoding." In *Handbook of Psycholinguistics*, edited by Morton Gernsbacher, 945–984. London: Academic Press.

Bourdin, Beatrice, and Michel Fayol. 2002. "Even in Adults, Written Production Is Still More Costly Than Oral Production." *International Journal of Psychology* 37 (4): 219–227. https://doi.org/10.1080/00207590244000070.

Brackett, Marc A., James L. Floman, Claire Ashton-James, Lillia Cherkasskiy, and Peter Salovey. 2013. "The Influence of Teacher Emotion on Grading Practices: A Preliminary Look at the Evaluation of Student Writing." *Teachers and Teaching* 19 (6): 634–646. https://doi.org/10.1080/13540602.2013.827453.

Briggs, Dennis. 1980. "A Study of the Influence of Handwriting upon Grades Using Examination Scripts." *Educational Review* 32 (2): 186–193. https://doi.org/10.1080/0013191800320207.

Burton, Allen W., and Michael J. Dancisak. 2000. "Grip Form and Graphomotor Control in Preschool Children." *American Journal of Occupational Therapy* 54: 9–17. https://doi.org/10.5014/ajot.54.1.9.

Cermak, Sharon A., and Julie Bissell. 2014. "Content and Construct Validity of Here's How I Write (HHIW): A Child's Self-Assessment and Goal Setting Tool." *American Journal of Occupational Therapy* 68: 296–306. https://doi.org/10.5014/ajot.2014.010637.

Chang, Shao-Hsia, and Nan-Ying Yu. 2013. "Handwriting Movement Analyses Comparing First and Second Graders with Normal or Dysgraphic Characteristics." *Research in Developmental Disabilities* 34 (9): 2433–2441. https://doi.org/10.1016/j.ridd.2013.02.028.

Chase, Clinton I. 1986. "Essay Test Scoring: Interaction of Relevant Variables." *Journal of Educational Measurement* 23 (1): 33–41. https://doi.org/10.1111/j.1745-3984.1986.tb00232.x.

Christensen, Carol A. 2005. "The Role of Orthographic–Motor Integration in the Production of Creative and Well-Structured Written Text for Students in Secondary School." *Educational Psychology* 25 (5): 441–453. https://doi.org/10.1080/01443410500042076.

Chung, Peter J., Dilip R. Patel, and Iman Nizami. 2020. "Disorder of Written Expression and Dysgraphia: Definition, Diagnosis, and Management." *Translational Pediatrics* 9 (1): 46–54. https://doi.org/10.21037%2Ftp.2019.11.01.

Cornhill, Heidi, and Jane Case-Smith. 1996. "Factors That Relate to Good and Poor Handwriting." *American Journal of Occupational Therapy* 50:732–739. https://doi.org/10.5014/ajot.50.9.732.

Cunningham, Anne E., and Yi-Jui Iva Chen. 2014. "Matthew Effects: The Rich Get Richer in Literacy." *Encyclopedia of Language Development.* Thousand Oaks, CA: SAGE Publications, Inc.

Danna, Jérémy, Vietminh Paz-Villagrán, and Jean-Luc Velay. 2013. "Signal-To-Noise Velocity Peaks Difference: A New Method for Evaluating the Handwriting Movement Fluency in Children with Dysgraphia." *Research in Developmental Disabilities* 34 (12): 4375–4384. https://doi.org/10.1016/j.ridd.2013.09.012.

Döhla, Diana, and Stefan Heim. 2016. "Developmental Dyslexia and Dysgraphia: What Can We Learn from the One About the Other?" *Frontiers in Psychology* 6:2045. https://doi.org/10.3389/fpsyg.2015.02045.

Engel-Yeger, Batya, Limor Nagauker-Yanuv, and Sara Rosenblum. 2009. "Handwriting Performance, Self-Reports, and Perceived Self-Efficacy among Children with Dysgraphia." *American Journal of Occupational Therapy* 63 (2): 182–192. https://doi.org/10.5014/ajot.63.2.182.

Erez, Nirit, and Shula Parush. 1999. *The Hebrew Handwriting Evaluation*. 2nd ed. Jerusalem: School of Occupational Therapy, Faculty of Medicine, Hebrew University.

Feder, Katya P., and Annette Majnemer. 2003. "Children's Handwriting Evaluation Tools and Their Psychometric Properties." *Physical & Occupational Therapy in Pediatrics* 23:65–84. https://doi.org/10.1080/J006v23n03_05.

———. 2007. "Handwriting Development, Competency, and Intervention." *Developmental Medicine & Child Neurology* 49 (4): 312–317. https://doi.org/10.1111/j.1469-8749.2007.00312.x.

Flower, Linda, and John R. Hayes. 1981. "A Cognitive Process Theory of Writing." *College Composition and Communication* 32 (4): 365–387. https://doi.org/10.2307/356600.

Fryburg, Estelle L. 1997. *Reading and Learning Disability: A Neuropsychological Approach to Evaluation and Instruction.* Springfield, IL: Charles C Thomas Pub. Limited.

Graham, Steve. 1990. "The Role of Production Factors in Learning Disabled Students' Compositions." *Journal of Educational Psychology* 82 (4): 781–791. https://psycnet.apa.org/doi/10.1037/0022-0663.82.4.781.

Graham, Steve, Karen R. Harris, and Barbara Fink. 2000. "Is Handwriting Causally Related to Learning to Write? Treatment of Handwriting Problems in Beginning Writers." *Journal of Educational Psychology* 92 (4): 620–663. https://psycnet.apa.org/doi/10.1037/0022-0663.92.4.620.

Graham, Steve, and Naomi Weintraub. 1996. "A Review of Handwriting Research: Progress and Prospects from 1980 to 1994." *Educational Psychology Review* 8:7–87. https://doi.org/10.1007/BF01761831.

Graham, Steve, Virginia Berninger, Robert D. Abbott, Sylvia P. Abbott, and Dianne Whitaker. 1997. "Role of Mechanics in Composing of Elementary School Students: A New Methodological Approach." *Journal of Educational Psychology* 89 (1): 170–182. https://psycnet.apa.org/doi/10.1037/0022-0663.89.1.170.

Hammerschmidt, Sandra L., and Pimjai Sudsawad. 2004. "Teachers' Survey on Problems with Handwriting: Referral, Evaluation, and Outcomes." *American Journal of Occupational Therapy* 58 (2): 185–192. https://doi.org/10.5014/ajot.58.2.185.

Hamstra-Bletz, Lisa, and Anke W. Blöte. 1990. "Development of Handwriting in Primary School: A Longitudinal Study." *Perceptual and Motor Skills* 70 (3): 759–770. https://doi.org/10.2466%2Fpms.1990.70.3.759.

Hamstra-Bletz, Lisa, J. DeBie, and Den Brinker. 1987. *Concise Evaluation Scale for Children's Handwriting*. Lisse, The Netherlands: Swets.

Hayes, J. R., and J. Gradwohl Nash. 1996. "On the Nature of Planning in Writing." In *The Science of Writing: Theories, Methods, Individual Differences and Applications*, edited by Levy C. Michael and Sarah Ransdell, 29–55. Mahwah, NJ: Lawrence Erlbaum.

Jones, Dian, and Carol A. Christensen. 1999. "Relationship between Automaticity in Handwriting and Students' Ability to Generate Written Text." *Journal of Educational Psychology* 91 (1): 44–49. https://psycnet.apa.org/doi/10.1037/0022-0663.91.1.44.

Kandel, Sonia, Ronald Peereman, Géraldine Grosjacques, and Michel Fayol. 2011. “For a Psycholinguistic Model of Handwriting Production: Testing the Syllable-Bigram Controversy.” *Journal of Experimental Psychology: Human Perception and Performance* 37 (4): 1310–1322. https://doi.org/10.1037/a0023094.

Kejřová, Kateřina, and Krejčová, Lenka. 2015. “Výskyt Dyslexie u Osob ve Výkonu Trestu Odnětí Svobody v České Republice” [The Incidence of Dyslexia among Prison Inmates in the Czech Republic]. *Psychologie pro praxi* 3: 75–87. https://karolinum.cz/data/clanek/2247/PPP_3_4_2015%20kejrova%2075-87.pdf.

Kellogg, Ronald T. 1996. “A Model of Working Memory in Writing.” In *The Science of Writing: Theories, Methods, Individual Differences and Applications*, edited by Levy C. Michael and Sarah Ransdell, 57–72. Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, Joseph, and David Taub. 2005. “The Effect of Variations in Handwriting and Print on Evaluation of Student Essays.” *Assessing Writing* 10 (2): 134–148. https://doi.org/10.1016/j.asw.2005.05.002.

Larsen, Stephen C., and Donald D. Hammill. 1989. *Test of Legible Handwriting: An Ecological Approach to Holistic Assessment*. Austin, TX: Pro-Ed.

Longstaff, Mitchell G., and Richard A. Heath. 1997. “Space-Time Invariance in Adult Handwriting.” *Acta Psychologica* 97 (2): 201–214. https://doi.org/10.1016/S0001-6918(97)00015-2.

Lurija, Aleksandr Romanovič. 1973. *The Working Brain: An Introduction to Neuropsychology*. New York: Basic Books.

Maples, Willis C. 2003. “The Wold Sentence Copy Test Academic Performance.” *Journal of Behavioral Optometry* 14 (3): 71–76. https://www.oepf.org/sites/default/files/journals/jbo-volume-14-issue-3/14-3%20Maples.pdf.

Margolin, David I. 1984. “The Neuropsychology of Writing and Spelling: Semantic, Phonological, Motor, and Perceptual Processes.” *The Quarterly Journal of Experimental Psychology* 36 (3): 459–489. https://doi.org/10.1080/14640748408402172.

McCloskey, Michael, and Brenda Rapp. 2017. “Developmental Dysgraphia: An Overview and Framework for Research.” *Cognitive Neuropsychology* 34 (3–4): 65–82. https://doi.org/10.1080/02643294.2017.1369016.

McCutchen, Deborah. 1996. “A Capacity Theory of Writing: Working Memory in Composition.” *Educational Psychology Review* 8 (3): 299–325. https://doi.org/10.1007/BF01464076.

Medwell, Jane, Steve Strand, and David Wray. 2009. “The Links between Handwriting and Composing for Y6 Children.” *Cambridge Journal of Education* 39 (3): 329–344. https://doi.org/10.1080/03057640903103728.

Mekyska, Jiří, Zoltán Galáž, Katarína Šafárová, Vojtěch Zvončák, Ján Mucha, Zdeněk Smékal, Anežka Onráčková, Tomáš Urbánek, Jana Marie Havigerová, Jiřina Bednářová, and Marcon Faundey-Zanuy. 2019. “Computerised Assessment of Graphomotor Difficulties in a Cohort of School-aged Children.” *2019 11th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT)*. IEEE. https://ieeexplore.ieee.org/abstract/document/8970767.

Meulenbroek, Ruud G. J., and Gerard P. Van Galen. 1989. “The Production of Connecting Strokes in Cursive Writing: Developing Co-Articulation in 8 to 12-Year-Old Children.” *Computer Recognition and Human Production of Handwriting*, 273–286. https://doi.org/10.1142/9789814434195_0019.

Milone, Michael. 2007. *THS-R: Test of Handwriting Skills, Revised*. Novato, CA: Academic Therapy Publications.

Palmis, Sarah, Jeremy Danna, Jean-Luc Velay, and Marieke Longcamp. 2017. “Motor Control of Handwriting in the Developing Brain: A Review.” *Cognitive Neuropsychology* 34 (3–4): 187–204. https://doi.org/10.1080/02643294.2017.1367654.

Parush, Shula, Nirit Lifshitz, Aviva Yochman, and Naomi Weintraub. 2010. “Relationships between Handwriting Components and Underlying Perceptual-Motor Functions Among Students During Copying and Dictation Tasks.” *OTJR: Occupation, Participation and Health* 30 (1): 39–48. https://doi.org/10.3928%2F15394492-20091214-06.

Paz-Villagrán, Vietminh, Jérémy Danna, and Jean-Luc Velay. 2014. “Lifts and Stops in Proficient and Dysgraphic Handwriting.” *Human Movement Science* 33: 381–394. https://doi.org/10.1016/j.humov.2013.11.005.

Phelps, Joanne, and Lynn Stempel. 1988. “The Children’s Handwriting Evaluation Scale for Manuscript Writing.” *Reading Improvement* 25 (4): 247–254. https://eric.ed.gov/?id=EJ386986.

Pokorná, Věra. 2015. *Teorie a Náprava Vývojových Poruch Učení a Chování*. [Theory and Remedy of Developmental Learning and Behavioral Disorders]. Praha: Portál.

Reisman, Judith E. 1993. “Development and Reliability of the Research Version of the Minnesota Handwriting Test.” *Physical & Occupational Therapy in Pediatrics* 13 (2): 41–55. https://doi.org/10.1080/J006v13n02_03.

———. 1999. *Minnesota Handwriting Assessment*. New York: Psychological Corporation.

Rey, André. 1959. *Test de Copie d’Une Figure Complexe* [The Complex Figure Test]. Paris: Centre de Psychologie Appliquée.

Richards, Todd L., Virginia W. Berninger, Pat Stock, Leah Altemeier, Pamala Trivedi, and Kenneth R. Maravilla. 2011. “Differences between Good and Poor Child Writers on fMRI Contrasts for Writing Newly Taught and Highly Practiced Letter Forms.” *Reading and Writing* 24 (5): 493–516. https://doi.org/10.1007/s11145-009-9217-3.

Rosenblum, Sara, and Gideon Dror. 2017. “Identifying Developmental Dysgraphia Characteristics Utilizing Handwriting Classification Methods.” *IEEE Transactions on Human-Machine Systems* 47 (2): 293–298. https://doi.org/10.1109/THMS.2016.2628799.

Rosenblum, Sara, and Liat Gafni-Lachter. 2015. “Handwriting Proficiency Screening Questionnaire for Children (HPSQ–C): Development, Reliability, and Validity.” *American Journal of Occupational Therapy* 69: 6903220030p1–p9. https://doi.org/10.5014/ajot.2015.014761.

Rosenblum, Sara, Patrice L. Weiss, and Shula Parush. 2003. “Product and Process Evaluation of Handwriting Difficulties.” *Educational Psychology Review* 15:41–81. https://doi.org/10.1023/A:1021371425220.

Rosenblum, Sara, Shula Parush, and Patrice L. Weiss. 2003a. “Computerized Temporal Handwriting Characteristics of Proficient and Non-Proficient Handwriters.” *American Journal of Occupational Therapy* 57 (2): 129–138. https://doi.org/10.5014/ajot.57.2.129.

———. 2003b. “The In-Air Phenomenon: Temporal and Spatial Correlates of the Handwriting Process.” *Perceptual and Motor Skills* 96 (3): 933–954. https://doi.org/10.2466%2Fpms.2003.96.3.933.

Roston, Karen Laurie, Jim Hinojosa, and Howard Kaplan. 2008. “Using the Minnesota Handwriting Assessment and Handwriting Checklist in Screening First and Second Graders’ Handwriting Legibility.” *Journal of Occupational Therapy, Schools, & Early Intervention* 1: 100–115. https://doi.org/10.1080/19411240802312947.

Šafárová, Katarína, Jiří Mekyska, Vojtěch Zvončák, Zoltán Galáž, Pavlína Francová, Barbora Čechová, Barbora Losenická, Zdeněk Smékal, Tomáš Urbánek, Jana Marie Havigerová, and Sara Rosenblum. 2020. “Psychometric Properties of Screening Questionnaires for Children with Handwriting Issues.” *Frontiers in Psychology* 10 (2020): 2937. https://doi.org/10.3389/fpsyg.2019.02937.

Sassoon, Rosemary, Ian Nimmo-Smith, and Alan M. Wing. 1986. “An Analysis of Children’s Penholds.” *Advances in Psychology* 37: 93–106. https://doi.org/10.1016/S0166-4115(09)60074-1.

Schwellnus, Heidi, Heather Carnahan, Azadeh Kushki, Helene Polatajko, Cheryl Missiuna, and Tom Chau. 2012. "Effect of Pencil Grasp on the Speed and Legibility of Handwriting in Children." *American Journal of Occupational Therapy* 66 (6): 718–726. https://doi.org/10.5014/ajot.2012.004515.

Smits-Engelsman, Bouwien C. M., and Gerard P. Van Galen. 1997. "Dysgraphia in Children: Lasting Psychomotor Deficiency or Transient Developmental Delay?" *Journal of Experimental Child Psychology* 67 (2): 164–184. https://doi.org/10.1006/jecp.1997.2400.

Stanovich, Keith E. 1993. "Does Reading Make You Smarter? Literacy and the Development of Verbal Intelligence." *Advances in Child Development and Behavior* 24:133–180. https://doi.org/10.1016/S0065-2407(08)60302-X.

Stott, Denis Herbert, Fred A. Moyes, and Sheila E. Henderson. 1985. *Diagnosis and Remediation of Handwriting Problems*. Guelph, ON, Canada: Brook Educational Publishing Ltd.

Van Galen, Gerard. P. 1991. "Handwriting: Issues for a Psychomotor Theory." *Human Movement Science* 10 (2–3): 165–191. https://doi.org/10.1016/0167-9457(91)90003-G.

Van Gemmert, Arend W. A., and Hans-Leo Teulings. 2006. "Advances in Graphonomics: Studies on Fine Motor Control, Its Development and Disorders." *Human Movement Science* 25 (4–5): 447–453. https://doi.org/10.1016/j.humov.2006.07.002.

Wallen, Margaret Anne, Mary-Ann Bonney, and Lyn Lennox. 1996. *The Handwriting Speed Test*. Adelaide: Helios Art and Book Company.

World Health Organization. 1992. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization.

Zelinková, Olga. 2015. *Poruchy Učení: Dyslexie, Dysgrafie, Dysortografie, Dyskalkulie, Dyspraxie, ADHD*. [Learning Disorders: Dyslexia, Dysgraphia, Dysortography, Dyscalculia, Dyspraxia, ADHD]. Praha: Portál.

Zesiger, Pascal, Pierre Mounoud, and Claude-Alain Hauert. 1993. "Effects of Lexicality and Trigram Frequency on Handwriting Production in Children and Adults." *Acta Psychologica* 82 (1–3): 353–365. https://doi.org/10.1016/0001-6918(93)90020-R.

Ziviani, Jenny, and John Elkins. 1986. "Effect of Pencil Grip on Handwriting Speed and Legibility." *Educational Review* 38 (3): 247–257. https://doi.org/10.1080/0013191860380305.

ABOUT THE AUTHORS

Katarína Šafárová: Doctor, Department of Psychology, Masaryk University, Brno, Czech Republic

Jiří Mekyska: Doctor, Department of Telecommunications, Brno University of Technology, Brno, Czech Republic

Vojtěch Zvončák: PhD Student, Department of Telecommunications, Brno University of Technology, Brno, Czech Republic

The International Journal of Assessment and Evaluation is one of ten thematically focused journals in the collection of journals that support The Learner Research Network—its journals, book series, conference, and online community. The journal investigates the dimensions of educational measurement.

As well as articles of a traditional scholarly type, this journal invites presentations of practice—including documentation of measurement practices and exegeses of the effects of those practices.

The International Journal of Assessment and Evaluation is a peer-reviewed, scholarly journal.

ISSN: 2327-7920

A.8 Psychometric Properties of Screening Questionnaires for Children With Handwriting Issues

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 January 2020
doi: 10.3389/fpsyg.2019.02937

Psychometric Properties of Screening Questionnaires for Children With Handwriting Issues

Katarína Šafářová[1], Jiri Mekyska[2], Vojtěch Zvončák[2], Zoltán Galáž[2], Pavlína Francová[1], Barbora Čechová[1], Barbora Losenická[1], Zdeněk Smékal[2], Tomáš Urbánek[1], Jana Marie Havigerová[1] and Sara Rosenblum[3]*

[1] Department of Psychology, Masaryk University, Brno, Czechia, [2] Department of Telecommunications, Brno University of Technology, Brno, Czechia, [3] Department of Occupational Therapy, University of Haifa, Haifa, Israel

Dysgraphia (D) is a complex specific learning disorder with a prevalence of up to 30%, which is linked with handwriting issues. The factors recognized for assessing these issues are legibility and performance time. Two questionnaires, the Handwriting Proficiency Screening Questionnaire (HPSQ) for teachers and its modification for children (HPSQ-C), were established as quick and valid screening tools along with a third factor – emotional and physical well-being. Until now, in the Czechia, there has been no validated screening tool for D diagnosis. A study was conducted on a set of 294 children from 3rd and 4th year of primary school (132 girls/162 boys; M_{age} 8.96 ± 0.73) and 21 teachers who spent most of their time with them. Confirmatory factor analysis based on the theoretical background showed poor fit for HPSQ [$\chi^2(32) = 115.07$, $p < 0.001$; comparative fit index (CFI) = 0.95; Tucker–Lewis index (TLI) = 0.93; root mean square error of approximation (RMSEA) = 0.09; standard root mean square residual (SRMR) = 0.05] and excellent fit for HPSQ-C [$\chi^2(32) = 31.12$, $p = 0.51$; CFI = 1.0; TLI = 1.0; RMSEA = 0.0; SRMR = 0.04]. For the HPSQ-C models, there were no differences between boys and girls [$\Delta\chi^2(7) = 12.55$, $p = 0.08$]. Values of McDonalds's ω indicate excellent (HPSQ, $\omega = 0.9$) and acceptable (HPSQ-C, $\omega = 0.7$) reliability. Boys were assessed as worse writers than girls based on the results of both questionnaires. The grades positively correlate with the total scores of both HPSQ ($r = 0.54$, $p < 0.01$) and HPSQ-C ($r = 0.28$, $p < 0.01$). Based on the results, for the assessment of handwriting difficulties experienced by Czech children, we recommend using the HPSQ-C questionnaire for research purposes.

Keywords: developmental dysgraphia, reliability, validity, HPSQ, HPSQ-C

OPEN ACCESS

Edited by:
Michael S. Dempsey, Boston University, United States

Reviewed by:
Ali Montazeri, Iranian Institute for Health Sciences Research, Iran
Chung-Ying Lin, The Hong Kong Polytechnic University, Hong Kong

****Correspondence:***
Katarína Šafářová
safarovakatarina@gmail.com

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *17 September 2019*
Accepted: *11 December 2019*
Published: *21 January 2020*

Citation:
Šafářová K, Mekyska J, Zvončák V, Galáž Z, Francová P, Čechová B, Losenická B, Smékal Z, Urbánek T, Havigerová JM and Rosenblum S (2020) Psychometric Properties of Screening Questionnaires for Children With Handwriting Issues. Front. Psychol. 10:2937. doi: 10.3389/fpsyg.2019.02937

INTRODUCTION

Handwriting is a complex task requiring a perfect combination of motor and cognitive skills (Feder and Majnemer, 2007; McCutchen, 2011). During childhood, children learn to write at both qualitative as well as quantitative levels, which in general spans a period of approximately 10 years (from the age of 5 years, when a child first encounters this task, up to the age of 15 years, when a child is supposed to be comfortable with writing on a daily basis), i.e., the handwriting should meet the expectations of being legible, fast enough, etc. (Ziviani and Wallen, 2006; Accardo et al., 2013). Handwriting forms the basis of a child's capability of being educated, the ability to express his/her ideas, and to communicate throughout his/her life (Graham, 1990). Therefore, writing issues could consistently cause problems in everyday life, they could lower self-esteem and reduce academic achievement (Blöte and Hamstra-Bletz, 1991; Dunford et al., 2005; Feder and Majnemer, 2007; Dinehart, 2015), e.g., teachers tend to give worse grades to children whose handwriting is poor (Briggs, 1980; Chase, 1986; Graham et al., 2000). Although children frequently use new technology, such as smartphones and tablets, handwriting is still an important part of their education process.

Dysgraphia (D) occurs in literature as a subtype of specific learning disorder (SpLD). It can be found in the 10th edition of International Statistical Classification of Diseases and Related Health Problems (ICD-10), a medical classification system established by the World Health Organization (WHO), in Specific developmental disorders of scholastic skills, more specifically, as a Disorder of written expression (F81.81; World Health Organization [WHO], 1992). This classification is used in the Czechia where the questionnaires were adapted. D is usually defined as a disturbance in the production of the written process. Döhla and Heim (2016) defined developmental D as a problem in the acquisition of writing skills and report that a child with D is below the expected level of writing performance in comparison with her/his peers. Children with this disorder are not identified as having neurological problems or mental retardation (Hamstra-Bletz and Blöte, 1990).

The prevalence of D ranges between 10 and 30% (Cermak and Bissell, 2014; Döhla and Heim, 2016). In the Czechia, the prevalence of SpLD is estimated to be 3–5% (Zelinková, 2003; Kejřová and Krejčová, 2015); nevertheless, there is a lack of sole statistics for D. Besides, boys are generally considered to be worse in legibility and quality of handwriting than girls (Hawke et al., 2009), which results in two to three times higher prevalence (Snowling, 2005; Katusic et al., 2009). The differences in prevalence percentage are due to different diagnosis criteria and a lack of information about D. In comparison, the keyword "dyslexia" has 12 120 search results according to the Web of Science, while "D" has only 832. If we want to provide children with D with better care, it is necessary to introduce better diagnostic and screening tools and to learn more about underlying processes and their manifestation.

Generally, two factors are used to assess and/or define poor handwriting: (1) legibility and (2) performance time (Graham et al., 1998; Koziatek and Powell, 2002; Rosenblum et al., 2003; Germano et al., 2016). For that reason, there are plenty of tests which have been designed to assess these two factors. Legibility is generally understood as an extent of readability of the text or as the ease with which the letters or words are recognized (Amundson, 1995). Rosenblum et al. (2003) distinguished between global and analytic types of tools used to assess legibility. The global scales are based on the overall judgment of the sole factor of legibility. The analytic ones focus on different aspects of handwriting (e.g., letter form, size, slant, spacing, alignment, spelling and grammatical mistakes, speed, and spatial organization). It is assumed that all the features are part of the legibility factor. Performance time, also referred to as speed, is usually measured as the number of letters or words per time unit (1–5 min). Recent reviews of these tests were conducted by several authors (Feder and Majnemer, 2003; Roston et al., 2008), with the same outcome: most of them do not have proper manuals or standardization, they have old norms, and they are problematic in terms of reliability and validity.

Moreover, since a single study does not provide enough evidence to validate a test, than a design of practically useful D diagnosis tool with appropriate psychometric properties must be based on several works. Finally, concerning the replication crisis in psychology, we could not neglect the impact of drawer effect (publishing only significant and positive results) on reported findings. To overcome some of the above-mentioned limitations, in 2008, Rosenblum (2008) introduced Handwriting Proficiency Screening Questionnaire (HPSQ) that is used to assess handwriting proficiency by teachers. Later, Rosenblum and Gafni-Lachter (2015) proposed its modification (HPSQ-C) that is used by children to assess themselves (more information about these questionnaires can be found in the section "Materials and Methods"). Since clinicians also reported fatigue or pain while writing and unwillingness to do homework in children with D (Benbow, 1995; Feder et al., 2000; Tseng and Chow, 2000), Rosenblum (2008) considered these factors as important signs of D and included a well-being factor into both questionnaires. This factor was omitted in previous tests and no data on that subject exists in the literature (Engel-Yeger et al., 2009). Author of both questionnaires reported sufficient reliability and validity.

In the Czechia, the D diagnostic process includes: (1) creation of family anamnesis based on interviews with parents and child her/himself; (2) teacher's evaluation of a child's performance at school, where marks and written homework are analyzed; (3) psychological examination, including assessment of intellect, working memory, and visual and spatial differentiation; (4) examination of graphomotor difficulties, motor skills, laterality, quantitative analysis of grammar mistakes in written text (dictation and transcription), and qualitative analysis of observation (pen grip, sitting position, subjective assessment of temporal, spatial, kinematic and dynamic handwriting characteristics). To assess the handwriting problems a team of experts (psychologists and a special educationist) is working together. Nevertheless, in our research, we are focusing on two types of evaluations: children and teachers. We perceive those groups are usually omitted, yet very important because they are in the front line in diagnosing D.

In Czech school practice, there is no screening tool for children or for teachers which could provide quick and efficient differentiation between children with/without handwriting problems. HPSQ and HPSQ-C could bridge this gap. They are focused on three domains of non-proficient handwriting issues, which are: (1) legibility; (2) performance time; and (3) physical and emotional well-being. Previous studies indicated that these tools could be reliable and valid for screening handwriting deficits (Rosenblum, 2008; Rosenblum and Gafni-Lachter, 2015). Moreover, its assessment could be extended by computerized analysis, which makes the overall process more objective (Mekyska et al., 2017). Nevertheless, until now there have been no norms for Czech pupils that use cursive handwriting.

To sum up, D diagnosis and rating is a complex task that nowadays relies mostly on experience of teachers, psychologists, and/or occupational therapists. There is no valid screening tool which could provide fast and reliable differentiation between dysgraphic and non-dysgraphic children in schools. Therefore, the general goal of this study is to adapt the HPSQ and HPSQ-C for Czech language and check their validity and reliability. In addition, none of the previous studies compared the results of both questionnaires. They were used as a research tool, but there is no evidence of their comparison in one context (Rosenblum, 2008; Cantero-Téllez et al., 2015). We perceive this information as missing one and this step as logical, because these questionnaires contain the same items, they are just adjusted to children or their teachers. To sum up, in the range of this study we focus on:

1. Construct validity – hypothesis: (1) Factor structure of HPSQ and HPSQ-C will correspond with its theoretical background, i.e., it should have a three-factor structure: legibility (items 1, 2, and 10), performance time (items 3, 4, and 9), and physical and emotional well-being (items 5, 6, 7, and 8) (Rosenblum, 2008; Rosenblum and Gafni-Lachter, 2015).
2. Reliability analysis – hypothesis: (2) Internal consistency (McDonald's ω) of both questionnaires will be >0.7, which is considered as an acceptable level.
3. Discriminant validity – hypotheses: (3) HPSQ and HPSQ-C will differentiate girls and boys; (4) the higher the total scores of HPSQ and HPSQ-C, the higher the average grade will be.
4. Exploration of differences between HPSQ and HPSQ-C – hypothesis: (5) There is no significant difference between total scores of HPSQ and HPSQ-C; (6) Pearson's correlations coefficient between the same items of each questionnaire will be positive and >0.6, which is considered as a strong relationship.

MATERIALS AND METHODS

Study Participants

In this study, we used two sources of data, i.e., data from children and their teachers, respectively. Each group filled in a related questionnaire (see the following section about the HPSQ and HPSQ-C instruments). We enrolled 294 Czech-speaking children (132 girls/162 boys; mean age 8.96 ± 0.73, HPSQ-C: m = 12.86, SD = 5.68) and 21 teachers who spent most of the school-time with the enrolled children (HPSQ: m = 11.55, SD = 6.79), in seven Czech schools (3rd and 4th class). Related demographic data for children can be found in **Table 1**. Thirty-three children (12.89%) were left-handed which is in line with 10–13% prevalence previously reported (Hardyck and Petrinovich, 1977; Raymond et al., 1996). Based on reports of teachers, 28.87% of children have handwriting difficulties (cf. 37.5% in Schwellnus et al., 2012). The parents of all children enrolled in the study and the teachers signed an informed consent form. Through the whole study the Ethical Principles of Psychologists and Code of Conduct released by the American Psychological Association (2019)[1] were followed.

TABLE 1 | Gender distribution in both classes.

	Third class	Fourth class	Total
Girls	73 (49.7%)	59 (40.1%)	132 (44.9%)
Boys	74 (50.3%)	88 (59.9%)	162 (55.1%)
Total	147 (100%)	147 (100%)	294 (100%)

Instruments: HPSQ and HPSQ-C

The original version of HPSQ and HPSQ-C is written in Hebrew and has been consequently translated into English (Rosenblum, 2008). Questionnaires contain the same questions which are modified for person's evaluating bias. In HPSQ, teacher is asked about her or his student's handwriting problems and in HPSQ-C children evaluate themselves. Both questionnaires comprise of 10 items that are grouped in three factors: legibility (items 1, 2, and 10), performance time (items 3, 4, and 9), and physical and emotional well-being (items 5, 6, 7, and 8) (Rosenblum, 2008; Rosenblum and Gafni-Lachter, 2015). An example of HPSQ legibility question is "*Is the child's handwriting readable?*" performance time question "*Does the child often erase while writing?*," and physical and emotional well-being question "*Does the child tire while writing?*." Every item is scored on a 5-point Likert scale ranging from 0 (never) to 4 (always). The final score (max. 40) is computed as a sum of all items, where higher sum means poorer handwriting performance. In addition, the questionnaires record information about age, class, and average grade (Czech language, English language, Maths and Fundamentals of social and natural science).

Rosenblum (2008) reports that Cronbach's α of the HPSQ and HPSQ-C is equal to 0.90 and 0.77, respectively, indicating high to moderate reliability (Rosenblum and Gafni-Lachter, 2015). Spanish colleagues (Cantero-Téllez et al., 2015) report internal consistency of HPSQ α = 0.78. First attempts to validate this method showed only two factors in HPSQ: (1) items 3 and 9 (performance time and well-being); (2) items 1, 2, and 10 (legibility); with 67% of the variance explained (Rosenblum, 2008). These results are similar to those reported by Cantero-Téllez et al. (2015): (1) items 1, 2, and 10 (legibility); (2) the rest of items (performance time and well-being together); with the 49% of the variance explained. Another study focused on factor analysis of HPSQ-C (Rosenblum and Gafni-Lachter, 2015) found two factors: (1) items 3 and 5–9 (performance time and well-being); (2) items 1, 2, 4, and 10 (legibility); these two factors together explain 45% of the variance. Rosenblum (2008) recommended further research.

Procedure

Translation Process

In the frame of this study, we performed the forward–backward translation process, where the English version was translated into Czech language (forward translation) and back into English

[1] https://www.apa.org/ethics/code/

(backward translation). As a first step the English version of both questionnaires was translated by two experts (an educational psychologist, as well as one of the authors of the study). Both had conceptual knowledge and were familiar with terminology covered by research topic. Two independent Czech versions were created and compared with minimum discrepancies.

Afterward, a third expert, a researcher in educational and school psychology, reviewed the Czech versions of the questionnaires collaboratively with one of the original translators from the previous step. The main goal was to identify inadequate concepts. In this part, the expert suggested that items 1–3, 6, and 10 should be reversed because they were negatively formulated (e.g., "*Does the child not do his/her homework?*"). This was perceived as an issue also by other researcher, e.g., Schwellnus et al. (2012) mentioned it as one of the limitations of HPSQ. In Czech language a negation could be created by prefix added to a verb, by special pronouns or adverbs. Moreover, the negation itself could make some difficulties while being cognitively processed by primary school children (Kaup et al., 2006; Lüdtke et al., 2008). Therefore, every negatively formulated item was rewritten into a positive way (in our example "*Does the child do his/her homework?*"). These items were reverse-scored during data transcription.

In the backward translation process, the Czech version of both questionnaires were translated by another researcher, who had no knowledge about the questionnaires. The final versions of HPSQ and HPSQ-C were discussed with the author Rosenblum of the both questionnaires.

Data Collection and Sample Size Justification

Recruitment of participants was done via e-mails to headmasters of 176 elementary schools in Brno, the capital of the east part of the Czechia. We got replies from seven schools. Two of the schools are attended by more than 500 pupils, three of them by more than 100 pupils, and two of the schools are attended by fewer than 50 pupils. Children and teachers were enrolled from both types of schools, both from larger schools in the city and its suburbs, and from smaller ones in villages.

Data were collected based on the convenience sampling method. Because there is no established rule of thumb for the sample size determination in the confirmatory factor analysis (CFA), or just with little empirical evidence (Guadagnoli and Velicer, 1988), we followed different recommendations. Some authors estimate that a sample of 100 participants would be sufficient for a measure with three or more indicators per factor (Anderson and Gerbing, 1984; MacCallum et al., 1999). Kline (1998) regards samples between 100 and 200 participants as medium sized. There are other researchers (e.g., Su et al., 2014; Fang et al., 2015) who argue that the minimum sample should be at least 200. For more information, we also refer to Osborne and Costello (2004) or Chang et al. (2018). Based on previous estimates, which consider minimum 200 participants in the sample, we justified our sample size as sufficient.

Both questionnaires HPSQ and HPSQ-C were administered in a paper–pencil form. At the beginning of testing we explained to the participants how to fill out the questionnaires, particularly in the case of the children. HPSQ was administered individually and HPSQ-C was administered to whole classes. Children were not aware of their teachers' evaluation.

Data Analysis

Both Kolmogorov–Smirnov (D_{294} = 0.96, p < 0.001) and Shapiro–Wilk (W_{294} = 0.96, p < 0.001) tests confirmed non-normal distribution of the HPSQ total score. Same conclusions were drawn in the case of HPSQ-C (D_{294} = 0.98, p < 0.001, W_{294} = 0.98, p = 0.001). **Table 2** shows the values of skewness (Sk) and kurtosis (Ku) for each item in both questionnaires. All values are in acceptable limits ± 2 (Trochim and Donnelly, 2006; Field, 2013; Gravetter and Wallnau, 2014) except for item 6 from HPSQ-C. Moreover, due to the fact that both overall distributions tend to be normal (**Figure 1**) and that a bigger sample size could cause that even a small deviation from normality can lead to a rejected null hypothesis of both tests (Field, 2013), in this study we decided to employ parametric tests. To analyze the data, we used IBM SPSS 25 (IBM Corp, 2017), IBM SPSS AMOS (Arbuckle, 2019), and R 3.2.2 (R Core Team, 2019).

Construct validity was tested using the CFA to measure the fit with the theoretical background of both questionnaires. As the estimation method, we used the maximum likelihood (ML) for both questionnaires. In general, there are several indices used in the literature to check the goodness fit of CFA. For continuous data Tucker–Lewis index (TLI) and comparative fit index (CFI) should be >0.95 threshold. In addition, root mean square error of approximation (RMSEA) < 0.6 and standard root mean square residual (SRMR) < 0.08 (Hu and Bentler, 1999; Schreiber et al., 2006). For computing CFA of both questionnaires, we used the lavaan package (Rosseel, 2012) and for computing sex invariance in HPSQ-C model we used the software IBM SPSS AMOS (Arbuckle, 2019).

TABLE 2 | Skewness (Sk) and kurtosis (Ku) values of each HPSQ and HPSQ-C items.

Item		Min.	Max.	*M*	SD	Sk	SD	Ku	SD
HPSQ	1	0	3	0.95	0.89	0.53	0.14	−0.63	0.28
	2	0	3	0.83	0.85	0.73	0.14	−0.28	0.28
	3	0	4	1.08	1.07	0.69	0.14	−0.51	0.28
	4	0	4	1.94	0.92	0.17	0.14	−0.40	0.28
	5	0	4	1.38	1.02	0.30	0.14	−0.71	0.28
	6	0	3	0.38	0.71	1.94	0.14	3.33	0.28
	7	0	3	0.50	0.69	1.15	0.14	0.55	0.28
	8	0	4	1.23	0.97	0.46	0.14	−0.32	0.28
	9	0	4	2.05	1.11	0.27	0.14	−0.79	0.28
	10	0	4	1.15	0.89	0.29	0.14	−0.59	0.28
HPSQC	1	0	4	1.28	0.95	0.41	0.14	−0.18	0.28
	2	0	4	0.61	0.96	1.56	0.14	1.77	0.28
	3	0	4	0.92	1.03	0.94	0.14	0.26	0.28
	4	0	4	2.10	1.06	−0.01	0.14	−0.49	0.28
	5	0	4	1.90	1.34	0.03	0.14	−1.08	0.28
	6	0	4	0.24	0.65	3.34	0.14	12.91	0.28
	7	0	4	0.93	1.17	1.04	0.14	0.07	0.28
	8	0	4	1.60	1.36	0.33	0.14	−1.01	0.28
	9	0	4	2.32	1.15	−0.14	0.14	−0.55	0.28
	10	0	4	0.97	1.13	1.06	0.14	0.34	0.28

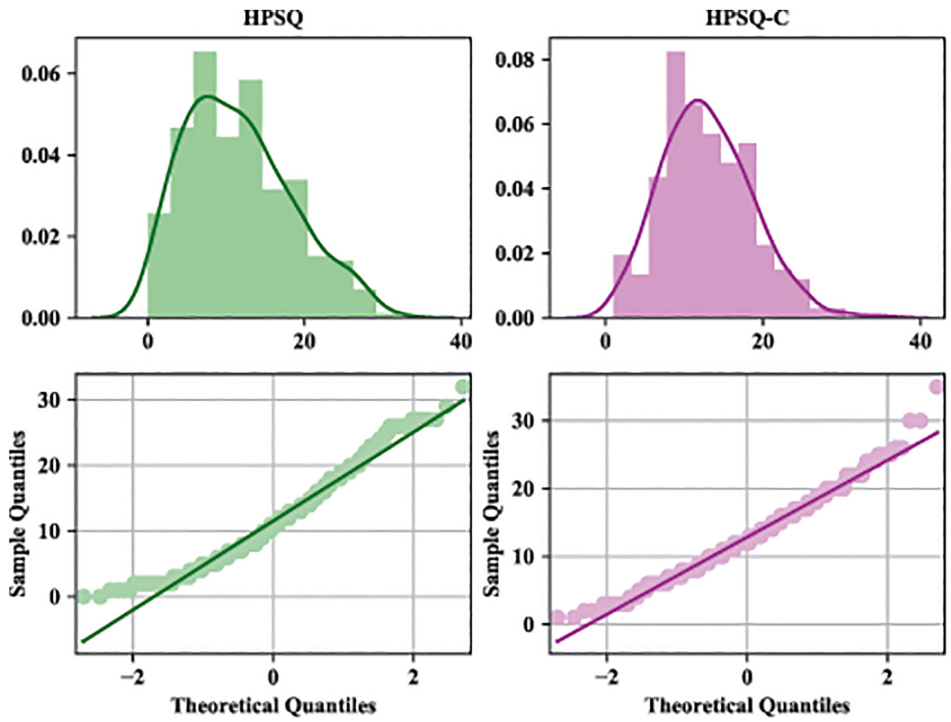


FIGURE 1 | Kernel density estimation, histogram, and Q–Q plot of HPSQ/HPSQ-C total score.

To assess the internal consistency of HPSQ and HPSQ-C, we calculated McDonalds's ω, item-total correlations, and ω coefficients in the case the items are deleted. Internal consistency of the theoretical factor structure was computed using the JASP Team (2019) and of CFA model fit using R 3.2.2 software (R Core Team, 2019) with the semTools package (Jorgensen et al., 2019).

The *t*-test (sex) and Pearson's correlation coefficient (grades) were computed for hypotheses related to the discriminant validity of HPSQ and HPSQ-C. As the last step in this article, we provide an exploration of differences and relationship between both questionnaires by computing Pearson's correlation coefficient between items of both questionnaires and *t*-test for the differences between total scores.

RESULTS

Construct Validity

Confirmatory Factor Analysis

The CFA was conducted with the three factors explaining the covariances of the HPSQ and HPSQ-C items separately (items 1, 2, and 10 loading on the legibility factor; items 3, 4, and 9 on the performance time factor; and items 5, 6, 7, and 8 on the physical and emotional well-being factor, which is the structure assumed by Rosenblum, 2008). All factor loadings were >0.4 and significant ($p < 0.01$) except these items: six HPSQ (0.36), three HPSQ-C (0.38) and six HPSQ-C (0.17). Parameter estimates, standardized error (SE), and standardized loadings (SLs) for both questionnaires with corresponding factors and their meanings are reported in **Table 3**.

The global model fit of HPSQ was statistically significant [$\chi^2(32) = 115.07$, $p < 0.001$] with indexes values CFI = 0.95, TLI = 0.93, RMSEA = 0.09 with 90% CI (0.076, 0.113), and SRMR = 0.05. The correlations among all three latent factors were all highly statistically significant ($p < 0.001$) and positive (see in **Table 4**), mostly between 0.52 and 0.66 indicating that teachers who evaluated child's issues as higher in one dimension were more likely to evaluate hers/his issues as high in the others as well. According to cut-off values mentioned in the section "Data Analysis" we do not consider these results as a good fit. The data did not support the theoretical structure.

The global model fit of HPSQ-C was not statistically significant [$\chi^2(32) = 31.12$, $p = 0.51$] with indexes values CFI = 1.0, TLI = 1.0, RMSEA = 0.0 with 90% CI (0.000, 0.042), and SRMR = 0.04. The correlations among the three factors were all highly statistically significant and positive, but weak. The range of correlation values was from 0.13 to 0.23 (see in **Table 4**). It indicates that children understand those latent variables as independent. According to cut-off values mentioned in the section "Data Analysis" we consider these results as an excellent fit. The data support the theoretical structure.

In addition, we performed the CFA invariance analysis for the HPSQ-C model, where the model has two different parts for girls ($N = 132$) and boys ($N = 162$). We used the ML estimation method, where the parameters were estimated freely in each group. We tested the model on the three levels: (1) configural invariance; (2) metric invariance; and (3) scalar invariance.

TABLE 3 | Estimates, standardized errors (SE), and standardized factor loadings (SL) from CFA for each item of HPSQ and HPSQ-C.

	Assumed factor	Item's meaning	Item	Estimate	SE	SL
HPSQ	Legibility	Legibility	1	1.00	0.00	0.78
	Legibility	Success with reading own handwriting	2	1.00	0.05	0.78
	Legibility	Satisfaction with own handwriting	10	0.81	0.06	0.63
	Performance time	Amount of time to copying	3	1.00	0.00	0.82
	Performance time	Erasing during writing	4	0.85	0.07	0.70
	Performance time	Child frequently looks at a blackboard during copying	9	0.84	0.08	0.70
	Physical and emotional well-being	Child does not want to write	5	1.00	0.00	0.90
	Physical and emotional well-being	Doing homework	6	0.41	0.04	0.36
	Physical and emotional well-being	Child feels pain (complains)	7	0.44	0.04	0.40
	Physical and emotional well-being	Tired while writing	8	0.96	0.05	0.86
HPSQ-C	Legibility	Legibility	1	1.00	0.00	0.69
	Legibility	Success with reading own handwriting	2	0.82	0.12	0.57
	Legibility	Satisfaction with own handwriting	10	0.94	0.14	0.65
	Performance time	Amount of time to copying	3	1.00	0.00	0.38
	Performance time	Erasing during writing	4	1.55	0.37	0.59
	Performance time	Child frequently looks at a blackboard during copying	9	1.39	0.35	0.53
	Physical and emotional well-being	Child does not want to write	5	1.00	0.00	0.71
	Physical and emotional well-being	Doing homework	6	0.25	0.07	0.17
	Physical and emotional well-being	Child feels pain (complains)	7	0.71	0.14	0.50
	Physical and emotional well-being	Tired while writing	8	1.37	0.22	0.97

All standardized loadings are statistically significant on the level $p < 0.01$.

TABLE 4 | Latent factor correlations in HPSQ and HPSQ-C.

Questionnaire	Factor 1	Factor 2	Correlation
HPSQ	Legibility	Performance time	0.53
	Legibility	Well-being	0.53
	Performance time	Well-being	0.67
HPSQ-C	Legibility	Performance time	0.13
	Legibility	Well-being	0.23
	Performance time	Well-being	0.19

All standardized loadings are statistically significant on the level $p < 0.01$.

A non-significant result means that the model has acceptable fit when a particular level of measured invariance is assumed (Bialosiewicz et al., 2013; Chakraborty, 2017).

Requirements of configural invariance are fulfilled when the basic factor structure is invariant for both groups (Chakraborty, 2017). The model fit for girls is not excellent, but acceptable [$\chi^2(32) = 37.38$, $p = 0.24$] with indexes CFI = 0.96, TLI = 0.94, and RMSEA = 0.04 [90% CI (0.00, 0.08)]. The model fit for boys is acceptable [$\chi^2(32) = 37.79$, $p = 0.22$] with indexes CFI = 0.97, TLI = 0.96, and RMSEA = 0.03 [90% CI (0.00, 0.07)]. The global model fit is acceptable and the obtained data for unconstrained factor structure fit well with the theoretical factor structure (see indexes in **Table 5**). Results indicate that there are no statistically significant differences between girls' and boys' models [$\Delta\chi^2(7) = 12.55$, $p = 0.08$, ΔTLI = −0.004]. Based on this results we can conclude that girls and boys conceptualized the factors in same way.

The metric invariance explains whether girls and boys answered to the items in a similar way. The obtained data for metric factor structure of boys or girls (Chakraborty, 2017) fit acceptable with the theoretical factor structure. Except the TLI value, other obtained values crossed the threshold for the rest of the goodness of fit measures (**Table 5**). Scalar estimates for every item in both groups were significant on the level $p < 0.001$. Based on this results we can conclude that there are no differences between girls and boys in the way how they answered to the items.

The scalar variance compares the means of the construct across gender groups to check if the observed scores and the latent scores are related (Chakraborty, 2017). On this level we found statistically significant differences ($p = 0.02$) and two of the indexes TLI and CFI do not cross requested threshold (**Table 5**). Those results showed that children with the same latent construct score did not have same observed scores with respect to the sex membership.

The differences between RMSEA values of configural, metric, and scalar levels are equal to 0.01 which is considered as a good indicator of invariance (Rutkowski and Svetina, 2014; Putnick and Bornstein, 2016). Usually the differences in CFI are reported and requested threshold for change is −0.01 (Cheung and Rensvold, 2010). In our study, there are bigger differences, but Rutkowski and Svetina (2014) permitted the difference −0.02, which is fulfilled between configural and metric level. Based on this results, we assumed that HPSQ-C is sex invariant on the configural and metric level, but the stricter level of scalar invariance is probably much less certain.

Internal Consistency

Firstly, we checked the internal consistency following the theoretical background. We employed the JASP Team (2019) to calculate the overall McDonald's ω as well as ω of three factors

TABLE 5 | Goodness of fit measures for different factor structure for boys and girls of HPSQ-C.

Level of factor structure/measure	χ^2	DF	p	CMIN/DF	RMSEA	TLI	CFI
Threshold			>0.05	<3	<0.06	>0.95	>0.95
Configural	75.31	65	0.18	1.16	0.02	0.96	0.97
Metric	88.14	72	0.10	1.22	0.03	0.94	0.95
Scalar	110.68	82	0.02	1.35	0.04	0.91	0.91

(legibility, performance time, and physical and emotional well-being) in each questionnaire. Based on McDonald's ω = 0.91 the reliability of overall HPSQ is considered as excellent. First subscale (legibility) had ω = 0.87, the second subscale (performance time) had ω = 0.77, and finally, the last subscale (physical and emotional well-being) had ω = 0.81. The lowest corrected correlation between item and total score was found in item 6, where $r = 0.49$. During the analysis of particular subscales, we did not find any items that should be removed.

In the case of HPSQ-C, the overall reliability is identified as acceptable (ω = 0.70). In this questionnaire the first subscale (legibility) had ω = 0.67, the second subscale (performance time) had ω = 0.46, and the third subscale (physical and emotional well-being) had ω = 0.57. Two corrected item-total correlations with the overall score were <0.3. More specifically, the corrected item-total correlation for item 3 was $r = 0.27$ and for item 6 was $r = 0.21$. Nevertheless, removal of item 3 did not increase the value of McDonald's ω of the second subscale (performance time). Only the third subscale McDonald's ω (physical and emotional well-being) increases to 0.59 when eliminating the item 6.

Additionally, the McDonald's ω was computed in the proposed models from the CFA analysis (we used the semTools package; Jorgensen et al., 2019). McDonald's ω of HPSQ for factor legibility was 0.87, for factor performance time it was 0.76, and for the last factor physical and emotional well-being it was 0.85. The overall ω for HPSQ was 0.93. McDonald's ω of HPSQ-C for factor legibility was 0.66, for factor performance time it was 0.46, and for the last factor physical and emotional well-being it was 0.60. The overall ω for HPSQ was 0.74.

Discriminant Validity

Sex Differences

Based on the independent t-test, sex differences were observed when assessed by both questionnaires. Leven's homogeneity tests were non-significant in both questionnaires: HPSQ ($F = 0.97$, $p = 0.33$) and HPSQ-C ($F = 1.44$, $p = 0.23$). Teachers evaluated boys as worse (HPSQ: $m = 12.53$, SD = 6.91, SE = 0.54) than girls (HPSQ: $m = 10.34$, SD = 6.47, SE = 0.56) with 2.2 difference [95% CI (0.64, 3.74)], which is significant [$t(292) = 2.78$, $p = 0.006$] and has medium effect size $d = 0.33$. Similarly, boys perceived themselves as worse (HPSQ-C: $m = 13.66$, SD = 5.90, SE = 0.46) than girls (HPSQ-C: $m = 11.89$, SD = 5.27, SE = 0.46) with 1.77 difference [95% CI (0.48, 3.07)], which is significant as well [$t(292) = 2.69$, $p = 0.008$] and has medium-sized effect $d = 0.32$.

Relationship to Grades

Using Pearson's correlation coefficients, we observed significant correlation between average grade and total score of HPSQ ($r = 0.54$, $p < 0.01$) and slightly weaker but still significant correlation with total score of HPSQ-C ($r = 0.28$, $p < 0.01$).

Differences Between Questionnaires

On average, children perceived themselves more strictly ($m = 12.86$; SE = 5.68) than teachers did ($m = 11.55$, SE = 6.79). This difference, 1.32, CI (0.30, 2.33) was significant $t(294) = 2.55$, $p = 0.011$ and it is represented by a small effect size $d = 0.21$.

Pearson's correlation coefficients were computed to assess the relationships between the items of both questionnaires. **Table 6** summarizes all correlation coefficients and related p-values and the correlation between the same items of both questionnaires are highlighted in bold. Items 7, 8 (physical and emotional well-being), and 9 (performance time) did not correlate significantly. The highest correlation is between the items with the number 1 (legibility; $r = 0.34$, $p < 0.01$) and 3 (performance time; $r = 0.32$, $p < 0.01$) of both questionnaires. There was a weak positive correlation between the total scores of HPSQ and HPSQ-C ($r = 0.37$, $p < 0.01$).

DISCUSSION

The primary purpose of this study was to adapt and evaluate the psychometric qualities of HPSQ and HPSQ-C as screening tools among children in the Czechia. The secondary purpose was to compare both questionnaires because there is no information about their psychometric qualities in one context. The questionnaires were designed as screening tools for identification of handwriting difficulties in children population. We replicated Rosenblum's (2008) studies (Rosenblum and Gafni-Lachter, 2015) and validated her screening questionnaires in a Czech cohort. As mentioned before, in the Czechia, there is no standardized assessment for D, nor a screening questionnaire, that would enable complex D examination. Moreover, there is no study which compares the psychometric properties of both questionnaires simultaneously.

Initially, both questionnaires were designed as follows: items 1, 2, and 10 should be grouped in factor legibility; items 3, 4, and 9 belong to factor called performance time; and finally, items 5, 7, 6, and 8 are part of physical and emotional well-being factor (Rosenblum, 2008; Rosenblum and Gafni-Lachter, 2015). For the CFA, we built the models based on the theoretical background for each questionnaire separately. The CFA showed poor model fit for the teachers' model (HPSQ) and excellent model fit for children's model (HPSQ-C). Correlation between latent factors showed that teachers had a tendency to evaluate children without discriminating

TABLE 6 | Pearson's correlation coefficients between the items of HPSQ and HPSQ-C.

		HPSQ-C											HPSQ										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Sum**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Sum**
HPSQ-C	1	1																					
	2	0.433**	1																				
	3	0.180**	0.128*	1																			
	4	0.149*	0.141*	0.211**	1																		
	5	0.227**	0.234**	0.098	0.213**	1																	
	6	0.146*	0.096	0.193**	0.070	0.111	1																
	7	0.202**	0.110	0.084	0.176**	0.177**	0.167**	1															
	8	0.196**	0.128*	0.195**	0.312**	0.399**	0.189**	0.310**	1														
	9	0.188**	0.194**	0.148*	0.270**	0.127*	0.016	0.191**	0.223**	1													
	10	0.417**	0.333**	0.115*	0.159**	0.189**	0.049	0.138*	0.185**	0.166**	1												
	Sum	0.580**	0.516**	0.434**	0.529**	0.578**	0.321**	0.507**	0.648**	0.501**	0.531**	1											
HPSQ	1	**0.336****	0.289**	0.104	0.204**	0.156**	0.139*	0.040	0.188**	0.108	0.202**	0.330**	1										
	2	0.290**	**0.279****	0.164**	0.226**	0.090	0.173**	0.009	0.145*	0.127*	0.200**	0.310**	0.805**	1									
	3	0.186**	0.223**	**0.332****	0.260**	0.139*	0.194**	−0.009	0.152**	0.095	0.169**	0.319**	0.528**	0.621**	1								
	4	0.261**	0.310**	0.150**	**0.261****	0.103	0.121*	0.015	0.153**	0.111	0.205**	0.313**	0.594**	0.562**	0.542**	1							
	5	0.184**	0.177**	0.198**	0.199**	**0.175****	0.125*	0.002	0.169**	0.087	0.121*	0.271**	0.558**	0.602**	0.559**	0.620**	1						
	6	0.170**	0.219**	0.187**	0.158**	0.027	**0.199****	0.015	0.097	0.047	0.097	0.213**	0.413**	0.450**	0.363**	0.343**	0.492**	1					
	7	0.079	0.181**	0.221**	0.190**	0.149*	0.120*	**0.090**	0.191**	0.058	0.084	0.260**	0.387**	0.400**	0.470**	0.452**	0.510**	0.186**	1				
	8	0.137*	0.175**	0.263**	0.254**	0.153**	0.090	−0.068	**0.114**	0.102	0.108	0.249**	0.527**	0.608**	0.645**	0.611**	0.787**	0.426**	0.519**	1			
	9	0.099	0.116*	0.205**	0.268**	0.070	0.111	−0.045	0.065	**0.056**	0.119*	0.194**	0.381**	0.422**	0.558**	0.489**	0.474**	0.260**	0.225**	0.534**	1		
	10	0.294**	0.238**	0.178**	0.191**	0.191**	0.150**	0.007	0.162**	0.076	**0.230****	0.321**	0.625**	0.626**	0.440**	0.572**	0.566**	0.389**	0.374**	0.537**	0.325**	1	
	Sum	0.271**	0.289**	0.278**	0.304**	0.176**	0.187**	−0.002	0.186**	0.119*	0.205**	**0.373****	0.775**	0.815**	0.787**	0.782**	0.835**	0.562**	0.589**	0.843**	0.659**	0.729**	1

***Correlation is significant at the 0.01 level (two-tailed). *Correlation is significant at the 0.05 level (two-tailed).*

based on the theorized factors. In contrast, children perceived those factors as relatively independent. In addition, we checked measurement properties varied by sex, which came out as not significant. We also conduct the invariance analysis for groups of girls and boys with results that suggest possibility of claiming configural and metric invariance for the HPSQ-C, but not scalar invariance.

Previous studies used exploratory factor analysis with different outcomes. Rosenblum (2008) reports two factors in HPSQ. The first factor includes items 3–9 and the second factor comprises items 1, 2, and 10. Her results were confirmed by a Spanish sample (Cantero-Téllez et al., 2015) with the same factor arrangement. Also, in this case, item 6 had the lowest factor score (0.18). Similarly, the original study of HPSQ-C mentions only two factors. The first factor includes item 3 and items from 5 to 9, and the second factor was formed by items 1, 2, 4, and 10 (Rosenblum and Gafni-Lachter, 2015). Based on our results we can conclude that our data for HPSQ did not support the theoretical structure. In contrast, our data support the proposed theoretical structure for HPSQ-C.

We think that the differences between our study and those published in the original study could be caused by the reversal of some items meaning. The same issue with the double negation in the HPSQ items was identified by Schwellnus et al. (2012). Another explanation could be based on the different number of evaluators in both questionnaires. Even when the number of evaluations was the same (N = 294), there was a difference in the numbers of independent evaluations of HPSQ (N = 21) and HPSQ-C (N = 294). Hammerschmidt and Sudsawad (2004) reported that the most important criterion which teachers used for evaluating handwriting issues is legibility (67.8%; N = 314). Furthermore, as a major method for handwriting evaluation, they compare student's handwriting to classroom peers (36.8%). These outcomes support our results. We understand that as an explanation of higher correlations between latent variables in HPSQ model. That is, teachers are probably better in the evaluation of the whole group than of individuals.

Values of McDonald's ω of the theoretical model indicate excellent (HPSQ, ω = 0.91) and acceptable (HPSQ-C, ω = 0.70) reliability of both questionnaires. Further analysis suggests deleting two items: items 3 and 6 from HPSQ-C. Nevertheless, without item 3, the reliability will not decrease. The total values of McDonald s ω in the proposed CFA model of HPSQ and HPSQ-C could be considered as nearly excellent (ω = 0.93 and ω = 0.74). Both values meet the condition of acceptable reliability for research purposes.

We have a common finding in all results points at item 6 (doing homework). The Sk value 3.34 and the Ku value 12.91 in HPSQ-C indicate that there are very few children who do not do homework. This item was also the less assessed one by both groups, children (m = 0.24) and teachers (m = 0.39), which again explains minimal problems with homework. We assume that this particular item has minimal discrimination information because almost every Czech child in 3rd and 4th grade do his/her homework. In addition, thanks to the sex invariance analysis in CFA, we know that this issue is related only to the group of girls, because the standardized factor loading of item 6 in HPSQ-C was not significant (p = 0.79). This means that this item does not detect differences among girls.

Reliability analysis of HPSQ-C also excludes item 3 (assessing whether a child has enough time to copy text from blackboard). Rosenblum and Gafni-Lachter (2015) wrote about the content validation process. They asked 10 children to complete HPSQ-C questionnaire and rate its items based on their clarity. Each item obtained 100% agreement, nevertheless, two children had a problem with item 3. They reported problems with meaning "repeatedly" in comparison with their classmates. Also, a few teachers from our study had the same difficulties. They complained that the question is not clear. According to them, the time of copying the text from blackboard depends on the length and complexity of sentence or paragraph. Even when internal consistency recommended to delete items 3, 6, and 9, we did not do this. Those items could be more efficient in other age cohorts and further analysis is needed. Moreover, the content of all items is meaningful considering the scope of the questionnaires.

Boys were assessed as worse writers than girls by both groups – teachers (HPSQ) and children (HPSQ-C). These gender differences are a well-known fact, which corresponds with previous research (Blöte and Hamstra-Bletz, 1991; Hawke et al., 2009; Shih et al., 2018). Next, we found out that grades positively correlate with scores of both questionnaires. In addition, worse school achievement is linked with worse handwriting. Similar findings could be found in other studies (Graham et al., 2000; Klein and Taub, 2005).

We found significant differences between total scores made by children (HPSQ-C) and their teachers (HPSQ). Children from our study were more critical during self-evaluation. Similar conclusions are reported by Rosenblum and Gafni-Lachter (2015). The authors write that: "*.children as a whole evaluated their handwriting as less proficient than did their teachers*." (p. 5). According to outcomes of correlation analysis, there were three items of HPSQ and HPSQ-C which did not correlate: item 7 focused on child's complaining to pain during the writing, item 8 aimed at fatigue during writing (both from physical and emotional well-being factor) and item 9 surveyed frequency of looking at a blackboard during copying (from performance time factor).

In accordance with the first research studies (Rosenblum, 2008; Rosenblum and Gafni-Lachter, 2015), we found out that children can better distinguish between questions in HPSQ-C questionnaire. Some studies indicate potential disparities between children and teachers/parents in the way of evaluation (e.g., Sturgess and Ziviani, 1996; Bouman et al., 1999; Petersson et al., 2013). Other studies report that children could be better judges of their performance than their parents or teachers (Petersson et al., 2013). Hammerschmidt and Sudsawad (2004) reported that teachers' evaluation of handwriting problems is not congruent with standardized tests (Sudsawad et al., 2001).

When adults assess children, their opinion could be influenced by their point of view. They are trying to figure out how the child should feel in a situation (Begley, 2000). Teachers cannot discriminate between items and understand them as well as children do. This finding is consistent with that of Germano et al. (2016) who reported that HPSQ did not distinguish dyslectic students. They stated that: "*...teachers perhaps do not have enough knowledge about the handwriting skills*." (p. 593) as an explanation for higher "never" and "rarely" answer frequency. Based on our results, we can infer that children's perception of their handwriting is quite different and more accurate.

This study has several limitations. First, we did not control the IQ variable. Nevertheless, the cohort was enrolled in elementary schools where children with mental retardation are usually not included. For that reason, we do not assume this limitation would have a significant impact on our results. Next, we did not compare the results with a baseline diagnosis of D; however, the diagnostic assessment of D in the Czechia is rather subjective and there are problems with establishing the correct diagnosis. On the other hand, the percentage of children with handwriting problems confirmed the estimated prevalence of D in foreign studies (Cermak and Bissell, 2014; Döhla and Heim, 2016). Another limitation could be the different number of independent evaluations collected in each questionnaire as mentioned above.

Another limitation is linked with the teachers' sample. Variables such are sex, age, or years of experience were not recorded for teachers, therefore they could not be controlled. In future research, especially for the HPSQ these variables should be part of the questionnaire to observe their potential influence on the results. Also, the sample of children consisted only of those enrolled from the 3rd and 4th grades, and this could be seen as non-representative. We chose this range for several reasons: (1) during the 3rd and 4th grades handwriting becomes automatic, (2) thus handwriting issues are more conspicuous and (3) the disorder of written expression (F81.81 in ICD-10) is diagnosed between the 3rd and 4th grades. Nevertheless, Rosenblum and Gafni-Lachter (2015) used HPSQ-C for younger children, i.e., from first grade. Further research in this field is needed.

The last limitation which we want to emphasize is the translation process. The original questionnaire was in Hebrew, but for forward and backward translation we used the English version. Moreover, we did not conduct any of recommended final steps (i.e., cognitive interview, expert panel, or pilot study). Given that the items of the questionnaires are quite simple in terms of comprehension, we did not assume any difficulty of understanding from the children's or teachers' points of view. This statement is supported by the fact that there were no conceptual or terminology discrepancies between versions, nor in the forward or in the backward phases of translation.

In the Czechia, there is no quick and reliable screening tool for D assessment which could help teachers, children, and their parents to recognize handwriting issues. Therefore, in the frame of this study, we adapted HPSQ and HPSQ-C questionnaires for the Czech population and evaluated their reliability and validity. Based on a statistical analysis we recommend the HPSQ-C questionnaire for further research use in the Czech population. We would like to emphasize that this questionnaire is only considered as an auxiliary screening tool which should help teachers to identify children with handwriting difficulties. But some additional and accurate diagnosis is still necessary. To sum up, based on the results we suggest that: (1) a child is a better evaluator of her/his issues and they should be seen through her/his eyes; (2) in the case of further research in other languages, inversion of items 1–3, 6, and 10 should be considered. To develop a full picture of psychometric characteristics of this questionnaire additional studies are needed.

As the next logical step in this research, we are going to extend the D assessment methodology based on HPSQ-C by applying a quantitative analysis of digitized handwriting/drawing and utilization of machine learning. This approach can enable us to identify underlying patterns and processes in children with graphomotor disabilities, which can make the general diagnosis more objective and accurate.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Board of the Department of Psychology from Masaryk University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin and by themselves.

AUTHOR CONTRIBUTIONS

SR is the author of both questionnaires, and with KS and JM contributed conception and design of the study. PF, BC, and BL collected the data and organized database. KS and TU performed the statistical analysis and interpretation of data. KS translated both questionnaires and wrote the first draft of the manuscript. JM, VZ, ZG, TU, PF, BC, and BL wrote sections of the manuscript. JH contributed with translation process and language correction of both questionnaires. ZS, JH, and SR provide revision of whole article critically for important intellectual content. All authors contributed to manuscript revision, and read and approved the submitted version of the manuscript.

FUNDING

This work was supported by the grant of the Czech Science Foundation 18-16835S (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing).

REFERENCES

Accardo, A. P., Mariangela, G., and Michela, B. (2013). Development, maturation and learning influence on handwriting kinematics. *Hum. Mov. Sci.* 32, 136–146. doi: 10.1016/j.humov.2012.10.004

American Psychological Association (2019). *Ethical Principles of Psychologists and Code of Conduct*. Washington, DC: American Psychological Association. Available at: https://www.apa.org/ethics/code/ (accessed February 22, 2019).

Amundson, S. J. (1995). *Evaluation Tool of Children's Handwriting: ETCH, Examiner's Manual*. Homer, PA: O.T. Kids.

Anderson, J. C., and Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika* 49, 155–173. doi: 10.1007/BF02294170

Arbuckle, J. L. (2019). *Amos (Version 26.0) [Computer Program]*. Chicago, IL: IBM SPSS.

Begley, A. (2000). *The Educational Self-Perceptions of Children With Down Syndrome*, eds A. Lewis, and G. Lindsay, (Buckingham: Open University Press), 98–111.

Benbow, M. (1995). *Principles and Practices of Teaching Handwriting*, eds A. Henderson, and C. Pehoski, (St. Louis, MO: Mosby), 255–281.

Bialosiewicz, S., Murphy, K., and Berry, T. (2013). *An Introduction to Measurement Invariance Testing: Resource Packet for Participants: Do our Measures Measure up? The Critical Role of Measurement Invariance*. Washington, DC: American Evaluation Association.

Blöte, A. W., and Hamstra-Bletz, L. (1991). A longitudinal study on the structure of handwriting. *Percept. Mot. Skills* 72, 983–994. doi: 10.2466/pms.1991.72.3.983

Bouman, N. H., Koot, H. M., Van Gils, A. P., and Verhulst, F. C. (1999). Development of a health-related quality of life instrument for children: the quality of life questionnaire for children. *Psychol. Health* 14, 829–846. doi: 10.1080/08870449908407350

Briggs, D. (1980). A study of the influence of handwriting upon grades using examination scripts. *Educ. Rev.* 32, 185–193. doi: 10.1080/0013191800320207

Cantero-Téllez, R., Porqueres, M. I., Piñel, I., and Orza, J. G. (2015). Cross-cultural Adaptation, internal consistency and validity of the handwriting proficiency screening questionnaire (HPSQ) for Spanish primary school-age children. *J. Nov. Physiother.* 5:278. doi: 10.4172/2165-7025.1000278

Cermak, S. A., and Bissell, J. (2014). Content and construct validity of Here's How I Write (HHIW): a child's self-assessment and goal setting tool. *Am. J. Occup. Therapy* 68, 296–306. doi: 10.5014/ajot.2014.010637

Chakraborty, R. (2017). Configural, metric and scalar invariance measurement of academic delay of gratification scale. *Int. J. Human. Soc. Stud.* 3:3.

Chang, C.-C., Lin, C.-Y., Gronholm, P. C., and Wu, T.-H. (2018). Cross-validation of two commonly used self-stigma measures, Taiwan versions of the internalized stigma mental illness scale and self-stigma scale-short, for people with mental illness. *Assessment* 25, 777–792. doi: 10.1177/1073191116658547

Chase, C. (1986). Essay test scoring: interaction of relevant variables. *J. Educ. Meas.* 23, 33–41. doi: 10.1111/j.1745-3984.1986.tb00232.x

Cheung, G. W., and Rensvold, R. B. (2010). Testing factorial invariance across groups: a reconceptualization and proposed new method. *Int. J. Psychol. Res.* 3, 111–121. doi: 10.1177/014920639902500101

Dinehart, L. H. (2015). Handwriting in early childhood education: current research and future implications. *J. Early Child. Liter.* 15, 97–118. doi: 10.1177/1468798414522825

Döhla, D., and Heim, S. (2016). Developmental dyslexia and dysgraphia: what can we learn from the one about the other? *Front. Psychol.* 6:2045. doi: 10.3389/fpsyg.2015.02045

Dunford, C., Missiuna, C., Street, E., and Sibert, J. (2005). Children's perceptions of the impact of developmental coordination disorder on activities of daily living. *Br. J. Occup. Therapy* 68, 207–214. doi: 10.1177/030802260506800504

Engel-Yeger, B., Jarus, T., Anaby, D., and Law, M. (2009). Differences in patterns of participation between youths with cerebral palsy and typically developing peers. *Am. J. Occup. Therapy* 63, 96–104. doi: 10.5014/ajot.63.1.96

Fang, S.-Y., Lin, Y.-C., Chen, T.-C., and Lin, C.-Y. (2015). Impact of marital coping on the relationship between body image and sexuality among breast cancer survivors. *Support. Care Cancer* 23, 2551–2559. doi: 10.1007/s00520-015-2612-1

Feder, K. P., and Majnemer, A. (2007). Handwriting development, competency, and intervention. *Dev. Med. Child Neurol.* 49, 312–317. doi: 10.1111/j.1469-8749.2007.00312.x

Feder, K. P., and Majnemer, A. (2003). Children's handwriting evaluation tools and their psychometric properties. *Phys. Occup. Therapy Pediatr.* 23, 65–84. doi: 10.1300/j006v23n03_05

Feder, K., Majnemer, A., and Synnes, A. (2000). Handwriting: current trends in occupational therapy practice. *Can. J. Occup. Ther.* 67, 197–204. doi: 10.1177/000841740006700313

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Thousand Oaks, CA: SAGE Publications Ltd.

Germano, G. D., Giaconi, C., and Capellini, S. A. (2016). Characterization of brazilians students with dyslexia in handwriting proficiency screening questionnaire and handwriting scale. *Psychol. Res.* 6, 590–597. doi: 10.17265/2159-5542/2016.10.004

Graham, S., Harris, K. R., and Fink, B. (2000). Is handwriting causally related to learning to write? Treatment of handwriting problems in beginning writers. *J. Educ. Psychol.* 92, 620–633. doi: 10.1037/0022-0663.92.4.620

Graham, S., Weintraub, N., and Berninger, V. W. (1998). The relationship between handwriting style and speed and legibility. *J. Educ. Res.* 91, 290–297. doi: 10.1080/00220679809597556

Graham, S. (1990). The role of production factors in learning disabled students' compositions. *J. Educ. Psychol.* 82, 781–791. doi: 10.1037/0022-0663.82.4.781

Gravetter, F., and Wallnau, L. (2014). *Essentials of Statistics for the Behavioral Sciences* 8th Edn. Belmont, CA: Wadsworth.

Guadagnoli, E., and Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychol. Bull.* 103, 265–275. doi: 10.1037//0033-2909.103.2.265

Hammerschmidt, S. L., and Sudsawad, P. (2004). Teachers' survey on problems with handwriting: referral, evaluation, and outcomes. *Am. J. Occup. Therapy* 58, 185–192. doi: 10.5014/ajot.58.2.185

Hamstra-Bletz, L., and Blöte, A. W. (1990). Development of handwriting in primary school: a longitudinal study. *Percept. Mot. Skills* 70, 759–770. doi: 10.2466/PMS.70.3.759-770

Hardyck, C., and Petrinovich, L. F. (1977). Left-handedness. *Psychol. Bull.* 84, 385–404. doi: 10.1037/0033-2909.84.3.385

Hawke, J. L., Olson, R. K., Willcutt, E. G., Wadsworth, S. J., and DeFries, J. C. (2009). Gender ratios for reading difficulties. *Dyslexia* 15, 239–242. doi: 10.1002/dys.389

Hu, L.-T., and Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

IBM Corp (2017). *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM Corp.

JASP Team, (2019). *JASP (Version 0.11.1)[Computer software]*.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., and Rosseel, Y. (2019). *semTools: Useful Tools for Structural Equation Modeling. R package version 0.5-2*. Available at: https://CRAN.R-project.org/package=semTools (accessed November 20, 2019).

Katusic, S. K., Colligan, R. C., Weavers, A. L., and Barbaresi, W. J. (2009). Forgotten learning disability – epidemiology of written language disorder in a population-based birth cohort (1976–1982), Rochester, Minnesota. *Pediatrics* 123, 1306–1313. doi: 10.1542/peds.2008-2098

Kaup, B., Lüdtke, J., and Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *J. Pragmat.* 38, 1033–1050. doi: 10.1016/j.pragma.2005.09.012

Kejřová, K., and Krejčová, L. (2015). Výskyt dyslexie u osob ve výkonu trestu odnětí svobody v České republice. *Psychol. Pro Praxis* 50, 75–87.

Klein, J., and Taub, D. (2005). The effect of variations in handwriting and print on evaluation of student essays. *Assess. Writ.* 10, 134–148. doi: 10.1016/j.asw.2005.05.002

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Press.

Koziatek, S. M., and Powell, N. J. (2002). A validity study of the evaluation tool of children's handwriting–cursive. *Am. J. Occup. Therapy* 56, 446–453. doi: 10.5014/ajot.56.4.446

Lüdtke, J., Friedrich, C. K., De Filippis, M., and Kaup, B. (2008). Event-related potential correlates of negation in a sentence–picture verification paradigm. *J. Cogn. Neurosci.* 20, 1355–1370. doi: 10.1162/jocn.2008.20093

MacCallum, R. C., Widaman, K. F., Zhang, S., and Hong, S. (1999). Sample size in factor analysis. *Psychol. Methods* 4, 84–99. doi: 10.1037/1082-989X.4.1.84

McCutchen, D. (2011). From novice to expert: implications of language skills and writing-relevant knowledge for memory during the development of writing skill. *J. Writ. Res.* 3, 51–68 doi: 10.17239/jowr-2011.03.01.3

Mekyska, J., Faundez-Zanuy, M., Mzourek, Z., Galaz, Z., Smekal, Z., and Rosenblum, S. (2017). Identification and rating of developmental dysgraphia by handwriting analysis. *IEEE Trans. Hum. Mach. Syst.* 47, 235–248. doi: 10.1109/THMS.2016.2586605

Osborne, J. W., and Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Pract. Assess. Res. Eval.* 9:8.

Petersson, C., Simeonsson, R. J., Enskar, K., and Huus, K. (2013). Comparing children's self-report instruments for health-related quality of life using the International Classification of Functioning, Disability and Health for Children and Youth (ICF-CY). *Health Qual. Life Outcomes.* 11:75. doi: 10.1186/1477-7525-11-75

Putnick, D. L., and Bornstein, M. H. (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev. Rev.* 41, 71–90. doi: 10.1016/j.dr.2016.06.004

R Core Team (2019). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Raymond, M., Pontier, D., Dufour, A. B., and Møller, A. P. (1996). Frequency-dependent maintenance of left handedness in humans. *Proc. R. Soc. Lond. B.* 263, 1627–1633. doi: 10.1098/rspb.1996.0238

Rosenblum, S. (2008). Development, reliability, and validity of the Handwriting Proficiency Screening Questionnaire (HPSQ). *Am. J. Occup. Ther.* 62, 298–307. doi: 10.5014/ajot.62.3.298

Rosenblum, S., and Gafni-Lachter, L. (2015). Handwriting proficiency screening questionnaire for children (HPSQ–C): development, reliability, and validity. *Am. J. Occup. Ther.* 69, 6903220030p1–6903220030p9.

Rosenblum, S., Weiss, P. L., and Parush, S. (2003). Product and process evaluation of handwriting difficulties. *Educ. Psychol. Rev.* 15, 41–81. doi: 10.1023/A:1021371425220

Rosseel, Y. (2012). lavaan: an r package for structural equation modeling. *J. Stat. Softw.* 48, 1–36.

Roston, K. L., Hinojosa, J., and Kaplan, H. (2008). Using the minnesota handwriting assessment and handwriting checklist in screening first and second graders' handwriting legibility. *J. Occup. Ther. Sch. Early Interven.* 1, 100–115. doi: 10.1080/19411240802312947

Rutkowski, L., and Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educ. Psychol. Meas.* 74, 31–57. doi: 10.1177/0013164413498257

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., and King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *J. Educ. Res.* 99, 323–338. doi: 10.3200/JOER.99.6.323-338

Schwellnus, H., Carnahan, H., Kushki, A., Polatajko, H., Missiuna, C., and Chau, T. (2012). Effect of pencil grasp on the speed and legibility of handwriting in children. *Am. J. Occup. Ther.* 66, 718–726. doi: 10.5014/ajot.2012.004515

Shih, H.-N., Tsai, W.-H., Chang, S.-H., Lin, C.-Y., Hong, R.-B., and Hwang, Y.-S. (2018). Chinese handwriting performance in preterm children in grade 2. *PLoS One* 13:e0199355. doi: 10.1371/journal.pone.0199355

Snowling, M. J. (2005). Specific learning difficulties. *Psychiatry* 4, 110–113. doi: 10.1383/psyt.2005.4.9.110

Sturgess, J., and Ziviani, J. (1996). A self-report play skills questionnaire: technical development. *Aust. Occup. Ther. J.* 43, 142–154. doi: 10.1111/j.1440-1630.1996.tb01850.x

Su, C. -T., Ng, H. -S., Yang, A. -L., and Lin, C. -Y. (2014). Psychometric evaluation of the Short Form 36 Health Survey (SF-36) and the World Health Organization Quality of Life Scale Brief Version (WHOQOL-BREF) for patients with schizophrenia. *Psychol. Assess.* 26, 980–989. doi: 10.1037/a0036764

Sudsawad, P., Trombly, C. A., Henderson, A., and Tickle-Degnen, L. (2001). The relationship between the evaluation tool of Children's handwriting and teachers perceptions of handwriting legibility. *Am. J. Occup. Ther.* 55, 518–523. doi: 10.5014/ajot.55.5.518

Trochim, W. M., and Donnelly, J. P. (2006). *The Research Methods Knowledge Base* (3rd ed). Cincinnati, OH: Thomson Learning.

Tseng, M. H, and Chow, S. M. (2000). Perceptual-motor function of school-age children with slow handwriting speed. *Am. J. Occup. Ther.* 54, 83–88. doi: 10.5014/ajot.54.1.83

World Health Organization [WHO] (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization.

Zelinková, O. (2003). *Poruchy Učení: Specifické Vývojové Poruchy Ètení, Psaní a Dalších Školních Dovedností*. Praha: Portál.

Ziviani, J. M. and Wallen, M. (2006). "The Development of Graphomotor Skills," in *Hand Function in the Child: Foundations for Remediation* eds A. Henderson, and C. Pehoski, (St Louis, MO: Mosby Elsevier), 217-236.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

A.9 Advanced Parametrization of Graphomotor Difficulties in School-Aged Children

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 13, 2020, accepted June 15, 2020, date of publication June 17, 2020, date of current version June 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003214

Advanced Parametrization of Graphomotor Difficulties in School-Aged Children

ZOLTAN GALAZ[1], JAN MUCHA[1], VOJTECH ZVONCAK[1], JIRI MEKYSKA[1], ZDENEK SMEKAL[1], (Senior Member, IEEE), KATARINA SAFAROVA[2], ANEZKA ONDRACKOVA[2], TOMAS URBANEK[2], JANA MARIE HAVIGEROVA[2], JIRINA BEDNAROVA[2], AND MARCOS FAUNDEZ-ZANUY[3]

[1]Department of Telecommunications, Brno University of Technology, 61600 Brno, Czech Republic
[2]Department of Psychology, Faculty of Arts, Masaryk University, 60200 Brno, Czech Republic
[3]Escola Superior Politecnica, Tecnocampus, 08302 Mataró, Spain

Corresponding author: Zoltan Galaz (galaz@feec.vutbr.cz)

This work was supported in part by the Czech Science Foundation (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing) under Grant 18-16835S, and in part by the Project under Grant TEC2016-77791-C4-2-R.

ABSTRACT School-aged children spend 31–60 % of their time at school performing handwriting, which is a complex perceptual-motor skill composed of a coordinated combination of fine graphomotor movements. As up to 30 % of them experience graphomotor difficulties (GD), timely diagnosis of these difficulties and therapeutic intervention are of great importance. At present, an objective, computerized decision support system for the identification and assessment of GD in school-aged children is still missing. In this study, we propose three novel advanced handwriting parametrization techniques based on modulation spectra, fractional order derivatives, and tunable Q-factor wavelet transform to improve the identification of GD using online handwriting. For this purpose, we analyzed signals acquired from 7 basic graphomotor tasks performed by 53 children attending 3rd and 4th grade at several primary schools around the Czech Republic. Combining the newly proposed features with the conventionally used ones, we were able to identify GD with 84 % accuracy. In this study, we showed that using advanced parametrization of basic graphomotor movements can be potentially used to improve our capabilities of quantifying problems with the development of legible, fast-paced handwriting, and help with the early diagnosis of handwriting difficulties frequently manifested in developmental dysgraphia.

INDEX TERMS Advanced parametrization, computerized analysis, graphomotor difficulties, machine learning, online handwriting.

I. INTRODUCTION

At present, every school-aged child is expected to master legible, well-coordinated and fast-paced handwriting, which is a complex perceptual-motor skill learned by instruction that quantifies a child's timely maturation and integration of psycho-motor, linguistic and mental abilities, and readiness for education [1]. It is known that it takes approximately 10 years to develop handwriting skills [2] on both quantitative (speed) and qualitative (legibility) level [3], [4]. However, before a child starts to write, she/he first needs to learn how to draw [5]. In general, until the age of 6, a child starts to develop a combination of motor and non-motor skills such as motor planning and execution, visual–perceptual abilities, orthographic coding, kinesthetic feedback, and visual–motor coordination, which eventually become automated at the age of 8–9 [6], [7]. These skills are referred to as graphomotor skills (GS) [8], [9], and form the foundation of drawing and consequently, handwriting abilities [2] that accompany every person throughout the life-time.

Even though modern technologies brought new ways of communication, self-expression, and education, handwriting is still an important part of a child's life [9]. In general, it has been estimated that children spend 31–60 % of their school-time performing handwriting [10]. Given that children at school need to write under time constraints, the acquisition of GS is crucial for a child's ability to write legibly, as well as quickly and efficiently. Basically,

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

the development of GS affects a child's academic success and professional career [11]. It has also been shown that approximately 10–30 % of children experience graphomotor difficulties (GD) [8], [9] such as motor-memory dysfunction (problems combining memory input with motor output), graphomotor production deficits (poor muscle coordination, unusual pen-grip and less precise graphomotor movements), motor feedback difficulties (over-activation of certain muscles and joints during handwriting as well as problems tracking the location of the pen's tip), etc. Such an impairment of the neuro-muscular system can have serious pedagogical and psychological consequences, and can greatly affect a child's every-day life [12] starting with slow and less-legible handwriting, lack of motivation to write, lower self-esteem combined with poor emotional well-being, bad attitude and behaviour, communication and social interaction problems, and in some cases going as far as being diagnosed with a serious neurodevelopmental disorder such as developmental dysgraphia (DD) [9], [13]–[15]. To provide children with both preventive as well as corrective therapeutic care, GD should be identified and treated as soon as possible [16], [17].

To identify and evaluate GD and handwriting difficulties (HD) in general, occupational therapists and/or special educational counsellors use specialized questionnaires or tests that aim at quantification of the quality of the handwritten product in multiple domains using its visual inspection. Some of the most commonly used questionnaires (rating scales) are the following: Concise Assessment Scale for Children's Handwriting (Brave Handwriting Kinder) (BHK) [18], Handwriting Proficiency Screening Questionnaire (HPSQ) [19] or Handwriting Proficiency Screening Questionnaire for Children (HPSQ–C) [20]. Even though these scales are a well-established way of identification and rating of GD and HD in school-aged children, its administration and coding are time-consuming, which limits the use of this type of evaluation on a regular day-to-day basis. Moreover, it is naturally limited by the perceptual capabilities, subjective judgement and experience of an examiner [21]. Finally, it is also a subject to inter-rater variability [22]. Due to the complexity and limitations associated with GD/HD identification, many children, especially those attending lower grades of a primary school, may remain undiagnosed or may be diagnosed later than appropriate.

To overcome the limitations of the perceptual analysis and search for a more robust view of various hidden complexities of the handwriting process, new methods based on digitization and signal processing techniques have been developed [23]–[28]. More specifically, instead of a conventional data acquisition using a pen and paper, digitizing tablets (digitizers) have been used to record a variety of signals describing the evolution of handwriting in time. Such a collection of data about handwriting (i. e. that one associated with timestamps) is referred to as online handwriting [29]. Using advanced digital signal processing algorithms a variety of handwriting parameters (commonly referred to as handwriting features) quantifying kinematic (velocity, acceleration, jerk) as well as dynamic (pen pressure, tilt and azimuth) components contributing to the execution of the handwriting process have been designed [6], [30]–[32]. Such characteristics are very hard to be perceived and precisely quantified by a human observer and are almost impossible to be extracted using only the final handwritten product.

In recent years, several studies focusing on computerized analysis, identification and assessment of HD, mostly associated with writing in children with developmental dysgraphia, have been conducted. In 2017, Pagliarini *et al.* [27] reported that the governing principles of rhythmic organization, namely homothety and isochrony, describe the handwriting process in school-aged children from the time where the very first handwritten products are made, i. e. before the handwriting is performed automatically. Moreover, they pointed out the potential of quantitative analysis to indicate the development of HD at a very early age. In the same year, Mekyska *et al.* [32] performed a study in a cohort of 27 school-aged children in which they introduced a new intra-writer normalisation method aiming at improving the discrimination capabilities of a large variety of conventional and non-conventional handwriting features. They also built a random forest classifier identifying the presence of DD with 96 % sensitivity and specificity. Next, Rosenblum and Dror [26] employed a study focusing on automatic identification and characterization of DD in a cohort of 99 third-grade children. Using various kinematic and dynamic features, they trained a linear support vector machines classifier achieving 90 % sensitivity and specificity. In 2018, Asselborn *et al.* [28] developed a diagnostic tool for DD evaluated on a cohort of 298 children (56 children with DD) performing the BHK test on a digitizing tablet covered with a sheet of paper. To identify the presence of DD, they computed 53 handwriting features and built a random forest classifier with 96 % sensitivity and 99 % specificity. In 2019, Mekyska *et al.* [33] employed a study that is the closest one to a study proposed in this work. They aimed at exploring the impact of specific elementary graphomotor tasks on the accuracy of computerized diagnosis of GD. For this purpose, they analysed 7 basic graphomotor elements performed by a cohort of 76 school-aged children. Using only conventional handwriting features, they trained an XGBoost [34] classifier and achieved 50 % sensitivity and 90 % specificity. In the same year, Zvoncak *et al.* [35] used features based on fractional order derivatives to enrich a set of conventional features and analysed their correlation with HPSQ–C in 55 children (19 third-grade children, and 36 fourth-grade children) performing an alphabet writing task. With this setup, they reported that features based on fractional order derivatives improved quantification and robustness of the description of in-air movements. And finally, in 2020 Asselborn *et al.* [36] proposed a data driven strategy for estimating handwriting quality in a battery of 448 school-aged children (390 typically developing children and 58 children with HD). They utilized principal component analysis to reduce 53 handwriting features also used in [28] to three dimensions that are independent of the BHK scores.

Next, they used the reduced feature space to cluster children into two groups (typical handwriting, HD), and evaluated how far a child's score is from the average score of children of the same age and gender. With this approach, they reported four specific handwriting scores for kinematics, pressure, pen tilt and static features to describe the handwriting profile of a child in a finer way that enables measuring the quality of handwriting in multiple domains.

Although there is a body of research dealing with computerized quantitative analysis of HD in school-aged children, several key points have not been fully investigated yet. First of all, most of the studies aimed at identifying HD and/or DD. Studies focusing on quantification and identification of GD are very sparse. This is an important topic as HD can have many forms and can vary even among typically developing children. As mentioned in one of the most recent publications dealing with computerized analysis of handwriting in school-aged children proposed by Asselborn *et al.* [36], dysgraphia is an umbrella term that describes a variety of handwriting difficulties. Therefore, GD play a crucial role in determining the handwriting profile of a child, and should be investigated as well. Moreover, most of the studies focused on writing signals such as writing words, sentences, etc., only. Finally, conventional handwriting features have been used to describe HD almost exclusively. To the best of our knowledge, a comprehensive study aiming at quantifying GD manifested during performing a battery of simple but important graphomotor elements (loops, spirals, etc.) using not only conventional but also more advanced graphomotor features is missing. For this purpose, in this study, we propose the use of seven graphomotor tasks and three novel types of handwriting features based on: a) modulation spectra; b) fractional order derivatives; and c) tunable Q-factor wavelet transform. We hypothesize that these features can bring more information about specific GD accompanying the handwriting process of children with GD in its very basis. In addition, we also hypothesize that a combination of conventional and more advanced parametrization of online handwriting can improve identification of GD and contribute to a development of a decision support system that can be used for diagnosis of HD and eventually DD.

II. MATERIALS AND METHODS

The methodology can be briefly summarized as follows: a) dataset description (cohort, acquisition protocol, data acquisition, etc.), b) presentation of the feature extraction methods (conventional, newly-proposed features), and c) statistical analysis and machine learning (normality testing and feature pre-processing, feature selection, correlation analysis, hypothesis testing, and binary classification). Finally, an overview of the methodology can also be seen in Fig. 1.

A. DATASET

Altogether, we enrolled 53 Czech-speaking children (22 girls and 31 boys) attending 3rd and 4th grade at several primary schools in the Czech Republic: 26 healthy children (HC) (2 3rd-grade girls, 12 4th-grade girls, and 12 4th-grade boys) and 27 children with GD (1 3rd-grade girl, 5 3rd-grade boys, 7 4th-grade girls, and 14 4th-grade boys). Description of the dataset can be seen in Table 1. During the data acquisition, all of the children were asked to perform a specifically designed drawing protocol consisting of 7 elementary graphomotor tasks (TSK) (for more information, see Fig. 2): TSK1 – Archimedean spiral (approximately 15 cm in height); TSK2 – half-sized version of TSK1; TSK3 – connected loops; TSK4 – flipped version of TSK3; TSK5 – sawtooth; TSK6 – rainbow; TSK7 – combination of TSK3 and TSK4. Each of the tasks was first shown to a child and then she/he replicated it on a blank sheet of paper with a comfortable speed. The protocol was designed in cooperation with psychologists and special educational counsellors so that it reflects all coordinated elementary movements that are

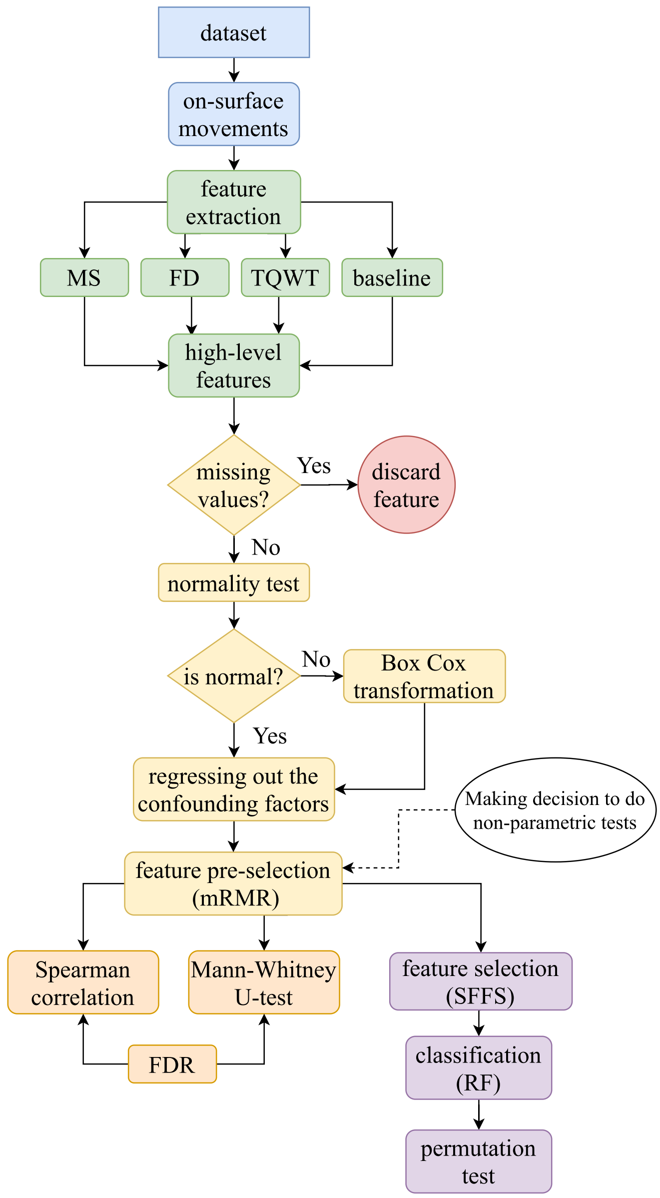


FIGURE 1. An overview of the methodology applied in the study.

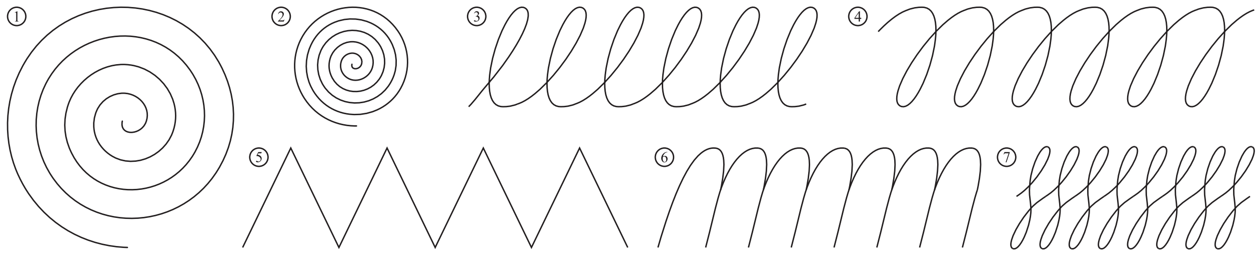


FIGURE 2. Drawing acquisition protocol with the selected graphomotor tasks.

TABLE 1. Description of the dataset.

	μ (σ)	min.	Q1	Q2	Q3	max.
all children (53 subjects)						
age [y]	10.92 (1.65)	8.46	10.73	11.33	11.67	12.32
class	3.84 (0.36)	3.00	4.00	4.00	4.00	4.00
HPSQ–C	13.66 (6.31)	4.00	9.00	12.00	19.00	27.00
HC (26 subjects)						
age [y]	11.23 (0.62)	9.77	10.99	11.43	11.66	12.32
class	3.92 (0.27)	3.00	4.00	4.00	4.00	4.00
HPSQ–C	12.50 (6.21)	4.00	9.00	10.50	14.00	27.00
GD (27 subjects)						
age [y]	10.57 (2.19)	8.46	10.52	10.95	11.66	12.27
class	3.77 (0.42)	3.00	4.00	4.00	4.00	4.00
HPSQ–C	14.44 (6.30)	6.00	10.00	13.00	19.50	25.00

[1] μ – mean estimate; σ – standard deviation estimate; HPSQ–C – Handwriting Proficiency Screening Questionnaire for Children [20] (only total score showing an overall degree of GD is shown); Qx – x-th quartile; y – years.

needed to successfully write cursive letters (i. e. cursive letters are constructed of these basic graphomotor elements, therefore, mastering these elements is a prerequisite for mastering legible handwriting). Examples of the final handwritten product for all graphomotor tasks performed by healthy children and children with GD can be seen in Fig. 3.

The protocol was printed on an A4 paper that was laid down and fixed to a digitising tablet. To acquire the handwriting data, we used Wacom Intuos Pro L (PHT-80) with the sampling frequency of 150 Hz, and the Wacom Inking pen. This set-up enabled us to take advantage of two facts: a) it provided the children as well as an examiner with immediate visual feedback and made it possible to simulate the feeling of using a conventional inking pen; and b) it allowed for recording of a variety of signals describing the drawing process: x and y position ($x[n]$ and $y[n]$); timestamp ($t[n]$); a binary variable ($b[n]$; 0 – in-air movement, i. e. movement of pen tip up to 1.5 cm above the tablet's surface, and 1 – on-surface movement, i. e. movement of pen tip on the paper), pressure exert on the tablet's surface during drawing/writing ($p[n]$); pen tilt ($a[n]$); and azimuth ($az[n]$). For more information, we refer to our previous works [32], [37].

Moreover, to assess legibility and performance time during handwriting as well as physical and emotional well-being, the children were asked to evaluate themselves using HPSQ–C (rating scale) [20], which is composed of 10 questions scored on a 5-point Likert scale (0 – never, i. e. no GD, 4 – always, i. e severe GD; total score, i. e. sum over all questions: 0 – min. value, 40 – max. value; legibility – items 1, 2, and 10, performance time – items 3, 4 and 9, and physical and emotional well-being – items 5–8). Using HPSQ–C brings two important advantages: a) the scale is language independent and therefore well-comparable across studies employed on cohorts coming from different language groups; b) it has already been validated in a couple of previous studies such as [8], [32], [38], [39]. The overall HPSQ–C scores, as well as the final handwritten product, were both examined by experienced psychologists and special educational counsellors. The decision about a child's assignment into HC or GD group was performed on a PC after the examination of a child's handwritten product, where an expert (remedial teacher) had no information about her/his sociodemographic information (e. g. sex, class, HPSQ–C, etc.). The description of HC/GD groups mentioned at the beginning of Section II presents the numbers after the final examination and assignment.

Parents of all children participating in this study signed an informed consent form approved by the Ethical Committee of the Masaryk University. Throughout the entire duration of this study, we strictly followed the Ethical Principles of Psychologists and Code of Conduct released by the American Psychological Association (https://www.apa.org/ethics/code/).

B. FEATURE EXTRACTION

To quantify GD, we extracted the following conventionally used graphomotor features (CONV) [25], [30], [40]: a) spatial features – width (WIDTH), height (HEIGHT), and length (LEN) of the signals (also referred to as writing). Even though the in-air movements can be used to capture a certain aspect of GD [25], [40], all graphomotor tasks proposed in this work should be performed using a single stroke. Since the number of multi-stroke signals analyzed in this study was only marginal, we did not distinguish between signals and strokes and used the stroke notation, i. e. stroke width (SWIDTH), height (SHEIGHT), and length (SLEN), as it is used in general; b) kinematic features (horizontal and vertical projection) – velocity (VEL), acceleration (ACC), and jerk (JERK); and c) dynamic features – pressure (PRESS), tilt (TILT), and azimuth (AZIM). These features were used as a baseline feature set. To build on top of these conventional features and to enhance their capability of describing GD in a more robust and complex way, we present three new feature-types aiming at improving the quantification and description of GD in school-aged children, namely:

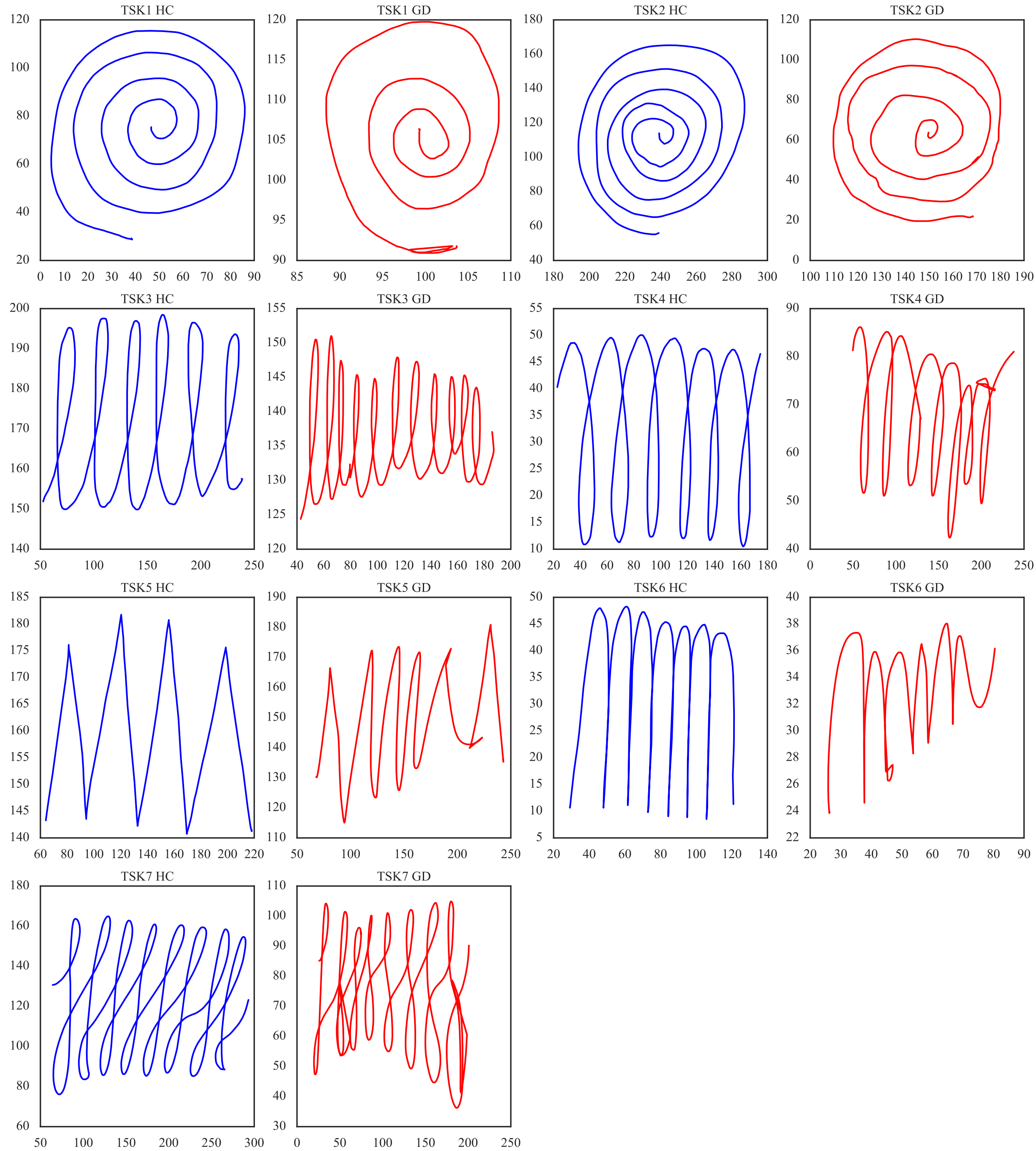


FIGURE 3. **Example of the final handwritten product for all graphomotor tasks performed by randomly selected healthy children (blue) and children with GD (red) (units are in millimeters).**

a) features based on modulation spectra (MS); b) features based on fractional order derivatives (FD); and c) features based on tunable Q-factor wavelet transform (TQWT). All vector-valued features were transformed to scalar values using mean and coefficient of variation (cv) estimates (some of the novel features used additional statistical functions that are described along with the features themselves).

An important fact to point out is that these features were designed not only to improve the robustness of the conventional features but also to maintain as much interpretability as possible. This is crucial especially for their real use in clinical practice because the complexity and great discrimination power without understanding the meaning of the features are not likely to bring trust and convenience. If psychologists and special educational counsellors are able to link the features with the specific physiological phenomena, the computerized quantitative analysis of GD can be finally deployed.

To present the features in a compact and easy to read way, we used the following naming convention: *TSK INF: DIR-FN (HL)*, where *TSK* denotes the specific graphomotor task, *INF* represent information about the movement (ON – on-surface, AIR – in-air), PRESS – pressure, TILT – tilt, and AZIM – azimuth), *DIR* stands for direction (H – horizontal and V – vertical), *FN* shows the feature name, and *HL* holds an applied statistic (if any). Moreover, each specific novel feature-type also sets *FN* accordingly (described in the section devoted to the proposed features). As all features presented in this work are computed from on-surface movements, the on-surface/in-air information is considered redundant and is not shown in the feature names.

1) MODULATION SPECTRA FEATURES

The first type of the novel features proposed in this work is based on modulation spectra as a non-parametric method for representing modulations in an analyzed biomedical signal. MS has already been used for parametrization of dysarthric speech in patients with Parkinson's disease (PD) [41]. These features however aimed at describing instability of vocal folds vibrations. The features proposed in this work aim at quantifying the ratio between the low and high-frequency movements present in a given handwriting signal of children attending a primary school.

To compute the modulation spectra features, Short-Time Fourier Transform (STFT) of the input handwriting signal $s[n]$ of length N is computed as

$$S[k,m] = \sum_{n=0}^{N-1} s[n]w[n-mL]\mathrm{e}^{-\mathrm{j}k\frac{2\pi}{N}n}, \quad (1)$$
$$k = 0, 1, \ldots, N-1,$$
$$m = 0, 1, \ldots, M-1,$$

where M denotes the number of segments obtained using a segmentation window $w[n]$ composed of L samples. In the frame of this work, we used Hamming segmentation windows with $L = 75$ samples ($f_s = 150$ Hz, windows of 0.5 s with the overlap of 50 %).

Next, power spectrum $|S[k,m]|^2$ of each segment is computed and filtered by a filer-bank P consisted of P_n filters. For this purpose, we used a filter bank of 50 linearly distributed triangular filters. After the filtration, the matrix $X[p,m]$ contains P_n sub-bands $p = 1, 2, \ldots, P_n$. Subsequently, each sub-band is normalized [42] as follows

$$\hat{X}[p,m] = \ln(X[p,m]) - \overline{\ln(X[p,m])}, \quad (2)$$

where $\overline{*}$ refers to the averaging operator applied over m.

To obtain a modulation spectra matrix, Discrete Fourier Transform (DFT) is applied on $\hat{X}[p,m]$.

$$\Psi[p,l] = \sum_{m=0}^{M-1} \hat{X}[p,m]\mathrm{e}^{-\mathrm{j}l\frac{2\pi}{M}m}, \quad (3)$$
$$l = 0, 1, \ldots, M-1,$$

where p and l denote the handwriting and modulation frequency, respectively. Finally, $\Psi[p,l]$ is normalized by the mean of each sub-band.

After obtaining the modulation spectra matrix, a vector of handwriting cut-off frequencies $f_c = 1, 2, \ldots, C$ [Hz] is defined. The values of f_c are subsequently converted to the filter indices c using their center frequencies. In this work, we used $f_c \in F_c$, where $F_c = 1, 2, \ldots, 10, 15, 20, 25$ Hz. Next, for each value of f_c, low (E_l) and high frequency (E_h) summation components of $\Psi[p,l]$ are computed as

$$E_{l(f_c)}[l] = \sum_{p=0}^{c} \Psi[p,l],$$
$$E_{h(f_c)}[l] = \sum_{p=c}^{P_n} \Psi[p,l], \quad (4)$$
$$l = 0, 1, \ldots, M-1,$$
$$fc = F_c. \quad (5)$$

Finally, $E_{l(f_c)}$ and $E_{h(f_c)}$ are used to compute the final energy ratio R_{f_c} between the low and high frequency movements in the analyzed handwriting signal. It is defined as

$$R_{f_c} = \frac{\sum_{l=0}^{M-1} E_l[l]^2}{\sum_{l=0}^{M-1} E_h[l]^2}. \quad (6)$$

We used the following naming convention for the MS features: $F\mathrm{R}f_c$, where F represents the name of the handwriting feature, R stands for ratio, and f_c holds the value of the specific handwriting cut-off frequency used to compute the energy ratio.

2) FRACTIONAL ORDER DERIVATIVE FEATURES

The second type of the novel features is based on the theory of fractional order derivatives. Handwriting features based on FD have already been explored in our previous studies focusing on the quantitative analysis of parkinsonian dysgraphia [43]–[46], where they brought a promising improvement in the power of the FD-based features to objectively discriminate between healthy and dysgraphic handwriting using machine learning. In this work, we aim at exploring the possibilities of utilizing FD to describe GD in school-aged children.

The most common approaches to compute FD are Riemann–Liouville, Caputo, and Grünwald–Letnikov formulations [47]–[49]. Parameterization of online handwriting using FD is performed by substituting the conventional differential derivative during the calculation of the basic kinematic features (velocity, acceleration, and jerk). The advantage of FDs lies in their wide range of settings (order α, kernel function, etc.). In this study, we followed the Grünwald–Letnikov approximation [48], [50] and used the implementation of FD by Jonathan Hadida. To decrease the computational requirements, we used a segmentation-based computation.

A direct definition of the $D^\alpha y(t)$ is based on the finite differences of an equidistant grid in $[0, \tau]$, assuming that the function $y(\tau)$ satisfies certain smoothness conditions in every finite interval $(0, t)$, $t \leq T$. Choosing the grid [48]

$$0 = \tau_0 < \tau_1 < \ldots < \tau_{n+1} = t = (n+1)h \tag{7}$$

with

$$\tau_{k+1} - \tau_k = h \tag{8}$$

and using the notation of finite differences

$$\frac{1}{h^\alpha} \Delta_h^\alpha y(t) = \frac{1}{h^\alpha} \left(y(\tau_{n+1}) - \sum_{v=1}^{n+1} c_v^\alpha y(\tau_{n+1-v}) \right), \tag{9}$$

where

$$c_v^\alpha = (-1)^{v-1} \binom{\alpha}{v}. \tag{10}$$

The Grünwald–Letnikov definition from 1867 is defined as

$$D^\alpha y(t) = \lim_{h \to 0} \frac{1}{h^\alpha} \Delta_h^\alpha y(t), \tag{11}$$

where $D^\alpha y(t)$ denotes a derivative with order α of a function $y(t)$, and h represents a sampling lattice. Following our previous works focused on optimization of α [43], [46], we used the ranges: from 0.1 to 0.4, and from 0.65 to 0.9, with iteration step of 0.05.

The naming convention for FD-based features can be described as: $F\alpha$, where F represents the name of the handwriting feature and α stands for the order of FD.

3) TUNABLE Q-FACTOR WAVELET TRANSFORM FEATURES

The last type of the novel features is based on tunable Q-factor wavelet transform [51]–[53]. Recently, we have shown that HD manifest themselves in higher energies of the residual component of the decomposed signal computed by TQWT [39]. Following our previous research, we aim at investigating the potential of TQWT to describe limited motor skills, poor dexterity and muscle tone or unspecified motor clumsiness in school-aged children suffering from GD.

TQWT is a non-linear discrete-time resonance-based signal decomposition technique that separates an input signal into high-resonance (sustained rhythmic oscillations), low-resonance (non-rhythmic and transient behaviour) and residual components (stochastic nature of the decomposed signal) [51]. It is parameterized by a tunable Q-factor and an oversampling rate (redundancy). In this study, we utilized the implementation of TQWT based on morphological component analysis (MCA) [54] and split augmented Lagrangian shrinkage algorithm (SALSA) [55] described in [52].

To decompose an input signal into high and low resonance components, an iterative J-level decomposition of its low-pass channel by a two-channel filter-bank composed of low- and high-pass filters is used [52]. The frequency responses of the low-pass $H_l(\omega)$ and the high-pass $H_h(\omega)$ filters are defined as

$$H_l(\omega) = \theta \frac{\omega + (\beta - 1)\pi}{\alpha + \beta - 1}, \tag{12}$$

$$H_h(\omega) = \theta \frac{\alpha\pi - \omega}{\alpha + \beta - 1}, \tag{13}$$

for $(1 - \beta)\pi < \omega < \alpha\pi$, where α and β are the low- and high-pass scaling parameters, and θ is the Daubechies frequency response [52] given as

$$\theta(\omega) = 0.5(1 + \cos\omega)\sqrt{2 - \cos\omega}, \tag{14}$$

for $|\ \omega\ | \leq \alpha$. More details can be found in [51], [52].

To describe the proposed features, we define the clean graphomotor signal $x_c[n]$ as

$$x_c[n] = x[n] - x_r[n], \tag{15}$$

where $x[n]$ is a handwriting signal, and $x_r[n]$ is a residual signal given as $x_r[n] = x[n] - x_h[n] - x_l[n]$ ($x_h[n]$ and $x_l[n]$ are the high- and low-resonance components).

With $x_c[n]$ and $x_r[n]$ being defined, the signal-to-noise ratio is computed as

$$SNR = 10\log_{10}\left(\frac{E(x_c[n])}{E(x_r[n])}\right)[\mathrm{dB}], \tag{16}$$

where E denotes energy computed as

$$E(s[n]) = \sum_{n=0}^{N-1} s[n]^2, \tag{17}$$

for s being a substitution for $x_c[n]$ and $x_r[n]$.

Next, absolute value of the first order derivative of $E(x_r[n])$ is computed as $E_d(x_r[n]) = |E'(x_r[n])|$. To quantify the variability of $E_d(x_r[n])$, a slope of its cumulative sum is computed as

$$E_\Delta = \Delta C(E_d), \tag{18}$$

where $C(E_d)[n]$ for $n = 0, 1, \ldots, N-1$ refers to the cumulative sum applied on E_d, and Δ denotes the slope of a function. Finally, to compute the number of significant changes in $E_d(x_r[n])$, the number of its peaks E_p above the median value is computed.

Naming convention for TQWT-based features can be described as: FN, where F represents the name of the handwriting feature and N stands for the specific TQWT feature: signal-to-noise ratio (SNR), E_Δ as RES (csum), and E_p as RES (npeaks).

C. STATISTICAL ANALYSIS

At first, the features with any missing values were discarded from the analysis. Consequently, normality of the features was tested using Shapiro-Wilk test [56]. All non-normally distributed features were adjusted using Box-Cox [57] transformation. After the normalization, the features were re-inspected. As not all of the features were fully-normalized, an entire feature set was considered as being non-normally distributed. As a result, only non-parametric statistical methods were employed during the subsequent statistical analysis. Next, to control for the effect of confounding factors (also known as covariates), we computed the Spearman's correlation between the values of the features and the following characteristics: age, gender, grade (these characteristics were chosen after the consultation with psychologists and special educational counsellors). With this approach, age and grade were identified as having a statistically significant effect on the feature values. The effect of children's gender on the features was only marginal. Therefore, during the statistical analysis, we controlled for the effect of age and grade only. After the feature-transformation, we reduced the size of the feature set using a feature pre-selection process independently for each analyzed feature-type. More specifically, we used a filter method named minimum Redundancy Maximum Relevance (mRMR) to select a relevant sub-set of the features with minimum redundancy and cross-correlation among the features. After the feature pre-selection, we obtained 15 features per feature-type. Having the same number of the features for each features-type is important especially for the classification analysis, where each classifier is built starting with the same feature-space complexity.

Next, to compare the distribution of the graphomotor features for healthy children and children with GD, we used Mann-Whitney U-test with the significance level of 0.05. Moreover, to assess the strength of a relationship between the features and the children's clinical status (HC/GD), we computed Spearman's correlation coefficient with the significance level of 0.05. To control for the issue of multiple comparisons, p-values were adjusted using the False Discovery Rate (FDR) method.

Subsequently, to identify the presence of GD, we built binary classification models using an ensemble learning algorithm named Random Forests (RF) [58]. This particular algorithm was chosen due to its robustness to outliers, ability to find complex interactions among features as well as the possibility of ranking their importance. Using a randomized search strategy, we selected the following model settings: number of estimators (500), maximum tree depth (10), minimum number of samples required for splitting (2), minimum number of samples at a leaf node (1). Additionally, to train the models using only a parsimonious, information-rich subset of the features, to considerably decrease the risk of overfitting, and to reduce the computational performance requirements, we employed a feature selection process using a wrapper method named Sequential Floating Forward Selection (SFFS). As shown previously, reduction of the feature space complexity can significantly improve the model's prediction power [59].

i To quantify the classification performance of the trained models as well as to control the addition and removal of the features during the feature selection, we used Matthew's correlation coefficient (MCC) [60]. This particular metric was chosen due to its ability to summarize the confusion matrix with the focus on obtaining a balance between the model's sensitivity and specificity [61]. The training and testing features were standardized before classification on a per-feature basis to have 0 mean and a standard deviation of 1. The trained models were evaluated conducting a stratified 5-fold cross-validation (we chose the 5-fold cross-validation scheme as a reasonable compromise between the number of samples in the training and validation folds) with 20 repetitions, and the classification test performance was determined using the following classification metrics: MCC, accuracy (ACC), sensitivity (SEN), and specificity (SPE).

Finally, to evaluate the statistical significance of the prediction performance obtained by the trained classification models, a non-parametric statistical method named permutation test was employed (exact p-values were computed to mitigate the type I error rate and the multiple testing issues) [62], [63]. In this work, we used 1 000 permutations and the significance level of 0.01 (to estimate the performance of the models on the permuted data, we used the same classification setup as in the training phase [64]).

III. RESULTS

At first, the cross-correlation matrices (using Pearson's correlation) of the 15 features per feature-type selected using feature pre-selection performed by the mRMR algorithm are visualized in Fig. 4. As can be seen, there are some features that can be considered redundant, i. e. having a strong correlation with one/more features, however, as we did not want to reduce the feature-space complexity too much (the redundancy is not the same in every feature-type, so by reducing the feature space complexity any further, some relevant features could be removed as well. This would most likely result in having sub-optimal feature space for some of the feature-types.), we decided to use all of the 15 features, and analyze them accordingly (having the possibility of cross-correlated features appearing in the results of the statistical analysis together in mind).

Results of the statistical analysis can be seen in Table 2. The table shows the top 5 features for each of the feature-types according to the p-value computed by the Mann-Whitney U-test (if some of the cross-correlated features appeared together, we selected only one of them and replace the other with the feature/s bellow the top 5). Regarding the p-values of the Mann-Whitney U-test, the following number of features can be considered as coming from a distribution that is significantly different for the two subject groups (threshold of 0.05): a) CONV features – 5/5 (prior adjustment), 1/5 (after adjustment); b) MS features – 5/5 (prior adjustment), 4/5 (after adjustment); c) FD features – 5/5 (prior adjustment), 1/5 (after adjustment); and d) TQWT features – 3/5 (prior adjustment), 1/5 (after adjustment). With respect to the Spearman's correlation, the following features were found to have the strongest correlation with the presence of GD (where $**$ denotes p-value < 0.01, and $*$ denotes p-value < 0.05): a) CONV features – TSK1 TILT (mean) $\rho = -0.42^{**}$; b) MS features – TSK2 V-JERKR25 $\rho = 0.41^{**}$; c) FD features – TSK1 TILTVEL0.3 (mean) $\rho = -0.41^{**}$; and d) TQWT features – TSK6 V-VELSNR $\rho = -0.39^{**}$. All of these features were found to have a statistically significant relationship with the presence of GD (both prior and after p-value adjustment). For better visualization, violin plots showing the distribution estimates of the best-discriminating

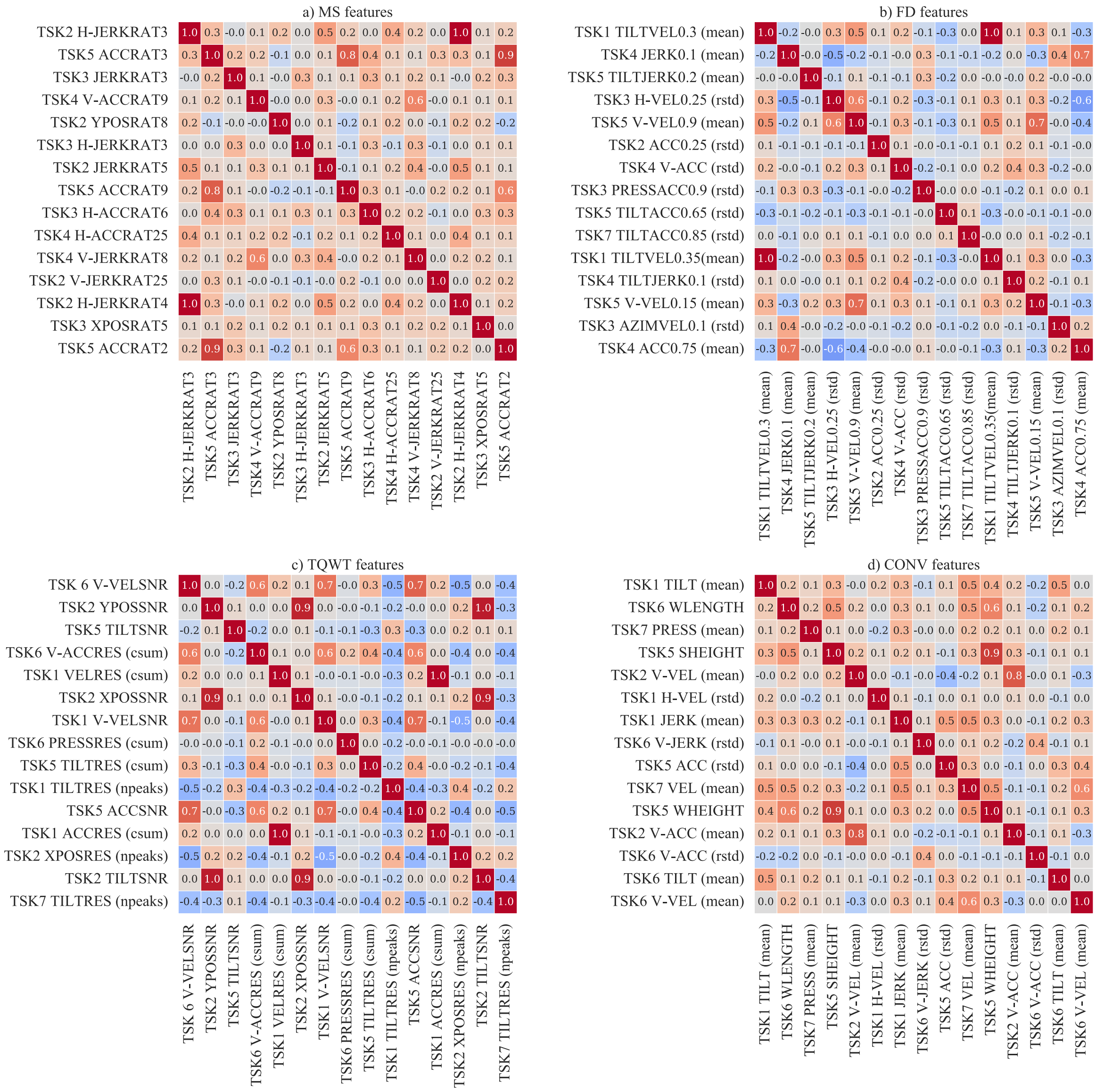


FIGURE 4. Cross-correlation matrices of the feature sets (Pearson's correlation coefficient (r); 15 features per feature-type) after the pre-selection. Color notation: linear scale in the range of < −1, 1 >, where the maximum positive correlation is denoted by the red color, and the maximum negative correlation is denoted by the blue color. More information about the features can be seen in Section II-B.

features of every feature-type for both healthy children and children with GD are presented in Fig. 5.

And finally, results of the classification analysis can be seen in Table 3. Regarding the individual feature-types, the following results were achieved (where $**$ denotes p-value $<$ 0.01, and $*$ denotes p-value $<$ 0.05): a) CONV features (7 features selected) – ACC $=$ 0.74**; b) MS features (8 features selected) – ACC $=$ 0.73**; c) FD features (3 features selected) – ACC $=$ 0.76**; and d) TQWT features (2 features selected) – ACC $=$ 0.71**. Features used to train these classification models for each feature-type are summarized in Table 4. With respect to an overall feature set (all 60 features combined), the classification performance was: ACC $=$ 0.84** using 10 features. All classification results were evaluated by the permutation test as being statistically significant.

IV. DISCUSSION

In the search for novel and more robust graphomotor features that can be used to improve the quantification and identification of GD in school-aged children, we introduced three non-conventional advanced types of features, specifically, features based on modulation spectra, features based on fractional order derivatives, and features based on tunable Q-factor wavelet transform. As each feature-type produced a different number of features, we employed feature pre-selection to reduce the feature-space complexity and minimize the effect of the curse of dimensionality occurring when the number of analyzed features greatly outnumbers the number of observations present in the dataset, as well as to unify the number of features among the feature sub-sets. With this approach, we reduced each feature-type to 15 features with minimal cross-correlation. An important

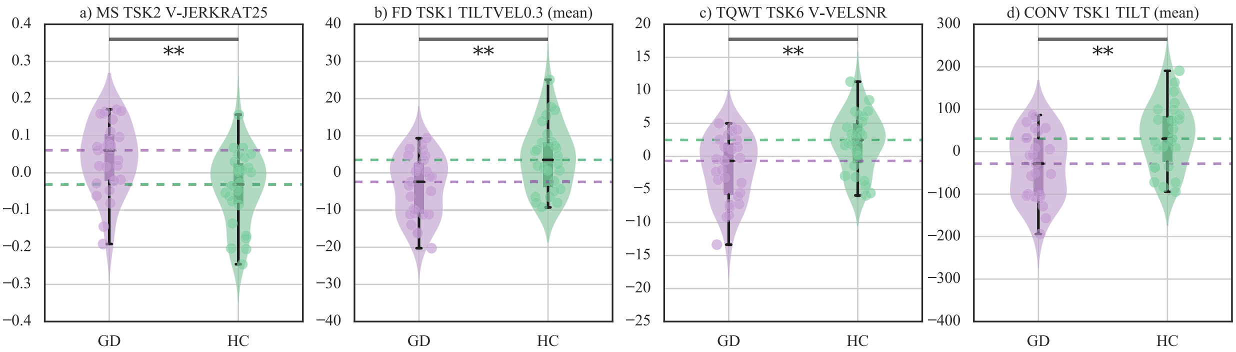


FIGURE 5. Violin plots of graphomotor features in both GD and HC groups (after removing the covariates). Figure notation: background of the box plots represents vertically mirrored kernel density estimations; horizontal dashed lines represent medians; and a star(s) between two violins mean(s) the p-value of Mann-Whitney U-test (** denotes p-value < 0.01, and * denotes p-value < 0.05).

TABLE 2. Results of the statistical analysis.

feat.	TSK	ρ	p(ρ)	p(ρ)*	p(U)	p(U)*
		CONV features				
TILT (mean)	TSK1	-0.42	0.001	0.027	0.001	0.019
TILT (mean)	TSK6	-0.32	0.017	0.129	0.009	0.072
SHEIGHT (mean)	TSK5	-0.31	0.028	0.142	0.015	0.076
WLENGTH	TSK6	-0.25	0.074	0.190	0.038	0.096
WHEIGHT	TSK5	-0.25	0.074	0.190	0.038	0.096
		MS features				
V-JERKR25	TSK2	0.41	0.002	0.024	0.001	0.016
XPOSR5	TSK3	0.40	0.003	0.024	0.002	0.016
ACCR3	TSK5	0.36	0.009	0.033	0.005	0.020
JERKR3	TSK3	0.36	0.009	0.033	0.005	0.020
H-ACCR25	TSK4	0.27	0.058	0.146	0.030	0.075
		FD features				
TILTVEL0.3 (mean)	TSK1	-0.41	0.002	0.031	0.001	0.020
H-VEL0.25 (cv)	TSK3	-0.32	0.021	0.094	0.011	0.050
V-VEL0.9 (mean)	TSK5	-0.31	0.028	0.094	0.015	0.050
ACC0.75 (mean)	TSK4	0.30	0.031	0.094	0.016	0.050
ACC0.25 (cv)	TSK2	-0.25	0.074	0.152	0.038	0.077
		TQWT features				
V-VELSNR	TSK6	-0.39	0.004	0.070	0.003	0.044
V-ACCRES (csum)	TSK6	-0.26	0.061	0.345	0.031	0.177
ACCSNR	TSK5	-0.26	0.069	0.345	0.035	0.177
TILTSNR	TSK2	-0.23	0.110	0.409	0.055	0.206
V-VELSNR	TSK1	-0.21	0.136	0.409	0.068	0.206

[1] feat – feature; TSK – graphomotor task; ρ – Spearman's correlation coefficient; p(ρ) – p-value of ρ; p(ρ)* – adjusted p(ρ); p(U) – p-value of Mann-Whitney U-test; p(U)* – adjusted p(U); for the feature naming convention, see Section II-B.

TABLE 3. Results of the classification analysis.

type	MCC	ACC	SEN	SPE	N	p
CONV	0.50 (0.26)	0.74 (0.12)	0.80 (0.19)	0.71 (0.21)	7	**
MS	0.48 (0.27)	0.73 (0.14)	0.75 (0.19)	0.73 (0.21)	8	**
FD	0.51 (0.30)	0.76 (0.13)	0.75 (0.20)	0.77 (0.20)	3	**
TQWT	0.42 (0.29)	0.71 (0.14)	0.74 (0.19)	0.68 (0.23)	2	**
ALL	0.65 (0.25)	0.84 (0.13)	0.83 (0.17)	0.81 (0.18)	10	**

[1] the results are shown as mean (standard deviation); type – specific type of graphomotor feature; MCC – Matthew's correlation coefficient; ACC – accuracy; SEN – sensitivity; SPE – specificity; N – Number of selected features; p – p-values computed by the permutation test (1 000 permutations); ALL (combination of all feature-types, i. e. 60 features); for the feature naming convention, see Section II-B.

TABLE 4. Features selected for the trained classification models.

CONV	MS	FD
TS6 V-ACC (cv)	TS5 ACCR2	TS3 H-VEL0.25 (cv)
TS1 H-VEL (cv)	TS3 H-ACCR6	TS7 TILTACC0.85 (cv)
TS7 VEL (mean)	TS2 YPOSR8	TS5 V-VEL0.9 (mean)
TS2 V-ACC (mean)	TS4 V-JERKR8	
TS1 TILT (mean)	TS2 JERKR5	
TS5 WHEIGHT	TS3 JERKR3	
TS2 JERK (mean)	TS2 V-JERKR25	
	TS5 ACCR3	
TQWT	**ALL**	
TS2 YPOSSNR	TSK1 H-VEL (cv)	
TS1 VELRES (csum)	TSK1 TILTVEL0.35 (mean)	
	TSK2 JERKR5	
	TSK3 JERKR3	
	TSK2 V-VEL (mean)	
	TSK6 V-ACC (cv)	
	TSK1 V-VELSNR	
	TSK3 PRESSACC0.1 (cv)	
	TSK7 TILTACC0.85 (cv)	
	TSK5 TILTRES (csum)	

[1] TSK – graphomotor task. For the feature naming convention, see Section II-B.

observation to note here is that in all cases, the selected features do not cover an entire spectrum of the graphomotor tasks (TSK1–TSK7) under investigation. Moreover, the distributions of the tasks per feature-type vary as well. This indicates that each individual type of the features can potentially be used to describe slightly different task-specific aspects of GD experienced by school-aged children supporting the use of a variety of specialized feature-types to provide a more robust and wide-scale description of the hidden complexities underlying GD in general.

Regarding the results of the statistical analysis, it can be seen that basic parameters such as mean tilt, height, and length of writing were found as the most statistically significant features in the case of the conventional (baseline) feature set. More specifically, mean tilt during the drawing of Archimedean spiral (TSK1) and rainbow (TSK6) showed the strongest relationship with the presence of GD. As can be seen, children with GD held the pen less steeply when performing such spiral- and rainbow shape-based movements. In addition, when compared with the cohort of healthy children, sawtooth (TSK5) and rainbow (TSK6) drawn by children with GD were found to be smaller in both height as well as length further underlining the difficulties associated with these tasks.

Another fact that can be observed in the results of the statistical analysis is that as opposed to the conventional features which consisted solely of the spatial (stroke length and height) and dynamic (tilt) parameters, the top-ranking non-conventional features mostly consisted of kinematic features (velocity, acceleration, and jerk) computed in both horizontal as well as vertical projections, and dynamic features (tilt). This observation is in line with the analysis performed by a variety of previous studies [6], [65]–[67] using kinematic features to quantify GD, and confirms the fact that kinematic features are an important measure of the quality of handwriting as well as drawing. Furthermore, such features are specific to computerized analysis as they are almost impossible to be quantified precisely using the human perception of the final handwritten product.

With respect to the features based on modulation spectra, all of the top-ranking features showed a positive correlation with the presence of GD indicating the existence of an increased low-frequency noise in the analyzed handwriting signals. This noise seems to be relatively task-independent as it appeared in all spiral-, loop- as well as sawtooth-based movements. Moreover, in four out of five cases, the features were based on acceleration or jerk, which points out to inability of children with GD to perform a given graphomotor task with steady and controlled velocity that is eventually reflected in an increased noise in the acquired kinematic signals (mathematical point of view) as well as in the lack of fluency and efficiency during handwriting (clinical point of view). Such observation is in line with the previous research reporting non-fluent handwriting as being present in children with HD (diagnosed with DD) [32], [68].

Regarding the top-ranking FD-based features, it may be noticed that all of them were extracted from different graphomotor tasks (TSK1–TSK5) further underlying the need for a variety of specifically-designed features to quantify GD. The most significant FD-based feature, the mean velocity of tilt extracted from TSK1, probably refers to the difficulties in changing the direction of the Archimedean spiral caused by hesitancy, distress, etc. This is an interesting finding as it is in line with the most significant conventional feature being the mean tilt, which highlights the importance of different tilt parametrizations. The rest of the most correlated FD-based features are derived from velocity and acceleration. This shows that FDs can be advantageously applied to both kinematic as well as dynamic features. Additionally, the values of α suggest that regular derivation is not optimal for kinematic handwriting features, which is in line with our previous research [43], [45].

Regarding the top-ranking TQWT features, the only statistically significant correlation was found for the signal-to-noise ratio of the vertical velocity extracted from the rainbow task (TSK6). This probably shows that maintaining steady velocity while performing this particular task is not causing problems to healthy children, but is challenging for children with GD, which is in line with the previous publication reporting problems in vertical movements in children with DD [6] caused by the psychological and muscular fatigue in the finger system. The vertical movement requires coordinated movement and finer flexions/extensions of more joints (interphalangeal and metacarpophalangeal) and therefore it is more complex than ulnar abductions of the wrist [69], [70], which plays a key role in the horizontal one, i. e. GD are more pronounced in the vertical projection of handwriting/drawing. Next, we assume, that children with GD are unable to quickly change the acceleration of their handwriting. On the other hand, healthy children have fewer problems with handwriting automation and therefore can change the acceleration more fluently. This can indirectly cause higher noise-level in the residual component of vertical acceleration in the handwritten product of healthy children, as can be seen in the second most significant TQWT feature.

Finally, concerning the results of the classification analysis, it can be seen that all of the three novel feature-types achieved similar classification performance in comparison to the conventional handwriting features. This shows that a single type of feature, even if more complex, is not likely to improve the identification of GD provided by the conventional features significantly. However, as the results suggest, when these features are combined, the classification performance can be increased by approximately 10 % in terms of accuracy, 3 % in terms of sensitivity and 10 % in terms of specificity. An important fact to note is that when compared with the previous research, the results proposed in this work might at first seem unsatisfactory as some of the recent publications reported over 90 % sensitivity [26], [28], [32]. However, those studies aimed at identifying HD in children with DD using a complex acquisition protocol comprising writing. The results proposed in this work are based solely on graphomotorics and aim at predicting the presence of GD that can lead to HD and possibly to DD. It is of great importance to also focus on simple graphomotor movements as they form the basis of handwriting, hence, a robust parametrization of GD has a potential to be used as an early marker of DD in children in pre-school age or first grades of a primary school. Another important fact to note is that all of the feature-types, as well as the conventional features, were selected when training the combined model. In addition, except TSK4 (flipped version of the connected loops in TSK3), all of the graphomotor tasks are present as well, This shows that all of the selected features extracted from almost all of the graphomotor tasks contributed to an improvement in the identification of GD confirming the hypothesis of enhancing the model's capability to model the relationship between the properties of the handwriting signals and the presence of GD in school-aged children.

V. LIMITATIONS OF THE STUDY

This work has several limitations. First, we need to be aware of the restricted statistical strength of the inference about the population of school-aged children given a relatively small sample size of 53 children. Next, only children attending 3rd and 4th grade of the primary school were enrolled in this study. To obtain a more complex spectrum of handwriting signals, i. e. to have additional information about the performance of the proposed graphomotor features and their relationship with children's age, grade, etc., handwriting signals of children attending 1st and 2nd grade of the primary school (possibly even pre-school children) as well as children attending the higher grades should also be analyzed. On the other hand, our cohort includes children from the 3rd and 4th grade of primary schools, where the handwriting should become automatic. Therefore a possibility to identify GD in this stage is critical for the consequent diagnosis and therapeutic care of DD. The results proposed in this work therefore laid the foundations (baseline) for future studies that should bring even more information about GD in various age profiles and their evolution in time. Next, deeper investigation and design of the features can be performed, e. g. additional tuning of the filter-banks to compute modulation spectra, other formulations of fractional order derivatives or sub-bands of the tunable Q-factor wavelet transform could be analyzed. Next, various machine learning models should be trained and compared in the future studies to get more information about the classification performance of the proposed features and to obtain the most robust models for GD identification. Finally, the relationship between the classification performance of the trained models with the feature space complexity as well as the cross-validation setup should be investigated to evaluate and confirm the robustness of the proposed methodology. To sum up, concerning the limitations mentioned above, this study should be considered as being rather exploratory and pilot in nature, and its results should be confirmed by the subsequent scientific research.

VI. CONCLUSION

In this study, we presented three novel types of graphomotor features providing more robust and complex quantification of GD in school-aged children. In each feature-type, we identified several features that significantly differentiate healthy children and children with GD. Of note is the fact that the novel features mostly quantified kinematic aspects of the handwriting process that are very hard to be perceived by a human examiner using only a final handwritten product. In addition, we also showed that combining the proposed graphomotor features with the set of conventionally used ones can increase the prediction capability of the trained binary classifier significantly. With respect to the acquisition protocol, all of the chosen graphomotor tasks but one appeared in the final selection of the features used to train the combined classification model. This confirms that using a variety of basic graphomotor tasks requires coordinated movement of fingers, wrist, elbow, shoulder as well as visuospatial cognitive functions that allow the more advanced features to quantify subtle and rather imperceptible manifestations of GD using online handwriting.

To the best of our knowledge, it is the first work exploring the possibilities of using modulation spectra, fractional order derivatives and tunable Q-factor wavelet transform to extract advanced graphomotor features for the purpose of quantification and identification of GD in school-aged children. Based on the reported results, we conclude that the proposed features have a great potential to improve the computerized identification and assessment of GD. However, to generalize the results, our findings should be confirmed by further scientific research.

ACKNOWLEDGMENT

For the research, the infrastructure of the SIX Center was used.

REFERENCES

[1] A. W. A. Van Gemmert and H.-L. Teulings, "Advances in graphonomics: Studies on fine motor control, its development and disorders," *Hum. Movement Sci.*, vol. 25, nos. 4–5, pp. 447–453, Oct. 2006, doi: 10.1016/j.humov.2006.07.002.

[2] A. P. Accardo, M. Genna, and M. Borean, "Development, maturation and learning influence on handwriting kinematics," *Hum. Movement Sci.*, vol. 32, no. 1, pp. 136–146, Feb. 2013.

[3] A. W. Blöte and L. Hamstra-Bletz, "A longitudinal study on the structure of handwriting," *Perceptual Motor Skills*, vol. 72, no. 3, pp. 983–994, Jun. 1991.

[4] S. M. Rueckriegel, F. Blankenburg, R. Burghardt, S. Ehrlich, G. Henze, R. Mergl, and P. Hernáiz Driever, "Influence of age and movement complexity on kinematic hand movement parameters in childhood and adolescence," *Int. J. Develop. Neurosci.*, vol. 26, no. 7, pp. 655–663, Nov. 2008.

[5] V. Matijević-Mikelić, T. Košiček, M. Crnković, Z. Trifunović-Maček, and S. Grazio, "Development of early graphomotor skills in children with neurodevelopmental risks," *Acta Clinica Croatica*, vol. 50, no. 3, pp. 317–321, 2011.

[6] A. Kushki, H. Schwellnus, F. Ilyas, and T. Chau, "Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia," *Res. Develop. Disabilities*, vol. 32, no. 3, pp. 1058–1064, May 2011.

[7] S. Rosenblum, "Inter-relationships between objective handwriting features and executive control among children with developmental dysgraphia," *PLoS ONE*, vol. 13, no. 4, Apr. 2018, Art. no. e0196098.

[8] A. A. Alhusaini, G. R. Melam, and S. Buragadda, "Short-term sensorimotor-based intervention for handwriting performance in elementary school children," *Pediatrics Int.*, vol. 58, no. 11, pp. 1118–1123, Jul. 2016.

[9] K. P. Feder and A. Majnemer, "Handwriting development, competency, and intervention," *Develop. Med. Child Neurol.*, vol. 49, no. 4, pp. 312–317, Apr. 2007.

[10] K. McHale and S. A. Cermak, "Fine motor activities in elementary school: Preliminary findings and provisional implications for children with fine motor problems," *Amer. J. Occupational Therapy*, vol. 46, no. 10, pp. 898–903, Oct. 1992.

[11] B. Engel-Yeger, L. Nagauker-Yanuv, and S. Rosenblum, "Handwriting performance, self-reports, and perceived self-efficacy among children with dysgraphia," *Amer. J. Occupational Therapy*, vol. 63, no. 2, pp. 182–192, Mar. 2009.

[12] S. Graham, K. R. Harris, and B. Fink, "Is handwriting causally related to learning to write? Treatment of handwriting problems in beginning writers.," *J. Educ. Psychol.*, vol. 92, no. 4, pp. 620–633, 2000, doi: 10.1037/0022-0663.92.4.620.

[13] B. C. M. Smits-Engelsman and G. P. Van Galen, "Dysgraphia in children: Lasting psychomotor deficiency or transient developmental delay?" *J. Experim. Child Psychol.*, vol. 67, no. 2, pp. 164–184, Nov. 1997.

[14] S. T. Peverly, J. K. Garner, and P. C. Vekaria, "Both handwriting speed and selective attention are important to lecture note-taking," *Reading Writing*, vol. 27, no. 1, pp. 1–30, Jan. 2014.

[15] J. Medwell and D. Wray, "Handwriting automaticity: The search for performance thresholds," *Lang. Edu.*, vol. 28, no. 1, pp. 34–51, Jan. 2014.

[16] C. A. Christensen, "The critical role handwriting plays in the ability to produce high-quality written text," in *The SAGE Handbook Writing Development*. Thousand Oaks, CA, USA: SAGE, 2009, pp. 284–299.

[17] M. R. I. Martins, J. A. Bastos, A. T. Cecato, M. D. L. S. Araujo, R. R. Magro, and V. Alaminos, "Screening for motor dysgraphia in public schools," *Jornal de Pediatria*, vol. 89, no. 1, pp. 70–74, Jan. 2013.

[18] L. Hamstra-Bletz, J. DeBie, and B. P. L. M. D. Brinker, *Concise Evaluation Scale for Children's Handwriting*, vol. 1, Lisse, The Netherlands: Swets, 1987.

[19] S. Rosenblum, "Development, reliability, and validity of the handwriting proficiency screening questionnaire (HPSQ)," *Amer. J. Occupational Therapy*, vol. 62, no. 3, pp. 298–307, May 2008.

[20] S. Rosenblum and L. Gafni-Lachter, "Handwriting proficiency screening questionnaire for children (HPSQ—C): Development, reliability, and validity," *Amer. J. Occupational Therapy*, vol. 69, no. 3, Apr. 2015, Art. no. 6903220030p1.

[21] S. Rosenblum, P. L. Weiss, and S. Parush, "Product and process evaluation of handwriting difficulties," *Educ. Psychol. Rev.*, vol. 15, no. 1, pp. 41–81, 2003.

[22] D. V. Cicchetti, "Assessing inter-rater reliability for rating scales: Resolving some basic issues," *Brit. J. Psychiatry*, vol. 129, no. 5, pp. 452–456, Nov. 1976.

[23] S. Rosenblum, P. L. Weiss, and S. Parush, "Handwriting evaluation for developmental dysgraphia: Process versus product," *Reading Writing*, vol. 17, no. 5, pp. 433–458, Jul. 2004.

[24] S. Rosenblum, S. Goldstand, and S. Parush, "Relationships among biomechanical ergonomic factors, handwriting product quality, handwriting efficiency, and computerized handwriting process measures in children with and without handwriting difficulties," *Amer. J. Occupational Therapy*, vol. 60, no. 1, pp. 28–39, Jan. 2006.

[25] S. Rosenblum, S. Parush, and P. L. Weiss, "The in air phenomenon: Temporal and spatial correlates of the handwriting process," *Perceptual Motor Skills*, vol. 96, no. 3, pp. 933–954, Jun. 2003, doi: 10.2466/pms.2003.96.3.933.

[26] S. Rosenblum and G. Dror, "Identifying developmental dysgraphia characteristics utilizing handwriting classification methods," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 2, pp. 293–298, Apr. 2017.

[27] E. Pagliarini, L. Scocchia, M. Vernice, M. Zoppello, U. Balottin, S. Bouamama, M. T. Guasti, and N. Stucchi, "Children's first handwriting productions show a rhythmic structure," *Sci. Rep.*, vol. 7, no. 1, p. 5516, Dec. 2017.

[28] T. Asselborn, T. Gargot, Ł. Kidziński, W. Johal, D. Cohen, C. Jolly, and P. Dillenbourg, "Automated human-level diagnosis of dysgraphia using a consumer tablet," *NPJ Digit. Med.*, vol. 1, no. 1, p. 42, Dec. 2018.

[29] C. Alonso-Martinez, M. Faundez-Zanuy, and J. Mekyska, "A comparative study of in-air trajectories at short and long distances in online handwriting," *Cognit. Comput.*, vol. 9, no. 5, pp. 712–720, Oct. 2017.

[30] S. Rosenblum, A. Y. Dvorkin, and P. L. Weiss, "Automatic segmentation as a tool for examining the handwriting process of children with dysgraphic and proficient handwriting," *Hum. Movement Sci.*, vol. 25, nos. 4–5, pp. 608–621, Oct. 2006.

[31] G. Luria, A. Kahana, and S. Rosenblum, "Detection of deception via handwriting behaviors using a computerized tool: Toward an evaluation of malingering," *Cognit. Comput.*, vol. 6, no. 4, pp. 849–855, Dec. 2014.

[32] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and rating of developmental dysgraphia by handwriting analysis," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 2, pp. 235–248, Apr. 2017.

[33] J. Mekyska, J. Bednarova, M. Faundez-Zanuy, Z. Galaz, K. Safarova, V. Zvoncak, J. Mucha, Z. Smekal, A. Ondrackova, T. Urbanek, and J. M. Havigerova, "Computerised assessment of graphomotor difficulties in a cohort of school-aged children," in *Proc. 11th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT)*, Oct. 2019, pp. 1–6.

[34] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[35] V. Zvoncak, J. Mucha, Z. Galaz, J. Mekyska, K. Safarova, M. Faundez-Zanuy, and Z. Smekal, "Fractional order derivatives evaluation in computerized assessment of handwriting difficulties in school-aged children," in *Proc. 11th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT)*, Oct. 2019, pp. 1–6.

[36] T. Asselborn, M. Chapatte, and P. Dillenbourg, "Extending the spectrum of dysgraphia: A data driven strategy to estimate handwriting quality," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, Dec. 2020.

[37] V. Zvoncak, J. Mekyska, K. Safarova, Z. Galaz, J. Mucha, T. Kiska, Z. Smekal, B. Losenicka, B. Cechova, P. Francova, M. Faundez-Zanuy, and S. Rosenblum, "Effect of stroke-level intra-writer normalization on computerized assessment of developmental dysgraphia," in *Proc. 10th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT)*, Nov. 2018, pp. 1–5.

[38] Z. WU, T. LIN, and M. LI, "Automated detection of children at risk of chinese handwriting difficulties using handwriting process information: An exploratory study," *IEICE Trans. Inf. Syst.*, vol. 102, no. 1, pp. 147–155, Jan. 2019.

[39] V. Zvoncak, J. Mekyska, K. Safarova, Z. Smekal, and P. Brezany, "New approach of dysgraphic handwriting analysis based on the tunable Q-Factor wavelet transform," in *Proc. 42nd Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2019, pp. 289–294.

[40] S. Rosenblum, S. Parush, and P. L. Weiss, "Computerized temporal handwriting characteristics of proficient and non-proficient handwriters," *Amer. J. Occupational Therapy*, vol. 57, no. 2, pp. 129–138, Mar. 2003, doi: 10.5014/ajot.57.2.129.

[41] J. Mekyska, E. Janousova, P. Gomez-Vilda, Z. Smekal, I. Rektorova, I. Eliasova, M. Kostalova, M. Mrackova, J. B. Alonso-Hernandez, M. Faundez-Zanuy, and K. López-de-Ipiña, "Robust and complex approach of pathological speech signal analysis," *Neurocomputing*, vol. 167, pp. 94–111, Nov. 2015.

[42] M. Markaki, Y. Stylianou, J. D. Arias-Londono, and J. I. Godino-Llorente, "Dysphonia detection based on modulation spectral features and cepstral coefficients," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 5162–5165.

[43] J. Mucha, K. Lopez-de-Ipina, M. Faundez-Zanuy, J. Mekyska, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Analysis of Parkinson's disease dysgraphia based on optimized fractional order derivative features," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2019, pp. 1–5.

[44] J. Mucha, V. Zvoncak, Z. Galaz, M. Faundez-Zanuy, J. Mekyska, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de-Ipina, "Fractional derivatives of online handwriting: A new approach of parkinsonic dysgraphia analysis," in *Proc. 41st Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2018, pp. 214–217.

[45] J. Mucha, J. Mekyska, M. Faundez-Zanuy, K. Lopez-De-Ipina, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Advanced Parkinson's disease dysgraphia analysis based on fractional derivatives of online handwriting," in *Proc. 10th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT)*, Nov. 2018, pp. 1–6.

[46] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, K. Lopez-de-Ipina, V. Zvoncak, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Identification and monitoring of Parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting," *Appl. Sci.*, vol. 8, no. 12, p. 2566, Dec. 2018.

[47] J. T. Machado, "Fractional coins fractional derivatives," *Abstract Appl. Anal.*, vol. 2013, May 2013, Art. no. 205097.

[48] I. Podlubny, *Fractional Differential Equations an Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*. San Diego, CA, USA: Academic, 1999.

[49] A. A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo, *Theory and Applications of Fractional Differential Equations*, vol. 204. Amsterdam, The Netherlands: Elsevier, 2006.

[50] R. Scherer, S. L. Kalla, Y. Tang, and J. Huang, "The Grünwald–Letnikov method for fractional differential equations," *Comput. Math. Appl.*, vol. 62, no. 3, pp. 902–917, 2011.

[51] I. W. Selesnick, "Resonance-based signal decomposition: A new sparsity-enabled signal analysis method," *Signal Process.*, vol. 91, no. 12, pp. 2793–2809, Dec. 2011.

[52] I. W. Selesnick, "Wavelet transform with tunable Q-Factor," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3560–3575, Aug. 2011.

[53] I. Bayram and I. W. Selesnick, "Frequency-domain design of overcomplete rational-dilation wavelet transforms," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2957–2972, Aug. 2009.

[54] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1570–1582, Oct. 2005.

[55] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[56] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, nos. 3–4, pp. 591–611, Dec. 1965.

[57] G. E. Box and D. R. Cox, "An analysis of transformations," *J. Roy. Stat. Soc., Ser. B, Methodol.*, vol. 26, no. 2, pp. 211–243, 1964.

[58] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[59] I. Guyon, S. Gunn, M. Nikravesh, L. A. Zadeh, and Eds, *Feature Extraction: Foundations and Applications*. New York, NY, USA: Springer, 2006.

[60] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochim. Biophys. Acta (BBA)*, vol. 405, no. 2, pp. 442–451, 1975.
[61] G. Jurman, S. Riccadonna, and C. Furlanello, "A comparison of MCC and CEN error measures in multi-class prediction," *PLoS ONE*, vol. 7, no. 8, Aug. 2012, Art. no. e41882.
[62] B. Phipson and G. K. Smyth, "Permutation P-values should never be zero: Calculating exact P-values when permutations are randomly drawn," *Stat. Appl. Genet. Mol. Biol.*, vol. 9, no. 1, pp. 1–12, Jan. 2010.
[63] E. Combrisson and K. Jerbi, "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy," *J. Neurosci. Methods*, vol. 250, pp. 126–136, Jul. 2015.
[64] M. Ojala and G. C. Garriga, "Permutation tests for studying classifier performance," in *Proc. 9th IEEE Int. Conf. Data Mining*, Dec. 2009, pp. 908–913.
[65] A. Accardo, A. Chiap, M. Borean, L. Bravar, S. Zoia, M. Carrozzi, and A. Scabar, "A device for quantitative kinematic analysis of children's handwriting movements," in *Proc. 11th Medit. Conf. Med. Biomed. Eng. Comput.*, 2007, pp. 445–448.
[66] S.-H. Chang and N.-Y. Yu, "Handwriting movement analyses comparing first and second graders with normal or dysgraphic characteristics," *Res. Develop. Disabilities*, vol. 34, no. 9, pp. 2433–2441, Sep. 2013.
[67] S. Rosenblum and H. E. Roman, "Fluctuation analysis of proficient and dysgraphic handwriting in children," *EPL (Europhys. Lett.)*, vol. 85, no. 5, p. 58007, Mar. 2009.
[68] J. Danna, V. Paz-Villagrán, and J.-L. Velay, "Signal-to-noise velocity peaks difference: A new method for evaluating the handwriting movement fluency in children with dysgraphia," *Res. Develop. Disabilities*, vol. 34, no. 12, pp. 4375–4384, Dec. 2013.
[69] N. Dounskaia, A. W. A. Van Gemmert, and G. E. Stelmach, "Interjoint coordination during handwriting-like movements," *Exp. Brain Res.*, vol. 135, no. 1, pp. 127–140, Oct. 2000.
[70] G. P. van Galen, "Handwriting: Issues for a psychomotor theory," *Hum. Movement Sci.*, vol. 10, nos. 2–3, pp. 165–191, May 1991.

ZOLTAN GALAZ was born in Galanta, Slovak Republic. He received the Ph.D. degree in telecommunications from the Brno University of Technology, Czech Republic, in 2018.

He is currently a Team Leader of the Brain Diseases Analysis Laboratory, Department of Telecommunications, Brno University of Technology. He is a coauthor of more than ten publications indexed by the Web of Science. He deals with the research of advanced digital signal processing and machine learning algorithms applied in the field of computerized non-invasive neurodegenerative and neurodevelopmental disorders diagnosis, analysis, and monitoring. In cooperation with neurologists/psychologists from different countries, he develops diagnostic/monitoring systems focused on disorders, such as Parkinson's disease and developmental dysgraphia.

JAN MUCHA was born in Liptovský Mikuláš, Slovak Republic. He received the M.Sc. degree in telecommunications from the Brno University of Technology, Czech Republic, in 2016, where he is currently pursuing the Ph.D. degree with the Department of Telecommunications.

He is also a Researcher at the Brain Diseases Analysis Laboratory, Brno University of Technology, where he is also a part of the Human–Machine Interaction Group, Signal Processing Laboratory. He deals with the research of non-invasive neurodegenerative and neurodevelopmental disorders analysis based on handwriting processing, especially with advanced parametrization techniques of developmental dysgraphia.

VOJTECH ZVONCAK was born in Zlin, Czech Republic. He received the M.Sc. degree in telecommunications from the Brno University of Technology, Czech Republic, in 2016, where he is currently pursuing the Ph.D. degree with the Department of Telecommunications, Faculty of Electrical Engineering and Communication.

He is also a Researcher at the Brain Diseases Analysis Laboratory, Brno University of Technology. He deals with handwriting and speech signal processing. He currently focuses on the diagnosis of developmental dysgraphia based on quantitative analysis of online handwriting.

JIRI MEKYSKA was born in Holesov, Czech Republic. He received the Ph.D. degree in telecommunications from the Brno University of Technology, Czech Republic, in 2014.

He is the Head of the Brain Diseases Analysis Laboratory and the Human–Machine Interaction Group, Signal Processing Laboratory, Department of Telecommunications, Brno University of Technology. He is a coauthor of more than 30 publications indexed by the Web of Science. He deals with the research of non-invasive neurodegenerative and neurodevelopmental disorders analysis based on speech, sleep, and handwriting processing. In cooperation with neurologists/psychologists from different countries, he develops diagnostic/monitoring systems focused on Lewy body diseases and developmental dysgraphia.

Dr. Mekyska is a member of the Cybernetics and Information Processing Expert Panel of the Czech Science Foundation and the Expert Advisory Body of Evaluators of the Development and Innovation Council of the Czech Government. In 2013, he received the Joseph Fourier Award for his work in the field of paraclinical neurodegenerative diseases analysis. He serves as a reviewer of EU (Innovative Medicines Initiative) and national (Czech Science Foundation and Czech Ministry of Health) projects. He also serves as a reviewer of many international journals, such as the IEEE Journal of Selected Topics in Signal Processing, the IEEE Transactions on Human-Machine Systems, and *Parkinsonism & Related Disorders*.

ZDENEK SMEKAL (Senior Member, IEEE) was born in Brno, Czech Republic. He received the M.Sc. and Ph.D. degrees from the Brno University of Technology, in 1973 and 1978, respectively.

In 1986, he was an Associate Professor in the field of electronics and communication engineering and a Professor in the field of theoretical electrical engineering, in 2000. He is currently a full-time Professor at the Brno University of Technology. He is the Chairman of the Board for Doctoral Study Programme of Teleinformatics and the Leading Researcher of the Signal Processing Laboratory (SPLab), Department of Telecommunications, Brno University of Technology. He is a coauthor of more than 90 publications indexed by the Web of Science and an author of five monographs on digital signal processing.

Prof. Smekal is a member of the Scientific Council at the Brno University of Technology. He was involved in the European COST research projects: OC 10 G6 Digital Audio Effects, from 1997 to 2001, OC 277 Nonlinear Signal Processing, from 2004 to 2006, and OC 2102 Cross-Modal Analysis of Verbal and Nonverbal Communication, from 2006 to 2010. He is also an Editor of *Cognitive Computation*, and he serves as a reviewer of many international journals, such as *Digital Signal Processing*, *Neural Networks*, and *Pattern Recognition*.

KATARINA SAFAROVA was born in Sala, Slovakia. She received the Ph.D. degree in general psychology from the Department of Psychology, Masaryk University, Brno, Czech Republic, in 2018.

Since 2018, she has been a Researcher with the Department of Psychology, Masaryk University, and a member of a team dealing with the topic of graphomotor issues, especially dysgraphia and its diagnosis. Since 2018, she has been working as an Assistant at the Institute of Social Sciences, Brno University of Technology. In 2017, she spent two months as a Researcher at the Department of Psychology, University of Essex, U.K. Her research interests include educational psychology, learning disabilities, diagnostic methods, and its psychometric properties and methodology.

ANEZKA ONDRACKOVA was born in Brno, Czech Republic. She received the M.Sc. degree in psychology from the Faculty of Arts, Masaryk University, Brno, Czech Republic, in 2019.

She currently works as a School Counselor at a nursery high school in Brno, and she also cooperates with archers on their mental preparation for archery competitions. Her research interests include school, developmental, and counseling psychology. She is involved in the research of dysgraphia and analysis based on handwriting processing utilizing advanced parametrization techniques.

TOMAS URBANEK was born in Hustopece, Czech Republic. He received the Ph.D. degree in psychology from the Faculty of Arts, Masaryk University, Brno, Czech Republic, in 1999.

Since 1999, he has been a Teacher at the Faculty of Arts, Masaryk University, Brno, where he currently works as a Professor of general psychology at the Department of Psychology. Since 1995, he has been working with the Institute of Psychology, Czech Academy of Sciences, as a Researcher and Senior Researcher, where he has also been the Director of the Institute of Psychology, since 2017. He published several books and about 70 articles, and he is the author or coauthor of several psychological tests.

JANA MARIE HAVIGEROVA was born in South Moravia, Czech Republic. She received the M.Sc. and Ph.D. degrees from the Faculty of Arts, Masaryk University, Brno, Czech Republic, in 1997 and 2004, respectively.

She was an Associate Professor of psychology at the Faculty of Arts, Masaryk University, Brno, in 2017. She has been working with the University of Hradec Kralove (part-time), since 2005. She has been cooperating on a long-term basis with the University of South Bohemia, Ceske Budejovice, on the research of the relationship between personality and language (the CPACT project). She is currently managing the DAQUAo Project focused on the study of psycholinguistic phenomena of therapeutic diaries. Within the Dysgraphy project, she participates in data collection and partial analysis, and focusing on pre-school children.

JIRINA BEDNAROVA was born in Jindrichuv Hradec, Czech Republic. She graduated in psychology from the Faculty of Arts, Masaryk University, Brno, Czech Republic.

Since 1995, she has been working as a Special Educational Counselor. She deals with the diagnosis and intervention in children with inequalities of cognitive and graphomotor development, and specific learning disabilities. She also deals with the education of adults working with such children. She is also an author of several diagnostic materials (standardized and dynamic diagnosis) and a number of publications aimed at targeted support of children's development.

MARCOS FAUNDEZ-ZANUY was born in Barcelona, Spain. He received the Ph.D. degree from the Polytechnic University of Catalonia, in 1998.

He is currently a Full Professor at ESUP Tecnocampus, Mataró, and heads the Signal Processing Group. From 2009 to 2018, he was the Dean of the Escola Superior Politecnica Tecnocampus (Polytechnic University of Catalunya, till 2015, and Pompeu Fabra University, after 2015). Since 2010, he is the Head of the Research at Tecnocampus. His research interests include the fields of biometrics applied to security and health. He is an author of more than 50 articles indexed in ISI Journal citation report, more than 100 conference papers, around ten books, and responsible of ten national and European research projects. He is a Spanish liaison of EURASIP. He was an Initiator and the Chairman of the European COST action 277 *Nonlinear speech processing*, and the secretary of COST action 2102 *Cross-Modal Analysis of Verbal and Non-Verbal Communication*.

A.10 Analysis of Various Fractional Order Derivatives Approaches in Assessment of Graphomotor Difficulties

Received November 26, 2020, accepted December 1, 2020, date of publication December 4, 2020, date of current version December 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042591

Analysis of Various Fractional Order Derivatives Approaches in Assessment of Graphomotor Difficulties

JAN MUCHA[1], JIRI MEKYSKA[1], ZOLTAN GALAZ[1], MARCOS FAUNDEZ-ZANUY[2], VOJTECH ZVONCAK[1], KATARINA SAFAROVA[3], TOMAS URBANEK[3], JANA MARIE HAVIGEROVA[3], JIRINA BEDNAROVA[3], AND ZDENEK SMEKAL[1], (Senior Member, IEEE)

[1]Department of Telecommunications, Brno University of Technology, 61600 Brno, Czech Republic
[2]Escola Superior Politécnica del TecnoCampus, 08302 Mataró, Spain
[3]Department of Psychology, Faculty of Arts, Masaryk University, 60200 Brno, Czech Republic

Corresponding author: Jan Mucha (muchajano@phd.feec.vutbr.cz)

This work was supported in part by the Czech Science Foundation (research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing) under Grant 18-16835S, and in part by the Ministry of Economic Affairs and Competitiveness of Spain the through the FEDER and MEC projects under Grant TEC2016-77791-C4-2-R.

ABSTRACT Graphomotor disabilities (GD) are present in up to 30 % of school-aged children and are associated with several symptoms in the field of kinematics. Although the basic kinematic features such as velocity, acceleration, and jerk were proved to effectively quantify these symptoms, a recent body of research identified that the theory of fractional calculus can be used to even improve the objective GD assessment. The goal of this study is to extend the current knowledge in this field and explore the abilities of several fractional order derivatives (FD) approximations to estimate the severity of GD in the children population. We enrolled 85 children attending the 3rd and 4th grade of primary school, who performed a combined loop task on a digitizing tablet. Their performance was rated by psychologists and the online handwriting signals were parametrised by kinematic features utilising three FD approximations: Grünwald-Letnikov's, Riemann–Liouville's, and Caputo's. In this study, we showed the differences across the employed FD approaches for the same kinematic handwriting features and their potential in GD analysis. The results suggest that the Riemann-Liouville's approximation in the field of quantitative GD analysis outperforms the other ones. Using this approach, we were able to estimate the overall score with a low error of 0.65 points, while the scale range is 4. In fact, the psychologists tend to make the error even higher.

INDEX TERMS Fractional calculus, fractional order derivatives, graphomotor difficulties, graphonomics, online handwriting.

I. INTRODUCTION

Fractional calculus (FC) is a name of the theory of integrals and derivatives of an arbitrary order [1]. The concept of fractional operators has been introduced almost simultaneously with the development of the classical differential, integral or other well-known calculus [2]. It attracted the interest of many famous mathematicians, including Euler, Liouville, Laplace, Riemann, Grünwald, and Letnikov. The principles of FC have been used in modeling of many physical and chemical processes, as well as in modern engineering and science in general [3]–[5]. It has been advantageously used during the modeling of different diseases such as the human immunodeficiency virus (HIV) [6] or malaria [7]. Recently, the FC has been significantly examined in computer vision, particularly in image restoration, super-resolution, image segmentation or motion estimation [8]. In line with this trend, in our recent research, we developed new parametrisation techniques of online handwriting (a handwritten signal with temporal information) based on the application of the fractional order derivatives (FD) [9]–[13].

It has been estimated that approximately 10–30 % of children experience graphomotor difficulties [20] such as graphomotor production deficits, motor feedback difficulties

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo.

(e. g. the pen's tip location tracking problems), motor-memory dysfunctions, etc. Considering that children spend 31–60 % of their school-time performing handwriting [21], the early identification of graphomotor disabilities (GD) is crucial in the prevention of serious pedagogical and psychological consequences [22]. Otherwise, a child's every-day life can be greatly affected starting with a lack of motivation to write, a decrease in self-esteem in combination with poor emotional well-being continuing to bad attitude and behaviour, communication and social interaction problems [23]. In some cases, it may result in being diagnosed with a serious neurodevelopmental disorder such as developmental dysgraphia (DD) [24], [25]. To identify and evaluate GD in school-aged children, several well-established questionnaires or tests based on a visual inspection of the handwritten product have been developed [26]. Though, their utilization on a day-to-day basis is still limited due to the fact that the administration and coding are very time-consuming. Furthermore, the perceptual abilities, experience, and subjective judgment of an examiner are limited as well.

To overcome the limitations of the perceptual GD analysis, researchers have been focusing on computerized quantitative analysis of online handwriting. Pen and paper have been replaced by digitizing tablets used to record a variety of signals describing the evolution of the handwritten product in time. It allowed quantification of kinematic (velocity, acceleration or jerk) as well as dynamic (pen pressure, tilt or azimuth) components of the handwritten signal. For instance, Pagliarini *et al.* [27] (2017) presented the potential of quantitative analysis to identify the development of handwriting difficulties (HD) at a very early age. Mekyska *et al.* [14] (2017) built a classifier (random forests; 54 children) identifying the presence of DD with 96 % sensitivity and specificity. Rosenblum and Dror [15] (2017) achieved 90 % sensitivity and specificity in DD diagnosis (support vector machines; 99 children) using various kinematic and dynamic features. Asselborn *et al.* [16] (2018) reported 96 % sensitivity and 99 % specificity (random forests; 268 children) using 53 handwriting features quantifying different dimensions of handwriting. Next, Mekyska *et al.* [17] (2019) proposed a model (based on XGBoost, 76 children) and achieved 50 % sensitivity and 90 % specificity in identification of GD presence using 7 basic graphomotor elements quantified by conventional temporal, spatial, kinematic, and dynamic parameters. In 2020, Galaz *et al.* [13] published a work dealing with advanced analysis (utilising modulation spectra, fractional order derivatives, and tunable-Q wavelet transform) of graphomotor elements in 53 children attending 3rd and 4th grade of elementary schools. Employing random forests they reached 83 % sensitivity and 81 % specificity. In the same year, Asselborn *et al.* [18] proposed new data-driven based approaches for an assessment of handwriting difficulties, that were divided into 4 dimensions: kinematic, pressure, tilt and static. This novel approach enables a detailed analysis in children having a very similar overall score of dysgraphia, but differing in specific difficulties. Finally, Garot *et al.* [19] (2020) enrolled 280 children who were performing the Concise Evaluation Scale for Children's Handwriting (BHK) while recorded by digitising tablets. Employing a cluster analysis, the authors were able to automatically discriminate among 3 groups of children associated with dysgraphia: 1) children with mild dysgraphia usually not identified in schools, 2) children with severe dysgraphia manifested in kinematics and pressure, and 3) children with severe dysgraphia manifested mainly in tilt. The overview of the mentioned current works and their achievemnts can be found in Table 1.

Considering the success of utilizing the FD (Grünwald-Letnikov approach) in Parkinson's disease dysgraphia analysis in our previous works [9]–[12], and in the assessment of GD in school-aged children [13], [28], this study, as a next logical step, has the following aims:

- to extend our previous research by the employment of several FD-approaches instead of one (Grünwald-Letnikov approach),
- to explore the differences of several FD approaches in the assessment of GD in the children population,
- to compare the power of the FD-based handwriting features computed by several FD approaches to estimate the severity of GD.

II. DATASET & METHODOLOGY

A. DATASET

For this study, we enrolled 85 children (31 girls and 54 boys) attending 3rd and 4th grade at several primary schools in the Czech Republic. The demographic data of the participants can be found in Table 2 and the resulting grade distribution in Table 3. Children were asked to perform drawings,

TABLE 1. Overview of current works.

Study	DB	Age	Tasks	Features	Results
Mekyska et al. [14] (2017)	27	8–9	Drawings	K, D, S, T, A	Classification: SEN = 98 %, SPE = 98 %; Regression: ERR = 10 %
Rosenblum and Dror [15] (2017)	99	8–9	Writing	D, S, T	Classification: SEN = 90 %, SPE = 90 %
Asselborn et al. [16] (2018)	298	6–10	BHK tasks	K, D, S, A	Classification: SEN = 96 %, SPE = 99 %
Mekyska et al. [17] (2019)	76	6–11	Drawings	K, D, S, T	Classification: SEN = 50 %, SPE = 90 %
Galaz et al. [13] (2020)	53	9–12	Drawings	K, D, S, A	Classification: SEN = 83 %, SPE = 81 %
Asselborn et al. [18] (2020)	448	5–12	BHK tasks	K, D, S, T	New data-driven based approaches for an assessment of GD (4 dimensions)
Garot et al. [19] (2020)	280	5–12	BHK tasks	K, D, S, T	Automatic discrimination among 3 groups of children with dysgraphia

[1] DB – database size; BHK – Concise Evaluation Scale for Children's Handwriting; D – dynamic handwriting features; K – kinematic handwriting features; S – spatial handwriting features; T – temporal handwriting features; A – advanced handwriting features; SEN – sensitivity; SPE – specificity; EER – estimation error rate; GD – graphomotor disabilities

writings, and several cognitive tests based on a protocol consisting of 31 tasks designed in cooperation with psychologists and special educational counselors. Every graphomotor task of the protocol has been evaluated by a well-experienced psychologist and rated on the scale from 0 to 4 where: 0 – no graphomotor difficulties; 1 – mild graphomotor difficulties; 2 – graphomotor difficulties; 3 – dysgraphia; 4 – severe dysgraphia. Finally, an overall score has been assigned to each child based on a complex analysis of all the 31 tasks in the protocol (i.e. including the cognitive tests). Although the protocol contains 7 graphomotor tasks such as Archimedean spiral, loops, sawtooth, or rainbow, in this study, we focused on one graphomotor task (combined loops), which has been proved to discriminate well between children with/without graphomotor difficulties [17].The distribution of scores (the overall and the sub-score for the combined loops task) is presented in Fig. 1. Correlation between the scores and the demographic data is visualized in Fig. 2. Parents of all children participating in this study signed an informed consent form approved by the Ethical Committee of the Masaryk University. Throughout the entire duration of this study, we strictly followed the Ethical Principles of Psychologists and Code of Conduct released by the American Psychological Association (https://www.apa.org/ethics/code/).

TABLE 2. Demographic data of the enrolled children.

	μ (σ)	min	Q1	Q2	Q3	max
all children (85 subjects)						
age [y]	9.79 (0.65)	8	9	10	10	11
grade	3.86 (0.35)	3	4	4	4	4
sub-score	1.46 (0.82)	0	1	1	2	3
overall score	1.75 (0.84)	0	1	2	2	4
girls (31 subjects)						
age [y]	9.77 (0.66)	8	9	10	10	11
grade	3.84 (0.37)	3	4	4	4	4
sub-score	1.16 (0.72)	0	1	1	2	2
overall score	1.35 (0.86)	0	1	1	2	3
boys (54 subjects)						
age [y]	9.80 (0.65)	8	9.25	10	10	11
grade	3.87 (0.34)	3	4	4	4	4
sub-score	1.63 (0.82)	0	1	1	2	3
overall score	1.98 (0.73)	1	1.25	2	2	4

[1] μ – mean; σ – standard deviation; Qx – x-th quartile; y – years.

TABLE 3. Grade distribution.

grade	girls	boys	together
3rd grade	5	7	12
4th grade	26	47	73

B. DATA ACQUISITION

At first, a template of the combined loop task was shown to a child and then he/she was asked to replicate it on an A4 paper that was laid down and fixed to a digitizing tablet. The drawing was acquired by the Wacom Intuos Pro L (PHT-80) digitizer with the sampling frequency of 150 Hz, and the Wacom Inking pen, which provides a feeling of writing by a regular pen and offers immediate visual feedback.

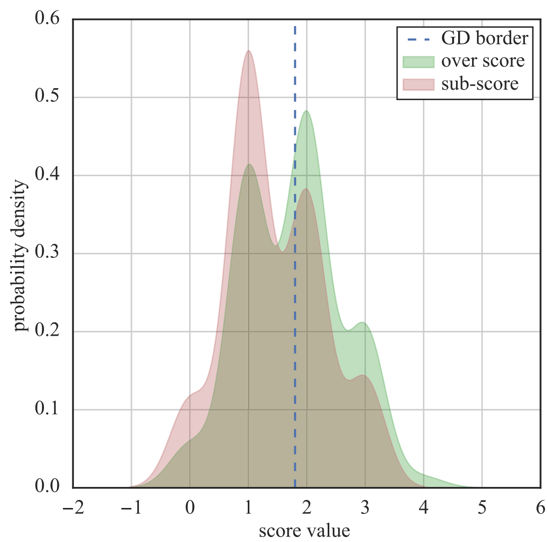


FIGURE 1. Distribution of the overall score and the sub-score. Blue dashed line represents imaginary threshold for the graphomotor difficulties (right of the line).

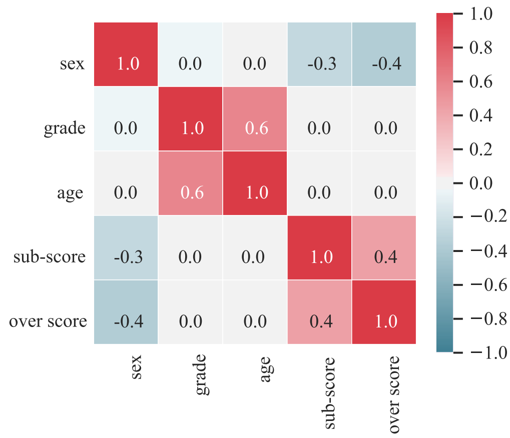


FIGURE 2. Correlation matrix between the scores and demographic data of the participants. A positive correlation is represented by red color and a negative correlation by blue color.

Moreover, this set-up enabled us to record a variety of signals describing the drawing process: x and y position ($x[n]$ and $y[n]$); timestamp ($t[n]$); a binary variable ($b[n]$; 0 – in-air movement, i. e. movement of the pen tip up to 1.5 cm above the tablet's surface, and 1 – on-surface movement, i. e. movement of the pen tip on the paper), pressure exerted on the tablet's surface during drawing ($p[n]$); pen tilt ($a[n]$); and azimuth ($az[n]$). For more information, see our previous works [12], [14], [28]. An example of the selected combined loop task performed by a child with/without GD can be seen in Fig. 3.

C. FRACTIONAL ORDER DERIVATIVES

The essential of this study is the investigation of the several (non-equivalent) FD approximations as a new advanced approach of drawing/handwriting parameterisation. We

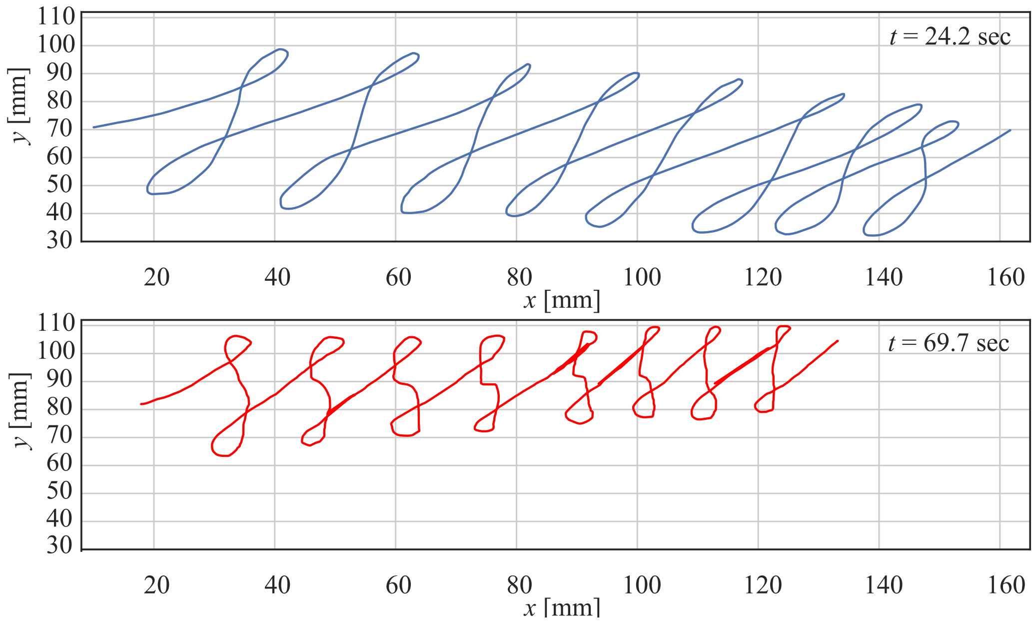


FIGURE 3. Example of the combined loop task performed by a child without graphomotor difficulties (upper part) and with graphomotor difficulties (bottom part). The thick parts of the red line represent the line-up places after the interruptions of the writing.

developed this method to substitute the conventional differential derivatives in the feature extraction process (see our previous works [9]–[12], [28]) in order to improve the quantitative analysis of the GD. In the scope of this study, we utilized three FD approximations: Grünwald-Letnikov (GL), Riemann–Liouville (RL), and Caputo (C), implemented by Valério Duarte in Matlab [29]–[31].

1) GRÜNWALD-LETNIKOV

The FD definition by Grünwald-Letnikov is one of the first and basic approaches [2]. A direct definition of the derivation of the function $y(t)$ by the order $\alpha - D^{\alpha}y(t)$ [1] is based on the finite differences of an equidistant grid in $[0, \tau]$, assuming that the function $y(t)$ satisfies certain smoothness conditions in every finite interval $(0, t), t \leq T$, where T denotes the period. Choosing the grid

$$0 = \tau_0 < \tau_1 < \ldots < \tau_{n+1} = t = (n+1)h, \tag{1}$$

with

$$\tau_{k+1} - \tau_k = h, \tag{2}$$

and using the notation of finite differences

$$\frac{1}{h^{\alpha}}\Delta_h^{\alpha}y(t) = \frac{1}{h^{\alpha}}\left(y(\tau_{n+1}) - \sum_{v=1}^{n+1} c_v^{\alpha} y(\tau_{n+1-v})\right), \tag{3}$$

where

$$c_v^{\alpha} = (-1)^{v-1}\binom{\alpha}{v}. \tag{4}$$

The Grünwald–Letnikov definition from 1867 is defined as

$${}^{GL}\mathrm{D}^{\alpha}y(t) = \lim_{h\to 0}\frac{1}{h^{\alpha}}\Delta_h^{\alpha}y(t), \tag{5}$$

where ${}^{GL}\mathrm{D}^{\alpha}y(t)$ denotes the Grünwald-Letnikov derivatives of order α of the function $y(t)$, and h represents the sampling lattice.

2) RIEMANN–LIOUVILLE

Another classical form of the FD has been given by Riemann-Liouville. The left-inverse interpretation of $D^{\alpha}y(t)$ by Riemann-Liouville [1], [3] from 1869 is defined as

$${}^{RL}\mathrm{D}^{\alpha}y(t) = \frac{1}{\Gamma(n-\alpha)}\left(\frac{d}{dt}\right)^n \int_0^t (t-\tau)^{n-\alpha-1}y(t)\,\mathrm{d}t, \tag{6}$$

where ${}^{RL}\mathrm{D}^{\alpha}y(t)$ denotes the Riemann–Liouville derivatives of order α of the function $y(t)$, Γ is the gamma function and $n-1 < \alpha \leq n, n \in \mathbf{N}, t > 0$.

3) CAPUTO

Nowadays, the most significant contributions to the field of FC are the results achieved by M. Caputo [32]. In contrast to the previous ones, the improvement hereabouts lie in the unnecessity to define the initial FD condition [1], [3]. The Caputo's definition from 1967 is

$${}^{C}\mathrm{D}^{\alpha}y(t) = \frac{1}{\Gamma(n-\alpha)}\int_0^t (t-\tau)^{n-\alpha-1}y^n(t)\,\mathrm{d}t, \tag{7}$$

where ${}^{C}\mathrm{D}^{\alpha}y(t)$ denotes the Caputo derivatives of order α of the function $y(t)$, Γ is the gamma function and $n-1 < \alpha \leq n, n \in \mathbf{N}, t > 0$.

D. HANDWRITING FEATURES

Altogether, we extracted 3 sets of handwriting features, one feature set per one employed FD approach. Basic kinematic

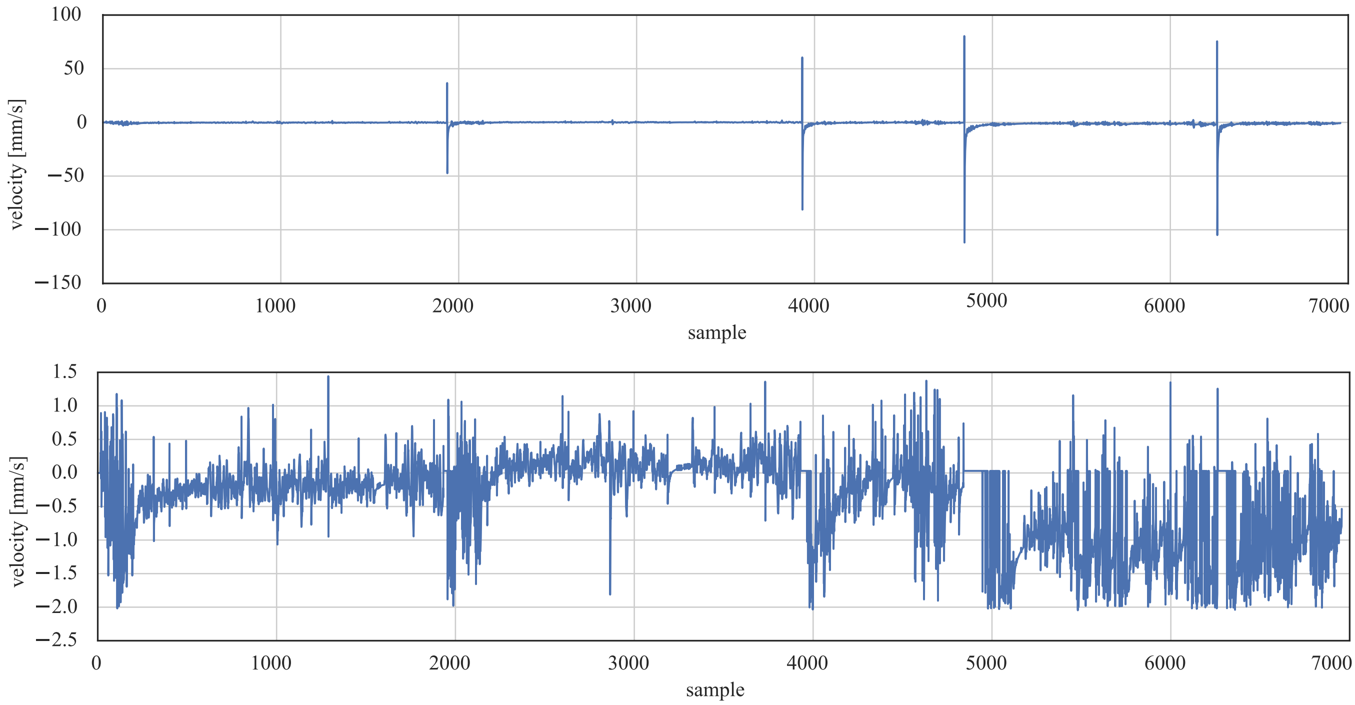


FIGURE 4. Illustration of the in-signal outlier removal, where the original handwritten signal before removing the outlier samples is placed in the upper part and after the outlier removal in the bottom part of the Figure. The velocity for $\alpha = 0.7$ computed by Caputo's approach from a sample of healthy children is used. The magnitude of the removed samples (peaks) is up to 100-times higher in comparison with the normal ones.

features from the input handwritten signal were extracted as well, namely velocity, acceleration, jerk and their horizontal and vertical variants. Due to rare omissions of 3–4 samples by the digitizing tablet during the acquisition, we performed the in-signal outliers removal (outliers were considered as elements more than three scaled median absolute deviations from the median). If not pre-processed, the differentiation of this gap would leave significant peaks in the output handwriting feature as illustrated in Figure 4. All handwriting features were computed for α in the range of 0.1–1.0 (with 0.1 step), where $\alpha = 1.0$ is equal to the full derivation. Finally, the statistical properties of all extracted handwriting features were described by the mean and the relative standard deviation (relstd). To sum up, each feature set consists of 180 computed kinematic features.

E. STATISTICAL ANALYSIS

At first, we performed the normality test of the handwriting features using the Shapiro-Wilk test [33]. In the case of non-normally distributed features, we utilised the Box-Cox transformation [34].

Next, to assess the strength of the relationship between the feature values and the scores (the overall score and the sub-score), Spearman's and Pearson's correlation coefficients were computed (we considered the level of significance 0.05). The p-values were adjusted using the False Discovery Rate (FDR) method to address the issue of multiple comparisons.

During the statistical analysis, we controlled for the effect of several confounding factors (covariates), namely age, grade, and sex.

Finally, to evaluate the power of the handwriting features to support the estimation of scores assessing the GD, we performed a multivariate analysis. For this purpose, we employed the state-of-the-art algorithm XGBoost [35] (10-fold cross-validation with 20 repetitions). The XGBoost algorithm was selected, because of its ability to achieve good performance on a small data set. Moreover, it is able to compete with the deep learning methods that are still not being used in the case of the small dataset as they require much larger data to be trained on [36]. Hyper-parameter space optimization was performed by a random search strategy with following parameter values:

- learning rate: [0.001, 0.01, 0.1, 0.2, 0.3];
- gamma: [0, 0.10, 0.15, 0.25, 0.5];
- maximum depth of a tree: [6, 8, 10, 12, 15];
- sub-sample ratio: [0.5, 0.6, 0.7, 0.8, 0.9, 1.0];
- sub-sample ratio of columns for level: [0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0];
- sub-sample ratio of columns for tree: [0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0];
- minimum child weight: [0.5, 1.0, 3.0, 5.0, 7.0, 10.0].

The model's performance was evaluated by the mean absolute error (MAE), the mean square error (MSE), the root mean square error (RMSE), and the estimation error rate (EER).

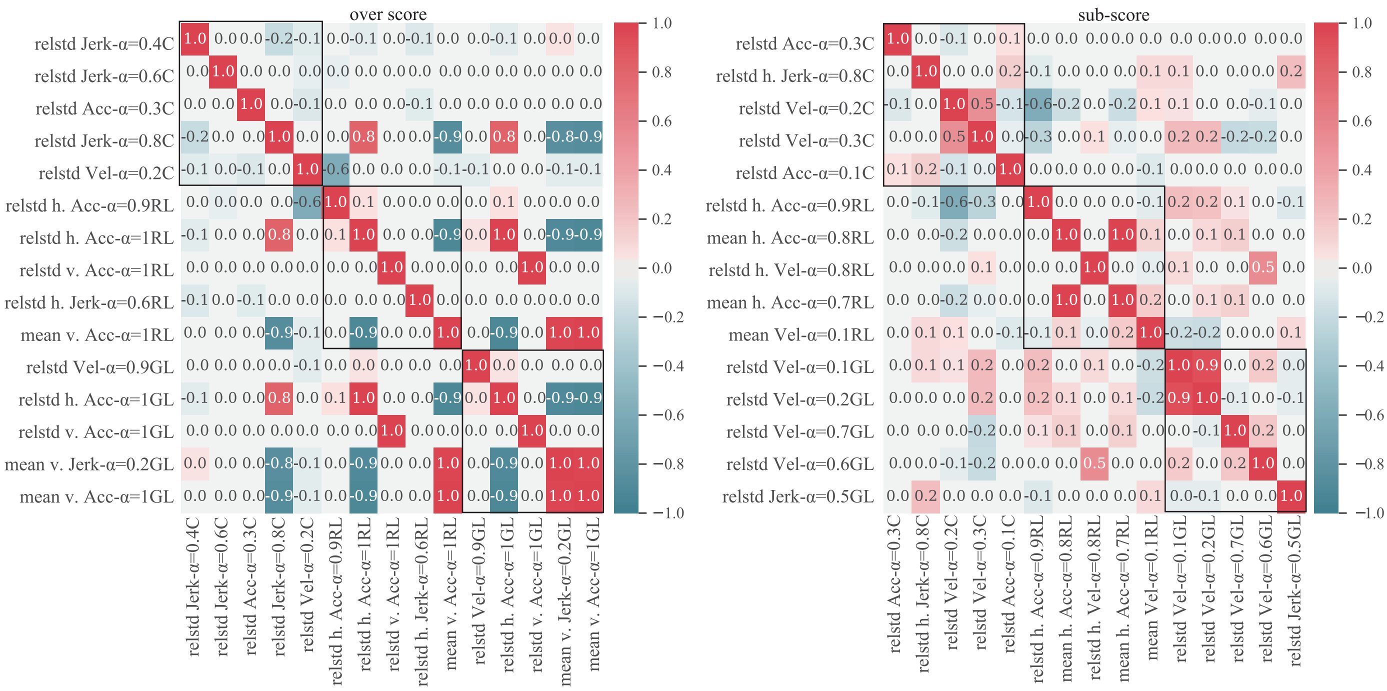


FIGURE 5. **Cross-correlation matrices of the most significant FD features as assessed by the Spearman's correlation (see Table 4).Framed sub-areas in each cross-correlation matrix visually isolates the handwriting features computed by the same FD approach.**

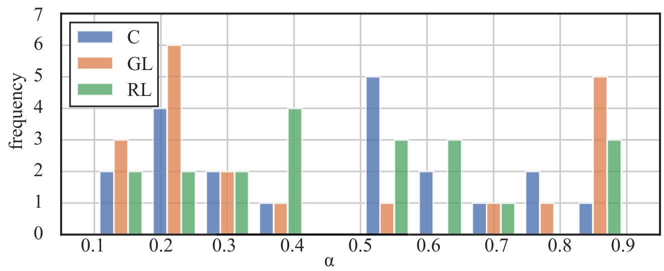


FIGURE 6. **Distribution of the FD order** α **for the features mostly correlating with the overall score, see Table 4 (GL – Grünwald-Letnikov; C – Caputo; RL – Riemann–Liouville).**

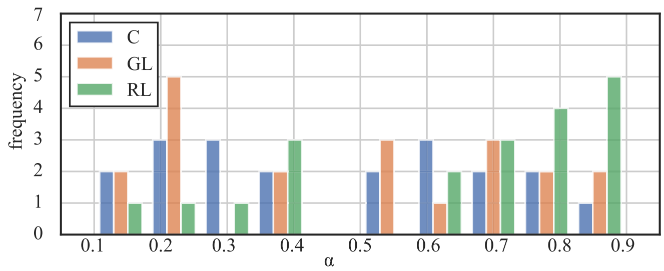


FIGURE 7. **Distribution of the FD order** α **for the features mostly correlating with the sub-score, see Table 4 (GL – Grünwald-Letnikov; C – Caputo; RL – Riemann–Liouville).**

III. RESULTS

The results of the correlation analysis can be seen in Table 4. The table shows the top 5 features per FD approximation according to p-values of the Spearman's correlation related to the overall score (upper part) and the sub-score (bottom part). The strongest correlation (after the FDR adjustment) with the overall score was identified in features extracted by the Caputo's FD. However, in the case of the sub-score, the Riemann-Liouville's FD arises as the most significant.

The correlation matrices (using the Spearman's correlation) are visualized in Fig. 5. Each matrix includes the top 5 features per FD approximation (i. e. 15 features in one matrix) identified in Table 4. The distribution of the FD order α of 20 best features regarding the Spearman's correlation per FD approximation is visualised in Fig. 6 for the overall score and in Fig. 7 for the sub-score.

Finally, the results of the multivariate analysis can be found in Table 5. In the case of the overall score estimation, the best results were achieved by the Riemann-Liouville FD. In the case of the sub-score estimation, the lowest error was achieved when combining features of all the approximations. Hyper-parameters of the best XGBoost models can be found in Table 6.

IV. DISCUSSION

The main goal of this study is to explore the differences across various FD approximations utilized in the analysis of the GD. A comparison of an identical feature (i. e. velocity for $\alpha = 0.2$) extracted from the handwritten product associated with the GD (the same sample as in the bottom part of Fig. 3) is shown in Fig. 8. It illustrates the differences across the involved FD approximations. The velocity function extracted by the Caputo's FD dominates by significant peaks in the positions, where a child interrupts the performance for a moment and then continues writing. These interruptions are also visible in the function computed by the Riemann-Liouville approach, though in the form of a constant line followed by elevated oscillations instead of peaks. On the

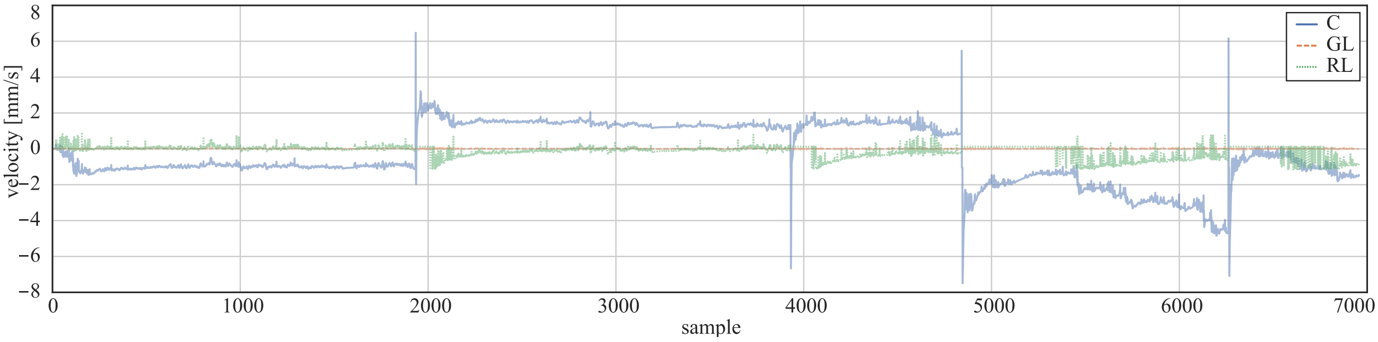


FIGURE 8. Comparison of the velocity function ($\alpha = 0.2$) across all the FD approximations (a child associated with graphomotor difficulties; C – Caputo; GL – Grünwald-Letnikov; RL – Riemann–Liouville).

TABLE 4. Results of the correlation analysis between the score values and computed handwriting features ranked by the adjusted p-value of Spearman's correlation.

	overall score					
feature name	ρ	p_s	p_s^*	r	p_p	p_p^*
	Caputo					
relstd jerk-α=0.4	**-0.3821**	**0.0003**	**0.0360**	**-0.1624**	**0.1377**	**0.5934**
relstd acceleration-α=0.3	-0.3649	0.0006	0.0360	-0.1204	0.2723	0.5934
relstd jerk-α=0.6	-0.3669	0.0006	0.0360	-0.1510	0.1678	0.5934
relstd jerk-α=0.8	0.3542	0.0009	0.0405	0.0702	0.5230	0.7298
relstd velocity-α=0.2	-0.3405	0.0014	0.0504	-0.1492	0.1729	0.5934
	Grünwald-Letnikov					
relstd velocity-α=0.9	-0.3435	0.0013	0.1106	0.0447	0.6843	0.9988
mean v. jerk-α=0.2	-0.3178	0.0030	0.1106	-0.0902	0.4118	0.9988
mean h. jerk-α=0.2	0.3113	0.0037	0.1106	0.2146	0.0486	0.6194
mean v. jerk-α=0.1	-0.3071	0.0043	0.1106	-0.0729	0.5071	0.9988
relstd v. jerk-α=0.1	-0.2811	0.0092	0.1998	-0.0180	0.8702	0.9988
	Riemann-Liouville					
relstd h. acceleration-α=0.9	0.3472	0.0011	0.0709	0.1304	0.2343	0.9521
relstd h. jerk-α=0.6	0.3154	0.0033	0.0709	0.0502	0.6481	0.9521
relstd velocity-α=0.4	-0.3144	0.0034	0.0709	0.0173	0.8753	0.9847
mean jerk-α=0.1	-0.3058	0.0044	0.0709	-0.0901	0.4122	0.9521
mean acceleration-α=0.6	-0.3047	0.0046	0.0709	-0.0880	0.4231	0.9521
	sub-score					
feature name	ρ	p_s	p_s^*	r	p_p	p_p^*
	Caputo					
relstd acceleration-α=0.3	-0.3353	0.0017	0.1560	-0.1294	0.2380	0.6211
relstd h. jerk-α=0.8	-0.3319	0.0019	0.1560	-0.1918	0.0787	0.4047
relstd velocity-α=0.2	-0.3230	0.0026	0.1560	-0.1888	0.0835	0.4175
relstd velocity-α=0.3	-0.2904	0.0070	0.2556	-0.0406	0.7121	0.8720
relstd acceleration-α=0.1	-0.2898	0.0071	0.2556	0.0003	0.9975	0.9975
	Grünwald-Letnikov					
relstd velocity-α=0.1	0.3475	0.0011	0.1980	0.3231	0.0026	0.2603
relstd velocity-α=0.2	0.3196	0.0029	0.1980	0.2784	0.0099	0.2603
relstd velocity-α=0.7	0.3157	0.0033	0.1980	0.2247	0.0387	0.2603
relstd velocity-α=0.6	0.2923	0.0066	0.2970	0.1150	0.2945	0.5049
relstd jerk-α=0.5	-0.2781	0.0100	0.3600	-0.0979	0.3729	0.5888
	Riemann-Liouville					
relstd h. acceleration-α=0.9	**0.4014**	**0.0001**	**0.0180**	**0.1548**	**0.1571**	**0.6672**
mean h. acceleration-α=0.8	0.3767	0.0004	0.0360	0.0833	0.4484	0.8302
relstd h. velocity-α=0.8	0.3649	0.0006	0.0360	0.0030	0.9786	0.9850
mean h. acceleration-α=0.7	0.3539	0.0009	0.0405	0.0678	0.5375	0.8346
mean h. acceleration-α=0.9	0.3394	0.0015	0.0411	0.0952	0.3859	0.8302

[1] ρ – Spearman's correlation coefficient; p_s – p-value of Spearman's correlation; p_s^* – adjusted p-value of Spearman's correlation; r – Pearson's correlation coefficient; p_p – p-value of Pearson's correlation; p_p^* – adjusted p-value of Pearson's correlation; relstd – relative standard deviation; h. – horizontal; v. – vertical.

other hand, the function based on the Grünwald-Letnikov approach seems to be a constant line, nevertheless after a scale normalization (min-max normalization), see Fig. 9, it is clear that the function has the oscillatory nature as well.

The differences across FD approaches are underlined by the comparison in Fig. 10, where the dependency of the relative standard deviation of the velocity on the FD order α is visualized. Feature values computed by the Grünwald-Letnikov approach are generally higher in comparison with the Caputo and Riemann-Liouville ones, which are more similar. On the other hand, the envelope of the velocity profile based on the Grünwald-Letnikov approach is more similar to the Riemann-Liouville one. Moreover, all functions meet at the point where $\alpha = 0.9$ and continue simultaneously to the full derivation ($\alpha = 1.0$), which is expected, because the full derivation has to be the same for all approaches.

Experts in the field of psychology need to understand and clearly interpret the results of the graphomotor analysis, i. e. to link them with specific symptoms or physiological processes. This is very challenging especially in the case of advanced signal parameterisation, which is also our case. Therefore, to bring credibility for a non-technical reader, we provide an illustration in Fig. 11. In this figure, we compare the vertical projection of the movement (y axis) and the vertical velocity (Grünwald-Letnikov approach, $\alpha = 0.8$) in a child without graphomotor difficulties (same as in the upper part of Fig. 3). The function extracted by FD for $\alpha = 0.8$ is difficult to be understood, but the relationship to the velocity is obvious.

Regarding the results of the correlation analysis (association with the overall score), the most significant features (after the FDR adjustment) are extracted by the Caputo's FD, where the top 5 have the p-value < 0.05. Most significant handwriting features are related to the variability of the jerk, which refers to the disturbances in the fluent handwriting performance of the child with GD. The values of the correlation coefficients are negative, which means that the handwriting performance of the subject is worse with the lower variability of the jerk. This may be confusing, because just the opposite effect may be expected. Nevertheless, this is specific for the combined loop task. A child without GD is less focused on the writing (the movement is more automatic), therefore the changes between loops are more dynamic, which results in higher jerk variability. Vice versa, a child with GD is more focused on his/her performance, therefore, the handwriting is associated with lower acceleration and jerk. In the case of Grünwald-Letnikov based features, 4 out of the 5 most significant ones are jerk

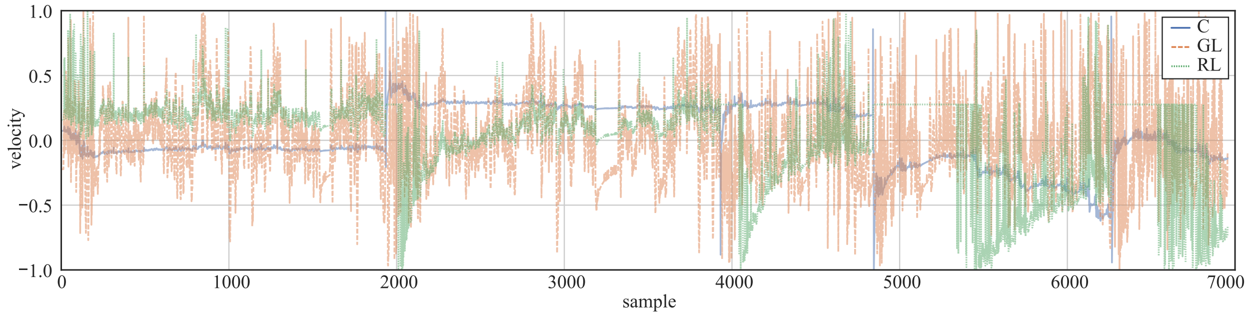


FIGURE 9. Comparison of the velocity function ($\alpha = 0.2$, normalized scale) across all the FD approximations (a child associated with graphomotor difficulties; C – Caputo; GL – Grünwald-Letnikov; RL – Riemann–Liouville).

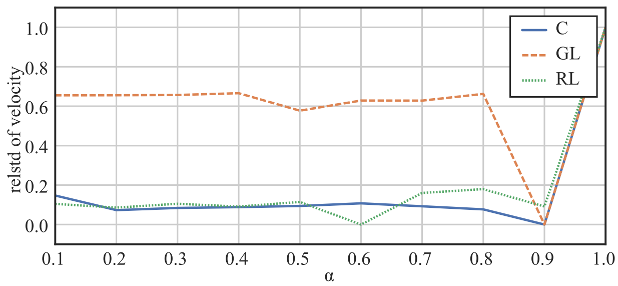


FIGURE 10. Relative standard deviation of velocity depending on FD order α (C – Caputo; GL – Grünwald-Letnikov; RL – Riemann–Liouville).

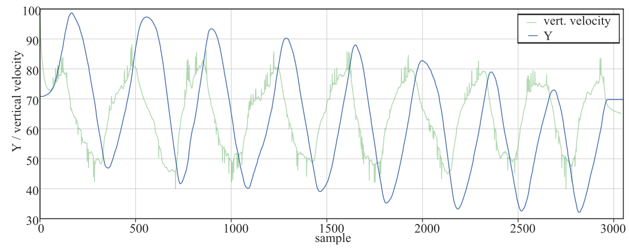


FIGURE 11. Comparison of the vertical projection of movement and the vertical velocity (Grünwald-Letnikov, $\alpha = 0.8$) in a child without graphomotor difficulties.

related too, what supports the results obtained by the Caputo's approach. In the view of the Riemann-Liouville FD, the most significant features are mostly acceleration and jerk related, this likewise supports the association with the smooth handwriting disabilities.

Considering the correlation with the sub-score, the most significant features (after the FDR adjustment) are extracted by the Riemann-Liouville FD, while 4 out of 5 features are acceleration-based. This again refers to the disruptions in continuous handwriting of a child with GD (i. e. less automatic and dynamic movements). In the case of the Grünwald-Letnikov approach, the variation of the velocity is observed to be the most significant, however, none of the features is significant after the p-value adjustment (similarly to the Caputo's approach). Due to the omission of the full derivations in best correlation results, the FD-based features outperform the conventional handwriting features in the scope of the sub-score correlation analysis for the connected loops task. In addition, this is in line with our previous results. [11], [12].

Regarding the cross-correlation of the top-ranked features strongly associated with the overall score (see the left matrix in Fig. 5), we did not observe any strong correlations among the features based on the Caputo's approach. In the case of the Riemann-Liouville's approximation, we identified a significant correlation between the mean of the vertical acceleration and the relstd of the horizontal acceleration, in both features $\alpha = 1$, which means full derivation. Similarly, in the Grünwald-Letnikov's approach, we identified a strong association between the relstd of the horizontal acceleration, and the mean vertical jerk and the mean vertical acceleration. The last two mentioned features are in fact very close to each other, because the acceleration with $\alpha = 1$ is very similar to the jerk with $\alpha = 0.2$ We assume that the above-mentioned association is linked with the fact that the vertical movement, contrary to the horizontal one, requires coordinated movement and finer flexions/extensions of more joints (interphalangeal and metacarpophalangeal) and therefore, it is more complex than ulnar abductions of the wrist [37], [38]. Since the vertical movement is complex, it is strongly affected by psychological and muscular fatigue [39], which could be manifested in lower vertical acceleration in children with GD. Nevertheless, low relstd in the horizontal direction could mean monotonous and less dynamic movement too.

In the case of the cross-correlation matrix linked with the sub-score, we can observe significant correlations only in features that express the same information, e. g. the mean of the horizontal acceleration, but differ only in α, e. g. the difference is 0.1. Since this difference is very low, it is obvious that these features significantly correlate. Except for this, the features do not correlate much among themselves which means that they are not redundant, but still relevant (see Table 4).

Based on the distribution of α in the 20 top-ranked features, we can observe that those based on the Caputo's approach are mostly concentrated around 0.2 and 0.5 for the overall score and almost evenly distributed in the case of sub-score correlation analysis. The Grünwald-Letnikov FD-based features associated with the overall score have α concentrated around 0.2 and 0.9. Those associated with the sub-score are mainly around 0.2, 0.5 and 0.7. Finally, in the case of the Riemann-Liouville's approach, we can observe a higher concentration in the range [0.4; 0.6] for the overall score, and in the range [0.7; 0.9] for the sub-score. Since the distribution of the α varies per FD approximation and rat-

TABLE 5. Results of the multivariate analysis.

overall score (range 4)					
APP	MAE	MSE	RMSE	EER [%]	N
GL	0.72 ± 0.17	0.76 ± 0.30	0.85 ± 0.18	17.93 ± 4.30	5
C	0.76 ± 0.21	0.95 ± 0.45	0.95 ± 0.24	19.01 ± 5.07	24
RL	$\mathbf{0.65 \pm 0.16}$	$\mathbf{0.66 \pm 0.28}$	$\mathbf{0.79 \pm 0.17}$	$\mathbf{16.25 \pm 3.96}$	**16**
ALL	0.68 ± 0.18	0.73 ± 0.33	0.83 ± 0.20	17.09 ± 4.50	17
sub-score (range 4)					
APP	MAE	MSE	RMSE	EER [%]	N
GL	0.68 ± 0.16	0.70 ± 0.29	0.82 ± 0.17	22.53 ± 5.23	18
C	0.65 ± 0.18	0.75 ± 0.33	0.85 ± 0.19	21.75 ± 6.10	15
RL	0.66 ± 0.15	0.66 ± 0.25	0.79 ± 0.16	22.03 ± 5.04	24
ALL	$\mathbf{0.64 \pm 0.15}$	$\mathbf{0.63 \pm 0.24}$	$\mathbf{0.78 \pm 0.16}$	$\mathbf{21.44 \pm 5.02}$	**17**

[1] APP – specific FD approximation; MAE – mean absolute error; MSE – mean squared error; RMSE – root mean squared error; EER – estimation error rate; N – number of selected features; GL – Grünwald-Letnikov; C – Caputo; RL – Riemann–Liouville; ALL (combination of all feature-types, i. e. 540 features).

TABLE 6. Hyper-parameters of the best XGBoost models.

hyper-parameter	overall score	sub-score
gamma	0.1	0.1
learning rate	0.1	0.1
maximum depth of a tree	15	8
minimum child weight	0.5	0.5
balance of positive and negative weights	1	1
sub-sample ratio	0.9	1
sub-sample ratio of columns for tree	0.9	0.5
sub-sample ratio of columns for level	0.9	0.4
number of estimators	500	500
seed	42	42

ing scale, we hypothesise that further and finer optimization of this parameter would bring even better quantification of the GD.

Concerning the multivariate analysis (Table 5), where we estimated the overall score, the best results were achieved by the Riemann-Liouville FD-based features. The resulting MAE was 0.65, and RMSE = 0.79. When estimating the sub-score, all approaches had a very similar MAE, nevertheless, the lowest RMSE (0.79) was reached by the Riemann-Liouville's approach too. A combination of all the approaches slightly decreased the error. These results suggest that the Riemann-Liouville's approximation in the field of quantitative GD analysis outperforms the other ones. In addition, using this approach we were able to estimate the scores with MAE = 0.65 and MAE = 0.66, respectively. If we take into account that the range of the first scale is 4, and of the second one 3, the error can be considered as very low. In fact, when assessing GD in children, psychologists tend to make the error even higher, e. g. two experts can frequently differ by 1 point (compare it to 0.65 or 0.66).

V. CONCLUSION

To the best of our knowledge, this is a unique study that performs an investigation of the various FD approaches in the computerized assessment of the GD in school-aged children. Therefore, it should be considered as being rather exploratory and pilot in nature. We can conclude that the employment of various FD approximations brings major differences in kinematic handwriting features. In the scope of the correlation analysis associated with the overall score, the Caputo's FD approach exceeds the rest of the analysed FD approximations. However, in the scope of the sub-score, the Riemann-Liouville gained the most significant features. Moreover, the results of the multivariate analysis suggest that the Riemann-Liouville's approximation in the field of the quantitative GD analysis outperforms the other ones (MAE = 0.65 for overall score and MAE = 0.66 for sub-score).

This study has several limitations and possible parts, that could be further improved. First of all, the dataset is relatively small in terms of the statistical validity of the results. To generalize the results, the larger dataset have to be acquired and more handwriting tasks should be included in the analysis. Next, a more granular FD α order search (step of 0.01 or even less) in order to find the optimal α range should be performed. Moreover, other feature types, such as temporal, spatial, and dynamic, should be included in future comparisons. The future study should be detailly focused on the comparison of the FD-based features with the conventionally used handwriting features. The different handwriting tasks have to be investigated separately for the best performing FD-based features. Besides, when comparing the several feature sets performance (regression, etc.) an ANOVA test should be performed in the future to analyze the differences between them. Finally, various machine learning models should be trained and compared in the future studies to get more information about the classification performance of the proposed features and to obtain the most robust models for GD identification.

ACKNOWLEDGMENT

For the research, infrastructure of the SIX Center was used.

REFERENCES

[1] I. Podlubny, *Fractional Differential Equations An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*. San Diego, CA, USA: Academic, 1999.

[2] M. Lazarević, "Further results on fractional order control of a mechatronic system," *Sci. Tech. Rev.*, vol. 63, no. 3, pp. 22–32, 2013.

[3] Y. Luchko and R. Gorenflo, "An operational method for solving fractional differential equations with the caputo derivatives," *Acta Mathematica Vietnamica*, vol. 24, no. 2, pp. 207–233, 1999.

[4] S. G. Samko, A. A. Kilbas, and O. Marichev, *Fractional integrals and derivatives, Theory and Applications*. Yverdon-les-Bains, Switzerland: Gordon Breach, 1993, p. 44.

[5] V. V. Uchaikin, *Fractional Derivatives for Physicists Engineers*, vol. 2. Berlin, Germany: Springer, 2013.

[6] S. Arshad, D. Baleanu, W. Bu, and Y. Tang, "Effects of HIV infection on CD4+ T-cell population based on a fractional-order model," *Adv. Difference Equ.*, vol. 2017, no. 1, p. 92, Mar. 2017, doi: 10.1186/s13662-017-1143-0.

[7] C. M. A. Pinto and J. A. Tenreiro Machado, "Fractional model for malaria transmission under control strategies," *Comput. Math. Appl.*, vol. 66, no. 5, pp. 908–916, Sep. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0898122112006785

[8] H. Sun, Y. Zhang, D. Baleanu, W. Chen, and Y. Chen, "A new collection of real world applications of fractional calculus in science and engineering," *Commun. Nonlinear Sci. Numer. Simulation*, vol. 64, pp. 213–231, Nov. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1007570418301308

[9] J. Mucha, V. Zvoncak, Z. Galaz, M. Faundez-Zanuy, J. Mekyska, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de-Ipina, "Fractional derivatives of online handwriting: A new approach of parkinsonic dysgraphia analysis," in *Proc. 41st Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2018, pp. 214–217.

[10] J. Mucha, J. Mekyska, M. Faundez-Zanuy, K. Lopez-De-Ipina, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Advanced Parkinson's disease dysgraphia analysis based on fractional derivatives of online handwriting," in *Proc. 10th Int. Congr. Ultra Mod. Telecommun. Control Syst. Workshops (ICUMT)*, Nov. 2018, pp. 1–6.

[11] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, K. Lopez-de-Ipina, V. Zvoncak, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Identification and monitoring of Parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting," *Appl. Sci.*, vol. 8, no. 12, p. 2566, Dec. 2018.

[12] J. Mucha, M. Faundez-Zanuy, J. Mekyska, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de-Ipina, "Analysis of Parkinson's disease dysgraphia based on optimized fractional order derivative features," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2019, pp. 1–5.

[13] Z. Galaz, J. Mucha, V. Zvoncak, J. Mekyska, Z. Smekal, K. Safarova, A. Ondrackova, T. Urbanek, J. M. Havigerova, J. Bednarova, and M. Faundez-Zanuy, "Advanced parametrization of graphomotor difficulties in school-aged children," *IEEE Access*, vol. 8, pp. 112883–112897, 2020.

[14] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and rating of developmental dysgraphia by handwriting analysis," *IEEE Trans. Human-Machine Syst.*, vol. 47, no. 2, pp. 235–248, Apr. 2017.

[15] S. Rosenblum and G. Dror, "Identifying developmental dysgraphia characteristics utilizing handwriting classification methods," *IEEE Trans. Human-Machine Syst.*, vol. 47, no. 2, pp. 293–298, Apr. 2017.

[16] T. Asselborn, T. Gargot, Ł. Kidzinski, W. Johal, D. Cohen, C. Jolly, and P. Dillenbourg, "Automated human-level diagnosis of dysgraphia using a consumer tablet," *NPJ Digit. Med.*, vol. 1, no. 1, p. 42, Dec. 2018.

[17] J. Mekyska, J. Bednarova, M. Faundez-Zanuy, Z. Galaz, K. Safarova, V. Zvoncak, J. Mucha, Z. Smekal, A. Ondrackova, T. Urbanek, and J. M. Havigerova, "Computerised assessment of graphomotor difficulties in a cohort of school-aged children," in *Proc. 11th Int. Congr. Ultra Mod. Telecommun. Control Syst. Workshops (ICUMT)*, Oct. 2019, pp. 1–6.

[18] T. Asselborn, M. Chapatte, and P. Dillenbourg, "Extending the spectrum of dysgraphia: A data driven strategy to estimate handwriting quality," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, Dec. 2020.

[19] T. Gargot, T. Asselborn, H. Pellerin, I. Zammouri, S. M. Anzalone, L. Casteran, W. Johal, P. Dillenbourg, D. Cohen, and C. Jolly, "Acquisition of handwriting in children with and without dysgraphia: A computational approach," *PLoS ONE*, vol. 15, no. 9, Sep. 2020, Art. no. e0237575.

[20] A. A. Alhusaini, G. R. Melam, and S. Buragadda, "Short-term sensorimotor-based intervention for handwriting performance in elementary school children," *Pediatrics Int.*, vol. 58, no. 11, pp. 1118–1123, Jul. 2016.

[21] K. McHale and S. A. Cermak, "Fine motor activities in elementary school: Preliminary findings and provisional implications for children with fine motor problems," *Amer. J. Occupational Therapy*, vol. 46, no. 10, pp. 898–903, Oct. 1992.

[22] S. Graham, K. R. Harris, and B. Fink, "Is handwriting causally related to learning to write? Treatment of handwriting problems in beginning writers," *J. Educ. Psychol.*, vol. 92, no. 4, pp. 620–633, 2000, doi: 10.1037/0022-0663.92.4.620.

[23] K. P. Feder and A. Majnemer, "Handwriting development, competency, and intervention," *Develop. Med. Child Neurol.*, vol. 49, no. 4, pp. 312–317, Apr. 2007.

[24] S. T. Peverly, J. K. Garner, and P. C. Vekaria, "Both handwriting speed and selective attention are important to lecture note-taking," *Reading Writing*, vol. 27, no. 1, pp. 1–30, Jan. 2014.

[25] B. C. M. Smits-Engelsman and G. P. Van Galen, "Dysgraphia in children: Lasting psychomotor deficiency or transient developmental delay?" *J. Exp. Child Psychol.*, vol. 67, no. 2, pp. 164–184, Nov. 1997.

[26] S. Rosenblum and L. Gafni-Lachter, "Handwriting proficiency screening questionnaire for children (HPSQ–C): Development, reliability, and validity," *Amer. J. Occupational Therapy*, vol. 69, no. 3, pp. 1–9, Apr. 2015.

[27] E. Pagliarini, L. Scocchia, M. Vernice, M. Zoppello, U. Balottin, S. Bouamama, M. T. Guasti, and N. Stucchi, "Children's first handwriting productions show a rhythmic structure," *Sci. Rep.*, vol. 7, no. 1, p. 5516, Dec. 2017.

[28] V. Zvoncak, J. Mucha, Z. Galaz, J. Mekyska, K. Safarova, M. Faundez-Zanuy, and Z. Smekal, "Fractional order derivatives evaluation in computerized assessment of handwriting difficulties in school-aged children," in *Proc. 11th Int. Congr. Ultra Mod. Telecommun. Control Syst. Workshops (ICUMT)*, Oct. 2019, pp. 1–6.

[29] D. Valério. *Variable Order Derivatives*. Accessed: Nov. 11, 2019. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/24444-variable-order-derivatives

[30] D. Valério and J. S. D. Costa, "Ninteger: A fractional control toolbox for MATLAB," in *Fractional Differentiation and Its Applications*. Bordeaux, France: Ubooks, 2004.

[31] D. Valério and J. S. D. Costa, Eds., *An Introduction to Fractional Control*. Stevenage, U.K.: IET, 2013.

[32] M. Caputo, "Linear models of dissipation whose Q is almost frequency independent-II," *Geophys. J. Int.*, vol. 13, no. 5, pp. 529–539, Nov. 1967, doi: 10.1111/j.1365-246X.1967.tb02303.x.

[33] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, nos. 3–4, pp. 591–611, Dec. 1965.

[34] G. E. Box and D. R. Cox, "An analysis of transformations," *J. Roy. Stat. Soc. B, Methodol.*, vol. 26, no. 2, pp. 211–243, 1964.

[35] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[36] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, Jul. 2018.

[37] N. Dounskaia, A. W. A. Van Gemmert, and G. E. Stelmach, "Interjoint coordination during handwriting-like movements," *Exp. Brain Res.*, vol. 135, no. 1, pp. 127–140, Oct. 2000.

[38] G. P. van Galen, "Handwriting: Issues for a psychomotor theory," *Hum. Movement Sci.*, vol. 10, nos. 2–3, pp. 165–191, May 1991.

[39] A. Kushki, H. Schwellnus, F. Ilyas, and T. Chau, "Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia," *Res. Develop. Disabilities*, vol. 32, no. 3, pp. 1058–1064, May 2011.

JAN MUCHA was born in Liptovsky Mikulas, Slovakia. He received the M.Sc. degree from telecommunications from the Brno University of Technology, Czech Republic, in 2016.

He is currently pursuing the Ph.D. degree (defense) with the Department of Telecommunications, Brno University of Technology. He is currently a Leading Researcher in graphonomics with the Brain Diseases Analysis Laboratory, Brno University of Technology, and where he is also a part of the Human–Machine Interaction Group, Signal Processing Laboratory. He has coauthored more than ten publications indexed by the Web of Science. He deals with the research of non-invasive neurodegenerative and neurodevelopmental disorders analysis based on online handwriting processing, with an emphasis on advanced parametrization techniques of developmental dysgraphia and Parkinson's disease dysgraphia.

JIRI MEKYSKA was born in Holesov, Czech Republic. He received the Ph.D. degree in telecommunications from the Brno University of Technology, Czech Republic, in 2014.

He is currently the Head of the Brain Diseases Analysis Laboratory, Brno University of Technology, where he is also the Head of the Human–Machine Interaction Group, Signal Processing Laboratory, Department of Telecommunications. In cooperation with neurologists/psychologists from different countries, he develops diagnostic/monitoring systems focused on Lewy body diseases and developmental dysgraphia. He has coauthored more than 30 publications indexed by the Web of Science. He deals with the research of non-invasive neurodegenerative and neurodevelopmental disorders analysis based on speech, sleep, and handwriting processing.

Dr. Mekyska is a member of the Cybernetics and Information Processing Expert Panel of the Czech Science Foundation and a member of the Expert Advisory Body of Evaluators of the Development and Innovation Council of the Czech government. In 2013, he received the Joseph Fourier Award for his work in the field of paraclinical neurodegenerative diseases analysis. He serves as a reviewer for the Europe (Innovative Medicines Initiative) and national (Czech Science Foundation, Czech Ministry of Health) projects. He serves as a Reviewer for many international journals, including IEEE JSTSP, IEEE THMS, and *Parkinsonism & Related Disorders*.

ZOLTAN GALAZ was born in Galanta, Slovakia. He received the Ph.D. degree in telecommunications from the Brno University of Technology, Czech Republic, in 2018.

He is currently the Team Leader of the Brain Diseases Analysis Laboratory, Department of Telecommunications, Brno University of Technology. He has coauthored more than ten publications indexed by the Web of Science. In cooperation with neurologists/psychologists from different countries, he develops diagnostic/monitoring systems focused on disorders, such as Parkinson's disease and developmental dysgraphia. He deals with the research of advanced digital signal processing and machine learning algorithms applied in the field of computerized non-invasive neurodegenerative and neurodevelopmental disorders diagnosis, analysis, and monitoring.

MARCOS FAUNDEZ-ZANUY was born in Barcelona, Spain. He received the Ph.D. degree from the Polytechnic University of Catalonia, in 1998. Since 2010, he has been the Head of research with Tecnocampus. From 2009 to 2018, he was the Dean of the Escola Superior Politecnica (ESUP) del Tecnocampus, Mataró (Polytechnic University of Catalonia till 2015 and Pompeu Fabra University till 2018), where he is currently a Full Professor and also the Head of the Signal Processing Group.

His research interests include in the fields of biometrics applied to security and health. He was the Initiator and the Chairman of the European COST action 277 *Nonlinear speech processing*, and the Secretary of COST action 2102 *Cross-Modal Analysis of Verbal and Non-Verbal Communication*. He has authored more than 50 articles indexed in the ISI Journal citation report, more than 100 conference papers, around ten books, and responsible of ten national and European research projects. He is a Spanish liaison of EURASIP.

VOJTECH ZVONCAK was born in Zlin, Czech Republic. He received the M.Sc. degree in telecommunications from the Brno University of Technology, Czech Republic, in 2016.

His is currently pursuing the Ph.D. degree (defense) with the Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology. He is currently a Researcher with the Brain Diseases Analysis Laboratory, Brno University of Technology. He deals with handwriting and speech signal processing. He currently focuses on the diagnosis of developmental dysgraphia based on quantitative analysis of online handwriting.

KATARINA SAFAROVA was born in Sala, Slovakia. She received the Ph.D. degree in general psychology from the Department of Psychology, Masaryk University, Brno, Czech Republic, in 2018.

Since 2018, she has been a Researcher with the Department of Psychology, Masaryk University, and a member of a team dealing with the topic of graphomotor issues, especially dysgraphia and its diagnosis. Since 2018, she has been as an Assistant with the Institute of Social Sciences of the Brno University of Technology. In 2017, she was a Researcher with the Department of Psychology, University of Essex, U.K., for two months. Her research interests include educational psychology, learning disabilities, diagnostic methods, and its psychometric properties and methodology.

TOMAS URBANEK was born in Hustopece, Czech Republic. He received the Ph.D. degree in psychology from the Faculty of Arts, Masaryk University, Brno, Czech Republic, in 1999.

He is currently a Professor of general psychology with the Department of Psychology, Faculty of Arts, Masaryk University, Brno. Since 1995, he has been with the Institute of Psychology, Czech Academy of Sciences, as a Researcher and a Senior Researcher, where he has been the Director since 2017. Since 1999, he has been teaching at the Faculty of Arts, Masaryk University, Brno. He has authored or coauthored several books and about 70 articles and several psychological tests.

JANA MARIE HAVIGEROVA was born in South Moravia, Czech Republic. She received the M.Sc. and Ph.D. degrees in 1997 and 2004, respectively. She has been an Associate Professor in psychology with the Faculty of Arts, Masaryk University, Brno, Czech Republic, since 2017.

She has been with the University of Hradec Kralove (part-time) since 2005. She has been cooperating on a long-term basis with the University of South Bohemia, Ceske Budejovice, on the research of the relationship between personality and language (the CPACT project). She is currently managing the DAQUAo project focused on the study of psycholinguistic phenomena of therapeutic diaries. Within the Dysgraphy project, she participates in data collection and partial analysis, focusing on pre-school children.

JIRINA BEDNAROVA was born in Jindrichuv Hradec, Czech Republic. She received the degree in psychology from the Faculty of Arts, Masaryk University, Brno, Czech Republic.

Since 1995, she has been a Special Educational Counselor. She deals with the diagnosis and intervention in children with inequalities of cognitive and graphomotor development, and specific learning disabilities. She also deals with the education of adults working with such children. She has also authored several diagnostic materials (standardized and dynamic diagnosis) and a number of publications aimed at targeted support of children's development.

ZDENEK SMEKAL (Senior Member, IEEE) was born in Brno, Czech Republic. He received the M.Sc. and Ph.D. degrees in 1973 and 1978, respectively. He is currently a full-time Professor with the Brno University of Technology. He has been an Associate Professor since 1986 in the field of electronics and communication engineering and a Professor since 2000 in the field of theoretical electrical engineering.

He is currently the Chairman of the board for doctoral study programme of teleinformatics and the a Leading Researcher with the Signal Processing Laboratory, Department of Telecommunications, University of Technology. He has coauthored more than 90 publications indexed by the Web of Science and authored five monographs on digital signal processing.

He is a member of Scientific Council with the Brno University of Technology. He was involved in European COST research projects, including OC 10 G6 Digital Audio Effects from 1997 to 2001, OC 277 Nonlinear Signal Processing from 2004 to 2006, and OC 2102 Cross-Modal Analysis of Verbal and Nonverbal Communication from 2006 to 2010. He is also an Editor of *Cognitive Computation*. He serves as a Reviewer of many international journals, including *Digital Signal Processing*, *Neural Networks*, and *Patter Recognition*.

A.11 Identification and Monitoring of Parkinson's Disease Dysgraphia Based on Fractional-Order Derivatives of Online Handwriting

applied sciences

MDPI

Article

Identification and Monitoring of Parkinson's Disease Dysgraphia Based on Fractional-Order Derivatives of Online Handwriting †

Jan Mucha [1], Jiri Mekyska [1], Zoltan Galaz [1,2], Marcos Faundez-Zanuy [3], Karmele Lopez-de-Ipina [4], Vojtech Zvoncak [1], Tomas Kiska [1], Zdenek Smekal [1], Lubos Brabenec [2] and Irena Rektorova [2,5,*]

1. Department of Telecommunications and SIX Research Centre, Brno University of Technology, Technicka 10, 61600 Brno, Czech Republic; muchajano@phd.feec.vutbr.cz (J.M.); mekyska@feec.vutbr.cz (J.M.) z.galaz@phd.feec.vutbr.cz (Z.G.); vojtech.zvoncak@phd.feec.vutbr.cz (V.Z.); kiskatomas@phd.feec.vutbr.cz (T.K.); smekal@feec.vutbr.cz (Z.S.)
2. Applied Neuroscience Research Group, Central European Institute of Technology, Masaryk University, Kamenice 5, 62500 Brno, Czech Republic; lubos.brabenec@ceitec.muni.cz
3. Escola Superior Politecnica, Tecnocampus Avda. Ernest Lluch 32, 08302 Mataro, Barcelona, Spain; faundez@tecnocampus.cat
4. Department of Systems Engineering and Automation, University of the Basque Country UPV/EHU, Av de Tolosa 54, 20018 Donostia, Spain; karmele.ipina@ehu.eus
5. First Department of Neurology, Masaryk University and St. Anne's University Hospital, Pekarska 53, 65691 Brno, Czech Republic

* Correspondence: irena.rektorova@fnusa.cz; Tel.: +420-543-182-639

† This paper is an extended version of our paper published in Mucha, J.; Zvoncak, V.; Galaz, Z.; Faundez-Zanuy, M.; Mekyska, J.; Kiska, T.; Smekal, Z.; Brabenec, L.; Rektorova, I.; Ipina, K.L. Fractional Derivatives of Online Handwriting: a New Approach of Parkinsonic Dysgraphia Analysis. In Proceedings of the 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, Greece, 4–6 July 2018; pp. 1–4.

Received: 30 October 2018; Accepted: 6 December 2018; Published: 11 December 2018

Abstract: Parkinson's disease dysgraphia affects the majority of Parkinson's disease (PD) patients and is the result of handwriting abnormalities mainly caused by motor dysfunctions. Several effective approaches to quantitative PD dysgraphia analysis, such as online handwriting processing, have been utilized. In this study, we aim to deeply explore the impact of advanced online handwriting parameterization based on fractional-order derivatives (FD) on the PD dysgraphia diagnosis and its monitoring. For this purpose, we used 33 PD patients and 36 healthy controls from the PaHaW (PD handwriting database). Partial correlation analysis (Spearman's and Pearson's) was performed to investigate the relationship between the newly designed features and patients' clinical data. Next, the discrimination power of the FD features was evaluated by a binary classification analysis. Finally, regression models were trained to explore the new features' ability to assess the progress and severity of PD. These results were compared to a baseline, which is based on conventional online handwriting features. In comparison with the conventional parameters, the FD handwriting features correlated more significantly with the patients' clinical characteristics and provided a more accurate assessment of PD severity (error around 12%). On the other hand, the highest classification accuracy (ACC = 97.14%) was obtained by the conventional parameters. The results of this study suggest that utilization of FD in combination with properly selected tasks (continuous and/or repetitive, such as the Archimedean spiral) could improve computerized PD severity assessment.

Keywords: Parkinson's disease dysgraphia; micrographia; online handwriting; kinematic analysis; fractional-order derivative; fractional calculus

1. Introduction

As a second most common neurodegenerative disorder, Parkinson's disease (PD) is expected to impose an increasing social and economic burden on societies as populations age [1]. Its prevalence rate is estimated to approximately 1.5% for people aged over 65 years [2]. The risk of being affected by PD strongly increases with age, and, in the next 15 years, the incidence of PD is expected to be doubled [3,4]. The rapid degeneration of dopaminergic cells in the substantia nigra pars compacta [5] arose as the most significant biological finding associated with the disease, but the exact pathophysiological cause of PD has not yet been discovered. PD cardinal motor symptoms involve bradykinesia (slowness of movement), tremor at rest, rigidity, gait impairment, and postural instability [6–8]. A variety of non-motor symptoms may emerge as well—for instance, cognitive impairment, dementia, depression, sleep disorders, or anxiety [6,9,10].

Handwriting requires cognitive, perceptual, and fine motor abilities. In conjunction with motor dysfunctions in people suffering from PD, it has been proven that disrupted handwriting may be used as a significant biomarker for PD diagnosis [11,12]. Micrographia, which is associated with the progressive decrease in letters' amplitude, is the most commonly observed handwriting abnormality in patients with PD [13,14]. Moreover, according to McLennan et al. [14], in approximately 5% of PD patients, micrographia may be observed even before the onset of the cardinal motor symptoms.

The recent advantage of new technologies coming hand-in-hand with Health 4.0 systems enables the acquisition of online handwriting signals, where temporal information is added to the x and y position. Therefore, by using a digitizing tablet, the analysis is not limited to spatial features which mainly quantify PD micrographia. In addition, we are able to quantify temporal, kinematic, and dynamic manifestations of PD dysgraphia, such as hesitations, pauses, and slow movement [7], which cannot be studied objectively using a classical paper-and-pen method. Due to this complexity, Letanneux et al. [15] started to refer to these manifestations using the generalized term PD dysgraphia.

Several research teams have explored the impact of quantitative PD dysgraphia analysis utilizing simple handwriting/drawing tasks (e.g., separate characters, a combination of two or three characters, repetitive loops, circles), as well as more complex ones (e.g., words, sentences, figures, 3D objects, and the Archimedean spiral) [8,16–20]. An overview of recent related works (2015–present) can be seen in Table 1. Most of them confirm the irreplaceability of kinematic features in PD dysgraphia analysis. Additionally, the researchers usually employ temporal, spatial, and dynamic features. Some more advanced parameters are reported too. For instance, Drotar et al. [8,16,17] demonstrated a combination of kinematic, pressure, energy, or empirical mode decomposition (EMD)-based features that resulted in a classification accuracy of up to 89% using several handwriting tasks. Kotsavasilogloua et al. [21] achieved an average prediction accuracy of 91% using simple horizontal lines and features describing the variability in the pen tip's velocity, a deviation from the horizontal plane, and the trajectory's entropy. Other works report even higher classification accuracies (approximately 97%), e.g., Loconsole et al. [18], who used computer vision and electromyography signal processing techniques, or Taleb et al. [22], who used a combination of features related to the correlation between kinematic and pressure characteristics (but, in this case, applied to a very small dataset). Another promising approach was published by Moetesum et al. [23], who reached an 83% classification accuracy by employing convolutional neural networks (CNN) that were used to extract discriminating visual features from handwriting data transformed into the offline mode. In 2018, Impedovo et al. reported the results of a study focused only on the early stages of PD; the best accuracy was 74.76% for a combination of three handwriting tasks. Finally, in our previous work [20], we proposed a new approach of advanced kinematic feature extraction that utilizes fractional-order derivatives (FD). This approach increased the classification accuracy by 10% (72.39%) for Archimedean spiral tasks in comparison with the baseline [20].

Table 1. Overview of related works focused on computerized analysis of Parkinson's disease (PD) dysgraphia.

First Author	Year	PD/HC	Handwriting Task	Analysis	Features	Conclusions
Drotar * [17]	2015	37/38	letters, words, sentences	differential analysis (SVM)	kinematic, temporal, spatial, entropy, EMD, signal energy	The highest classification accuracy after feature selection approach was 88.13%.
Drotar * [16]	2015	37/38	letters, words, sentences	differential analysis (SVM)	kinematic, temporal, spatial, entropy, EMD, pressure	Classification performance was at its peak with on-surface features equal to AUC = 89.09%.
Heremans [24]	2015	34/10	up/down strokes at varying amplitudes	ANOVA	spatial and kinematic	Significant difference between groups was in spatial ($F(2.41) = 3.97$; $p = 0.03$).
Pereira [25]	2015	37/18	Archimedean spiral	differential an. (SVM, NB, OPF)	mean relative tremor and spatial parameters	The best results were obtained by NB classifier that provided around 79% classification accuracy.
Drotar * [8]	2016	37/38	letters, words, Archimedean spiral, sentences	differential an. (SVM, K-NN, ADA)	kinematic, temporal, spatial, entropy, EMD, pressure	Combining all exercises, SVM proved to be the best classifier with 82.5% accuracy.
Heremans [26]	2016	30/15	repetitive cursive loops	ANOVA, correlation an.	writing amplitude and velocity	PD dysgraphia is more severe in patients with freezing of gait.
Pereira [27]	2016	14/21	Archimedean spiral, meander	differential an. (CNN, OPF)	pen-based features	The best result was obtained by CNN with 87.14% classification accuracy using meander task.
Kotsavasil [21]	2017	24/20	horizontal lines	differential analysis (NB)	kinematic	Average classification accuracy was 91%.
Loconsole [18]	2017	4/7	sentence, repetitive loops	differential analysis (ANN)	temporal, kinematic, spatial	Highest classification accuracy (96.81%) was achieved using all the extracted features.
Taleb [22]	2017	16/16	letters, waves, words	differential analysis (SVM)	kinematic, stroke, pressure, entropy, energy, EMD	The highest classification accuracy was 96.88% for 12 kinematic and pressure features.
Moetesum * [23]	2018	37/38	Archimedean spiral, letters, words, sentence, loops	differential analysis (SVM)	CNN-based features	Extraction of features using CNN applied on raw handwriting data resulted in 83% classification accuracy.
Mucha * [20]	2018	30/36	Archimedean spiral	differential analysis (RF, SVM)	fractional derivatives based kinematic features	Improvement of classification accuracy by 10% (72.38%) in comparison to the baseline.
Impedovo * [28]	2018	37/38	Archimedean spiral letters, words, sentence	differential an. (RF, SVM, K-NN, NB, LDA, ADA)	kinematic, temporal, spatial, entropy, EMD, pressure	Analysis focused on PD diagnosis at earlier stages resulted in 74.76% classification accuracy.

SVM—support vector machine; EMD—empirical mode decomposition; K-NN—K-nearest neighbors; ANOVA—analysis of variance; NB—naïve Bayes classifier; OPF—optimum path forest; ANN—artificial neural network; CNN—convolutional neural network; RF—random forests; LDA—Linear Discriminant Analysis; ADA—AdaBoost; AUC—area under the receiver operating characteristics (ROC) curve; articles are sorted by the year of release and then alphabetically; * analyzes performed on the same database (Parkinson's disease handwriting database (PaHaW) [8]).

Although the authors of the previously mentioned studies reported high classification accuracies, further signal processing and machine learning pipeline improvements are expected to make the differential analysis even more accurate. One possible approach could involve an advanced feature extraction methodology based on fractional calculus (FC) [29,30], which enables the use of an arbitrary order of derivatives and/or integrals. Generally, FC has many applications in different fields of science [31–33]. For instance, it has been advantageously used during the modeling of different diseases, such as human immunodeficiency virus (HIV) [34] and malaria [35]. In addition, FC-based analytical tools have outperformed classical techniques in geology [36,37], economics and finance [38,39], etc. Moreover, in our recent paper [20], we identified a high potential for the use of FC in the kinematic analysis of PD drawings. Based on these preliminary results, we assume that FD-based handwriting features may bring improvements to PD diagnosis and assessment. In the frame of this article, we would like to go further and deeply explore the impact of FD on the PD dysgraphia diagnosis and its monitoring. More specifically, we aim to:

- investigate the relationship between newly designed FD handwriting features and a patient's clinical data and compare these results with a baseline (i.e., results based on conventional parameters),
- evaluate the discrimination power of the FD features in terms of binary classification accuracy and compare the results to the baseline,
- use the newly designed features to establish regression models that will estimate the severity of PD and compare its performance to that of a baseline.

The rest of this paper is organized as follows: Section 2 describes the cohort of patients and the methodology, and Section 3 includes the results. A discussion is presented in Section 4, and, finally, conclusions are drawn in Section 5.

2. Materials and Methods

2.1. Dataset

For the purpose of this work, the Parkinson's disease handwriting database (PaHaW) [8], which consists of multiple handwriting/drawing samples from 37 PD patients and 38 age- and gender-matched healthy controls (HC), was used. Since the Archimedean spiral drawing task is missing for some participants, we reduced the analyzed cohort to 33 PD patients and 36 HC. Demographic and clinical data of the participants can be found in Table 2. The participants were enrolled at the First Department of Neurology, St. Anne's University Hospital in Brno, Czech Republic. All participants reported the Czech language as their native language and were right-handed. The patients completed their tasks approximately 1 h after their regular dopaminergic medication (L-dopa). All participants signed an informed consent form approved by the local ethics committee. Unified Parkinson's disease rating scale, part V (UPDRS V): Modified Hoehn and Yahr staging score [40], was used to assess clinical symptoms of PD. In the frame of this work, the duration of the disease was considered as well. Descriptive visualization (histograms, regression, and residual plots) of the clinical data for the subjects participating in this study can be seen in Figure 1.

Table 2. Demographic and clinical data of the enrolled participants.

Gender	N	Age [years]	PD dur [years]	UPDRS V	LED [mg/day]
Parkinson's disease patients					
Females	17	71.76 ± 10.93	9.88 ± 5.27	2.18 ± 0.86	1146.03 ± 543.89
Males	16	66.50 ± 13.44	7.44 ± 4.04	2.31 ± 0.75	1673.38 ± 616.66
All	33	69.21 ± 11.10	8.70 ± 4.82	2.24 ± 0.80	1401.72 ± 630.71
Healthy controls					
Females	17	61.59 ± 10.17	-	-	-
Males	19	63.32 ± 13.14	-	-	-
All	36	62.50 ± 11.70	-	-	-

PD—Parkinson's disease; N—number of subjects; PD dur—PD duration; UPDRS V—Unified Parkinson's disease rating scale, part V: Modified Hoehn and Yahr staging score [40]; LED—L-dopa equivalent daily dose [41].

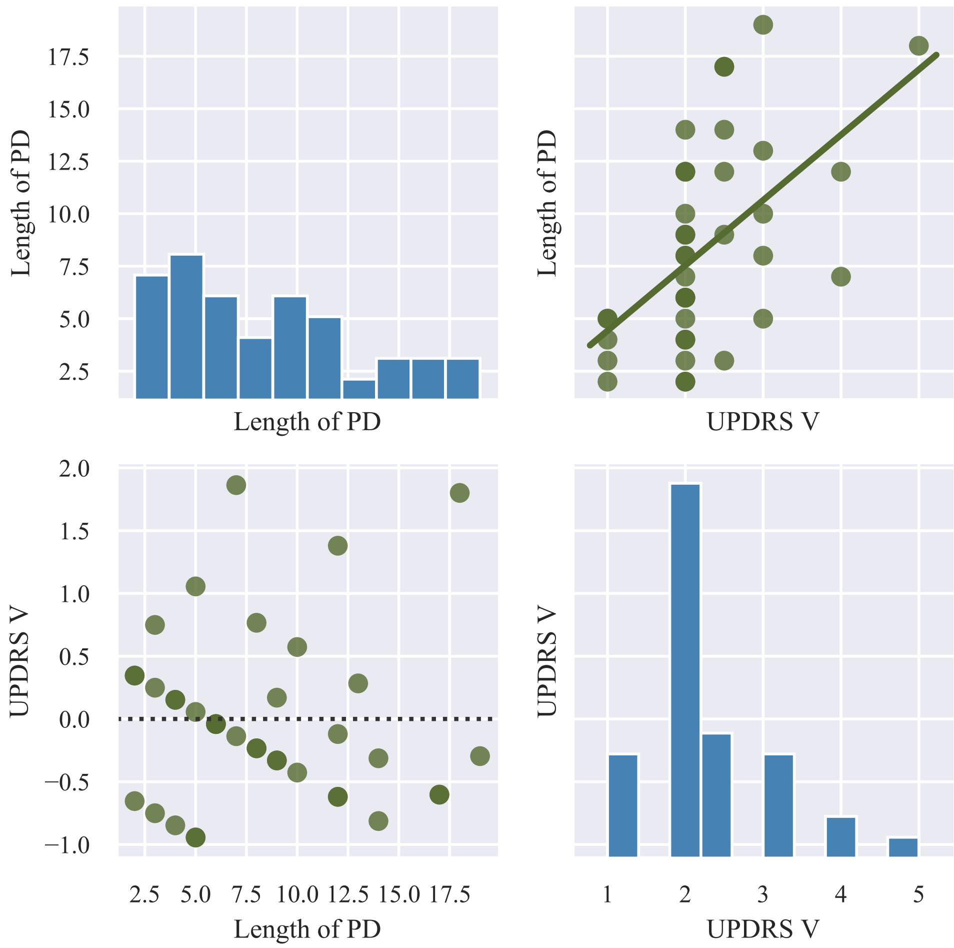


Figure 1. Descriptive graphs of patients' clinical characteristics: Unified Parkinson's disease rating scale (UPDRS V) and Parkinson's disease (PD) duration (in years). Histograms are visualized on the diagonal. A scatterplot with a line fitted using linear regression is visualized in the top-right corner. Residuals of the trained linear model are visualized in the bottom-left corner.

2.2. Data Acquisition

The PaHaW database [8] includes nine different handwriting tasks written in the Czech language. Their description and translation to English can be found in Table 3. During all handwriting tasks, the participants were rested and seated in a comfortable position with the possibility to look at the prefilled template (see Figure 2). A digitizing tablet (Wacom Intuos 4M, Wacom, Kazo, Saitama, Japan) was overlaid with an empty paper template and participants were asked to perform all tasks using a special Wacom inking pen that gave the patients immediate visual feedback. Online handwriting signals were recorded with a sampling frequency of $f_s = 150$ Hz. The following time sequences were acquired: x and y coordinates ($x[t]$, $y[t]$); time-stamp (t); in-air/on-surface (on-surface movement is a movement of a pen when its tip is touching the surface, e.g., paper (i.e., it provides the information about the pen writing/drawing on the paper); vice versa, in-air movement is a movement of a pen when its tip is up to 1.5 cm above the surface [42,43]) status ($b[t]$); pressure ($p[t]$); azimuth ($az[t]$); and altitude ($al[t]$).

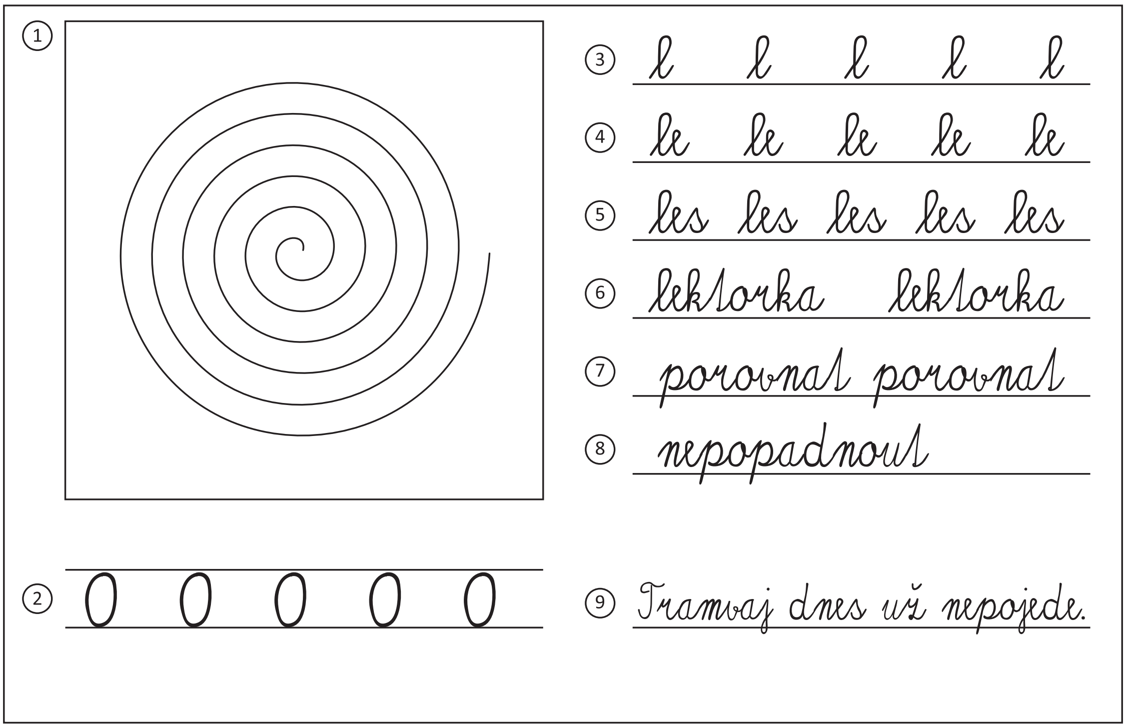


Figure 2. Filled template of the PaHaW database.

Table 3. Description of the PaHaW handwriting tasks.

N	Task	Czech (Original)	English (Translation)
1	Archimedean spiral	-	-
2	repetitive loops	-	-
3	letter	l	l
4	syllable	le	le
5	word	les	forest
6	word	lektorka	lecturer
7	word	porovnat	compare
8	word	nepopadnout	not grasped
9	sentence	Tramvaj dnes už nepojede.	The tram will no longer go today.

2.3. Feature Extraction

The main goal of this work is to compare a set of commonly used kinematic features with newly proposed FD-based features in terms of quantitative PD dysgraphia analysis. All of the handwriting features were computed using both on-surface as well as in-air movements. The two movements were quantified separately using *velocity* (rate at which the position of the pen changes with time [mm/s]), *acceleration* (rate at which the velocity of the pen changes with time [mm/s^2]), *jerk* (rate at which the acceleration of the pen changes with time [mm/s^3]), and their horizontal and vertical variants [8,44,45]. FD-based features were extracted for different values of α. In the frame of this work, α ranging from 0.1

to 1.0 with a step of 0.1 was used. Subsequently, the statistical properties of the computed handwriting features were described using the mean, median, standard deviation (std), and maximum (max). Finally, all of the extracted features were divided into nine different feature sets according to the type of the movement (on-surface, in-air, and combined) and the calculation approach, i.e., the type of feature (FD-based, conventional, and combined). For more information, see Table 4.

Table 4. Feature sets matrix.

Movement	FD (Count)	Conventional (Count)	Together (Count)
on-surface	4536	618	5154
in-air	2916	404	3320
together	7452	1022	8474

Fractional-Order Derivatives

Utilization of the FD as a substitution for the conventional differential derivative during calculation of the basic kinematic features provides a new advanced approach. The advantage of FDs is in their wide range of settings and many different approaches to approximation, e.g., Riemann–Liouville, Caputo, or Grünwald–Letnikov formulations [31,46,47]. For the purpose of this work, Jonathan Hadida's FD Matlab implementation was used following the Grünwald–Letnikov approximation [31,48]. A direct definition of the FD $D^{\alpha}y(t)$ is based on the finite differences of an equidistant grid in $[0, \tau]$, assuming that the function $y(\tau)$ satisfies certain smoothness conditions in every finite interval $(0, t), t \leq T$. Choosing the grid [31],

$$0 = \tau_0 < \tau_1 < ... < \tau_{n+1} = t = (n+1)h \tag{1}$$

with

$$\tau_{k+1} - \tau_k = h \tag{2}$$

and using the notation of finite differences

$$\frac{1}{h^{\alpha}}\Delta_h^{\alpha}y(t) = \frac{1}{h^{\alpha}}\left(y(\tau_{n+1}) - \sum_{v=1}^{n+1} c_v^{\alpha} y(\tau_{n+1-v})\right), \tag{3}$$

where

$$c_v^{\alpha} = (-1)^{v-1}\tbinom{\alpha}{v}. \tag{4}$$

The Grünwald–Letnikov implementation is defined as

$$D^{\alpha}y(t) = \lim_{h \to 0} \frac{1}{h^{\alpha}}\Delta_h^{\alpha}y(t), \tag{5}$$

where $D^{\alpha}y(t)$ denotes a derivative with order α of function $y(t)$, and h represents a sampling lattice.

2.4. Statistical Analysis

Prior to providing a description of the analytical setup, it is important to note that the effect of well-known confounding factors, also known as covariates, was controlled for in all of the analytical steps described below. In the frame of this work, we controlled for the effect of participants' age, gender, and L-dopa [41] (dopaminergic medication).

To assess the strength of the relationship between the computed handwriting features and patient's clinical data (UPDRS V and PD duration), we computed the partial Pearson's correlation coefficient (assessment of a linear relationship), as well as the partial Spearman's correlation coefficient (assessment of a monotonic relationship). With this approach, we aimed to identify the handwriting features that are significantly correlated with the clinical measures under focus and also to compare the FD features with conventional ones. A significance level of correlation (p) of 0.05 was selected for both of the correlation types. Only the results with a p-value below the significance level in both correlation coefficients were considered statistically significant.

Next, to evaluate and compare the power of the handwriting features to discriminate PD patients and HC, multivariate binary classification analysis was performed. For this purpose, state-of-the-art gradient boosted trees were employed. Specifically, we used the famous XGBoost algorithm [49]. The XGBoost algorithm was chosen for its ability to achieve a good performance, even for small datasets; its inherent robustness to outliers; its ability to model complex interdependencies in the data; and also its recent successes in the field of machine learning (e.g., the winning algorithm in many www.kaggle.com competitions). To train and evaluate the models, we used the following supervised learning setup: stratified 10-fold cross-validation with 20 repetitions. The performance of the trained classification models was evaluated by Matthew's correlation coefficient (MCC) [50], classification accuracy (ACC), sensitivity (SEN), and specificity (SPE), which are defined as follows:

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}, \tag{6}$$

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \cdot 100\ [\%], \tag{7}$$

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \cdot 100\ [\%], \tag{8}$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \cdot 100\ [\%], \tag{9}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number false negatives.

Finally, to evaluate and compare the power of the handwriting features' ability to predict the values of the selected clinical characteristics (UPDRS V and PD duration), multivariate regression analysis was performed. For this purpose, the same boosting tree algorithm (XGBoost) and the supervised learning setup were used. The performance of the trained regression models was evaluated by the mean absolute error (MAE), root mean square error (RMSE), and estimated error rate (EER), which are defined as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{10}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{11}$$

$$\mathrm{EER} = \frac{1}{n \cdot r}\sum_{i=1}^{n}|y_i - \hat{y}_i| \cdot 100\ [\%], \tag{12}$$

where y_i represents the true label of the ith observation, $\hat{y}_i$ denotes the predicted label of the ith observation, n is the number of observations, and r is the range of the values of the predicted clinical characteristic (not the range that can be theoretically reached, but the actual range of the values in the dataset). Therefore, the EER describes a percentage of error predictions in regard to the statistical properties of the data.

3. Results

In Table 5, the results of partial correlation analysis between the handwriting features (FD-based features, conventional features) and patients' clinical characteristics (UPDRS V, PD duration) are summarized. The table shows the five best features according to Spearman's correlation coefficient for each movement (on-surface, in-air).

In the case of UPDRS V (on-surface movement), the following FD-based features achieved a statistical significance of correlation: the median of jerk ($\alpha = 0.3$, $\alpha = 0.4$) and horizontal velocity (α = 0.1) for the repetitive letter *l*, the mean of vertical acceleration ($\alpha = 0.7$) for repetitive loops, and the standard deviation of the vertical velocity ($\alpha = 0.3$) for the sentence. The following conventional features achieved a statistical significance of correlation (p-value of only one of the coefficients was below the threshold): the maximum of horizontal jerk and velocity for the repetitive letters *le*, the maximum of horizontal jerk and horizontal velocity for the repetitive letter *l*, and the maximum of horizontal velocity for the letter *l*. Regarding UPDRS V (in-air movement), the following FD-based features achieved a statistical significance of correlation: the median of vertical velocity ($\alpha = 0.9$, $\alpha = 0.8$, $\alpha = 0.7$) for the sentence and the median of horizontal velocity ($\alpha = 0.5$) and vertical jerk ($\alpha = 0.3$) for the repetitive letters *le*. The following conventional features achieved a statistical significance of correlation (p-value of only one of the coefficients was below the threshold): the mean of acceleration for the repetitive word *lektorka*, the maximum of horizontal jerk for the word *porovnat*, the median of the vertical velocity for the repetitive letter *l*, and the median of the horizontal velocity of the repetitive letters *le*.

Table 5. Results of partial correlation analysis between handwriting features and clinical data.

UPDRS V								
FD on-suface					Conventional on-surface			
feature name	α	task	r_p	r_s	r_s	r_p	task	feature name
jerk (median)	0.3	r. letters l	0.37 *	0.48 **	−0.45 *	−0.24	r. letters le	h. jerk (max)
jerk (median)	0.4	r. letters l	0.43 *	0.46 *	−0.43 *	−0.2	r. letters le	velocity (max)
h. velocity (std)	0.1	r. letters l	−0.42 *	−0.41 *	−0.42 *	0.25	r. letters l	h. jerk (max)
v. acceleration (mean)	0.7	r. loops	0.48 **	0.40 *	−0.42 *	−0.16	r. letters l	h. velocity (max)
v. velocity (std)	0.3	sentence	0.40 *	0.40 *	−0.41 *	−0.15	letter l	h. velocity (max)
FD in-air					Conventional in-air			
feature name	α	task	r_p	r_s	r_s	r_p	task	feature name
v. velocity (median)	0.9	sentence	0.44 *	0.53 **	0.43 *	0.28	r. word lektorka	acceleration (mean)
v. velocity (median)	0.8	sentence	0.40 *	0.52 **	−0.37 *	−0.31	word porovnat	h. jerk (max)
h. velocity (median)	0.5	r. letters le	−0.38 *	−0.49 **	0.36 *	0.25	r. letters l	v. velocity (median)
v. jerk (median)	0.3	r. letters le	−0.43 *	−0.49 **	0.35	0.41 *	r. letters le	h. velocity (median)
v. velocity (median)	0.7	sentence	0.37 *	0.48 **	0.35	0.19	r. word lektorka	acceleration (median)
PD Duration								
FD on-surface					Conventional on-surface			
feature name	α	task	r_p	r_s	r_s	r_p	task	feature name
velocity (max)	0.1	spiral	0.54 **	0.55 **	−0.46 *	−0.40 *	r. letters l	h. velocity (max)
acceleration (max)	0.8	spiral	0.54 **	0.54 **	−0.40 *	−0.37 *	r. letters l	h. jerk (max)
acceleration (max)	0.6	spiral	0.54 **	0.54 **	−0.38 *	−0.37 *	r. letters l	velocity (max)
acceleration (max)	0.2	spiral	0.54 **	0.54 **	0.46 **	0.34	spiral	v. velocity (mean)
acceleration (max)	0.7	spiral	0.54 **	0.53 **	0.40 *	0.14	r. loops	h. acceleration (mean)
FD in-air					Conventional in-air			
feature name	α	task	r_p	r_s	r_s	r_p	task	feature name
jerk (median)	0.4	sentence	−0.37 *	−0.49 **	−0.44 *	−0.38 *	word lektorka	h. jerk (median)
jerk (max)	0.1	r. word les	0.57 **	0.46 *	0.38 *	0.40 *	word nepopad.	velocity (max)
jerk (max)	0.3	r. word les	0.57 **	0.45 *	0.37 *	0.42 *	word lektorka	h. n. jerk (mean)
velocity (max)	0.1	r. word les	0.57 **	0.45 *	−0.47 **	−0.13	r. word lektorka	h. velocity (mean)
jerk (max)	0.2	r. word les	0.57 **	0.45 *	−0.42 *	−0.13	word nepopad.	h. velocity (mean)

α—order of FD; r_p—Pearson's correlation coefficient; r_s—Spearman's correlation coefficient; v.—vertical; h.—horizontal; r.—repetitive task; *—$p < 0.05$; **—$p < 0.01$; rows are ordered by the absolute value of Spearman's correlation coefficient.

For PD duration (on-surface movement), the following FD-based features achieved a statistical significance of correlation (of note: all of these features satisfied the stronger threshold for statistical significance of correlation $p < 0.01$): the maximum of the velocity ($\alpha = 0.1$) and acceleration ($\alpha = 0.8$, $\alpha = 0.7$, $\alpha = 0.6$, $\alpha = 0.2$) for the Archimedean spiral. The following conventional features achieved a statistical significance of correlation (p-value of only one of the coefficients was below the threshold):

the maximum of horizontal velocity, horizontal jerk, and velocity for the repetitive letter *l*; the mean of the vertical velocity for the Archimedean spiral; and the mean of horizontal acceleration for repetitive loops. For PD duration (in-air movement), the following FD-based features achieved a statistical significance of correlation: the median of jerk ($\alpha = 0.4$) for sentence, the maximum of jerk ($\alpha = 0.1$, $\alpha = 0.2$, $\alpha = 0.3$) and velocity ($\alpha = 0.1$) for repetitive word *les*. The following conventional features achieved a statistical significance of correlation (p-value of only one of the coefficients was below the threshold): the median and mean of horizontal jerk for the word *lektorka*, the maximum of the velocity for the word *nepopadnout*, and the mean of horizontal velocity for the repetitive word *lektorka* and the word *nepopadnout*.

The results of the multivariate binary classification analysis are summarized in Table 6. In total, we built and evaluated nine different classification models. These models were selected according to the following criteria: movement type (on-surface, in-air, all), feature type (FD features, conventional features, all). We built models based on the combinations of these criteria as well. For more information, see Table 4.

Table 6. Results of multivariate binary classification analysis (PD/HC).

Feature Set	MCC	ACC [%]	SEN [%]	SPE [%]	Feat
conventional on-surface	0.83 ± 0.18	91.19 ± 9.65	93.00 ± 15.52	70.00 ± 0.46	1
conventional in-air	0.95 ± 0.10	**97.14 ± 5.71**	95.50 ± 9.07	100.00 ± 0.00	1
conventional together	0.95 ± 0.11	**97.14 ± 5.71**	95.50 ± 9.07	100.00 ± 0.00	1
FD on-surface	0.95 ± 0.12	87.14 ± 13.48	82.00 ± 21.24	90.00 ± 30.00	1
FD in-air	0.95 ± 0.13	81.43 ± 12.86	71.50 ± 30.83	60.00 ± 48.99	3
FD together	0.95 ± 0.14	81.43 ± 15.71	69.50 ± 32.13	70.00 ± 45.83	2
all on-surface	0.95 ± 0.15	88.33 ± 14.06	89.00 ± 22.11	70.00 ± 45.83	2
all in-air	0.95 ± 0.16	**97.14 ± 5.71**	95.50 ± 9.07	100.00 ± 0.00	1
all together	0.95 ± 0.17	**97.14 ± 5.71**	95.50 ± 9.07	100.00 ± 0.00	1

MCC—Matthew's correlation coefficient; ACC—accuracy; SEN—sensitivity; SPE—specificity; feat.—number of features important for the trained model (i.e., feature importance of the feature > 0.0); The feature importances, as well as the exact names of these features, are summarized in the text.

With respect to the classification performance, the highest MCC achieved was 0.95 was for eight out of the total nine feature sets (with the exception being the feature set composed of conventional handwriting features computed for the on-surface movements). An interesting fact to note is that for all models based on conventional handwriting features, only a single feature was capable of providing the classification models with such a high discrimination power. In terms of the specific features important for the trained models, the following feature importances were returned by the models (feature importance quantifies the relative importance of the features in the ensemble of the trained XGBoost model [49]; therefore, the higher the value of the feature importance, the more important the feature for the prediction of the dependent variable): conventional on-surface (horizontal jerk (median) of repetitive loops), conventional in-air (horizontal velocity (median) of the sentence), conventional together (horizontal velocity (median) of the sentence), FD on-surface (jerk (max) $\alpha = 0.3$ of the letters *le*), FD in-air (vertical acceleration (mean) $\alpha = 0.6$ of the word *nepopadnout* (FI $= 0.33$), horizontal jerk (mean) $\alpha = 0.9$ of the word *nepopadnout* (FI $= 0.33$), horizontal jerk (mean) $\alpha = 0.2$ of the repetitive word *lektorka* (FI $= 0.33$)), FD together (jerk (max) $\alpha = 0.3$ of the letters *le* (on-surface; FI $= 0.67$), horizontal jerk (mean) $\alpha = 0.9$ of the word *nepopadnout* (in-air; FI $= 0.33$)), all on-surface (horizontal jerk (median) of repetitive loops (FI $= 0.50$), jerk (max) $\alpha = 0.3$ of the letters *le* (FI $= 0.50$)), all in-air (horizontal velocity (median) of the sentence), and all together (horizontal velocity (median) of the sentence (in-air)).

The results of multivariate regression analysis are summarized in Table 7. For this purpose, we used UPDRS V and PD duration as our target variables. As in the case of binary classification, we built and evaluated nine different regression models according to the same criteria. For each of the rating scales, the table shows the results achieved using the trained models and the associated feature importance values. All obtained results are discussed in the following section.

Table 7. Results of regression analysis for clinical data.

Feature Set	MAE	RMSE	EER [%]	Feat
		UPDRS V		
conventional on-surface	0.59 ± 0.29	0.71 ± 0.41	13.82 ± 6.71	1
conventional in-air	0.60 ± 0.30	0.72 ± 0.42	14.01 ± 6.98	1
conventional together	0.60 ± 0.31	0.73 ± 0.42	14.05 ± 6.90	1
FD on-surface	0.60 ± 0.32	0.65 ± 0.45	**12.51** ± **7.55**	1
FD in-air	0.60 ± 0.33	0.68 ± 0.43	13.49 ± 7.29	1
FD together	0.60 ± 0.34	0.66 ± 0.45	13.06 ± 7.55	2
all on-surface	0.60 ± 0.35	0.65 ± 0.45	**12.51** ± **7.55**	1
all in-air	0.60 ± 0.36	0.71 ± 0.43	13.72 ± 7.36	1
all together	0.60 ± 0.37	0.66 ± 0.45	13.06 ± 7.55	2
		PD duration		
conventional on-surface	4.29 ± 0.94	5.03 ± 1.09	24.52 ± 5.39	18
conventional in-air	4.91 ± 1.38	5.56 ± 1.50	28.03 ± 7.85	16
conventional together	4.14 ± 1.32	4.85 ± 1.52	**23.64** ± **7.55**	16
FD on-surface	4.45 ± 0.66	5.06 ± 0.85	25.40 ± 3.75	14
FD in-air	4.79 ± 0.73	5.48 ± 0.72	27.36 ± 4.20	19
FD together	4.55 ± 0.68	5.32 ± 0.78	26.00 ± 3.88	21
all on-surface	4.48 ± 0.86	5.12 ± 0.96	25.62 ± 4.92	16 (12 F, 4 C)
all in-air	4.95 ± 1.18	5.59 ± 1.17	28.30 ± 6.75	17 (13 F, 4 C)
all together	4.70 ± 1.10	5.45 ± 1.23	26.82 ± 6.30	17 (12 F, 6 C)

UPDRS V—Unified Parkinson's disease rating scale, part V: Modified Hoehn and Yahr staging score [40]; MAE—mean absolute error; RMSE—root mean squared error; EER—estimation error rate; F—FD-based features; C—conventional handwriting features; feat.—number of features important for the trained model (i.e., feature importance of the feature > 0.0); The feature importances, as well as the exact names of these features for models built to assess UPDRS V, are summarized in the text. In the case of PD duration, this data can be found in Table S1 provided in the Supplementary Material.

Considering EER as our performance evaluation metric, the following results are worth pointing out. In the case of UPDRS V, the lowest EER was achieved using a single FD-based feature—specifically, the standard deviation of vertical velocity ($\alpha = 0.1$) computed for the on-surface movements ($12.51 \pm 7.55\%$). The same feature was selected when both FD and conventional features were considered while building the model. In general, all models achieved an EER of around 12–13%. In comparison with the conventional features, the FD-based features performed better, with a difference of about 1%. In terms of the specific features important for the trained models, the following feature importances were returned by the models: conventional on-surface (vertical normalized jerk (mean) of the repetitive word *lektorka*), conventional in-air (vertical velocity (mean) of the sentence), conventional together (vertical velocity (mean) of the sentence), FD on-surface (vertical velocity (std) $\alpha = 0.1$ of the sentence), FD in-air (vertical velocity (median) $\alpha = 0.3$ of the sentence), FD together (vertical velocity (std) $\alpha = 0.1$ of the sentence (on-surface; FI = 0.50), vertical velocity (median) $\alpha = 0.3$ of the sentence (in-air; FI = 0.50)), all on-surface (vertical velocity (std) $\alpha = 0.1$ of the sentence), all in-air (vertical velocity (median) $\alpha = 0.3$ of the sentence), and all together (vertical velocity (std) $\alpha = 0.1$ of the sentence (on-surface; FI = 0.50), vertical velocity (median) $\alpha = 0.3$ of the sentence (in-air; FI = 0.50)). With respect to PD duration, the lowest EER was achieved using conventional handwriting features computed for both on-surface as well as in-air movements ($23.64 \pm 7.55\%$).

4. Discussion

To the best of our knowledge, except for our pilot work [20], there are no prior studies which integrate FD into a handwriting parameterization for quantitative PD dysgraphia analysis. Therefore, the results published in this paper are exploratory in nature.

In comparison with the conventional kinematic features, FD-based ones correlate more significantly with the clinical characteristics (UPDRS V and PD duration). We observed especially strong correlations for handwriting tasks based on the periodic repetition of specific movements (Archimedean spiral; repetitive letter *l*, syllable *le*, or word *les*). Although the levels of significance based on the conventional handwriting parameters are lower, similar handwriting tasks are involved in the most significant results. We hypothesize that this is due to their ability to highlight or better quantify the cardinal motor symptoms of PD. For example, the most significant relationship between handwriting performance and PD duration was identified in acceleration extracted from the Archimedean spiral. Rigidity combined with tremor and/or bradykinesia makes a PD patient's handwriting/drawing less fluent (increased changes in velocity and higher acceleration). This is highlighted in a task such as the spiral, where the proper coordination of the fingers, wrist, and arm is required. Generally, the observed problems with coordination are in line with the work of Dounskaia et al. [51] and Teulings et al. [52]. To better illustrate these manifestations, Figure 3 plots the velocity profiles of repetitive loops for a healthy control and a PD patient. As can be seen, the patient introduced more changes in velocity, and their drawing became much more non-fluent. To summarize these findings, FD features in combination with properly selected tasks provide a stronger relationship with the severity and progress of PD.

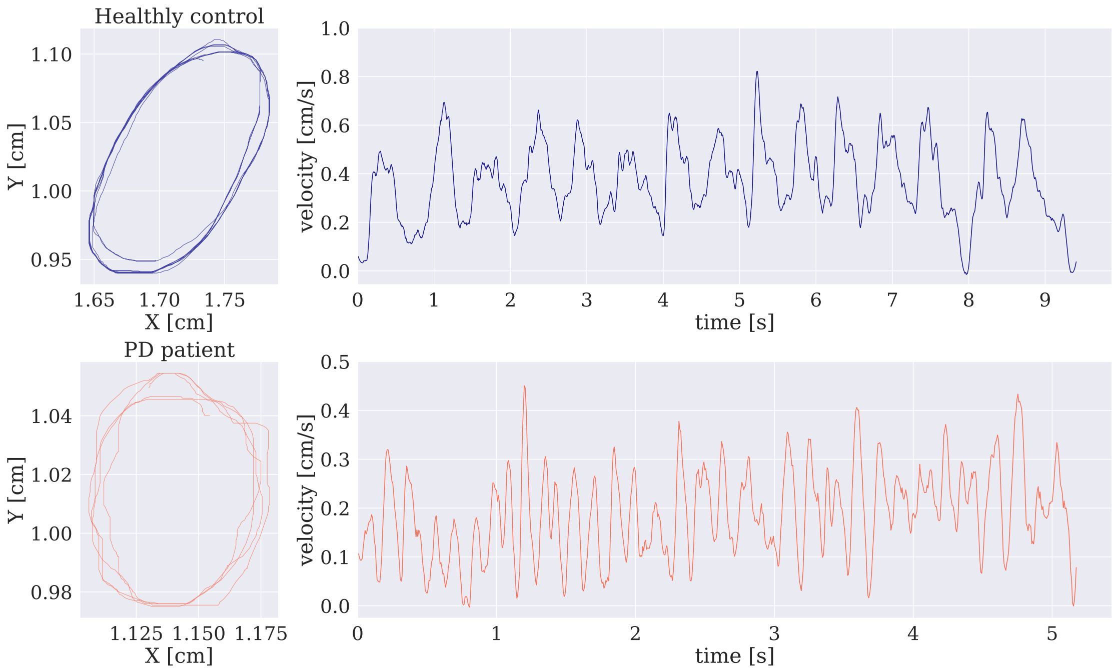


Figure 3. Handwriting samples of the repetitive loop task for HC and PD patients are on the left, and the resulting velocity profiles are on the right.

On the other hand, in terms of binary classification, the conventional parameters provided the best results. The classification performance is remarkable: ACC = 97.74%, SEN = 95.50%, and SPE = 100%. In fact, our results represent the highest classification accuracy that has ever been reported based on the PaHaW database (see Table 1). We hypothesize that the improvement was caused by the inclusion of the state-of-the-art XGBoost algorithm into our machine learning pipelines. As already mentioned, the result is based on one in-air feature: median horizontal velocity of a sentence. In comparison with the HC cohort, the PD patients exhibited much lower values of this measure, i.e., while writing the sentence, the PD patients were not able to perform horizontal transitions (movement between neighboring letters or words) as quickly as the HC could. This finding is in line with the work of Ma et al. [53], who observed that wrist extension stiffness in PD patients makes the handwriting in the horizontal direction more problematic. Therefore, scientists started to use the term *horizontal* *dysgraphia* [13]. Generally, vertical or horizontal dysgraphia may be considered a presymptomatic neurobehavioral biomarker of PD with possible significance in early PD diagnosis [13].

In [20], we proved that the FD features improved the accuracy of PD dysgraphia diagnosis in the Archimedean spiral drawing task by 10%. Contrary to our pilot results, in the frame of this work, these features did not lead to any improvements. After a deeper analysis, we found that this was caused by a combined task approach. Performance of the Archimedean spiral is a quasiparticle and continuous task with some repetitive patterns. It looks as though the FD features work especially well in these specific cases. Nevertheless, when combining these tasks with a complex handwriting task (such as a sentence), the measures quantifying in-air movement tend to be more discriminative (in our case, the median in-air horizontal velocity of a sentence). This brings us to the same conclusion that was given during the correlation analysis—the FD features advance the PD dysgraphia diagnosis only in some specific cases.

The best regression model, estimating the UPDRS V score with a 12.51% error, is based only on the standard deviation of on-surface vertical velocity ($\alpha = 0.1$) extracted from the sentence. This FD-based parameter was selected from the feature set combining all on-surface measures; therefore, we can confirm the positive influence of FC on the regression analysis performance. In fact, the FD features outperformed the conventional ones in all scenarios. To better understand this result, we plotted vertical velocity patterns of the sentence task for different orders of FD (see Figure 4). We can observe a big difference between $\alpha = 0.1$ and the rest of the orders, including the full derivative. This large distance means that we are working with completely new information that is far from that contained in the full derivative. Although it is difficult to clinically interpret this information, it is clear that FC opens new possibilities for monitoring PD severity.

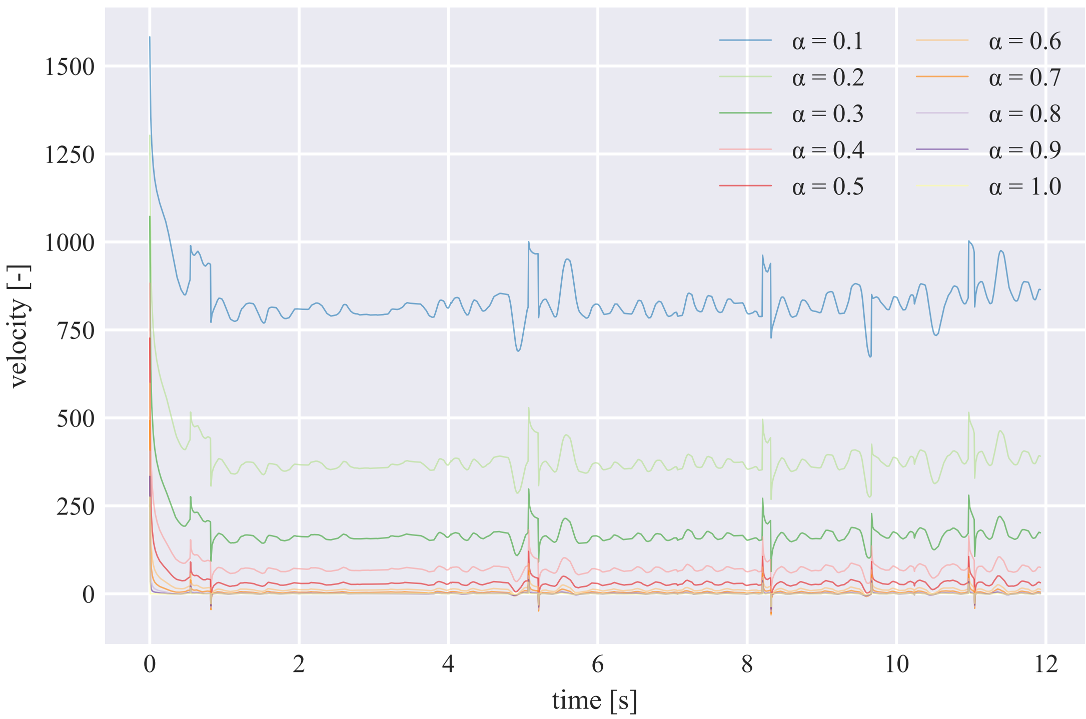


Figure 4. Vertical velocity patterns of the sentence task for different orders of fractional-order derivatives (FD).

Regarding the PD duration estimation results, the most successful model (EER = 23.46%) consists of 16 conventional on-surface/in-air features (all features' importance values can be found in Supplementary Table S1). The most frequent feature with the highest feature importance is the jerk extracted from several handwriting tasks. This probably means that as PD progresses, handwriting becomes more jerky and irregular. Vertical velocity is the second most frequent feature involved in the models, which is probably linked with micrographia. Generally, in the case of PD duration estimation, the FD-based features did not yield any improvement.

In conclusion, the FD-based features are better for modeling PD severity (in terms of UPDRS V score estimation), but they do not lead to an improvement in PD duration modeling. The progress of PD is nonlinear and very individual. This means that patients with the same PD duration can be in different stages of the disease. This fact supports our results: the estimation error of PD duration was generally much worse than the estimation error of the UPDRS V score. Since PD duration estimation is a difficult task with poor results, fine improvements based on FD parameters play no role.

5. Conclusions

This study deals with advanced approaches to PD dysgraphia diagnosis and monitoring based on FC integrated with online handwriting/drawing parameterization. To the best of our knowledge, it is the first work that performs a complex investigation into the possibilities for FC in online handwriting processing and proposes new advances in kinematic analyses based on FD. Although the conventional features provided better and very high classification accuracy, which is at the top of the state-of-the-art analyses based on the PaHaW database (ACC = 97.74%, SEN = 95.50%, and SPE = 100%), the newly designed parameters were proven to work better for specific tasks (continuous and/or repetitive, such as the Archimedean spiral) and for specific applications, i.e., PD severity estimation (EER = 12.51%). However, our results need to be confirmed by subsequent scientific research.

This study has several limitations and suggestions for further improvements. Since the dataset is small, to be able to generalize the results, bigger databases should be involved. On the other hand, it is common to have such small numbers of PD patients and HC samples in PD dysgraphia analysis, e.g., see our review in Table 1. Next, we considered only the kinematic measures. To better evaluate the discrimination power of the FD features and better evaluate their ability to estimate PD severity or progress, other feature types, such as temporal, spatial, and dynamic, should be included in future comparisons. Finally, the FD-based parameters could be further explored. For instance, we can consider other approximations (e.g., Caputo) or employ FC for other measures (e.g., entropies).

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-3417/8/12/2566/s1, Table S1: Feature relevance from multivariate regression (modeling PD duration).

Author Contributions: Conceptualization, J.M. (Jan Mucha), J.M. (Jiri Mekyska) and M.F.-Z.; methodology, J.M. (Jan Mucha), M.F.-Z. and J.M. (Jiri Mekyska); software, J.M. (Jan Mucha) and J.M. (Jiri Mekyska); validation, J.M. (Jan Mucha); formal analysis, J.M. (Jan Mucha) and J.M. (Jiri Mekyska); investigation, J.M. (Jan Mucha), M.F.-Z., J.M. (Jiri Mekyska), I.R. and L.B.; resources, I.R. and L.B.; data curation, J.M. (Jan Mucha), J.M. (Jiri Mekyska), Z.G., V.Z., T.K., L.B. and I.R.; writing—original draft preparation, J.M. (Jan Mucha), J.M. (Jiri Mekyska), Z.G. and M.F.-Z.; visualization, J.M. (Jan Mucha) and Z.G.; supervision, J.M. (Jiri Mekyska), Z.S., I.R., M.F.-Z. and K.L.-d.-I.; project administration, J.M. (Jiri Mekyska), Z.S., I.R., M.F.-Z. and K.L.-d.-I.; funding acquisition, J.M. (Jiri Mekyska), Z.S., I.R., M.F.-Z. and K.L.-d.-I.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 734718 (CoBeN). In addition, this work was supported by the grant of the Czech Science Foundation 18-16835S (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing) and the following projects: LO1401, FEDER and MEC, and TEC2016-77791-C4-2-R from the Ministry of Economic Affairs and Competitiveness of Spain. This article is based upon work from COST Action CA15225, a network supported by COST (European Cooperation in Science and Technology), and, for the research, infrastructure of the SIX Center was used.

Conflicts of Interest: The authors declare no conflict of interest.

Disclaimer: The article reflects only the authors' view, and the Research Executive Agency (REA) is not responsible for any use that may be made of the information it contains.

Abbreviations

The following abbreviations are used in this manuscript:

ACC	accuracy
ADA	AdaBoost
ANN	artificial neural network
ANOVA	analysis of variance
AUC	area under the ROC curve
CNN	convolutional neural network
EMD	empirical mode decomposition
EER	estimated error rate
FN	false negatives
FP	false positives
FC	fractional calculus
FD	fractional-order derivative
FI	feature importance
K-NN	K-nearest neighbors
LED	L-dopa equivalent daily dose
LDA	linear discriminant analysis
MCC	Matthew's correlation coefficient
max	maximum
MAE	mean absolute error
NB	naïve Bayes classifier
OPF	optimum path forest
PD	Parkinson's disease
RF	random forests
RMSE	root mean squared error
SEN	sensitivity
r_p	Pearson's correlation coefficient
r_s	Spearman's correlation coefficient
SPE	specificity
std	standard deviation
TN	true negatives
TP	true positives
SVM	support vector machine
UPDRS V	unified Parkinson's disease rating scale, part V: Modified Hoehn and Yahr staging score

References

1. Bellou, V.; Belbasis, L.; Tzoulaki, I.; Evangelou, E.; Ioannidis, J.P. Environmental risk factors and Parkinson's disease: An umbrella review of meta-analyses. *Parkinsonism Relat. Disord.* **2016**, *23*, 1–9. [CrossRef] [PubMed]
2. Berg, D.; Postuma, R.B.; Adler, C.H.; Bloem, B.R.; Chan, P.; Dubois, B.; Gasser, T.; Goetz, C.G.; Halliday, G.; Joseph, L.; et al. MDS research criteria for prodromal Parkinson's disease. *Mov. Disord.* **2015**, *30*, 1600–1611. [CrossRef] [PubMed]
3. Sapir, S.; Ramig, L.; Fox, C. Speech and swallowing disorders in Parkinson disease. *Curr. Opin. Otolaryngol. Head Neck Surg.* **2008**, *16*, 205–210. [CrossRef] [PubMed]
4. Hirsch, L.; Jette, N.; Frolkis, A.; Steeves, T.; Pringsheim, T. The incidence of Parkinson's disease: A systematic review and meta-analysis. *Neuroepidemiology* **2016**, *46*, 292–300. [CrossRef] [PubMed]
5. Hornykiewicz, O. Biochemical aspects of Parkinson's disease. *Neurology* **1998**, *51*, S2–S9. [CrossRef] [PubMed]
6. Elbaz, A.; Carcaillon, L.; Kab, S.; Moisan, F. Epidemiology of Parkinson's disease. *Rev. Neurol.* **2016**, *172*, 14–26, doi:10.1016/j.neurol.2015.09.012. [CrossRef] [PubMed]
7. Contreras-Vidal, J.L.; Poluha, P.; Teulings, H.L.; Stelmach, G.E. Neural dynamics of short and medium-term motor control effects of levodopa therapy in Parkinson's disease. *Artif. Intell. Med.* **1998**, *13*, 57–79. [CrossRef]

8. Drotar, P.; Mekyska, J.; Rektorova, I.; Masarova, L.; Smekal, Z.; Faundez-Zanuy, M. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artif. Intell. Med.* **2016**, *67*, 39–46. [CrossRef]
9. Brabenec, L.; Mekyska, J.; Galaz, Z.; Rektorova, I. Speech disorders in Parkinson's disease: Early diagnostics and effects of medication and brain stimulation. *J. Neural Transm.* **2017**, *124*, 303–334. [CrossRef]
10. Mucha, J.; Galaz, Z.; Mekyska, J.; Kiska, T.; Zvoncak, V.; Smekal, Z.; Eliasova, I.; Mrackova, M.; Kostalova, M.; Rektorova, I.; et al. Identification of hypokinetic dysarthria using acoustic analysis of poem recitation. In Proceedings of the 2017 40th International Conference on Telecommunications and Signal Processing (TSP), Barcelona, Spain, 5–7 July 2017; pp. 739–742.
11. De Stefano, C.; Fontanella, F.; Impedovo, D.; Pirlo, G.; di Freca, A.S. Handwriting analysis to support neurodegenerative diseases diagnosis: A review. *Pattern Recognit. Lett.* **2018**, in press. [CrossRef]
12. Rosenblum, S.; Samuel, M.; Zlotnik, S.; Erikh, I.; Schlesinger, I. Handwriting as an objective tool for Parkinson's disease diagnosis. *J. Neurol.* **2013**, *260*, 2357–2361. [CrossRef]
13. Thomas, M.; Lenka, A.; Kumar Pal, P. Handwriting Analysis in Parkinson's Disease: Current Status and Future Directions. *Mov. Disord. Clin. Pract.* **2017**, *4*, 806–818. [CrossRef] [PubMed]
14. McLennan, J.; Nakano, K.; Tyler, H.; Schwab, R. Micrographia in Parkinson's Disease. *J. Neurol. Sci.* **1972**, *15*, 141–152. [CrossRef]
15. Letanneux, A.; Danna, J.; Velay, J.L.; Viallet, F.; Pinto, S. From micrographia to Parkinson's disease dysgraphia. *Mov. Disord.* **2014**, *29*, 1467–1475. [CrossRef] [PubMed]
16. Drotar, P.; Mekyska, J.; Smekal, Z.; Rektorova, I.; Masarova, L.; Faundez-Zanuy, M. Contribution of different handwriting modalities to differential diagnosis of Parkinson's Disease. In Proceedings of the 2015 IEEE International Symposium Medical Measurements and Applications (MeMeA), Turin, Italy, 7–9 May 2015; pp. 1–5.
17. Drotar, P.; Mekyska, J.; Rektorova, I.; Masarova, L.; Smekal, Z.; Zanuy, M.F. Decision support framework for Parkinson's disease based on novel handwriting markers. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2015**, *23*, 508–516. [CrossRef] [PubMed]
18. Loconsole, C.; Trotta, G.F.; Brunetti, A.; Trotta, J.; Schiavone, A.; Tato, S.I.; Losavio, G.; Bevilacqua, V. Computer Vision and EMG-Based Handwriting Analysis for Classification in Parkinson's Disease. In *Intelligent Computing Theories and Application*; Springer International Publishing: Cham, Switzerland, 2017; pp. 493–503, doi:10.1007/978-3-319-63312-1_43.
19. Nackaerts, E.; Broeder, S.; Pereira, M.P.; Swinnen, S.P.; Vandenberghe, W.; Nieuwboer, A.; Heremans, E. Handwriting training in Parkinson's disease: A trade-off between size, speed and fluency. *PLoS ONE* **2017**, *12*, e0190223. [CrossRef] [PubMed]
20. Mucha, J.; Zvoncak, V.; Galaz, Z.; Faundez-Zanuy, M.; Mekyska, J.; Kiska, T.; Smekal, Z.; Brabenec, L.; Rektorova, I.; Lopez-de Ipina, K. Fractional Derivatives of Online Handwriting: A New Approach of Parkinsonic Dysgraphia Analysis. In Proceedings of the 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, Greece, 4–6 July 2018; pp. 214–217.
21. Kotsavasiloglou, C.; Kostikis, N.; Hristu-Varsakelis, D.; Arnaoutoglou, M. Machine learning-based classification of simple drawing movements in Parkinson's disease. *Biomed. Signal Process. Control* **2017**, *31*, 174–180. [CrossRef]
22. Taleb, C.; Khachab, M.; Mokbel, C.; Likforman-Sulem, L. Feature selection for an improved Parkinson's disease identification based on handwriting. In Proceedings of the 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR), Nancy, France, 3–5 April 2017; pp. 52–56.
23. Moetesum, M.; Siddiqi, I.; Vincent, N.; Cloppet, F. Assessing visual attributes of handwriting for prediction of neurological disorders: A case study on Parkinson's disease. *Pattern Recognit. Lett.* **2018**. [CrossRef]
24. Heremans, E.; Nackaerts, E.; Vervoort, G.; Vercruysse, S.; Broeder, S.; Strouwen, C.; Swinnen, S.P.; Nieuwboer, A. Amplitude Manipulation Evokes Upper Limb Freezing during Handwriting in Patients with Parkinson's Disease with Freezing of Gait. *PLoS ONE* **2015**, *10*, e0142874. [CrossRef] [PubMed]
25. Pereira, C.R.; Pereira, D.R.; da Silva, F.A.; Hook, C.; Weber, S.A.; Pereira, L.A.; Papa, J.P. A step towards the automated diagnosis of parkinson's disease: Analyzing handwriting movements. In Proceedings of the 2015 IEEE 28th International Symposium on Computer-Based Medical Systems (CBMS), Sao Carlos, Brazil, 22–25 June 2015; pp. 171–176.

26. Heremans, E.; Nackaerts, E.; Broeder, S.; Vervoort, G.; Swinnen, S.P.; Nieuwboer, A. Handwriting Impairments in People With Parkinson's Disease and Freezing of Gait. *Neurorehabil. Neural Repair* **2016**, *30*, 911–919. [CrossRef]
27. Pereira, C.R.; Weber, S.A.; Hook, C.; Rosa, G.H.; Papa, J.P. Deep Learning-Aided Parkinson. In Proceedings of the 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Sao Paulo, Brazil, 4–7 October 2016; pp. 340–346.
28. Impedovo, D.; Pirlo, G.; Vessio, G. Dynamic Handwriting Analysis for Supporting Earlier Parkinson's Disease Diagnosis. *Information* **2018**, *9*, 247. [CrossRef]
29. Baleanu, D.; Jajarmi, A.; Bonyah, E.; Hajipour, M. New aspects of poor nutrition in the life cycle within the fractional calculus. *Adv. Differ. Equ.* **2018**, *2018*, 230. [CrossRef]
30. Khalili Golmankhaneh, A.; Baleanu, D. New Derivatives on the Fractal Subset of Real-Line. *Entropy* **2016**, *18*, 1. [CrossRef]
31. Podlubny, I. *Fractional Differential Equations an Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of Their Solution and Some of Their Applications*; Academic Press: San Diego, CA, USA, 1999.
32. Uchaikin, V.V. *Fractional Derivatives for Physicists and Engineers*; Springer: Berlin, Germany, 2013; Volume 2.
33. Samko, S.G.; Kilbas, A.A.; Marichev, O.I. *Fractional Integrals and Derivatives: Theory and Applications*; Gordon and Breach: Yverdon, Switzerland, 1993; p. 44.
34. Arshad, S.; Baleanu, D.; Bu, W.; Tang, Y. Effects of HIV infection on CD4+ T-cell population based on a fractional-order model. *Adv. Differ. Equ.* **2017**, *2017*, 92. [CrossRef]
35. Pinto, C.M.; Machado, J.T. Fractional model for malaria transmission under control strategies. *Comput. Math. Appl.* **2013**, *66*, 908–916. doi:10.1016/j.camwa.2012.11.017. [CrossRef]
36. Lopes, A.M.; Machado, J.T. Integer and fractional-order entropy analysis of earthquake data series. *Nonlinear Dyn.* **2016**, *84*, 79–90. [CrossRef]
37. Lopes, A.M.; Machado, J.T. Application of fractional techniques in the analysis of forest fires. *Int. J. Nonlinear Sci. Numer. Simul.* **2016**, *17*, 381–390. [CrossRef]
38. Scalas, E.; Gorenflo, R.; Mainardi, F. Fractional calculus and continuous-time finance. *Phys. A Stat. Mech. Its Appl.* **2000**, *284*, 376–384. [CrossRef]
39. Baleanu, D. *Fractional Calculus: Models and Numerical Methods*; World Scientific: Singapore, 2012; Volume 3.
40. Fahn, S.; Elton, R.L. UPDRS Development Committee (1987) Unified Parkinson's Disease Rating Scale. In *Recent Developments in Parkinson's Disease*; Macmillan: Florham Park, NJ, USA, 1987.
41. Lee, J.Y.; Kim, J.W.; Lee, W.Y.; Kim, J.M.; Ahn, T.B.; Kim, H.J.; Cho, J.; Jeon, B.S. Daily dose of dopaminergic medications in Parkinson's disease: Clinical correlates and a posteriori equation. *Neurol. Asia* **2010**, *15*, 137–143.
42. Sesa-Nogueras, E.; Faundez-Zanuy, M.; Mekyska, J. An information analysis of in-air and on-surface trajectories in online handwriting. *Cogn. Comput.* **2012**, *4*, 195–205. [CrossRef]
43. Alonso-Martinez, C.; Faundez-Zanuy, M.; Mekyska, J. A comparative study of in-air trajectories at short and long distances in online handwriting. *Cogn. Comput.* **2017**, *9*, 712–720. [CrossRef] [PubMed]
44. Drotar, P.; Mekyska, J.; Rektorova, I.; Masarova, L.; Smekal, Z.; Faundez-Zanuy, M. A new modality for quantitative evaluation of Parkinson's disease: In-air movement. In Proceedings of the 13th IEEE International Conference on BioInformatics and BioEngineering (BIBE 2013), Chania, Greece, 10–13 November 2013; pp. 1–4.
45. Mekyska, J.; Faundez-Zanuy, M.; Mzourek, Z.; Galaz, Z.; Smekal, Z.; Rosenblum, S. Identification and rating of developmental dysgraphia by handwriting analysis. *IEEE Trans. Hum.-Mach. Syst.* **2017**, *47*, 235–248. [CrossRef]
46. Tenreiro Machado, J. Fractional coins and fractional derivatives. *Abstr. Appl. Anal.* **2013**, *2013*, 205097. [CrossRef]
47. Kilbas, A.A.; Srivastava, H.M.; Trujillo, J.J. *Theory and Applications of Fractional Differential Equations*; Elsevier: Amsterdam, The Netherlands, 2006.
48. Scherer, R.; Kalla, S.L.; Tang, Y.; Huang, J. The Grünwald–Letnikov method for fractional differential equations. *Comput. Math. Appl.* **2011**, *62*, 902–917. [CrossRef]

49. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
50. Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **1975**, *405*, 442–451. [CrossRef]
51. Dounskaia, N.; Van Gemmert, A.W.; Leis, B.C.; Stelmach, G.E. Biased wrist and finger coordination in Parkinsonian patients during performance of graphical tasks. *Neuropsychologia* **2009**, *47*, 2504–2514. [CrossRef]
52. Teulings, H.L.; Contreras-Vidal, J.L.; Stelmach, G.E.; Adler, C.H. Parkinsonism Reduces Coordination of Fingers, Wrist, and Arm in Fine Motor Control. *Exp. Neurol.* **1997**, *146*, 159–170. [CrossRef]
53. Ma, H.I.; Hwang, W.J.; Chang, S.H.; Wang, T.Y. Progressive micrographia shown in horizontal, but not vertical, writing in Parkinson's disease. *Behav. Neurol.* **2013**, *27*, 169–174. [CrossRef]

A.12 Advanced Analysis of Online Handwriting in a Multilingual Cohort of Patients with Parkinson's Disease

Oral Topic: *Biomedical Signal Processing*

Advanced Analysis of Online Handwriting in a Multilingual Cohort of Patients with Parkinson's Disease

J.Mucha [1], J.Mekyska [1], M.Faundez-Zanuy [2], P.Sanz-Cartagena [3], Z.Galaz [1,5], V.Zvoncak [1], T.Kiska [1], Z.Smekal [1], K.Lopez-de-Ipina [4] and I.Rektorova [5]

[1] Department of Telecommunications and SIX Research Centre, Brno University of Technology, Technicka 10, 61600 Brno, Czech Republic

[2] Escola Superior Politecnica, Tecnocampus Avda. Ernest Lluch 32, 08302 Mataro, Barcelona, Spain

[3] Hospital de Mataró, Neurology Unit, Consorci Sanitari del Maresme, Mataró, Spain

[4] Department of Systems Engineering and Automation, University of the Basque Country UPV/EHU, Av de Tolosa 54, 20018 Donostia, Spain;

[5] Applied Neuroscience Research Group, Central European Institute of Technology, Masaryk University, Kamenice 5, 62500 Brno, Czech Republic

Tel.: + 421 915 420 540, +420 54114 6968

E-mail: muchajano@phd.feec.vutbr.cz

Abstract: The majority of Parkinson's disease (PD) patients suffer from handwriting abnormalities commonly called as Parkinsonic dysgraphia. Several approaches of PD dysgraphia analysis exist, e.g. based on online handwriting processing. However, a small and unilingual cohort of PD patients is often an issue in quantitative PD dysgraphia analysis studies. Therefore, in this work, we aim to perform a discrimination analysis in a multilingual cohort of 73 PD patients and 48 healthy controls (Spanish and Czech). For this purpose, we extracted advanced handwriting features based on fractional order derivatives (FD). Discrimination power of the advanced FD-based features was evaluated by Mann-Whitney U test and random forests classifier. We reached 82 % classification accuracy (86 % sensitivity, 77 % specificity) in the multilingual cohort. In addition, we observed high discrimination power of the FD-based parameters and proofed the high impact of online handwriting processing in cross-cultural PD dysgraphia analysis studies.

Keywords: Parkinsonic dysgraphia; micrographia; online handwriting; fractional order derivative; fractional calculus; multilingual cohort

1. Introduction

Parkinson's disease (PD), as the second most frequent neurodegenerative disorder, affects approximately 1.5 % of the world population aged over 65 years [1]. A rapid degeneration of dopaminergic cells in substancia nigra pars compacta emerged as the most important biological finding accompanying the disease [2]. Considering the cardinal motor symptoms of PD (tremor in rest, bradykinesia and rigidity) in conjunction with cognitive, perceptual and motor requirements of handwriting, the disrupted handwriting of PD patients may be used as a significant biomarker for PD diagnosis [3]. The most commonly observed handwriting abnormality in PD patients is micrographia (progressive decrease of letters amplitude) [4], which may be noticed even before the onset of PD motor symptoms in approximately 5 % of PD patients.

Nowadays, by utilizing digitizing tablets, which brings an ability to acquire x and y position with temporal information, we have the opportunity to process online handwriting signals. Therefore, we are not limited to analyze the spatial features only, but we are able to quantify more manifestations of PD appearing in patients handwriting data (temporal, kinematic or dynamic), generally named as *PD dysgraphia* [5].

The impact of quantitative PD dysgraphia analysis employing several handwriting or drawing tasks (e. g. characters, loops, sentences, figures) has been explored in [6] [7] [8] [9]. Researchers usually use kinematic, temporal, spatial or dynamic handwriting features in PD dysgraphia analysis. However, more advanced parameters (based on entropy, energy operators or empirical mode decomposition) have been reported too. PD dysgraphia classification accuracies reported by recent works vary in the range of 85 and 97 %. In our previous works [6] [10] [11], we proposed and evaluated a new advanced approach of kinematic analysis based on fractional order derivatives (FD). Using this approach, we were able to identify PD with almost 90 % accuracy employing only 5 basic kinematic features.

The most common issue in PD differential analysis (cause by complicated and time-consuming patient examination process), which researchers are encountering with, is a small and unilingual cohort of patients. This may result into poor generalization. Especially, the size of examining dataset has a significant influence on results reliability. The smallest the dataset is, the more misleading results may be. Therefore, in this study, we aimed to analyze a multilingual cohort involving two PD handwriting databases (Czech and Spanish) in order to train a more robust classification model. To our best knowledge, this is the first study considering multilingual cohort in PD dysgraphia analysis.

2. Datasets and Methodology

2.1. Datasets

For the purpose of this study, we used two PD handwriting databases. The Czech (PaHaW [8]) database consists of 37 PD patients and 38 healthy controls (HC). It includes 9 different handwriting tasks (Archimedean spiral, repetitive loops, repetitive letter *l*, syllable, words and sentence). The Spanish database (recorded in Mataró Hospital, Spain) consists of 36 PD patients and 10 HC. It includes 2 handwriting tasks (repetitive and continuously written letter *l* and sentence). Demographic and clinical data of both cohorts can be found in Table 1. All patients were examined on their regular dopaminergic medication approximately 1 hour after the L-dopa dose. All participants were right-handed, and all participants signed an informed consent form approved by the local ethics committees.

Table 1. Demographic and clinical data of all participants.

Cohort	Number	Age [y]	PD dur [y]
Parkinson's disease patients			
Czech	37	69.21 ± 11.10	8.70 ± 4.82
Spanish	36	68.25 ± 10.46	6.10 ± 3.78
Healthy Controls			
Czech	38	62.50 ± 11.70	-
Spanish	10	57.50 ± 6.36	-

[1] **y – years; dur - duration**

For the purpose of this study, sentence handwriting task was selected from the databases. Even the tasks are different due to language, we hypothese that pathological characteristics in the handwritten signals will be similar. Sentences in their original language and with resulting English translations are listed below:

a) Czech: *"Tramvaj dnes už nepojede."*
English: The tram will no longer go today.

b) Spanish *"La casa de Barcelona es preciósa."*
English: The house in Barcelona is beautiful.

Samples of PD patients' sentences can be found in Figure 1. In Figure 2, descriptive statistics of both datasets are visualized. Handwriting data were acquired using a digitizing tablet Wacom Intuos 4M (both datasets). Following time sequences were sampled with frequency fs = 150 Hz: x and y coordinates (x[t], y[t]); time-stamp (t); in-air/on-surface status (b[t]); pressure (p[t]); azimuth (az[t]); and tilt (al[t]).

2.2. Methodology

Firstly, each handwritten signal was split into on-surface and in-air movements [12] (see Figure 1). Next, basic kinematic features such as velocity, acceleration and jerk were extracted. Instead of conventional differential derivative, we utilized FD as an advanced approach of kinematic features calculation. For this purpose, the Grünwald-Letnikov approximation was used [13] [14]. The advantage of FD is based on their extensive range of settings and several approaches of approximation. Moreover, we also applied FD on pen pressure, azimuth and tilt signals. All features were extracted for different values of α (order of FD). In the frame of this work, a range from 0.1 to 1.0 with a step of 0.1 was used. Finally, statistical properties of the features were described by: mean, median, standard deviation (std), and maximum (max). Altogether, 1188 handwriting features were extracted for each dataset.

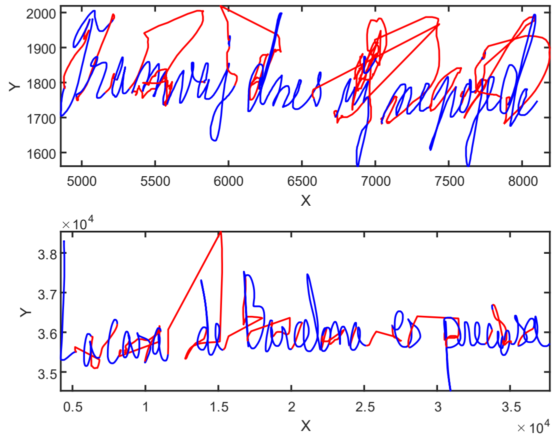


Figure 1. PD patient's sentences examples. Czech sentence in the upper part and Spanish in the bottom part of the figure. On-surface (blue) and in-air (red) movement are visualized.

We were considering 3 following feature sets: Czech, Spanish and multilingual (mixed – 73 PD, 48 HC). In order to identify features that discriminate HC and PD we employed the Mann-Whitney U test. The significance level was set to $\alpha = 0.001$.

Next, to evaluate the discrimination power of handwriting features, we performed multivariate classification analysis based on random forests (RF) [15]. In order to reduce the number of handwriting features entering into the classification analysis, we designed fast and efficient 2-stage feature selection. Firstly, each feature set was reduced by minimum redundancy maximum relevance [16] (mRMR) feature selection algorithm to 50 best features. Secondly, to obtain the most appropriate combination of the features, the sequential floating forward selection [17] (SFFS) algorithm was employed. To achieve the most accurate results for each dataset, we used different types of model validation techniques. In the case of Czech and Spanish feature sets we used leave-one-out cross-validation (due to small sample size). For the multilingual feature set, 10-fold cross-validation with 20 repetition was used. Classification performance was evaluated by the Matthew's correlation coefficient [18] (MCC), classification accuracy (ACC), sensitivity (SEN) and specificity (SPE).

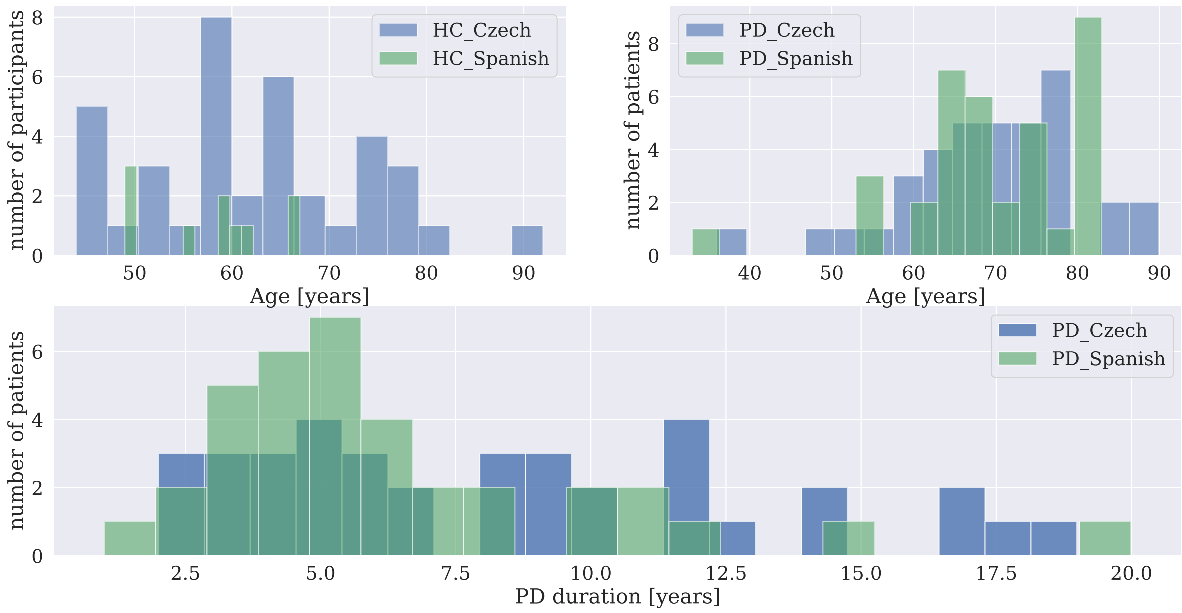


Figure 2. Descriptive statistics of examined datasets. In the top left part of the figure, the HC age distribution is visualized. The PD age distribution is in the top right part and in the bottom part, the distribution of PD duration is shown.

3. Results

The results of the Mann-Whitney U test can be found in the upper part of Table 2. Three most discriminative features which passed through the test are reported for each feature set. Features are sorted by significance level p, while all reported features obtained $p < 0.0001$. The most discriminative feature from the Spanish feature set is velocity (on-surface). In the case of Czech and multilingual feature set, it is its vertical variant, which is probably linked with the vertical micrographia [19]. As can be noticed, the results of the Czech and multilingual feature sets are quite similar, in comparison with Spanish one. This is probably caused by the size of the Spanish HC cohort (10 participants).

Next, the results of the multivariate classification analysis can be found in the bottom part of Table 2. The highest classification performance was obtained in the Spanish feature set (ACC = 95.65 %), nevertheless, due to the imbalanced cohort (36 PD patients and 10 HC), these results may be misleading. Number of HC in the Spanish database is 3.6 times lower than number of PD patients. By mixing the Spanish and Czech (well balanced) databases we have reduced the imbalance of the Spanish one (PD ≈ 1.5 x HC). Also, the distribution of PD duration for the Czech cohort is more uniform (see Figure 2). In the Spanish cohort, patients with shorter disease duration (less than 6 years) outweigh, however the distribution of PD patient's age is quite similar for both cohorts. Thus, by combining the datasets, we also improved non-uniformity of the final cohort. Although the accuracy of the multilingual feature set is the lowest one (82.29 %), credibility of the results may be considered as higher in comparison to the Spanish feature set.

Table 2. Results of Mann-Whitney U test and classification analysis

Mann-Whitney U test			
Feat. Set	**Feature Name**	**α**	***p***
Spanish	velocity^s (median)	0.1	0.000069
	velocity^s(median)	0.2	0.000069
	velocity^a (mean)	0.1	0.000077
Czech	vertical velocity^s (mean)	0.2	0.000012
	vertical velocity^s (mean)	0.2	0.000014
	vertical velocity^s (median)	0.4	0.000014
Multi-lingual	vertical velocity^s (mean)	0.1	0.000001
	vertical velocity^s (median)	0.4	0.000001
	vertical velocity^s (median)	0.3	0.000001

Multivariate classification analysis					
Feat. Set	**N**	**MCC**	**ACC [%]**	**SEN [%]**	**SPE [%]**
Spanish	2	0.87	95.65	97.22	90.00
Czech	9	0.71	85.33	89.19	81.58
Multi-lingual	8	0.63	82.29	85.99	77.22

[1] Feat. Set – feature set; α – order of FD; p – significance level; ^{s} – on-surface movement; ^{a} – in-air movement; N – number of features

4. Conclusions

This study deals with the advanced analysis of PD dysgraphia in a multilingual cohort. First of all, since the most significant features identified in the Mann-Whitney U test and features selected by the SFFS have a non-integer value of the FD order, we suppose that the FD based parameters play significant role in PD dysgraphia quantification. Next, we achieved more than 80 % classification accuracy in all scenarios, which suggests the high impact of online handwriting processing in cross-cultural clinical studies focused on PD dysgraphia diagnosis.

This study has several limitations and suggestions for further research. Firstly, the Spanish dataset is not balanced (PD/HC, PD duration). In addition, the overall sample size is not big. On the other hand, to the best of our knowledge, it is the first and therefore the biggest multilingual online handwriting PD dataset, that has ever been analyzed. Finally, the FD-based features may be more explored and extended (e.g. by Caputo approximation approach). To sum it up, this study has a pilot character and further research should be done to be able to generalize the results.

Acknowledgements

This work was supported by the grant of the Czech Ministry of Health 16-30805A (Effects of non-invasive brain stimulation on hypokinetic dysarthria, micrographia, and brain plasticity in patients with Parkinson's disease), grant of the Czech Science Foundation 18-16835S (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing) and the following projects: LO1401, FEDER and MEC, TEC2016-77791-C4-2-R, from the Ministry of Economic Affairs and Competitiveness of Spain. For the research, infrastructure of the SIX Center was used.

Reference

[1] D. Berg, R. B. Postuma, C. H. Adler, B. R. Bloem, P. Chan, B. Dubois, T. Gasser, C. G. Goetz, G. Halliday, L. Joseph and others, "MDS research criteria for prodromal Parkinson's disease," *Movement Disorders,* vol. 30, pp. 1600-1611, 2015.

[2] A. Elbaz, L. Carcaillon, S. Kab and F. Moisan, "Epidemiology of Parkinson's disease," *Revue Neurologique,* vol. 172, pp. 14-26, 2016.

[3] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo and A. S. Freca, "Handwriting analysis to support neurodegenerative diseases diagnosis: A review," *Pattern Recognition Letters,* 2018.

[4] J. E. McLennan, K. Nakano, H. R. Tyler and R. S. Schwab, "Micrographia in Parkinson's Disease," *Journal of the Neurological Sciences,* vol. 15, pp. 141-152, 1972.

[5] A. Letanneux, J. Danna, J.-L. Velay, F. Viallet and S. Pinto, "From micrographia to Parkinson's disease dysgraphia," *Movement Disorders,* vol. 29, pp. 1467-1475, 2014.

[6] J. Mucha, V. Zvoncak, Z. Galaz, M. Faundez-Zanuy, J. Mekyska, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova and K. Lopez-de-Ipina, "Fractional Derivatives of Online Handwriting: A New Approach of Parkinsonic Dysgraphia Analysis," in *2018 41st International Conference on Telecommunications and Signal Processing (TSP),* 2018.

[7] E. Nackaerts, S. Broeder, M. P. Pereira, S. P. Swinnen, W. Vandenberghe, A. Nieuwboer and E. Heremans, "Handwriting training in Parkinson's disease: A trade-off between size, speed and fluency," *PLOS ONE,* vol. 12, p. e0190223, 12 2017.

[8] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," *Artificial Intelligence in Medicine,* vol. 67, pp. 39-46, 2016.

[9] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis and M. Arnaoutoglou, "Machine learning-based classification of simple drawing movements in Parkinson's disease," *Biomedical Signal Processing and Control,* vol. 31, pp. 174-180, 2017.

[10] J. Mucha, J. Mekyska, M. Faundéz-Zanuy, K. Lopez-de-Ipina, V. Zvončák, Z. Galáž, T. Kiska, Z. Smékal, L. Brabenec and I. Rektorová, "Advanced Parkinson's Disease Dysgraphia Analysis Based on Fractional Derivatives of Online Handwriting.," in *10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT).* , 2018.

[11] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, K. Lopez-de-Ipina, V. Zvoncak, T. Kiska, Z. Smekal, L. Brabenec and I. Rektorova, "Identification and Monitoring of Parkinson's Disease Dysgraphia Based on Fractional-Order Derivatives of Online Handwriting," *Applied Sciences,* vol. 8, p. 2566, 2018.

[12] E. Sesa-Nogueras, M. Faundez-Zanuy and J. Mekyska, "An Information Analysis of In-Air and On-Surface Trajectories in Online Handwriting," *Cognitive Computation,* pp. 1-11, 2012.

[13] I. Podlubny, Fractional differential equations an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications, San Diego: Academic Press, 1999.

[14] A. A. Kilbas, H. M. Srivastava and J. J. Trujillo, "Theory and applications of fractional differential equations," 2006.

[15] L. Breiman, "Random Forests," *Mach. Learn.,* vol. 45, pp. 5-32, 2001.

[16] H. Peng, F. Long and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal.,* vol. 27, pp. 1226-1238, 2005.

[17] J. Pohjalainen, O. Rasanen and S. Kadioglu, "Feature Selection Methods and Their Combinations in High-Dimensional Classification of Speaker Likability, Intelligibility and Personality Traits," *Comput. Speech Lang.,* 2014.

[18] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochim. Biophys. Acta (BBA),* vol. 405, pp. 442-51, 1975.

[19] M. Thomas, A. Lenka and P. Kumar Pal, "Handwriting Analysis in Parkinson's Disease: Current Status and Future Directions," *Movement Disorders Clinical Practice,* vol. 4, pp. 806-818.

A.13 Computerised Assessment of Graphomotor Difficulties in a Cohort of School-aged Children

Computerised Assessment of Graphomotor Difficulties in a Cohort of School-aged Children

Jiri Mekyska*, Zoltan Galaz*, Katarina Safarova†, Vojtech Zvoncak*, Jan Mucha*, Zdenek Smekal*, Anezka Ondrackova†, Tomas Urbanek†, Jana Marie Havigerova†, Jirina Bednarova† and Marcos Faundez-Zanuy‡
*Department of Telecommunications, Brno University of Technology, 61600 Brno, Czech Republic
Email: mekyska@feec.vutbr.cz
†Department of Psychology, Faculty of Arts, Masaryk University, Arne Novaka 1, 60200 Brno, Czech Republic
‡Escola Superior Politecnica, Tecnocampus, Avda. Ernest Lluch 32, 08302 Mataro, Barcelona, Spain

***Abstract*—Although graphomotor difficulties (GD) are present in up to 30 % of school-aged children, the field of GD diagnosis and assessment is not fully explored and several research gaps can be identified. This study aims to explore the impact of specific elementary graphomotor tasks analysis on the accuracy of computerised diagnosis and assessment of GD. We analysed seven basic graphomotor tasks from 76 children (assessed by special educational counsellors and using the handwriting proficiency screening questionnaire for children HPSQ–C). Employing a differential analysis, we observed that the most discriminative tasks are based on combined loops, sawtooth and small Archimedean spiral drawings. Features with the highest discrimination power quantify kinematics, especially in the vertical projection. Using a multivariate mathematical model, we were able to identify GD with 50 % sensitivity and 90 % specificity, and to estimate the total score of HPSQ–C with 31 % error.**

***Index Terms*—computerised analysis, digitizer, graphomotor difficulties, graphomotor elements, machine learning, online handwriting**

I. Introduction

A combination of motor planning and execution, visual–perceptual abilities, orthographic coding, kinesthetic feedback, and visual–motor coordination is referred to as graphomotor skills [1], [2]. These skills start to develop in kindergarten at the age < 6 years, level off at the age of 7–8 years (typically 1st and 2nd class of an elementary school) and become automatic at the age of 8–9 years (3rd and 4th class) [3], [4]. It is estimated that children spend 31–60 % of their school day performing handwriting [5], therefore the acquisition of these skills is crucial for the consequent academic success and children' self-esteem [6]. Nevertheless, 10–30 % of them are associated with graphomotor difficulties (GD) or disturbance in the production of written language, i.e. they are considered as poor writers or as having dysgraphia [1], [2].

Dysgraphia or GD are nowadays diagnosed mainly subjectively based on experiences of special educational counsellors or following some tests/questionnaires such as Concise Assessment Scale for Children's Handwriting (BHK) [7], [8] or Handwriting Proficiency Screening Questionnaire (HPSQ) [9]. The latter one was also modified to HPSQ–C, where children assess themselves in three domains: legibility, performance time, and physical and emotional well-being [10]. In addition, all three tests we also extended by computerised analysis allowing advanced assessment of GD [8], [11], [12].

The computerised analysis is usually based on the processing of online handwriting signals acquired by digitising tablets. The online handwriting is handwriting associated with time series [13], i.e. it allows to capture kinematics and usually also dynamics in terms of pen pressure, tilt and azimuth [11]. It enables to go beyond the limitations of human perception and accurately assess handwriting characteristics such as velocity, acceleration, strokes' duration, etc. This technology has been advantageously used in several research studies linked with GD. E.g. Asselborn et al. modelled online handwriting data of 298 children (56 with dysgraphia) with 96.6 % sensitivity and 99.2 % specificity [8]. Mekyska et al. introduced a new intra-writer normalisation method and trained a model diagnosing dysgraphia with 96 % sensitivity/specificity (in a cohort of 27 school-aged children) and estimating the HPSQ total score with 10 % error [11]. Finally, in a cohort of 99 third-class students, Rosenblum et al. trained a model achieving 90 % sensitivity and specificity [14].

Drawing is an early form of a child's graphomotor skill [15]. Its acquisition is necessary for further development, therefore we hypothesize that identification of disturbances in graphomotor elements (contained in most of the alphabet letters) could lead to a general diagnosis of GD as well. Nevertheless, to the best of our knowledge, there is no complex research focused on the utilisation of specific graphomotor tasks in this field of science. Therefore, the general goal of this study is to explore the impact of specific elementary graphomotor tasks analysis on the accuracy of computerised diagnosis and assessment of GD. More specifically, we aim to:

1) identify online handwriting features that significantly differentiate children with and without GD, as assessed by special educational counsellors or children themselves,
2) identify an elementary graphomotor task that provides high discrimination power,
3) train and evaluate multivariate mathematical models that diagnose and/or assess GD automatically.

This work was supported by the grant of Czech Science Foundation 18-16835S (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing) and project LO1401. For the research, infrastructure of the SIX Center was used.

978-1-7281-5764-1/19/$31.00 ©2019 IEEE

II. Dataset and Methods

A. Participants

In total 76 children were enrolled in several elementary schools spread around the Czech Republic. They attended 1st class (3 girls, 12 boys), 2nd class (5 girls, 11 boys), 3rd class (11 girls, 6 boys), and 4th class (17 girls, 11 boys) and reported Czech as their native language. The children were split into experimental (E; children with GD) and comparative (C; children without GD) groups based on two approaches: 1) diagnosis made by a special educational counsellor (SEC); 2) diagnosis using total score of the HPSQ–C questionnaire (based on the 18-point cut-off score validated for Czech cohort). Demographic information of these groups can be found in Table I. Parents of all children signed an informed consent form. The study was approved by the Ethics Committee of Masaryk University. Besides, we followed the Ethical Principles of Psychologists and Code of Conduct released by the American Psychological Association (see https://www.apa.org/ethics/code/).

TABLE I
Demographic Characteristics of the Participants

	SEC criterion		HPSQ–C criterion		
	N	Age [y]	N	Age [y]	HPSQ–C
Experimental group					
Girls	13	9.46±0.97	12	9.77±0.80	21.67±2.67
Boys	2	10.08±0.57	7	10.22±0.80	22.00±3.61
Comparative group					
Girls	23	9.73±1.13	24	9.18±2.24	10.88±3.40
Boys	38	8.53±1.44	33	8.26±1.31	10.15±3.43

B. Data Acquisition

The children were asked to perform a drawing protocol on an A4 paper, that was laid down and fixed to a digitising tablet Wacom Intuos Pro L (PHT-80). For this purpose, they used a Wacom Inking pen that enabled them to have immediate visual feedback and feeling like they write with a conventional inking pen. Online drawings were sampled with frequency $f_s = 150\,\text{Hz}$.

The protocol contains 7 elementary graphomotor tasks (see Fig. 1): TSK1 – Archimedean spiral (approximately 15 cm heigh); TSK2 – Archimedean spiral (half the size of TSK1); TSK3 – connected loops; TSK4 – flipped TSK3; TSK5 – saw-tooth; TSK6 – rainbow; TSK7 – a combination of TSK3 and TSK4. Each task was shown to a child and then she/he replicated it on a blank sheet of paper with a comfortable speed. This protocol was designed in cooperation with psychologists and special educational counsellors. It was designed in a way so that it reflects all coordinated elementary movements that are needed to successfully write cursive letters. In other words, cursive letters are based on these graphomotor elements. An example of TSK7 performed by a child with and without GD can be seen in Fig. 2 (both children attend the 2nd class). One can immediately observe that the child with GD cannot keep the same height and vertical position of individual loops, he/she is not able to draw it without pen elevations (there is in-air movement; for more information see Section II-C) and he/she has difficulties in transitions from upper to lower loops.

C. Drawing Analysis

Although we collected drawings, we can still consider them as online handwriting signals. More specifically, the digitizer captures this information: x and y position ($x[n]$ and $y[n]$); timestamp ($t[n]$); a binary variable ($b[n]$), being 0 for in-air movement (i.e. movement of pen tip up to 1.5 cm above the tablet's surface) and 1 for on-surface movement (i.e. movement of pen tip on the paper), respectively; pressure exert on the tablet's surface during writing ($p[n]$); pen tilt ($a[n]$); azimuth ($az[n]$). For more information about these signals, we refer to e.g. [11], [12].

During parameterisation of the drawings, we focused on the most commonly used online handwriting features, that could be split into five categories:

- spatial – width (WIDTH), height (HEIGHT), and length (LEN) of the whole product, as well as its particular strokes, i.e. stroke width (SWIDTH), height (SHEIGHT), and length (SLEN).
- temporal – duration of drawing (DUR).
- kinematic – velocity (VEL), acceleration (ACC), and jerk (JERK).
- dynamic – pressure (PRESS), tilt (TILT), and azimuth (AZIM).
- other – number of interruptions (pen elevations; NINT) and relative number of interruptions (RNINT).

Spatial, temporal and kinematic features were extracted from both on-surface and in-air movements. Moreover, kinematic features were also analysed in horizontal and vertical projection. Features that are represented by time series, e.g. the velocity profile, were consequently transformed to a scalar value using mean and relative standard deviation (rstd). To make clear how a particular feature was calculated, we will use a notation in format *INF: DIR-FN (HL)*, where *INF* stands for processed information (ON for on-surface, AIR for in-air, PRESS for pressure, TILT for tilt, and AZIM for azimuth), *DIR* denotes direction (H for horizontal and V for vertical), *FN* contains feature name, and *HL* a statistic, that has been used for transformation to a scalar value. For example, AIR: V-ACC (mean) means mean of vertical acceleration during in-air movement.

D. Statistical Analysis

Since several handwriting features did not have the normal distribution (as assessed by the Kolmogorov-Smirnov test) we identified features with the highest discrimination power based on the Mann-Whitney U (Wilcoxon rank sum) test. Handwriting/drawing abilities are very dependent on class a child attends. Therefore, we performed this analysis in 3rd and 4th class separately (we have not done this for children attending 1st and 2nd class, because children with GD are underrepresented in these groups). To evaluate the association

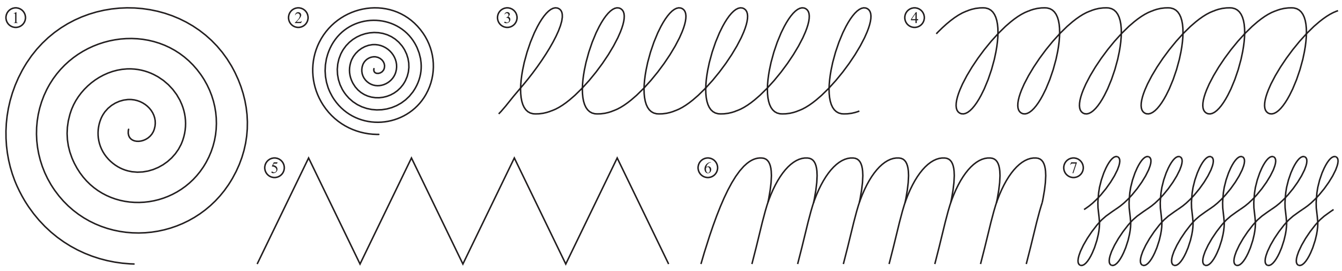


Fig. 1. Drawing acquisition protocol.

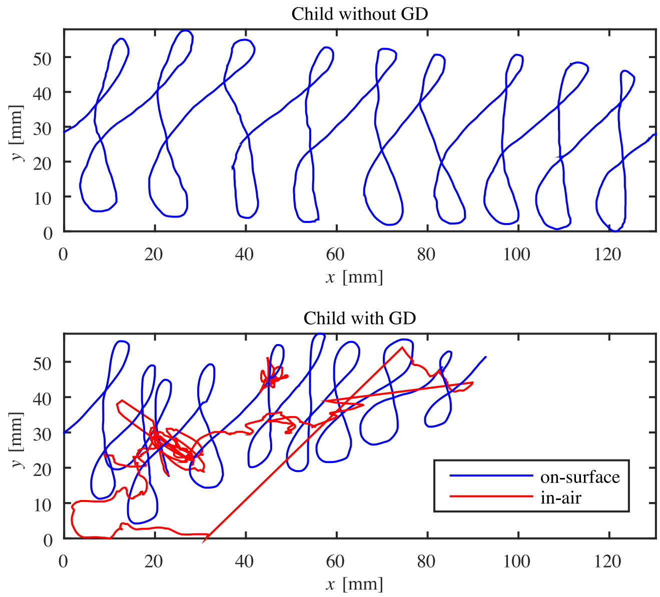


Fig. 2. Example of TSK7 performed by a child without GD (HPSQ–C = 13) and a child diagnosed with GD (HPSQ–C = 23). The blue line represents on-surface movement and the red line the in-air one.

between handwriting features and the total score of HPSQ–C we used the Spearman's correlation.

During the multivariate analysis, we employed an ensemble machine learning algorithm called XGBoost [16]. It is the state-of-the-art algorithm that has been used to win many data science challenges on www.kaggle.com. It is robust to outliers, it performs implicit variable selection, it captures non-linear relationship in data, it is able to capture higher-order interactions among data, and finally, in contrary to e.g. deep learning algorithms, it works well in small datasets too [16]. It was utilised in a 10-fold (with 20 repetitions) cross-validation setup. Information about age, gender, and class were used as an input to the model. Hyper-parameters of XGBoost were selected based on a random search algorithm. Discriminative models were evaluated using accuracy (ACC), sensitivity (SEN), specificity (SPE), and Matthews correlation coefficient (MCC). Regression model (estimating the total score of HPSQ–C) was evaluated based on mean absolute error (MAE), mean squared error (MSE), and estimation error rate (EER, [17]).

III. Results

Results of the Mann-Whitney U test are reported in Table II. For each criterion/class the table provides up to 5 most discriminative features sorted by their p values. In the case of the 3rd class and SEC criteria, the first two features with the highest discrimination power are based on vertical kinematics in the Archimedean spiral (generally instability, as expressed by the relative standard deviation, is higher in the E group), the next three features are spatial characteristics of combined loops (E group performed this task bigger). In the case of HPSQ–C criterion, we identified only two significant features: height of the Archimedean spiral (higher in the E group) and relative standard deviation of on-surface velocity (higher in the C group). In the case of the 4th class and SEC criterion, the first three features are linked with the instability of kinematic characteristics (higher in the E group), and the last two with the number of pen elevations (increased in the E group) during the performance of the combined loop task. Finally, based on the HPSQ–C criterion, all top five most discriminative features are based on kinematics, nevertheless, with inconsistent relative differences between E and C groups.

Results of the correlation analysis are summarised in Table III. In this case, for each class, we selected the top 3 features with the strongest association. Except for the number of interruptions reported in the 1st class, all of them are kinematic parameters. Directions of correlations are not consistent and they depend on a specific task. E.g. in TSK5 (sawtooth) the HPSQ–C total score was increased (more significant handwriting difficulties) with decreased vertical velocity; decreased variation of global velocity, vertical velocity, and vertical jerk; decreased number of elevations; increased variation of acceleration.

Results of the multivariate differential analysis can be found in Table IV. In both scenarios (SEC/HPSQ–C) TSK7 (combined loops) provided the highest discrimination power (based on MCC), nevertheless, sensitivity and specificity are not well balanced. In the SEC scenario, although the model reached 90 % specificity, it has only 47 % sensitivity. Moreover, the sensitivity has a very high standard deviation (45 %). Similarly, in the case of HPSQ–C, we reached SPE = 89 %, but SEN = 50±35 %. The most discriminative TSK7 was in the SEC scenario followed by Archimedean spirals, loops, sawtooth, flipped loops, and rainbow. In the HPSQ–C scenario, TSK7

TABLE II
RESULTS OF MANN-WHITNEY U TEST

SEC criterion				HPSQ–C criterion			
Task	Feature	E > C	p	Task	Feature	E > C	p
3rd class							
TSK1	ON: V-VEL (rstd)	+	0.0007	TSK1	ON: HEIGHT	+	0.0365
TSK1	ON: V-JERK (rstd)	+	0.0007	TSK5	ON: VEL (rstd)	−	0.0365
TSK7	ON: WIDTH	+	0.0007				
TSK7	ON: LEN	+	0.0012				
TSK7	ON: HEIGHT	+	0.0031				
4th class							
TSK5	ON: ACC (rstd)	+	0.0027	TSK5	ON: V-JERK (rstd)	−	0.0003
TSK6	ON: JERK (rstd)	+	0.0083	TSK7	ON: VEL (rstd)	+	0.0076
TSK7	ON: H-JERK (rstd)	+	0.0083	TSK5	ON: V-JERK (mean)	−	0.0118
TSK7	RNINT	+	0.0091	TSK5	ON: ACC (rstd)	+	0.0270
TSK7	NINT	+	0.0142	TSK6	ON: JERK (mean)	+	0.0446

[1] E – experimental group; C – comparative group; + – median of E is greater than median of C; − – median of E is lower than median of C

TABLE III
RESULTS OF CORRELATION ANALYSIS

Task	Feature	ρ	p
1st class			
TSK5	ON: V-VEL (mean)	-0.66	0.0079
TSK5	NINT	-0.60	0.0184
TSK5	ON: V-JERK (rstd)	-0.59	0.0214
2nd class			
TSK7	ON: V-VEL (rstd)	-0.58	0.0192
TSK4	ON: V-JERK (rstd)	-0.58	0.0196
TSK7	ON: V-VEL (mean)	-0.57	0.0216
3rd class			
TSK1	ON: V-ACC (mean)	0.66	0.0036
TSK1	ON: V-JERK (mean)	0.64	0.0058
TSK5	ON: VEL (rstd)	-0.61	0.0093
4th class			
TSK5	ON: ACC (rstd)	0.58	0.0013
TSK5	ON: V-JERK (rstd)	-0.55	0.0026
TSK6	ON: V-VEL (rstd)	-0.53	0.0037

[1] ρ – Spearman's correlation coefficient; p – significance level

TABLE IV
RESULTS OF MULTIVARIATE DIFFERENTIAL ANALYSIS

Task	ACC	SEN	SPE	MCC
SEC criterion				
TSK1	0.83±0.10	0.36±0.41	0.95±0.09	0.32±0.40
TSK2	0.84±0.11	0.34±0.41	0.96±0.08	0.33±0.44
TSK3	0.80±0.12	0.31±0.40	0.92±0.11	0.24±0.42
TSK4	0.79±0.13	0.33±0.41	0.89±0.12	0.22±0.42
TSK5	0.82±0.10	0.28±0.39	0.95±0.08	0.23±0.39
TSK6	0.79±0.10	0.18±0.34	0.94±0.09	0.12±0.36
TSK7	**0.82±0.12**	**0.47±0.45**	**0.90±0.12**	**0.35±0.43**
HPSQ–C criterion				
TSK1	0.69±0.13	0.18±0.26	0.86±0.15	0.05±0.33
TSK2	0.77±0.12	0.37±0.35	0.90±0.12	0.29±0.40
TSK3	0.73±0.13	0.28±0.34	0.88±0.14	0.16±0.37
TSK4	0.79±0.12	0.41±0.36	0.92±0.11	0.35±0.39
TSK5	0.78±0.11	0.32±0.34	0.93±0.11	0.28±0.38
TSK6	0.71±0.13	0.23±0.29	0.87±0.14	0.12±0.37
TSK7	**0.79±0.12**	**0.50±0.35**	**0.89±0.12**	**0.40±0.37**

[1] ACC – accuracy; SEN – sensitivity; SPE – specificity; MCC – Matthews correlation coefficient

was followed by flipped loops, small Archimedean spiral, sawtooth, loops, rainbow, and big Archimedean spiral.

Results of the multivariate regression analysis (estimation of the HPSQ–C total score) are summarised in Table V. Also, in this case, the lowest estimation error rate (EER = 31 %) was observed in TSK7. It was then followed by sawtooth, flipped loops, rainbow, loops, and Archimedean spirals.

TABLE V
RESULTS OF MULTIVARIATE REGRESSION ANALYSIS

Task	MAE	MSE	EER
TSK1	5.58±1.36	47.42±20.26	0.36±0.11
TSK2	5.62±1.43	47.81±20.84	0.36±0.11
TSK3	5.30±1.28	42.60±17.76	0.34±0.12
TSK4	5.10±1.22	38.32±15.34	0.33±0.11
TSK5	4.58±1.13	33.48±14.86	0.30±0.10
TSK6	5.24±1.24	42.23±16.80	0.34±0.11
TSK7	**4.78±1.12**	**33.79±14.30**	**0.31±0.12**

[1] MAE – mean absolute error; MSE – mean squared error; EER – estimation error rate

IV. DISCUSSION

Although the in-air movement has been reported as advantageous information during graphomotor difficulties analysis [18]–[20], in this study, beside the number of interruptions, the in-air features did not play a significant role. The rationale behind this fact is simple. The tasks considered in our protocol are very specific in a way that they can be theoretically performed as one stroke.

Considering horizontal and vertical projections of drawing, the majority of discriminative measures, or features associated with the HPSQ–C total score, is based on the vertical movement. This supports findings of Kushki et al. [3] who suggested that the finger system, which is mainly responsible for the vertical movement, may be more affected by psychological and muscular fatigue than the wrist system, which is more involved in the horizontal one. This can be explained from an anatomical point of view because the vertical movement requires coordinated movement and finer flexions/extensions of more joints (interphalangeal and metacarpophalangeal) and therefore it is more complex than ulnar abductions of the wrist [21], [22].

Except for TSK7 (combined loops), the spatial, temporal and dynamic features were in both differential and correlation analysis outperformed by the kinematic ones. The finding is in line with previous studies utilising these features during the analysis of handwriting difficulties [3], [23]–[25]. It again accents the impact of digitizers during the assessment of GD, because such measures are very difficult to be accurately quantified and perceived by humans.

Concerning specific tasks, we can conclude that the most discriminative one is based on combined loops. In fact, it is the most complex task in our protocol, which requires coordinated movement of fingers, wrist, elbow and shoulder. In addition, it is demanding in terms of visuospatial cognitive functions. The results also suggest that the task, where children draw sawtooth, can also work well during the differential analysis. This task requires a precise change in direction when hitting the top of each tooth. Children in the experimental group were associated with higher instability of acceleration when performing this task. We assume that the children were unstable especially in acceleration between upward and downward strokes, that is again linked with the vertical movement of the finger system. The rest of the conclusions regarding the tasks are dependent on a specific split criterion (SEC/HPSQ–C). E.g. the big Archimedean spiral (TSK1) had a high discrimination power based on the SEC criterion, but very low in the case of HPSQ–C. On the other hand, the small spiral worked well in both cases. We hypothesize that this originates from a fact, that smaller and denser spiral requires more precise coordination of fingers and wrist.

The observed results differ between groups as stratified by SEC or HPSQ–C criteria. Nevertheless, this was expected. The experimental groups have different sample sizes (see Table I), i.e. some children diagnosed with GD by one criterion can be diagnosed as without GD using the second one. The highest classification scores (SEN = 50 %, SPE = 89 %, MCC = 0.40) were reached in the case of HPSQ–C criterion suggesting that this approach is easier to be mathematically modelled (and probably provides lower miss-classification than assessment by special educational counsellors, who could be inconsistent in their decision), however, more studies must be conducted to be able to generalise this conclusion.

As already mentioned, we got a poor trade-off between sensitivity and specificity. Although the specificity was high (usually around 90 %), the sensitivity was low. We identified two possible explanations: 1) The dataset is highly unbalanced, i.e. there are 3–4 times more children in the comparative group, than in the experimental one. Although we employed the Matthews correlation coefficient during model training, it did not prevent possible overfitting to the control group. 2) The dataset has some discrepancies, especially in the case of SEC criterion. Fig. 3 provides a visualisation of high-dimensional data space embedded into two dimensions using the t-distributed stochastic neighbor embedding (t-SNE) method [26]. As can be seen, it is difficult to find a simple hyperplane that would differentiate C and E groups. This could be caused by incorrect classification made by the special educational counsellors. Based on this finding, it would be interesting to perform unsupervised machine learning, cluster datasets into two groups and explore their differences. To sum the issue of poor trade-off up, the model is able to identify children without GD with high probability. The children classified by the model as those having GD should be further examined to confirm this diagnosis.

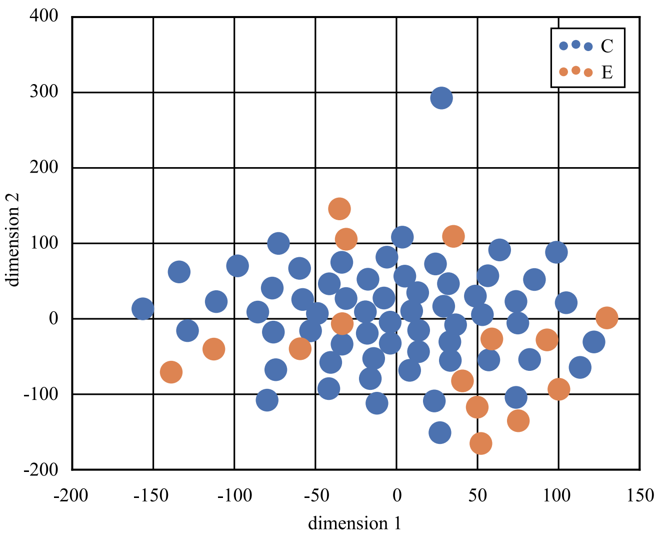


Fig. 3. Visualisation of high-dimensional data split by the SEC criterion using t-distributed stochastic neighbor embedding (C – comparative group; E – experimental group).

Regarding the multivariate regression analysis, the results are still challenging. In our recent study, we modelled the HPSQ–C total score with approximately 15 % error [12], while in this work we reached EER = 31 %. Nevertheless, both studies used different datasets and different tasks (graphomotor elements vs. handwritten paragraph). The results thus cannot be directly compared. We believe that with a larger dataset, more complex acquisition protocol and advanced parameterisation techniques we will be able to further improve the model's accuracy.

V. Conclusion

In the frame of this study, we identified online handwriting features that significantly differentiate children with and without GD, as assessed by special educational counsellors or children themselves (using the HPSQ–C questionnaire). They mostly quantify kinematics, especially in the vertical projection, which requires finer flexions/extensions of interphalangeal and metacarpophalangeal joints. Based on the observed results, to diagnose or assess GD in elementary school children we recommend to utilise the combined loops, sawtooth or small Archimedean spiral drawing tasks. Using the state-of-the-art machine learning approach, we were able to identify GD with 50 % sensitivity and 90 % specificity, which are still challenging numbers.

This work has several limitations. The analysed dataset has a small sample size and it is highly unbalanced, therefore further studies must be conducted to be able to generalise the results. Next, although we have tried the false discovery rate correction in the differential and correlation analysis, no significant features appeared after this adjustment. Finally, we employed just conventional parameterisation. We suppose that the inclusion of more advanced features (e.g. based on fractional-order derivatives, signal decomposition techniques or modulation spectra) could further improve the results. To sum up, concerning the limitations mentioned above, the study should be considered as a pilot one. On the other hand, it bridges a research gap in the field of computerised GD analysis, and to the best of our knowledge, it is the first work exploring the impact of simple graphomotor elements quantification on diagnosis and assessment of GD in school-aged children.

References

[1] A. A. Alhusaini, G. R. Melam, and S. Buragadda, "Short-term sensorimotor-based intervention for handwriting performance in elementary school children," *Pediatrics International*, vol. 58, no. 11, pp. 1118–1123, 2016.

[2] K. P. Feder and A. Majnemer, "Handwriting development, competency, and intervention," *Developmental Medicine & Child Neurology*, vol. 49, no. 4, pp. 312–317, 2007.

[3] A. Kushki, H. Schwellnus, F. Ilyas, and T. Chau, "Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia," *Research in Developmental Disabilities*, vol. 32, no. 3, pp. 1058–1064, 2011.

[4] S. Rosenblum, "Inter-relationships between objective handwriting features and executive control among children with developmental dysgraphia," *PLOS ONE*, vol. 13, no. 4, p. e0196098, 2018.

[5] K. McHale and S. A. Cermak, "Fine motor activities in elementary school: Preliminary findings and provisional implications for children with fine motor problems," *American Journal of Occupational Therapy*, vol. 46, no. 10, pp. 898–903, 1992.

[6] B. Engel-Yeger, L. Nagauker-Yanuv, and S. Rosenblum, "Handwriting performance, self-reports, and perceived self-efficacy among children with dysgraphia," *American Journal of Occupational Therapy*, vol. 63, no. 2, pp. 182–192, 2009.

[7] L. Hamstra-Bletz, J. DeBie, B. Den Brinker *et al.*, "Concise evaluation scale for children's handwriting," *Lisse: Swets*, vol. 1, 1987.

[8] T. Asselborn, T. Gargot, Ł. Kidziński, W. Johal, D. Cohen, C. Jolly, and P. Dillenbourg, "Automated human-level diagnosis of dysgraphia using a consumer tablet," *npj Digital Medicine*, vol. 1, no. 1, p. 42, 2018.

[9] S. Rosenblum, "Development, reliability, and validity of the handwriting proficiency screening questionnaire (HPSQ)," *American Journal of Occupational Therapy*, vol. 62, no. 3, pp. 298–307, 2008.

[10] S. Rosenblum and L. Gafni-Lachter, "Handwriting proficiency screening questionnaire for children (HPSQ–C): Development, reliability, and validity," *American Journal of Occupational Therapy*, vol. 69, no. 3, pp. 6 903 220 030p1–6 903 220 030p9, 2015.

[11] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and rating of developmental dysgraphia by handwriting analysis," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, 2017.

[12] V. Zvoncak, J. Mekyska, K. Safarova, Z. Galaz, J. Mucha, T. Kiska, Z. Smekal, B. Losenicka, B. Cechova, P. Francova *et al.*, "Effect of stroke-level intra-writer normalization on computerized assessment of developmental dysgraphia," in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2018, pp. 1–5.

[13] C. Alonso-Martinez, M. Faundez-Zanuy, and J. Mekyska, "A comparative study of in-air trajectories at short and long distances in online handwriting," *Cogn Comput*, vol. 9, no. 5, pp. 712–20, 2017.

[14] S. Rosenblum and G. Dror, "Identifying developmental dysgraphia characteristics utilizing handwriting classification methods," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 293–298, 2017.

[15] V. Matijević-Mikelić, T. Košiček, M. Crnković, Z. Trifunović-Maček, and S. Grazio, "Development of early graphomotor skills in children with neurodevelopmental risks," *Acta clinica Croatica*, vol. 50, no. 3, pp. 317–321, 2011.

[16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[17] J. Mekyska, Z. Galaz, T. Kiska, V. Zvoncak, J. Mucha, Z. Smekal, I. Eliasova, M. Kostalova, M. Mrackova, D. Fiedorova *et al.*, "Quantitative analysis of relationship between hypokinetic dysarthria and the freezing of gait in Parkinson's disease," *Cognitive computation*, vol. 10, no. 6, pp. 1006–1018, 2018.

[18] S. Rosenblum, S. Parush, and P. Weiss, "The in air phenomenon: Temporal and spatial correlates of the handwriting process," *Percept Mot Skills.*, vol. 96, pp. 933–954, 2003.

[19] S. Rosenblum, S. Parush, and P. L. Weiss, "Computerized temporal handwriting characteristics of proficient and non-proficient handwriters," *American Journal of Occupational Therapy*, vol. 57, no. 2, pp. 129–138, 2003.

[20] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease," *Computer Methods and Programs in Biomedicine*, vol. 117, no. 3, pp. 405–411, 2014.

[21] N. Dounskaia, A. Van Gemmert, and G. Stelmach, "Interjoint coordination during handwriting-like movements," *Experimental Brain Research*, vol. 135, no. 1, pp. 127–140, 2000.

[22] G. P. Van Galen, "Handwriting: Issues for a psychomotor theory," *Human movement science*, vol. 10, no. 2-3, pp. 165–191, 1991.

[23] A. Accardo, A. Chiap, M. Borean, L. Bravar, S. Zoia, M. Carrozzi, and A. Scabar, "A device for quantitative kinematic analysis of children's handwriting movements," in *11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007*, 2007, pp. 445–448.

[24] S.-H. Chang and N.-Y. Yu, "Handwriting movement analyses comparing first and second graders with normal or dysgraphic characteristics," *Research in Developmental Disabilities*, vol. 34, no. 9, pp. 2433–2441, 2013.

[25] S. Rosenblum and H. Roman, "Fluctuation analysis of proficient and dysgraphic handwriting in children," *EPL (Europhysics Letters)*, vol. 85, no. 5, p. 58007, 2009.

[26] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, pp. 2579–2605, 2008.

A.14 New Approach of Dysgraphic Handwriting Analysis Based on the Tunable Q-Factor Wavelet Transform

New Approach of Dysgraphic Handwriting Analysis Based on the Tunable Q-Factor Wavelet Transform

Vojtech Zvoncak*, Jiri Mekyska*, Katarina Safarova†, Zdenek Smekal*, Peter Brezany‡
*Department of Telecommunications, Brno University of Technology, Technicka 10, 61600 Brno, Czech Republic
†Department of Psychology, Faculty of Arts, Masaryk University, Arne Novaka 1, 60200 Brno, Czech Republic
‡Faculty of Computer Science, University of Vienna, Whringer Str. 29, 1090 Wien, Austria

***Abstract*—Developmental dysgraphia is a neurodevelopmental disorder present in up to 30 % of elementary school pupils. Since it is associated with handwriting difficulties (HD), it has detrimental impact on children's academic progress, emotional well-being, attitude and behaviour. Nowadays, researchers proposed a new approach of HD assessment utilizing digitizing tablets. I.e. that handwriting of children is quantified by a set of conventional parameters, such as velocity, duration of handwriting, tilt, etc. The aim of this study is to explore a potential of newly designed online handwriting features based on the tunable Q-factor wavelet transform (TQWT) in terms of computerized HD identification. Using a digitizing tablet, we recorded a written paragraph of 97 children who were also assessed by the Handwriting Proficiency Screening Questionnaire for Children (HPSQ–C). We evaluated discrimination power (binary classification) of all parameters using random forest and support vector machine classifiers in combination with sequential floating forward feature selection. Based on the experimental results we observed that the newly designed features outperformed the conventional ones (accuracy = 79.16 %, sensitivity = 86.22 %, specificity = 73.32 %). When considering the combination of all parameters (including the conventional ones) we reached 84.66 % classification accuracy (sensitivity = 88.70 %, specificity = 82.53 %). The most discriminative parameters were based on vertical movement and pressure, which suggests that children with HD were not able to maintain stable force on pen tip and that their vertical movement is less fluent. The new features we introduced go beyond the state-of-the-art and improve discrimination power of the conventional parameters by approximately 20.0 %.**

***Index Terms*—Handwriting difficulties, developmental dysgraphia; online handwriting; digitizing tablet; tunable Q-factor wavelet transform; machine learning**

I. Introduction

Despite the rapid technological evolution in society, handwriting is still one of the most important life skills that children have to manage in the first years of their school attendance. Fluent and legible handwriting is important for expressing, communicating and recording their ideas [1], [2]. There are many underlying component skills that may interfere with handwriting performance, such as fine motor control, motor planning, in-hand manipulation, visual perception, sustained attention, etc. [3], [4]. It is estimated that 10–30 % of school-aged children suffer from a neurodevelopmental disorder called developmental dysgrafia, which is associated with difficulties mastering handwriting [5]. Handwriting difficulties (HD) can negatively impact academic success, self esteem, emotional well-being, behavior, and attitude [6], [7]. For a benefitial and effective therapy, it is necessary to have an objective methodology that would enable HD identification and complex assessment [8].

For identification and rating of the gravity of possible HD, Rosenblum et al. [9] developed the Handwriting Proficiency Screening Questionnaire for Children (HPSQ–C). The questionnaire apprehends the most significant indicators of dysgraphic handwriting [10], such as legibility, performance time, and physical and emotional well-being. It was already used in several studies considering different language groups [11]–[13]. Although Rosenblum et al. also developed a questionnaire, where the assessment is done by teachers (HPSQ – Handwriting Proficiency Screening Questionnaire for Children) [14], in this study we have decided to focus on HPSQ–C, because children are able to evaluate their handwriting skills better than anyone else. We proved this fact in our recent paper, where we compared results of HPSQ and HPSQ-C scales [15].

Current methods for identification of HD in children are outdated in comparison with methods dealing with reading disorders [16]. Today's trend in clinical assessment of children's handwriting tends to point toward examination of global legibility or specific letter's criteria, such as shape, spacing, position, number of errors, etc. Even though these criteria provide valuable information about handwriting, it's assessment together with administration and pattern searching is time consuming, expensive, and subjective. Moreover, they are usually limited to handwriting product, while the process of handwriting itself is less analysed.

Recent ongoing advancement in biomedical and IT technology enabled a new approach of handwriting analysis based on digitizing tablets (sometimes called digitizers). The digitizers record various signals during handwriting (see Fig. 1): x and y position of a pen when it touches paper's surface (on-surface movement), same coordinates of the pen when it is up to 1.5 cm above the

This work was supported by the grant of Czech Science Foundation 18-16835S (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing) and project LO1401. For the research, infrastructure of the SIX Center was used.

surface (in-air movement), pressure, azimuth and altitude (tilt). We are used to call this kind of handwriting as online, because samples of these signals carry information about time [17].

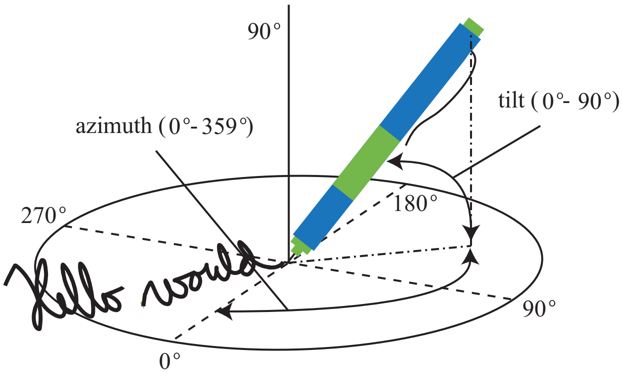


Fig. 1. Online handwriting signals: position (on-surface/in-air), pressure, azimuth, altitude(tilt).

Quantitative analysis of online handwriting is cost-effective, non-invasive, and recently proven diagnostic tool [5], [12], [18]–[23]. For instance, Asselborn et al. [24] reported 96.6 % sensitivity and 92.2 % specificity when diagnosing dysgraphia in a cohort of 298 pupils using BHK test in combination with random forest classifier [24]. Zhiming et al. reached 85.7 % HD identification accuracy employing modified support vector machine classifier in a database of 300 Chinese pupils assessed by HPSQ [11]. Next, in our recent study we were able to estimate total scores of HPSQ–C with approximately 14.77 % error using gradient boosted tree classifier in combination with only 6 specific in-air features (tested on a database of 97 Czech pupils) [15].

Almost all up-to-date scientific works are based on rather basic handwriting features. The advantage of this approach is, that the features could be easily clinically interpretable, which means that they can be linked with specific manifestations of HD. Nevertheless, since the developmental dysgraphia is associated with deficient fine motor skills, poor dexterity, poor muscle tone, or unspecified motor clumsiness, which generally manifests in higher complexity of handwriting, we assume that the conventional features are not able to sufficiently quantify these complexities. This led us to research of new parameterisation methods, that could significantly improve accuracy of computerized HD assessment. We hypothesize, that features based on the tunable Q-factor wavelet transform (TQWT) [25] could better quantify the hidden complexities in dysgraphic handwriting by residual of the decomposition, which should exhibit higher energy for dysgraphic (i.e. more irregular/complex) handwriting. Therefore, the specific aims of this study are:

- to introduce new TQWT based features quantifying handwriting complexity,
- to compare these features to the conventional ones (considered as a baseline) in terms of HD identification accuracy.

Further organization of the paper is as follows. Section II describes dataset and its acquisition process. In addition, it introduces the new TQWT based features and defines a baseline. Finally, this section provides information about employed statistical analysis. The results we achieved are presented in Section III. Discussion of the results and conclusions are given in Section IV and V, respectively.

II. Study & Methods

A. Dataset

For the purpose of this study we enrolled 65 pupils, who were attending 3rd and 4th grade of an elementary school. Almost all of them were right handed (only 2 kids were left handed). Children were asked to fill in the HPSQ–C questionnaire, which consists of 10 questions. Consequently, they were separated into two groups on the basis of a cut-off value derived from the HPSQ–C total score. The experimental group consists of children with higher values of HPSQ–C (i.e. children with HD). Children without HD are considered in the comparative group. Demographic information of children in both groups can be found in Table I. All children used the cursive handwriting and in all cases their parents signed an informed consent form. Thorough the whole study we followed the Ethical Principles of Psychologists and Code of Conduct released by the American Psychological Association (see https://www.apa.org/ethics/code/).

TABLE I
Dataset structure

Gender	N	Age [y]	Mark [-]	HPSQ–C [-]
Experimental group				
girls	16	9.19 ± 0.75	1.61 ± 0.47	20.88 ± 2.16
boys	17	9.18 ± 0.73	1.46 ± 0.38	22.82 ± 4.29
Comparative group				
girls	14	9.21 ± 0.70	1.11 ± 0.21	6.57 ± 2.38
boys	18	9.06 ± 0.73	1.01 ± 0.06	7.83 ± 2.28

N – number; y – years; HPSQ–C – HPSQ–C total score, Mark – mean mark of four major schools subjects (Czech language, Mathematics, English language, Fundamentals of civics and natural science)

B. Data Acquisition

The enrolled children were asked to copy a short paragraph (63 words, 371 characters including spaces), which was selected from a book for 3rd grade. During the acquisition, they were writing on a lined A4 paper, that was laid down and fixed to a digitizing tablet. For this purpose we used Wacom Intuos Pro L (PHT-80) digitizer with Wacom Inking pen. This pen is providing a valuable visual feedback during handwriting, which is entirely similar to the response of a regular inking pen. All signals of online handwriting were sampled with frequency f_s = 150 Hz. An example of the paragraph copy task performed by a pupil with and without HD can be found in Fig. 2.

C. Baseline Handwriting Features

Online handwriting signals were parameterized on a global level (i.e. the whole handwriting), as well as on the stroke one. We extracted the following set of

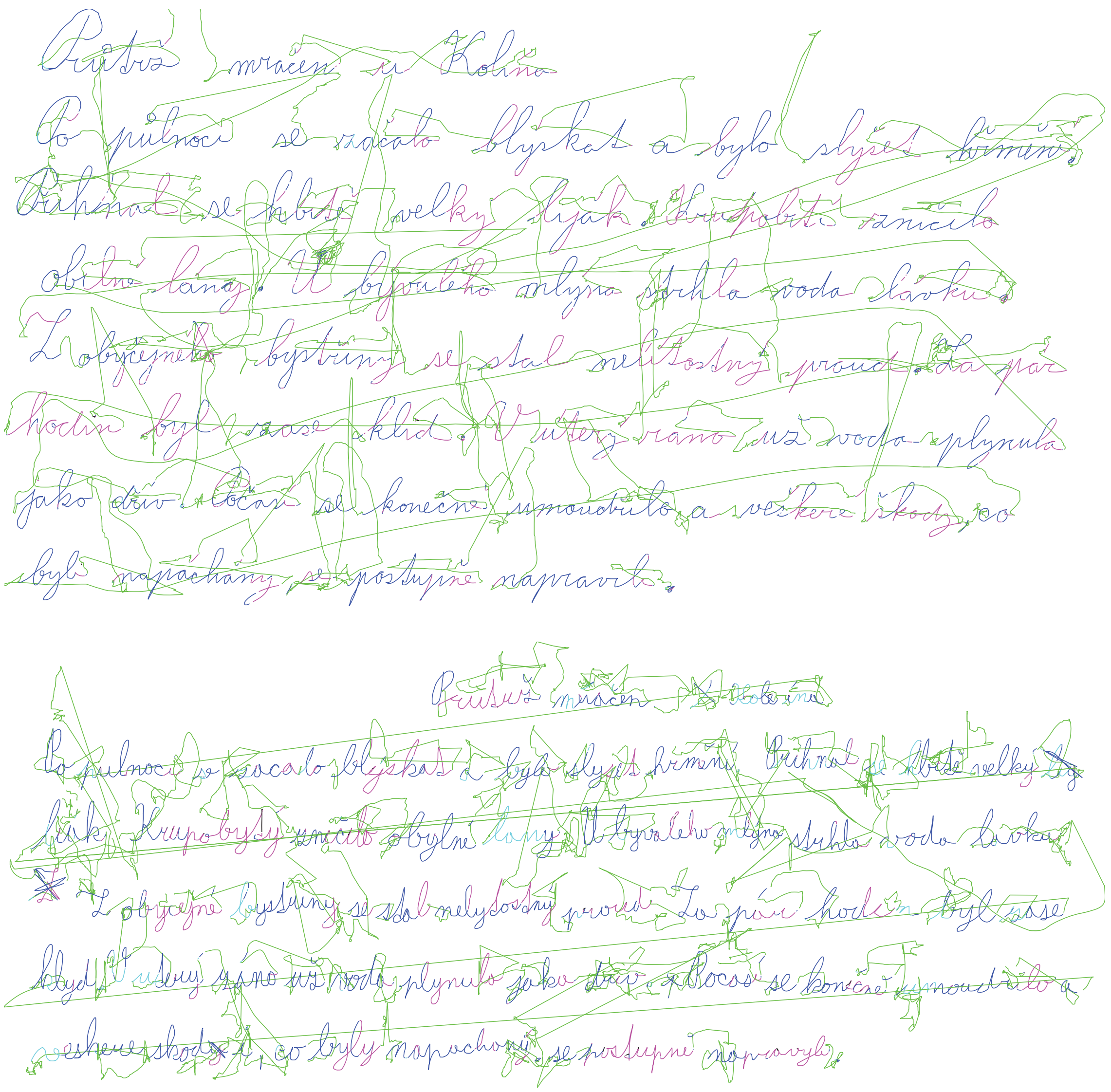


Fig. 2. The paragraph written by a child without HD (HPSQ-C = 3, upper part of the picture) and with HD (HPSQ-C = 35, lower part of the picture). The color of letters is given by the tip pressure of the pen (cyan: 0 – 25 %, blue: 25 – 50 %, purple: 50 – 75 %, black: 75 – 100 %). Green strokes around letters represent the in-air trajectories.

baseline parameters: kinematic (velocity, acceleration, jerk), temporal (duration), spatial (width, height, length of stroke), and dynamic (pressure). Kinematic, temporal and spatial features we calculated from both on-line and in-air movements. We also considered their horizontal and vertical projection. Features represented by a vector were consequently transformed into a scalar value using statistics such as mean, standard deviation, median, relative standard deviation, etc. For more information see [12], [18], [26].

D. Tunable Q-factor Wavelet Transform

The TQWT is a flexible fully-discrete wavelet transform, that can decompose a signal into two components, which represent its oscillatory behavior. With fine tuning we can decompose any signal into a high q-factor component $x_{\mathrm{HQ}}[n]$ and a low q-factor component $x_{\mathrm{LQ}}[n]$ [25]. If we consider the online handwriting signal as $x[n]$, then its TQWT residual part $x_{\mathrm{RES}}[n]$ can be calculated as:

$$x_{\mathrm{RES}}[n] = x[n] - x_{\mathrm{HQ}}[n] - x_{\mathrm{LQ}}[n]. \tag{1}$$

As already mentioned in the introduction, we assume that $x_{\mathrm{RES}}[n]$ contains information, that is linked with deficient fine motor skills, poor dexterity, poor muscle tone, or unspecified motor clumsiness [12]. Therefore, a signal-to-noise ratio (SNR) measure based on $x_{\mathrm{RES}}[n]$ could hypothetically differentiate handwriting associated with or without difficulties.

A clear part of $x[n]$, i.e. without the strong effect of HD, can be calculated using the following formula:

$$x_{\mathrm{CL}}[n] = x[n] - x_{\mathrm{RES}}[n]. \tag{2}$$

Next, we calculated SNR based on the three approaches published at [27]: SNR based on the Teager-Kaiser Energy Operator ($\mathrm{SNR}_{\mathrm{TEO}}$), SNR based on the Conventional

Energy Operator E ($\mathrm{SNR_{CON}}$), and SNR as the energy ratio of $x_{\mathrm{RES}}[n]$ and $x_{\mathrm{CL}}[n]$, i.e.:

$$\mathrm{SNR_E} = 10 \cdot \log_{10}\left(\frac{E(x_{\mathrm{CL}}[n])}{E(x_{\mathrm{RES}}[n])}\right)[\mathrm{dB}]. \quad (3)$$

We applied the TQWT on all raw online handwriting signals (see Fig. 1) and also on velocity, acceleration, and jerk profiles.

Extracting around eight hundred features from each pupil was computationally highly demanding, so we designed our program to work on various computers at the same time. Each computer can be set to extract features only for selected pupils and save its results to a remote server.

E. Statistical Analysis

We considered three scenarios in our study:

1) Baseline: experiments based on the baseline handwriting features (81 in total).
2) Scenario 1: based on the TQWT features (665 in total).
3) Scenario 2: combination of the baseline and TQWT features (774 in total).

Firstly, we performed an exploratory analysis using Pearson's and Spearman's correlation between handwriting features and HPSQ–C total score. In addition, we employed univariate classification analysis, where we tested discrimination power of individual features using support vector machine (SVM with linear kernel) [28] and random forest (RF including 40 trees) [29] classifiers. We considered 10-fold cross-validation with 100 repetitions. The discrimination power was evaluated by accuracy (ACC), sensitiviy (SEN), specificity (SPE), and the Matthew correlation coefficient (MCC).

In the second step, we performed multivariate classification analysis, using a combination of the sequential floating forward selection (SFFS) algorithm [30] and SVM/RF classifiers. We followed the same cross-validation settings and evaluation measures. Due to large computational demands of SFFS, especially in Scenario 2, where a feature space has big dimension, we additionally included minimum redundancy maximum relevance feature selection (mRMR) [31] into the machine learning pipeline (i.e. feature pre-selection). After this pre-selection each model was trained in approximately 1.5 day on a computer with CPU Intel i5 6500 3.2 GHz and 16 GB 1600 MHz DDR3 RAM.

III. Results

Results of the Baseline scenario are shown in Table II. Strongest relationship with the HSPQ–C total score exhibited mean velocity (in-air) with $\rho = -0.38, p = 0.0016$. The best discrimination power had standard deviation of altitude, that was modelled by RF (MCC $= 0.35$, ACC $= 67\,\%$, SEN $= 70\,\%$, SPE $= 65\,\%$). In the multivariate classification analysis we reached 73 % accuracy (MCC $= 0.49$, SEN $= 80\,\%$, SPE $= 68\,\%$) using 3 features modelled by SVM.

TABLE II
BASELINE SCENARIO

	COR / feature name	Spearman's r		Pearson's ρ	
	M of velocity††	-0.36^{**}		-0.38^{**}	
	M of height of stroke‡‡	0.37^{**}		0.34^{**}	
	Duration of writing††	0.31^{*}		0.28^{*}	
C	**UCA / feature name**	**ACC [%]**	**SEN [%]**	**SPE [%]**	**MCC [-]**
S	M of velocity††	64.1±17.4	70.6±25.7	59.3±28.9	0.3±0.4
S	M of height of stroke‡‡	63.5±18.2	65.0±28.1	63.7±28.2	0.3±0.4
S	M of jerk††	61.6±17.2	84.3±20.9	40.7±28.5	0.3±0.4
R	Std of altitude	66.7±17.6	69.6±27.4	65.0±28.1	0.3±0.4
R	Length of writing‡‡	65.2±18.5	59.8±28.5	71.7±27.7	0.3±0.4
R	M of vertical jerk‡‡	63.8±18.0	59.0±28.7	70.5±28.1	0.3±0.4
	MCA / feature name	**ACC [%]**	**SEN [%]**	**SPE [%]**	**MCC [-]**
S	SHOS	66.7±17.4	75.6±26.0	58.1±29.3	0.3±0.4
S	M of velocity††	66.6±18.9	76.1±26.4	59.1±29.8	0.3±.4
S	M of duration of stroke‡‡	73.5±17.1	80.8±23.5	68.8±27.4	0.5±0.3
R	Std of altitude	66.7±17.9	69.1±27.3	65.4±29.0	0.3±0.4
R	SVnJ††	65.6±18.4	69.2±27.7	64.3±29.3	0.3±0.4
R	Std of length of stroke‡‡	69.5±17.7	71.8±27.2	68.3±27.9	0.4±0.4

COR – Correlation analysis, C – Classificator, UCA – Univariate Classification Analysis, MCA – Multivariate Classification Analysis, $\dagger\dagger$ – In-air, $\ddagger\ddagger$ – On-surface, R – Random Forest Classifier, S – Support Vector Machine, M – Mean, Std – Standart deviation, SHOS – Std of height of the stroke $\ddagger\ddagger$, SVnJ – Std the of vertical normalized jerk, * – $p < 0.05$, ** – $p < 0,01$, *** – $p < 0.001$.

Table III reports results of Scenario 1. In correlation analysis $\mathrm{SNR_E}$ of vertical normalized jerk (in-air) achieved the highest value of $\rho = 0.37$ ($p = 0.0022$). The highest discrimination power was observed in 4th moment of $\mathrm{SNR_{CON}}$ extracted from pressure profile and modelled by RF, where ACC $= 68\,\%$ (MCC $= 0.38$, SEN $= 68\,\%$, SPE $= 71\,\%$). Regarding the multivariate analysis, we reached ACC $= 79\,\%$ using 6 TQWT features, that were modelled by SVM (MCC $= 0.58$, SEN $= 86\,\%$, SPE $= 73\,\%$).

Finally, results of Scenario 2 can be found in Table IV. In this case the classification accuracy was further improved to ACC $= 85\,\%$ (MCC $= 0.70$, SEN $= 89\,\%$, SPE $= 83\,\%$), where a feature space containing 9 parameters was modelled by SVM.

IV. Discussion

On the basis of our correlation analysis, we can confirm, that HD manifest in higher energies of TQWT residual signals of online handwriting. This finding is supported, for example, by the positive correlation ($\rho = 0.37, p = 0.002$) between $\mathrm{SNR_E}$ of vertical normalized jerk (in-air) and the HPSQ–C total score. Moreover, the vertical movement during handwriting involves activation of more muscles than in the horizontal case. Therefore, this movement is more complex, requires better handwriting proficiency, and better differentiates children with and without HD. Another finding based on the correlation analysis is, that children with lower values of mean velocity (in-air) exhibit higher HPSQ–C scores. This probably means that children with HD have slower transitions between strokes, which could be linked with cognitive functions.

Since MCC $= 0.49$ in the Baseline scenario was lower than MCC $= 0.58$ in Scenario 1, we can conclude that the TQWT based parameters outperformed the conventional ones. Nevertheless, a combination of both feature sets brought even better results (MCC $= 0.70$), which suggests

TABLE III
SCENARIO 1

COR / feature name	Spearman's r	Pearson's ρ
SNR_{E} of VNJ	0.36^{**}	0.37^{**}
SNR_{TEO} of RIRP	-0.31^{*}	-0.31^{*}
SNR_{CON} of 95PXP	-0.26^{*}	-0.30^{*}

C	UCA / feature name	ACC [%]	SEN [%]	SPE [%]	MCC [-]
S	SNR_{TEO} of STY	65.3±17.7	71.4±25.7	63.3±28.9	0.3±0.4
S	SNR_{CON} of SEY	65.3±17.9	67.7±28.4	66.2±27.5	0.3±0.4
S	SNR_{E} of HS	65.3±17.6	71.6±26.1	60.4±29.5	0.3±0.4
R	SNR_{CON} of 4MP	68.1±18.1	68.0±27.8	70.6±27.5	0.4±0.4
R	SNR_{E} of 1OEX	67.9±17.5	69.0±27.7	68.6±27.6	0.4±0.4
R	SNR_{E} of IRY	68.3±16.9	66.7±26.9	70.7±27.2	0.4±0.4

C	MCA / feature name	ACC [%]	SEN [%]	SPE [%]	MCC [-]
S	SNR_{E} of IRX	69.2±18.0	75.4±25.3	64.2±29.0	0.4±0.4
	SNR_{E} of RIXA	76.9±16.2	81.0±23.6	73.9±26.1	0.5±0.3
	SNR_{E} of 95XO	77.0±16.4	82.0±23.1	74.3±26.1	0.5±0.3
	SNR_{E} of 90XO	78.5±15.7	84.3±22.1	73.4±25.8	0.6±0.3
	SNR_{CON} of RRY	78.6±16.1	85.2±20.8	73.0±26.2	0.6±0.3
	SNR_{CON} of RIRYP	79.2±15.6	86.2±20.2	73.3±26.2	0.6±0.3
R	SNR_{E} of IRXS	66.3±17.8	66.3±28.1	67.3±27.9	0.3±0.4
	SNR_{CON} of 1OEYP	74.0±17.0	73.0±26.7	75.8±26.7	0.5±0.4
	SNR_{CON} of RINTQ	74.8±17.3	76.3±25.9	74.5±26.2	0.5±0.4
	SNR_{E} of 90PXP	75.1±16.8	77.3±25.0	74.0±26.8	0.5±0.4
	SNR_{TEO} of 1OEYP	76.5±16.2	78.5±23.7	76.9±24.8	0.5±0.3

COR – Correlation analysis, C – Classificator, UCA – Univariate Classification Analysis, MCA – Multivariate Classification Analysis, †† – In-air, ‡‡ – On-surface, R – Random Forest Classifier, S – Support Vector Machine, M – Mean, Std – Standart deviation, VNJ – vertical normalized jerk ††, STY – Shannon entropy of TEO of y position ‡‡, RIRP – relative interdecile range of pressure p, 95PXP – 95th percentile of x position ††, SEY – Shannon entropy of y position ‡‡, HS – height of stroke ‡‡, 4MP – 4th moment of pressure p, 1OEX – 1st order entropy of x position ††, IRY – interdecile range of y position ‡‡, IRX – interdecile range of x position ‡‡, RIXA – relative interpercentile range of x position ††, 95XO – 95th percentile of x position ‡‡, 90XO – 90th percentile of x position ‡‡, RRY – relative interdecile range of y position ††, RIRYP – relative interdecile range of y position ‡‡, 1OEYP – 1st order entropy of y position ‡‡, IRXS – interdecile range of x position ‡‡, RINTQ – relative interquartile range of y position ‡‡, 90PXP – 90th percentile of x position ‡‡, * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$.

TABLE IV
SCENARIO 2

C	MCA / feature name	ACC [%]	SEN [%]	SPE [%]	MCC [-]
S	SNR_{E} of 90XO	69.5±17.8	76.9±25.7	65.5±27.7	0.4±0.4
	M of velocity ††	77.8±15.5	80.6±22.6	76.1±24.9	0.6±0.3
	Std of jerk ‡‡	82.7±15.0	84.7±22.1	82.2±21.8	0.7±0.3
	SNR_{CON} of 90XO	84.6±14.3	87.7±20.5	82.0±22.5	0.7±0.3
	SNR_{CON} of MADX	84.6±14.1	88.5±18.1	82.1±22.7	0.7±0.3
	SNR_{TEO} of FCY	84.6±14.4	89.0±18.5	82.1±21.7	0.7±0.3
	SNR_{TEO} of FCX	84.5±14.4	89.0±18.0	82.0±22.7	0.7±0.3
	SNR_{CON} of FCX	84.2±14.6	88.7±17.9	82.5±22.3	0.7±0.3
	SNR_{CON} of FCY	84.7±14.3	88.7±18.7	82.5±22.4	0.7±0.3
R	SNR_{E} of IRX	66.0±17.1	66.5±27.4	67.5±27.1	0.3±0.4
	MOSW	75.3±16.1	74.0±26.4	77.7±24.7	0.5±0.3
	SNR_{TEO} of PXY	77.3±16.6	74.3±26.5	80.9±23.8	0.5±0.4
	SNR_{E} of 90XO	78.4±15.7	76.9±24.8	81.2±22.8	0.6±0.3
	SHS	79.6±15.7	79.1±23.7	81.0±24.9	0.6±0.3
	SNR_{E} RINTERP	79.7±15.1	79.2±23.6	81.9±23.6	0.6±0.3
	SNR_{E} of 95XO	80.3±15.1	79.4±24.4	81.9±23.5	0.6±0.3

COR – Correlation analysis, C – Classificator, UCA – Univariate Classification Analysis, MCA – Multivariate Classification Analysis, †† – In-air, ‡‡ – On-surface, R – Random Forest Classifier, S – Support Vector Machine, M – Mean, Std – Standart deviation, 90XO – 90th percentile of x position ‡‡, MADX – mean absolute deviation of x position ††, FCY – first correlation coefficient of y position ††, FCX – first correlation coefficient of x position ††, IRX – interdecile range of x position ‡‡, PXY – position of max. of y position ††, RINTERP – relative interpercentile range of y position ††, 95XO – 95th percentile of x position ‡‡, MOSW – Mean of speed of writing ††, SHS – Std of height of stroke ††, * – $p < 0.05$, ** – $p < 0,01$, *** – $p < 0.001$.

that the conventional parameters still play their significant role in HD analysis.

Regarding the computerized identification of HD, Asselborn et al. [24] reported SEN = 96.6 % and SPE = 99.2 % using 53 features. Zhimming et al. [11] reported SEN = 77 % and SPE = 77 % based on 7 features. Finally, in our recent article we reached SEN = 96 % and SPE = 97 % based on 7 features [12]. We received SEN = 89 % and SPE = 83 % in this study, which is lower than in the recent ones. Nevertheless, since each team used a different set of parameters, tasks and different cohorts, the results are hardly comparable.

Finally, we can observe that the most discriminative TQWT features (see Table III and Table IV) are extracted mainly from the on-surface/in-air x and y trajectories and from the pressure profile. Therefore, we assume that the higher handwriting complexities associated with HD are not that much manifested in tilt and azimuth.

V. CONCLUSION

The general goal of this study is to introduce new TQWT based parameters quantifying handwriting complexity and evaluate these features in terms of HD identification accuracy. The results suggest that the residual signal of TQWT decomposition contains some information about irregularities/complexities linked with HD. More specifically, the major occurrence of these complexities was observed in the x/y trajectories and in the pressure profile. Finally, we found out that the newly introduced parameters (those based on TQWT) improve HD identification accuracy by approximately 20 % (in comparison to the baseline feature set). Unfortunately, we cannot compare our results with achievements of other research teams, because any experiments based on the considered dataset have not been published yet. Nevertheless, our goal was to compare TQWT based parameters with a conventional baseline features.

This work has a couple of limitations. First of all, the dataset does not have a large number of samples, which means that the results cannot be well generalized, nevertheless, it provides an intuition and some pilot conclusions, that can be further developed and evaluated on larger cohorts. Next, the TQWT decomposition can be tuned using several parameters. In the frame of our experiments we followed the settings recommended in [25], however, some kind of optimization could further increase the discrimination accuracy. The children assessed themselves by HPSQ–C, which means that our trained models are dependent on a subjective rating. Finally, the children were recorded only in one session, therefore we are not able to monitor intra-writer variability, and its effect on models' sensitivity and specificity.

REFERENCES

[1] R. P. Erhardt and V. Meade, "Improving handwriting without teaching handwriting: The consultative clinical reasoning process," *Australian Occupational Therapy Journal*, vol. 52, no. 3, pp. 199–210, 2005.

[2] K. McHale and S. A. Cermak, "Fine motor activities in elementary school: Preliminary findings and provisional implications for children with fine motor problems," *American Journal of Occupational Therapy*, vol. 46, no. 10, pp. 898–903, oct 1992.

[3] K. P. Feder and A. Majnemer, "Handwriting development, competency, and intervention," *Developmental Medicine & Child Neurology*, vol. 49, no. 4, pp. 312–317, apr 2007.

[4] Z. Dankovicova, P. Drotar, J. Gazda, and L. Vokorokos, "Overview of the handwriting processing for clinical decision support system," in *2017 IEEE 14th International Scientific Conference on Informatics*. IEEE, nov 2017.

[5] A. Kushki, H. Schwellnus, F. Ilyas, and T. Chau, "Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia," *Research in Developmental Disabilities*, vol. 32, no. 3, pp. 1058–1064, may 2011.

[6] S. T. Peverly, P. C. Vekaria, L. A. Reddington, J. F. Sumowski, K. R. Johnson, and C. M. Ramsay, "The relationship of handwriting speed, working memory, language comprehension and outlines to lecture note-taking and test-taking among college students," *Applied Cognitive Psychology*, vol. 27, no. 1, pp. 115–126, nov 2012.

[7] V. Connelly, S. Campbell, M. MacLean, and J. Barnes, "Contribution of lower order skills to the written composition of college students with and without dyslexia," *Developmental Neuropsychology*, vol. 29, no. 1, pp. 175–196, feb 2006.

[8] M. R. I. Martins, J. A. Bastos, A. T. Cecato, M. de Lourdes Souza Araujo, R. R. Magro, and V. Alaminos, "Screening for motor dysgraphia in public schools," *Jornal de Pediatria*, vol. 89, no. 1, pp. 70–74, jan 2013.

[9] S. Rosenblum and L. Gafni-Lachter, "Handwriting proficiency screening questionnaire for children (HPSQ–c): Development, reliability, and validity," *American Journal of Occupational Therapy*, vol. 69, no. 3, p. 6903220030p1, apr 2015.

[10] L. Hamstra-Bletz and A. W. Blöte, "A longitudinal study on dysgraphic handwriting in primary school," *Journal of Learning Disabilities*, vol. 26, no. 10, pp. 689–699, dec 1993.

[11] Z. WU, T. LIN, and M. LI, "Automated detection of children at risk of chinese handwriting difficulties using handwriting process information: An exploratory study," *IEICE Transactions on Information and Systems*, vol. E102.D, no. 1, pp. 147–155, jan 2019.

[12] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and rating of developmental dysgraphia by handwriting analysis," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, apr 2017.

[13] A. A. Alhusaini, G. R. Melam, and S. Buragadda, "Short-term sensorimotor-based intervention for handwriting performance in elementary school children," *Pediatrics International*, vol. 58, no. 11, pp. 1118–1123, jul 2016.

[14] S. Rosenblum, "Development, reliability, and validity of the handwriting proficiency screening questionnaire (hpsq)," vol. 62, pp. 298–307, 05 2008.

[15] V. Zvoncak, J. Mekyska, K. Safarova, Z. Galaz, J. Mucha, T. Kiska, Z. Smekal, B. Losenicka, B. Cechova, P. Francova *et al.*, "Effect of stroke-level intra-writer normalization on computerized assessment of developmental dysgraphia," in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. IEEE, 2018, pp. 1–5.

[21] S. Rosenblum, A. Y Dvorkin, and P. Weiss, "Automatic segmentation as a tool for examining the handwriting process of children with dysgraphic and proficient handwriting," vol. 25, pp. 608–21, 11 2006.

[16] J. Medwell and D. Wray, "Handwriting automaticity: the search for performance thresholds," *Language and Education*, vol. 28, no. 1, pp. 34–51, mar 2013.

[17] C. Alonso-Martinez, M. Faundez-Zanuy, and J. Mekyska, "A comparative study of in-air trajectories at short and long distances in online handwriting," *Cognitive Computation*, vol. 9, no. 5, pp. 712–720, 2017.

[18] S. Rosenblum and G. Dror, "Identifying developmental dysgraphia characteristics utilizing handwriting classification methods," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 293–298, apr 2017.

[19] S. Rosenblum, S. Parush, L. EPSTAIN, and P. L. Weiss, "Process versus product evaluation of poor handwriting among children with developmental dysgraphia and adhd," 2004.

[20] S. Rosenblum, S. Parush, and P. Weiss, "The In Air Phenomenon: Temporal and Spatial Correlates of the Handwriting Process," *Percept Mot Skills.*, vol. 96, pp. 933–954, 2003.

[22] P. D'Antrassi, I. Perrone, A. Cuzzocrea, and A. Accardo, "A composite methodology for supporting early-detection of handwriting dysgraphia via big data analysis techniques," in *Intelligent Interactive Multimedia Systems and Services 2017*. Springer International Publishing, may 2017, pp. 241–253.

[23] A. Richardson, S. Kraus, P. L. Weiss, and S. Rosenblum, "Coach-cumulative online algorithm for classification of handwriting deficiencies." in *AAAI*, 2008, pp. 1725–1730.

[24] T. Asselborn, T. Gargot, Ł. Kidziński, W. Johal, D. Cohen, C. Jolly, and P. Dillenbourg, "Automated human-level diagnosis of dysgraphia using a consumer tablet," *npj Digital Medicine*, vol. 1, no. 1, aug 2018.

[25] I. W. Selesnick, "Resonance-based signal decomposition: A new sparsity-enabled signal analysis method," *Signal Processing*, vol. 91, no. 12, pp. 2793–2809, dec 2011.

[26] M. Faundez-Zanuy and J. Mekyska, *User-Centric Privacy and Security in Biometrics*. Institution of Engineering and Technology, 2017, ch. Privacy of online handwriting biometrics related to biomedical analysis, pp. 17–39.

[27] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. F. Zanuy, "Decision support framework for parkinson's disease based on novel handwriting markers," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, pp. 1–1, 2015. [Online]. Available: https://doi.org/10.1109/tnsre.2014.2359997

[28] J. Suykens and J. Vandewalle, *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.

[29] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[30] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal.*, vol. 27, no. 8, pp. 1226–1238, 2005.

[31] A. Tsanas, M. A. Little, and P. E. McSharry, *A Methodology for the Analysis of Medical Data, Chapter 7 in Handbook of Systems and Complexity in Health*. Springler, 2013.

A.15 Analysis of Parkinson's Disease Dysgraphia Based on Optimized Fractional Order Derivative Features

Analysis of Parkinson's Disease Dysgraphia Based on Optimized Fractional Order Derivative Features

Jan Mucha*, Marcos Faundez-Zanuy†, Jiri Mekyska*, Vojtech Zvoncak*, Zoltan Galaz*‡, Tomas Kiska*, Zdenek Smekal*, Lubos Brabenec‡, Irena Rektorova‡§ and Karmele Lopez-de-Ipina¶
*Department of Telecommunications and SIX Research Centre, Brno University of Technology
Technicka 10, 61600 Brno, Czech Republic
†Escola Superior Politecnica, Tecnocampus
Avda. Ernest Lluch 32, 08302 Mataro, Barcelona, Spain
‡Applied Neuroscience Research Group, Central European Institute of Technology
Masaryk University, Kamenice 5, 62500 Brno, Czech Republic
§First Department of Neurology, Masaryk University and St. Anne's University Hospital
Pekarska 53, 65691 Brno, Czech Republic
¶Department of Systems Engineering and Automation
University of the Basque Country UPV/EHU, Av de Tolosa 54, 20018 Donostia, Spain

***Abstract*—Parkinson's disease (PD) is a common neurodegenerative disorder with prevalence rate estimated to 1.5 % for people age over 65 years. The majority of PD patients is associated with handwriting abnormalities called PD dysgraphia, which is linked with rigidity and bradykinesia of muscles involved in the handwriting process. One of the effective approaches of quantitative PD dysgraphia analysis is based on online handwriting processing. In the frame of this study we aim to deeply evaluate and optimize advanced PD handwriting quantification based on fractional order derivatives (FD). For this purpose, we used 37 PD patients and 38 healthy controls from the PaHaW (PD handwriting database). The FD based features were employed in classification and regression analysis (using gradient boosted trees), and evaluated in terms of their discrimination power and abilities to assess severity of PD. The results suggest that the most discriminative and descriptive information provide FD based features extracted from a repetitive loop task or a sentence copy task (maximum sensitivity/specificity = 76 %, error in severity assessment = 14 %, error in PD duration estimation = 22 %). Next, we identified two optimal ranges for the order of fractional derivative, α = 0.05 – 0.45 and α = 0.65 – 0.80. Finally, we observed that inclusion of pressure, azimuth, and tilt together with kinematic features into mathematical modeling has no influence (positive or negative) on classification performance, however, there was a notable improvement in the estimation of PD duration.**

***Index Terms*—online handwriting; Parkinson's disease; dysgraphia; fractal calculus; fractional derivatives; classification; regression**

This work was supported by the grant of the Czech Ministry of Health 16-30805A (Effects of non-invasive brain stimulation on hypokinetic dysarthria, micrographia, and brain plasticity in patients with Parkinsons disease), grant of the Czech Science Foundation 18-16835S (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing) and the following projects: LO1401, FEDER and MEC, TEC2016-77791-C4-2-R from the Ministry of Economic Affairs and Competitiveness of Spain. For the research, infrastructure of the SIX Center was used.

I. Introduction

Parkinson's disease (PD) is a common neurodegenerative disorder affecting approximately 1.5 % of the world population aged over 65 years [1]. The risk of being affected by PD increases with age. Therefore, as populations age, the incidence rate is expected to be doubled in the next 15 years [2]. The exact pathophysiological cause of PD has not yet been discovered, though a rapid degeneration of dopaminergic cells in the substantia nigra pars compacta is the most significant biological finding linked with PD. Tremor at rest, rigidity, bradykinesia and postural instability are considered as the primary motor symptoms of PD [3]. Non-motor symptoms such as cognitive impairment, sleep disturbances, depression, etc. may also arise [4], [5]. Moreover, PD patients usually develop additional axial motor symptoms, e.g. hypokinetic dysarthria, dysphagia, and gait freezing [5].

Considering the primary motor symptoms of PD to be in line with cognitive, perceptual and motor requirements of handwriting, the disrupted handwriting of PD patients may be used as a significant biomarker in PD diagnosis [6]. Especially, by detecting micrographia (progressive decrease of letter's amplitude or width), which is the most commonly observed handwriting abnormality in PD patients [7]. Nevertheless, some PD patients never develop micrographia, but they still exhibit some other handwriting disabilities. Due to this complexity, Letanneux et al. [8] started to use the term PD dysgraphia. To be able to effectively quantify manifestations of PD in handwriting, more advanced approaches were introduced [9], [10]. They are based on digitizing tablets that are able to acquire x and y trajectories along with temporal information (this kind of signal is called online handwriting). Therefore, we are not limited to analyze the spatial features only, but we can process temporal, kinematic or dynamic characteristics.

Researchers have been exploring the influence of many

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

handwriting/drawing tasks in PD dysgraphia analysis, from the simplest ones (loops, circles, lines, Archimedean spiral, etc.) to more complex (words, sentences, drawings, etc.) [10]–[15]. The importance of kinematic features was confirmed by most of the recent works, however, temporal, spatial, dynamic or other more advanced features play their significant role as well. For instance, Drotar et al. [10]–[12] achieved PD classification accuracy up to 89 % using a combination of kinematic, pressure, energy or empirical mode decomposition (EMD) features. Average accuracy of 91 % was achieved by Kotsavasilogloua et al. [16] using kinematic and entropy based features extracted from simple horizontal lines. Some other works reported even higher classification accuracies ($\approx$97 %) [17], [18], but based on a very small dataset. Moetesum et al. [19] published a promising advanced approach by applying convolutional neural networks (CNN) on handwriting data transformed into the offline mode, which resulted in 89 % accuracy. Next, Taleb et al. [9] reported up to 94 % accuracy of PD severity prediction using kinematic and pressure features in combination with adaptive synthetic sampling approach (ADASYN) for model training. Finally, in our recent works [14], [15], [20] we introduced and evaluated a new advanced approach of PD dysgraphia analysis exploiting a fractional order derivative (FD) as a substitution of conventional differential derivative during basic kinematic feature extraction (i.e. velocity, acceleration, and jerk parameters). We achieved up to 90 % classification accuracy employing only 5 FD-based kinematic parameters in these works. Nevertheless, in comparison to conventional parameters, the newly proposed FD-based features yielded better performance only in specific tasks (continuous and/or repetitive movement) and in specific applications such as PD severity estimation.

Therefore, the main objective of this study is to extend our previous findings and perform a deeper and more sensitive analysis of FD-based features, especially in terms of their discrimination power and descriptive abilities. More specifically, we aim to:

- explore the utilization of FD in the other dimensions of online handwriting (i.e. pressure, azimuth, and tilt),
- identify an optimal combination of handwriting/drawing tasks and the FD-based features in terms of discrimination power and descriptive abilities,
- identify an optimal range of FD order α for classification and regression analysis.

The rest of this paper is organized as follows. Section II describes the used dataset and methodology. Results are summarized in Section III. In Section IV the discussion related to the results can be found and the conclusions are drawn in Section V.

II. Dataset and Methodology

A. Dataset

For the purpose of this work, we used the Parkinson's disease handwriting database (PaHaW) [11]. The database consists of several handwriting or drawing tasks acquired in 37 PD patients and 38 age- and gender-matched healthy controls (HC). Demographic and clinical data of the participants can be found in Table I. The participants were enrolled at the First Department of Neurology, St. Anne's University Hospital in Brno, Czech Republic. All participants reported Czech language as their native language and they were right-handed. The patients completed their tasks approximately 1 hour after their regular dopaminergic medication (L-dopa). All participants signed an informed consent form approved by the local ethics committee.

TABLE I
Demographic and clinical data of the enrolled participants.

Gender	N	Age [y]	PD dur [y]	UPDRS V	LED [mg/day]
Parkinson's disease patients					
Females	18	71.23 ± 8.03	9.55 ± 5.29	2.17 ± 0.84	1124.03 ± 535.84
Males	19	67.52 ± 13.15	7.26 ± 4.12	2.37 ± 0.86	1724.12 ± 733.03
All	37	69.32 ± 10.97	8.38 ± 4.80	2.27 ± 0.85	1432.19 ± 704.78
Healthy controls					
Females	18	61.44 ± 9.89	-	-	-
Males	20	63.30 ± 12.79	-	-	-
All	38	62.42 ± 11.39	-	-	-

[1] N – number of subjects; y – years; PD dur – PD duration; UPDRS V – Unified Parkinson's disease rating scale, part V: Modified Hoehn & Yahr staging score [21]; LED – L-dopa equivalent daily dose.

B. Data Acquisition

The PaHaW database [11] includes multiple handwriting tasks, namely: Archimedean spiral; repetitive loops; letter *l*; syllable *le*; Czech words *les*, *lektorka*, *porovnat*, and *nepopadnout*; Czech sentence *Tramvaj dnes už nepojede.* During handwriting tasks performance, the participants were rested and seated in a comfortable position with a possibility to look at a pre-filled template. In case of some mistakes, they were allowed to repeat the task. A digitizing tablet (Wacom Intuos 4M) was overlaid with an empty paper and the participants wrote on that using the Wacom Inking pen. Online handwriting signals were recorded with $f_s = 150$ Hz sampling rate. The following time sequences were acquired: x and y coordinates – $x[t]$, $y[t]$; time-stamp – t; on-surface (i.e. on paper movement) and in-air (i.e. movement up to 1.5 cm above the paper) status – $b[t]$; pressure – $p[t]$; azimuth $az[t]$; and tilt $al[t]$.

C. Fractional Derivative

We discovered the potential of FD-based kinematic features in PD dysgraphia analysis in our previous works [14], [15], [20]. By substitution of the conventional differential derivative during feature calculation, we have developed a new advanced approach of handwriting parametrization. Generally, FDs can have wide range of settings and several approaches of approximation (e.g. Caputo, Grünwald-Letnikov) [22]. In this work, we utilized the Grünwald-Letnikov approximation implemented by Jonathan Hadida. A direct definition of FD $D^{\alpha}y(t)$ is based on finite differences of an equidistant grid in $[0, \tau]$ assuming that the function $y(\tau)$ satisfies certain smoothness conditions in every finite interval $(0, t), t \leq T$. Choosing the grid [22]

$$0 = \tau_0 < \tau_1 < ... < \tau_{n+1} = t = (n+1)h \quad (1)$$

with

$$\tau_{k+1} - \tau_k = h \quad (2)$$

and using the notation of the finite differences

$$\frac{1}{h^\alpha}\Delta_h^\alpha y(t) = \frac{1}{h^\alpha}\left(y(\tau_{n+1}) - \sum_{v=1}^{n+1} c_v^\alpha y(\tau_{n+1-v})\right), \quad (3)$$

where

$$c_v^\alpha = (-1)^{v-1}\binom{\alpha}{v}. \quad (4)$$

The Grünwald-Letnikov implementation is defined as:

$$D^\alpha y(t) = \lim_{h\to 0}\frac{1}{h^\alpha}\Delta_h^\alpha y(t), \quad (5)$$

where $D^\alpha y(t)$ denotes a derivative with order α of function $y(t)$, and h represents sampling lattice.

D. Handwriting Features

The first set of parameters consists of conventional kinematic features extracted from all tasks of the PaHaW database for both on-surface and in-air movement. It means we calculated: *velocity* (rate at which a position of pen changes with time [mm/s]), *acceleration* (rate at which the velocity of pen changes with time [mm/s^2]), *jerk* (rate at which the acceleration of pen changes with time [mm/s^3]), and their horizontal and vertical variants [11], [23]. Next, we calculated the kinematic features based on FD. Moreover, to further extend and improve our previous research, FD was also similarly applied to pressure, azimuth and tilt.

In the first step, the FD-based features were calculated for different values of α in range from 0.1 to 1.0 with the step of 0.1. Next, the most discriminative handwriting tasks were selected and deeper analysed with a finer step of α (0.01). This selection was made in order to reduce computational cost of the analysis. Statistical properties of all extracted handwriting features were expressed using mean, median, standard deviation (std), and maximum (max).

E. Statistical Analysis

To evaluate the discriminative power of the handwriting features, a multivariate binary classification analysis based on the state-of-the-art Gradient Boosted Trees (10-fold cross-validation with 50 repetitions) was employed. More specifically, the famous XGBoost algorithm [24] was used in light of its ability to achieve good performance on a small dataset. Classification performance was evaluated by the Matthew's correlation coefficient (MCC), classification accuracy (ACC), sensitivity (SEN), and specificity (SPE). Next, in order to evaluate the power of handwriting features to estimate values of PD duration and UPDRS V, regression analysis was performed. The same boosting tree algorithm (XGBoost) with the same supervised learning setup was used. Regression performance was evaluated by mean absolute error (MAE), root mean square error (RMSE), and estimation error rate (EER).

III. Results

The results of classification and regression analysis for the FD-based handwriting features extracted from all tasks can be found in Table II. Selection of the most discriminative/descriptive handwriting tasks for the consequent optimization of FD was performed based on feature importances of trained models (feature importance quantifies the relative importance of the feature in an ensemble of the trained XGBoost model [24]). Distribution of particular tasks and derived features for all classification/regression scenarios can be found in Figure 1. Results of the classification/regression analysis after the fine tuning of FD are reported in Table III. Finally, distributions of the FD order α among the fine-tuned parameters are visualized in Figure 2.

TABLE II
RESULTS OF CLASSIFICATION AND REGRESSION ANALYSIS BASED ON ALL TASKS

Classification				
MCC	ACC [%]	SEN [%]	SPE [%]	Feat
0.62 ± 0.14	80.60 ± 9.87	79.41 ± 14.52	80.56 ± 7.25	18
Regression				
Scale	EER [%]	MAE	RMSE	Feat
UPDRS V	12.98 ± 7.01	0.55 ± 0.29	0.66 ± 0.42	3
PD duration	25.23 ± 3.65	4.42 ± 0.64	5.33 ± 0.89	30

[1] MCC – Matthew's correlation coefficient; ACC – accuracy; SEN – sensitivity; SPE – specificity; Feat – number of features important for the trained model; MAE – mean absolute error; RMSE – root mean squared error; EER – estimation error rate; UPDRS V – Unified Parkinson's disease rating scale, part V: Modified Hoehn & Yahr staging score [21].

TABLE III
RESULTS OF CLASSIFICATION AND REGRESSION ANALYSIS FOR SELECTED TASKS

Classification					
Task	MCC	ACC [%]	SEN [%]	SPE [%]	Feat
Sentence	0.34 ± 0.18	66.67 ± 12.45	65.79 ± 18.12	65.79 ± 21.58	21
Rep. loops	0.52 ± 0.11	76.00 ± 11.98	75.68 ± 12.36	76.32 ± 19.54	11
Regression					
Task	Scale	EER [%]	MAE	RMSE	Feat
Sentence	UPDRS V	14.67 ± 7.44	0.63 ± 0.32	0.78 ± 0.40	1
Rep. loops	UPDRS V	13.94 ± 7.61	0.61 ± 0.33	0.75 ± 0.41	2
Sentence	PD duration	23.73 ± 10.67	4.05 ± 1.82	4.62 ± 1.83	33
Rep. loops	PD duration	21.97 ± 8.97	3.75 ± 1.53	4.36 ± 1.60	39

[1] MCC – Matthew's correlation coefficient; ACC – accuracy; SEN – sensitivity; SPE – specificity; Feat – number of features important for the trained model; MAE – mean absolute error; RMSE – root mean squared error; EER – estimation error rate; UPDRS V – Unified Parkinson's disease rating scale, part V: Modified Hoehn & Yahr staging score [21].

IV. Discussion

Firstly, we performed the analysis using all tasks of the PaHaW database utilizing features calculated for α from 0.1 to 1.0 with step 0.1 (10 FD-based features for one handwriting parameter). As can be seen in the upper part of Table II, ACC (80.60 %) corresponds with our previous results (81.43 %) [14], while SEN and SPE were improved by approximately 10 %. Number of features involved in the trained model is 18, and as can be seen in Figure 1 (bottom part of column a), besides the kinematic features the pressure

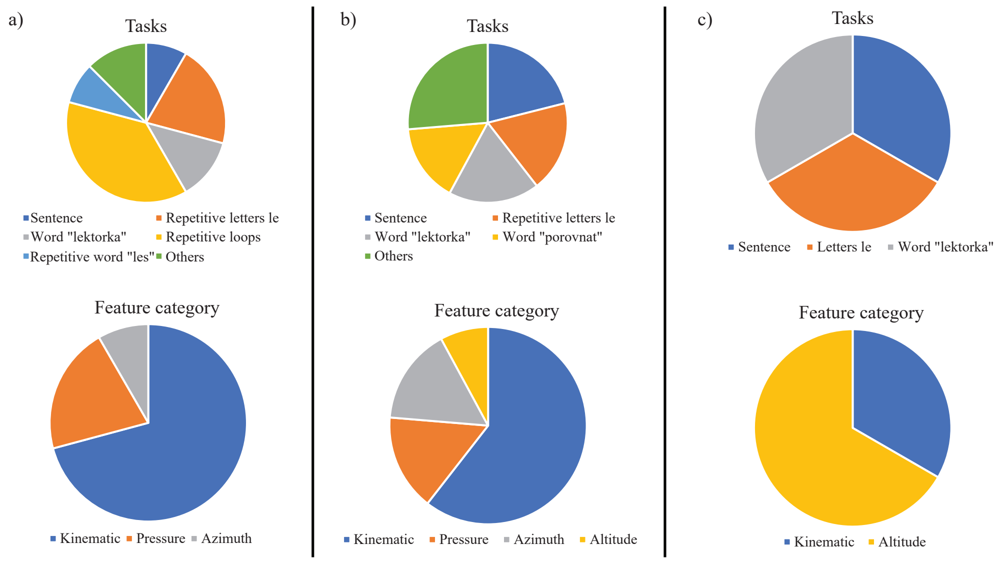


Fig. 1. Distribution of particular tasks and derived features in the trained XGBoost models: a) classification analysis; b) regression analysis (PD duration); c) regression analysis (UPDRS V).

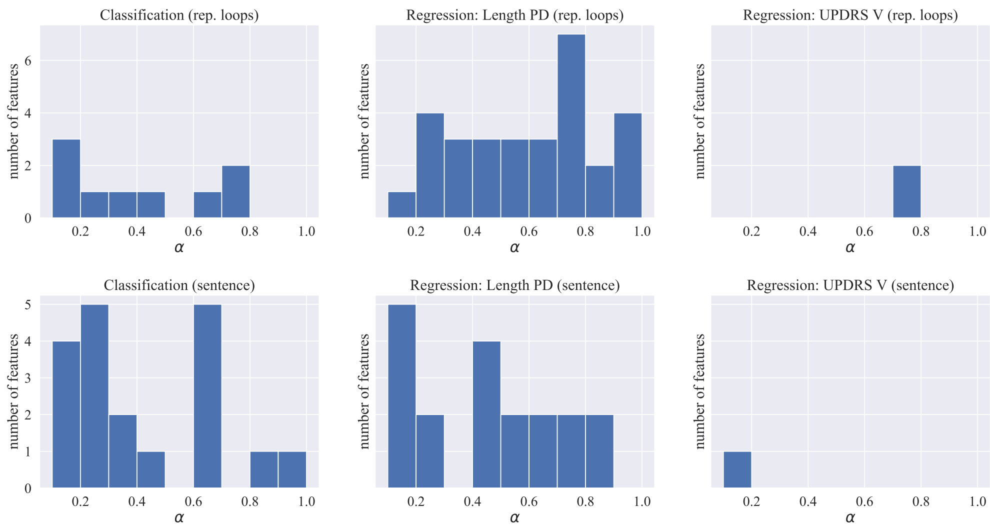


Fig. 2. Distributions of FD order α among the fine-tuned parameters.

and azimuth parameters are also modeled. Based on the distribution reported in the upper part of column a) (see Figure 1), it is noticeable that the highest discriminative power provide repetitive loops. Regarding the results of regression analysis, the most suitable task for further optimization of the FD-based features is the sentence (see the upper part of column b) and c) in Figure 1). In comparison with our previous results [14], the estimation error of PD duration differs minimally, however, the resulted models include parameters coming from all feature categories. In the case of UPDRS V, the value of EER is similar again, but in this case, most of the features are tilt-based instead of kinematic-based. Considering the facts mentioned above, we can conclude that utilizing FD analysis of pressure, azimuth and tilt does not have any noticeable effect on model's performance.

Secondly, we performed the optimization of FD-based features extracted from the repetitive loops and sentence. We re-calculated these features for α from 0.01 to 1.00 with 0.01 step (100 FD-based features for one time sequence) in order to identify the optimal values of α. As can be seen in the upper part of Table III, ACC for both tasks is lower in comparison with the all task classification. It is the consequence of using just a single task for classification, and it corresponds with previous works [10], [11], [14], [20]. Nevertheless, we have to point out that the main objective of this step is not to increase the classification accuracy but to identify the optimal values of α. It is visible from the first column of Figure 2 that the optimal α for PD classification is in ranges from 0.05 to 0.35 and 0.60 to 0.75. Regarding the results of regression analysis, in the case of UPDRS V estimation, EER is slightly worse in comparison with the first step. In the case of PD duration estimation, EER is slightly better (by $2-3.5\,\%$) than in the first step and also in comparison with our previous work [14] it was improved by 5 %. These results are probably caused by the usage of fine-tuned FD-based features. From the middle and last column in Figure 2, we may conclude that the optimal value of α for PD severity assessment and duration estimation is in ranges from 0.05 to 0.45 and from 0.65 to 0.80. By intersectioning optimal α ranges of classification and regression analysis, we created a final optimal range of α from 0.05 to 0.45 and from 0.60 to 0.80, that is recommended to be used in the field of PD dysgraphia analysis.

V. Conclusion

Based on the results we can conclude that applying FD on pressure, azimuth and tilt profiles has no influence (negative or positive) on classification performance. However, there was a notable improvement in the estimation of PD duration by 19 %. Next, in the field of PD dysgraphia analysis, we identified the optimal values of the FD order, which should be in the range from 0.05 to 0.45 or from 0.60 to 0.80. Identification of these ranges enables significant reduction of computational cost (by approximately 50 %), because researchers do not have to explore the full range of possible values of the FD order during quantitative analysis of PD dysgraphia.

This study has several limitations and possible parts, that could be further improved/explored. Since the processed dataset is small, further studies on this topic should be held in order to generalize the results. Next, the FD order could be further tuned for horizontal and vertical movement separately. And finally, some other approximations of FD (e.g. Caputo's) can further improve classification or regression performance.

References

[1] D. Berg, R. B. Postuma, C. H. Adler, B. R. Bloem, P. Chan, B. Dubois, T. Gasser, C. G. Goetz, G. Halliday, L. Joseph *et al.*, "Mds research criteria for prodromal Parkinson's disease," *Movement Disorders*, vol. 30, no. 12, pp. 1600–1611, 2015.

[2] L. Hirsch, N. Jette, A. Frolkis, T. Steeves, and T. Pringsheim, "The incidence of parkinson's disease: a systematic review and meta-analysis," *Neuroepidemiology*, vol. 46, no. 4, pp. 292–300, 2016.

[3] O. Hornykiewicz, "Biochemical aspects of Parkinson's disease," *Neurology*, vol. 51, no. 2 Suppl 2, pp. S2–S9, 1998.

[4] M. M. Hoehn and M. D. Yahr, "Parkinsonism: Onset, Progression, and Mortality," *Neurology*, vol. 17, pp. 427–42, 1967.

[5] J. Mekyska, Z. Galaz, T. Kiska, V. Zvoncak, J. Mucha, Z. Smekal, I. Eliasova, M. Kostalova, M. Mrackova, D. Fiedorova *et al.*, "Quantitative analysis of relationship between hypokinetic dysarthria and the freezing of gait in parkinsons disease," *Cognitive computation*, vol. 10, no. 6, pp. 1006–1018, 2018.

[6] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. S. di Freca, "Handwriting analysis to support neurodegenerative diseases diagnosis: A review," *Pattern Recognition Letters*, 2018.

[7] J. McLennan, K. Nakano, H. Tyler, and R. Schwab, "Micrographia in Parkinson's Disease," *Journal of the Neurological Sciences*, vol. 15, no. 2, pp. 141–152, 1972.

[8] A. Letanneux, J. Danna, J.-L. Velay, F. Viallet, and S. Pinto, "From micrographia to Parkinson's disease dysgraphia," *Movement Disorders*, vol. 29, no. 12, pp. 1467–1475, 2014.

[9] C. Taleb, M. Khachab, C. Mokbel, and L. Likforman-Sulem, "A reliable method to predict parkinsons disease stage and progression based on handwriting and re-sampling approaches," in *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)*. IEEE, 2018, pp. 7–12.

[10] P. Drotar, J. Mekyska, Z. Smekal, I. Rektorova, L. Masarova, and M. Faundez-Zanuy, "Contribution of different handwriting modalities to differential diagnosis of Parkinson's disease," in *Medical Measurements and Applications (MeMeA), 2015 IEEE International Symposium*, 2015, pp. 1–5.

[11] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," *Artificial Intelligence in Medicine*, vol. 67, pp. 39 – 46, 2016.

[12] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. F. Zanuy, "Decision support framework for Parkinson's disease based on novel handwriting markers," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, pp. 1–1, 2015.

[13] E. Nackaerts, S. Broeder, M. P. Pereira, S. P. Swinnen, W. Vandenberghe, A. Nieuwboer, and E. Heremans, "Handwriting training in Parkinson's disease: A trade-off between size, speed and fluency," *PLOS ONE*, vol. 12, no. 12, p. e0190223, dec 2017.

[14] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, K. Lopez-de Ipina, V. Zvoncak, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Identification and monitoring of parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting," *Applied Sciences*, vol. 8, no. 12, p. 2566, 2018.

[15] J. Mucha, J. Mekyska, M. Faundez-Zanuy, K. Lopez-de Ipina, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Advanced parkinson's disease dysgraphia analysis based on fractional derivatives of online handwriting." in *10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT).*, 2018.

[16] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, and M. Arnaoutoglou, "Machine learning-based classification of simple drawing movements in parkinson's disease," *Biomedical Signal Processing and Control*, vol. 31, pp. 174–180, 2017.

[17] C. Loconsole, G. F. Trotta, A. Brunetti, J. Trotta, A. Schiavone, S. I. Tato, G. Losavio, and V. Bevilacqua, "Computer vision and EMG-based handwriting analysis for classification in Parkinson's disease," in *Intelligent Computing Theories and Application*. Springer International Publishing, 2017, pp. 493–503.

[18] C. Taleb, M. Khachab, C. Mokbel, and L. Likforman-Sulem, "Feature selection for an improved parkinson's disease identification based on handwriting," in *Arabic Script Analysis and Recognition (ASAR), 2017 1st International Workshop on*. IEEE, 2017, pp. 52–56.

[19] M. Moetesum, I. Siddiqi, N. Vincent, and F. Cloppet, "Assessing visual attributes of handwriting for prediction of neurological disorders: A case study on parkinson's disease," *Pattern Recognition Letters*, 2018.

[20] J. Mucha, V. Zvoncak, Z. Galaz, M. Faundez-Zanuy, J. Mekyska, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de Ipina, "Fractional derivatives of online handwriting: A new approach of parkinsonic dysgraphia analysis," in *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2018, pp. 214–217.

[21] S. Fahn and R. L. Elton, *UPDRS Development Committee (1987) Unified Parkinson's Disease Rating Scale*. Recent developments in Parkinson's Disease. Macmillan, Florham Park, 1987.

[22] I. Podlubny, *Fractional differential equations an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*. San Diego: Academic Press, 1999.

[23] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and rating of developmental dysgraphia by handwriting analysis," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, 2017.

[24] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.

A.16 Fractional Order Derivatives Evaluation in Computerized Assessment of Handwriting Difficulties in School-aged Children

Fractional Order Derivatives Evaluation in Computerized Assessment of Handwriting Difficulties in School-aged Children

Vojtech Zvoncak*, Jan Mucha*, Zoltan Galaz*, Jiri Mekyska*, Katarina Safarova†, Marcos Faundez-Zanuy‡, Zdenek Smekal*
*Department of Telecommunications and SIX Research Centre, Brno University of Technology
Technicka 10, 61600 Brno, Czech Republic
†Department of Psychology, Faculty of Arts, Masaryk University
Arne Novaka 1, 60200 Brno, Czech Republic
‡Escola Superior Politecnica, Tecnocampus, Universitat Pompeu Fabra
Avda. Ernest Lluch 32, 08302 Mataro, Barcelona, Spain

***Abstract*—Handwriting difficulties (HD) affects some of the school-aged children and its current prevalence rate is between 5–34 %. Children at primary schools have to face rising cognitive demands that the handwriting represents, and some of them are not able to do so. As a result, they tend to make mistakes and their written product is dysfluent and has poor legibility. HD can also lead them to lower self-esteem, learning difficulties and ultimately to less academic achievements. For this reason occupational therapists are trying to identify HD through examination as early as possible. We extracted online handwriting signals of children using digitizing tablets. Handwriting Proficiency Screening Questionnaire for Children (HPSQ–C) was used to score severity of HD in children's written product. To advance current computerized analysis of online handwriting, we employed fractional order derivative features (FD) together with conventional measures. We selected significant features for HD identification and utilized correlation analysis together with Mann-Whitney U-test to evaluate their discriminative power. We can conclude that FD-based features bring benefits of more robust quantification of in-air movements as opposed to the conventionally used ones. Finally, we have shown that utilization of FD can be beneficial for computerized assessment of HD but should be further optimized and evaluated with advanced statistical or machine learning methods.**

***Index Terms*—fractal calculus; fractional derivative; handwriting difficulties; kinematic analysis; online handwriting; school-aged children; digitizer; developmental dysgraphia**

I. Introduction

In childhood, mastering legible handwriting is an important skill [1]. During this life period, a child has to develop adequate cognitive and motor abilities, such as fine motor control, stroke formation, thumb-to-finger sequencing, visual processing, formulation of an idea, planing a syntax of a sentence, achieving orthographic-motor integration to produce text, and evaluation of the outcome [2]. In fact, many children have problems to withstand rising cognitive demands that the handwriting represents, and are not able to comprehend simultaneous tasks such as grammar, spelling, composition [3], etc. As a result, their written product is dysfluent, it has poor legibility, and the in-air time (time spent above the writing surface) is generally longer [4]. Moreover, these children spend too much effort during handwriting, which leads to low dexterity [5] as well as the lack of fine motor control [6]. This phenomenon is commonly referred to as *handwriting difficulties* (HD) and its prevalence range between 5–34 % [7].

At present, occupational therapists examine HD based on the following criteria [8]: legibility and speed of writing, performance time, quality of letter formation, alignment, number of errors, spacing and sizing of letters, etc. Although the clinical assessment of HD provides valuable information about handwriting, it is still limited to a visual inspection of the written product, which does not provide complete information about the process itself. Besides, such an assessment is also dependent on the examiner's experience, level of expertise, physical and emotional state, etc. These factors result in inter-rater variability and less objectivity of the examination [9].

To overcome the limitations of conventional clinical evaluation and diagnosis of HD, researchers have been focusing on computerized quantitative analysis of online handwriting (where each sample is associated to its timestamp [10]) taking advantage of a variety of signal processing and machine learning techniques [1], [11]–[14]. In terms of the HD quantification, previous studies [6], [15]–[19] have been using conventional feature extraction methods aiming at stroke duration, velocity, acceleration, tilt, pressure, etc.

In our previous works [20]–[22], the potential of fractional order derivatives (FD) for development and application of robust and complex kinematic feature extraction methods in the field of Parkinson's disease dysgraphia analysis was uncovered and evaluated. Therefore, we hypothesize that the utilization of FD for the analysis of HD in children population may also bring a noticeable improvement. With this hypothesis

This work was supported by the grant of Czech Science Foundation 18-16835S (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing), project LO1401, and TEC2016-77791-C4-2-R from the Ministry of Economic Affairs and Competitiveness of Spain. For the research, infrastructure of the SIX Center was used.

978-1-7281-5764-1/19/$31.00 ©2019 IEEE

in mind, we aim at:

- exploring the utilization of FD in the field of computerized analysis of HD in children population,
- comparing the power of the FD-based features with the set of conventionally used ones to discriminate children without HD and children with HD,
- identifying the optimal range of FD α order for robust and complex quantification of HD.

II. Materials & Methods

A. Dataset

In this study, we enrolled 55 children (19 attending 3th grade, and 36 attending 4th grade of primary schools), see Table I for more information. To assess legibility and performance time during handwriting as well as physical and emotional well-being, the children were asked to fill a self-evaluating Handwriting Proficiency Screening Questionnaire for Children (HPSQ–C) [23]. It contains 10 questions scored on a 5-point Likert scale (0 – no difficulties, 4 – severe difficulties; total score, i. e. sum over all questions: 0 – no HD, 40 – severe HD). The important advantage of HPSQ–C is its language independence and the fact that it has already been validated in a couple of previous studies [11], [12], [19], [24]. Based on the HPSQ–C cut-off scores, the children were separated into two groups: a) children with HPSQ–C < 7 were considered as healthy controls (HC, i. e. no HD); b) children with HPSQ–C $>= 19$ were considered as children with handwriting difficulties (HD). Some of the children, that obtained HSPQ–C scores between these two values, had to be moved into HC or HD group based on the visual inspection of their handwritten product by an independent therapists.

Parents of all the children participating in this study signed an informed consent form, and trough the entire duration of the study, we followed the Ethical Principles of Psychologists and Code of Conduct released by the American Psychological Association (https://www.apa.org/ethics/code/).

B. Data Acquisition

To record the handwriting process, the children were asked to write all 34 letters of the Czech alphabet using cursive lower-case letters on a lined A4 paper attached to an active area of digitizing tablet Wacom Intuos Pro L (PTH-80) (sampling frequency $f_s = 150\,\text{Hz}$), which enabled us to not only inspect the written product but also to record a variety of signals describing the handwriting process: x and y position ($x[n]$ and $y[n]$); timestamp ($t[n]$); a binary variable ($b[n]$; 0 – in-air movement, i. e. movement of pen tip up to $1.5\,\text{cm}$ above the tablet's surface, and 1 – on-surface movement, i. e. movement of pen tip on the paper), pressure exert on the tablet's surface during writing ($p[n]$); pen tilt ($a[n]$); and azimuth ($az[n]$). For more information, we refer to [11], [13]. During data acquisition, all children were also using the Wacom Inking pen, which provides visual feedback as well as a feeling of writing by a regular inking pen. An example of the written product of the alphabet performed by a HC and a child with HD can be seen in Figure 1.

TABLE I
Demographic Characteristics of the Cohort.

clin	mean ± std	min	Q1	Q2	Q3	max
	healthy children					
age	9.13 ± 0.68	8.00	9.00	9.00	10.00	10.00
class	3.67 ± 0.48	3.00	3.00	4.00	4.00	4.00
HPSQ–C	7.07 ± 2.29	3.00	5.25	7.00	8.00	12.00
	children with HD					
age	9.20 ± 0.65	8.00	9.00	9.00	10.00	10.00
class	3.64 ± 0.49	3.00	3.00	4.00	4.00	4.00
HPSQ–C	21.88 ± 3.80	19.00	20.00	20.00	22.00	35.00
	all children					
age	9.16 ± 0.66	8.00	9.00	9.00	10.00	10.00
class	3.65 ± 0.48	3.00	3.00	4.00	4.00	4.00
HPSQ–C	13.80 ± 8.04	3.00	7.00	10.00	20.00	35.00

[1] age is expressed in years
[2] dataset consists of 30 healthy children and 25 children with HD

C. Fractional Order Derivatives

FD is used as a substitution of the conventional differential derivative during feature extraction. Hereby, we have developed a new advanced approach of handwriting parametrization. FDs have a wide range of settings and several approaches of approximation (e. g. Caputo, Grünwald-Letnikov) [25]. In this work, we utilized the Grünwald-Letnikov approximation implemented by Jonathan Hadida. A direct definition of FD $D^\alpha y(t)$ is based on finite differences of an equidistant grid in $[0, \tau]$ assuming that the function $y(\tau)$ satisfies certain smoothness conditions in every finite interval $(0, t), t \leq T$. Choosing the grid [25]

$$0 = \tau_0 < \tau_1 < ... < \tau_{n+1} = t = (n+1)h \tag{1}$$

with

$$\tau_{k+1} - \tau_k = h \tag{2}$$

and using the notation of the finite differences

$$\frac{1}{h^\alpha}\Delta_h^\alpha y(t) = \frac{1}{h^\alpha}\left(y(\tau_{n+1}) - \sum_{v=1}^{n+1} c_v^\alpha y(\tau_{n+1-v})\right), \tag{3}$$

where

$$c_v^\alpha = (-1)^{v-1}\binom{\alpha}{v}. \tag{4}$$

The Grünwald-Letnikov implementation is defined as:

$$D^\alpha y(t) = \lim_{h \to 0} \frac{1}{h^\alpha}\Delta_h^\alpha y(t), \tag{5}$$

where $D^\alpha y(t)$ denotes a derivative with order α of function $y(t)$, and h represents sampling lattice.

D. Handwriting Features

To quantify HD, we used two sets of handwriting features: a) conventional features [2], [11], [26], [27] (used as a baseline feature set); b) features utilizing FD (FD-based features) [20]–[22]. Concerning the conventional features, we extracted kinematic (velocity, acceleration, jerk), temporal (duration) and dynamic (azimuth, altitude) ones from both global as well as stroke-based movements. We also used vertical/horizontal projections together with on-surface/in-air trajectories. Finally,

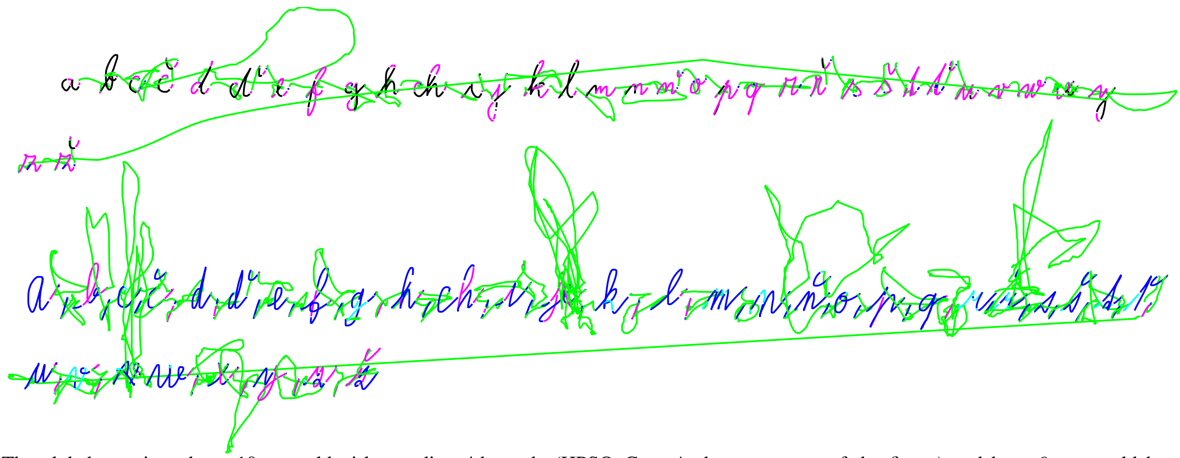


Fig. 1. The alphabet written by a 10-year-old girl attending 4th grade (HPSQ–C = 4, the upper part of the figure) and by a 9 years old boy with HD attending 3th grade (HPSQ–C = 30, the lower part of the figure). The four colors represents the actual tip pressure of the inking pen (cyan: 0 – 25 %, blue: 25 – 50 %, purple: 50 – 75 %, black: 75 – 100 %). The in-air trajectories (inking pen above the tablet's surface) of the inking pen are visualized using the light green colour.

we computed number of interruptions in writing and normalized jerk according to [28]. In terms of FD-based features, we extracted basic kinematic features only, namely velocity, acceleration, jerk and their horizontal and vertical variants. All of the features were computed for α in the range of 0.1–0.9 (step of 0.1) except for $\alpha = 1.0$ as it is covered by the conventional feature set (it equals the full-derivation). Finally, the statistical properties of all the features from both of the feature sets were described by mean and relative standard deviation (relstd).

E. Statistical Analysis

At first, normality of the handwriting features was tested using the Shapiro-Wilk test [29]. Features that were not found to be normally distributed were adjusted using Box-Cox [30] transformation. After that, the distributions of such features were visually re-inspected (some of the features were not fully normalized, however, we hypothesized that such features will not pass the subsequent statistical analysis).

Next, to select only a parsimonious, information-rich subset of the features, we applied a two-step feature selection (FS) before the analysis: a) we used Minimum Redundancy Maximum Relevance (mRMR) [31] algorithm to discard the most redundant features that bring no/very little information; b) we visualized the cross-correlation matrices of the features to discard the ones that have high correlation among each other. With this approach, we reduced the dimension of our feature sets by the following amount: a) conventional feature set: 63 (prior FS), 40 (after FS); and b) FD-based feature set 324 (prior FS), 40 (after FS). The cross-correlation matrices of the best 15 features according to mRMR for both feature sets are visualized in Figure 2.

Subsequently, to assess the strength of a relationship between the values of the handwriting features and the clinical status of the children (HC/HD), and the values of the HPSQ–C (severity of HD), Spearman's correlation coefficient [32] with the significance level of 0.05 was computed. Due to the exploratory nature of this study as well as a relatively small number of the features under investigation, no adjustment for multiple comparisons was made.

Finally, to quantify the ability of the handwriting features to discriminate healthy children and children with HD, Mann-Whitney U-test[1] [33] with the significance level of 0.05 between the handwriting features and clinical status of the children (HC/HD) was used.

III. Results

The results of the correlation analysis can be seen in Table II. In this table, only statistically significant correlations (i. e. those with the p-value bellow 0.05) are shown. As can be seen, the strongest correlations between the conventional handwriting features and the clinical characteristics of the children were found for the following pair/s[2]: a) $\rho = 0.3220^*$ relstd v. acc._s (HC/HD); and b) $\rho = 0.3191^*$ mean h. n. jerk_s (HPSQ–C). The strongest correlation for the FD-based features: a) $\rho = -0.3105^*$ relstd acc. $\alpha = 0.2$_a (HC/HD); and b) $\rho = -0.3405^*$ relstd v. vel. $\alpha = 0.5$_a (HPSQ–C) being the strongest correlated feature-clin. char. pair.

Next, the kernel density estimation plots of the 4 best features selected according to the power to distinguish healthy and impaired handwriting assessed by Mann-Whitney U-test are shown in Figure 3 (as well as in the case of the correlation analysis, only the features with the p-value bellow 0.05 were considered). The figure shows both conventional features and FD-based ones: a) conventional feature with the greatest discrimination power: mean l. stroke_a and mean v. n. jerk_s ($p = 0.0110$); b) FD-based feature with the greatest discrimination power: relstd acc. $\alpha = 0.3$_a ($p = 0.0169$).

Finally, the distribution of the order of FD (α) across the best 40 FD-based features selected according to mRMR is drawn in Figure 4.

[1]We did not use Student's t-test because not all features were normally distributed.

[2]Correlation with $p < 0.05$ (*), correlation with $p < 0.01$ (**).

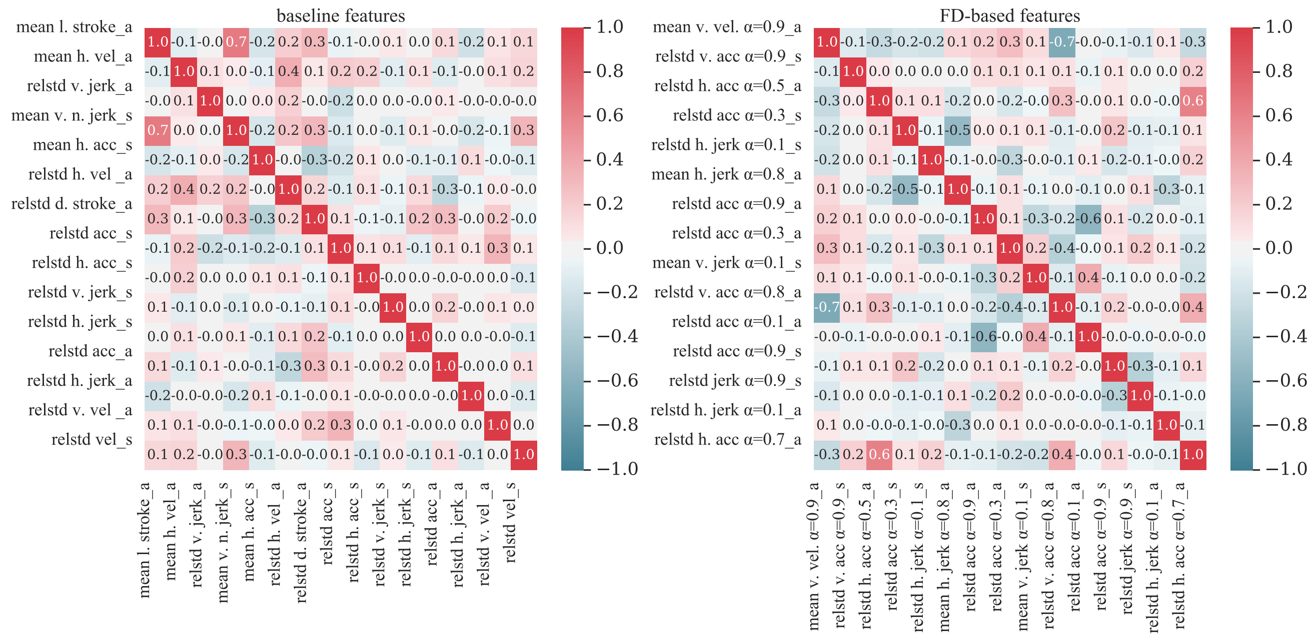


Fig. 2. Cross-correlation matrices of the 15 best features selected according to mRMR: a) conventional features (left side); b) FD-based features (right side). Table convention: vel. – velocity; acc. – acceleration; v. – vertical; h. – horizontal; l. – length; n. – normalized; d. – duration; _a – in air movement; _s – on surface movement.

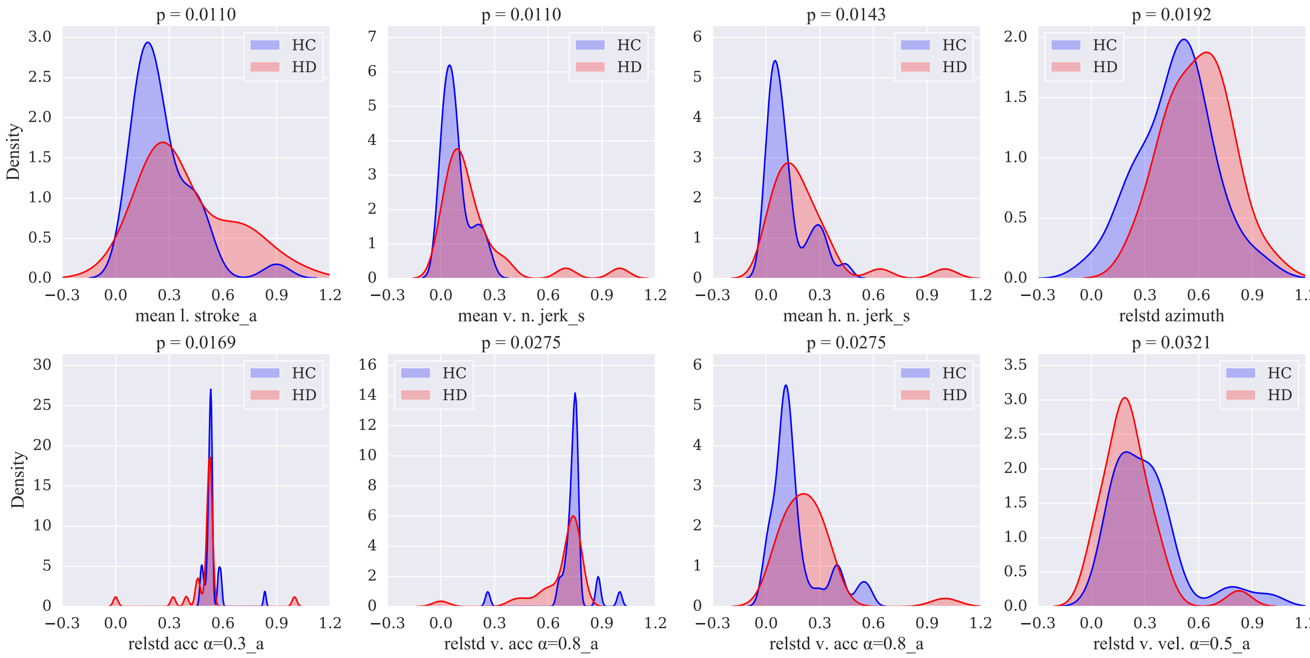


Fig. 3. Kernel density estimation plots of the 4 best features ranked by Mann-Whitney U-test: a) baseline (conventional) features (upper part); b) FD-based features (bottom part). Features are visualized separately for healthy children (HC) and children with HD. On top of each figure, the corresponding p-value is shown. All features were normalized using min-max normalization (min = 0, max = 1) prior the plotting. Figure convention: vel. – velocity; acc. – acceleration; v. – vertical; h. – horizontal; l. – length; n. – normalized; d. – duration; _a – in air movement; _s – on surface movement; p – p-value of Mann-Whitney U-test.

IV. Discussion

The correlation analysis (see Table II) for FD-based features shown that there is a statistically significant relationship between HPSQ–C and relative standard deviation of vertical velocity with α of 0.5, which is in line with the results of [16] reporting that vertical in-air velocity might be a potential

TABLE II
RESULTS OF THE CORRELATION ANALYSIS.

handwriting feature	feature type	clin	ρ	p
relstd v. acc._s	Conv	HC/HD	0.3220	0.0165
mean v. n. jerk_s	Conv	HC/HD	0.3128	0.0200
mean l. stroke_a	Conv	HC/HD	0.3128	0.0200
mean h. n. jerk_s	Conv	HC/HD	0.2990	0.0266
relstd azimuth	Conv	HC/HD	0.2829	0.0363
mean h. n. jerk_s	Conv	HPSQ–C	0.3191	0.0176
mean v. n. jerk_s	Conv	HPSQ–C	0.3058	0.0232
mean d. stroke_a	Conv	HPSQ–C	0.3054	0.0234
relstd azimuth	Conv	HPSQ–C	0.3040	0.0241
relstd v. acc._s	Conv	HPSQ–C	0.2798	0.0385
relstd acc. $\alpha=0.2$_a	FD-based	HC/HD	-0.3105	0.0210
relstd acc. $\alpha=0.3$_a	FD-based	HC/HD	-0.2898	0.0318
relstd v. vel. $\alpha=0.5$_a	FD-based	HPSQ–C	-0.3405	0.0110
relstd acc. $\alpha=0.3$_a	FD-based	HPSQ–C	-0.3150	0.0192
relstd acc. $\alpha=0.2$_a	FD-based	HPSQ–C	-0.2990	0.0266

[1] clin – clinical characteristics, i. e. dependent variable (clinical state: HC/HD, values of HPSQ–C); ρ – Spearman's correlation coefficient; p – p-value of ρ; Conv – conventional feature set; FD-based – FD-based feature set; vel. – velocity; acc. – acceleration; v. – vertical; h. – horizontal; l. – length; n. – normalized; d. – duration; _a – in air movement; _s – on surface movement.

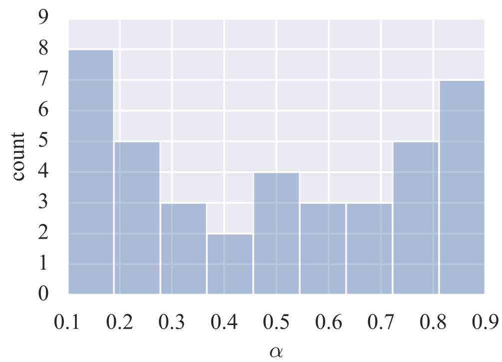


Fig. 4. Distribution of the FD order (α) across the best 40 FD-based features selected according to mRMR (feature selection step applied prior to the analysis).

biomarker for HD identification. With respect to the comparison between the two feature sets, it can be seen that all relevant FD-based features are related with in-air trajectories, more specifically with acceleration and velocity that probably points out to their capability of quantifying hesitating and dysfluent movements during stroke interruptions, which is also coherent with our previous studies [11], [12]. In contrast, three out of the total number of four selected conventional features were computed from the on-surface movements. An important observation to note is the presence of relative standard deviation of azimuth showing that even for a relatively automated task such as the Alphabet writing, the lack of fine motor control together with redundant wrist movements are present in children with HD [4], [5]. Finally, all selected FD-based features have α between 0.2 and 0.5, which suggest that regular derivation is not optimal for temporal handwriting features (acceleration and vertical stroke velocity) and that FD is likely to improve their ability to describe HD.

Regarding the results of Mann-Whitney U-test (see Figure 3), they suggest that the alphabet handwriting task is not very suitable for discrimination of HC and children with HD. When looking at the shape of the probability density function for the 4 selected features in both feature groups, it is obvious that a single feature will not have sufficient discrimination power. With respect to FD-based features, those derived from the acceleration of in-air movements emerged as the most significant ones. This may refer to the difficulties in writing of particular characters of the alphabet, such as long preparation, hesitancy, distress, etc., which can also be seen when inspecting the shapes of the particular characters in the example provided in Figure 1. It is evident that for the child with HD, the on-surface movements are more or less without visible corruptions. However, the difference is eminent for the movements above the tablet's surface.

According to the distribution of the FD α order across the handwriting features that passed the FS (see Figure 4), we were able to identify its optimal range for HC/HD discrimination: 0.1–0.3, and 0.7–0.9, which is also supported by the results of the statistical analysis (the features with the greatest discrimination power and the most statistically significant correlation were computed using the α values from one of those two ranges) and is also in line with our previous study [34] in which we focused on the FD optimization for Parkinson's disease dysgraphia and obtained similar α ranges (0.05–0.45, and 0.6–0.8). Altogether, we can hypothesize there exists some universal optimal range of α suitable for the analysis of corrupted handwriting performance via online handwriting quantification that we need to search for.

V. CONCLUSION

To the best of our knowledge, this is the first study that performs an investigation of the possibilities of using FD in the computerized assessment of HD in school-aged children. We can conclude that FD-based features bring benefits of more robust quantification of in-air movements as opposed to the conventionally used ones. These movements are likely to describe inter-stroke hesitation/s, uncertainty during writing, stiffness of hand/fingers, etc., which can definitely be linked with HD and are imperceptible to an examiner that only sees the written product (even computerized approaches, if not sensitive enough, can be incapable of the precise description of such phenomena).

Although we have shown that utilization of FD can be beneficial for a computerized assessment of HD, several limitations need to be pointed out too. First of all, the alphabet task does not seem to be optimal for the differential analysis, as some of the children's handwriting capabilities and habits are not fully quantified (e. g. copying/writing of words, sentences or paragraphs requires continuous writing, simple graphomotor elements require the application of children's drawing skills, etc. [7], [19]). Next, our dataset consists of only 55 subjects, which is a relatively small number in terms of the statistical validity of the results. Moreover, grouping children in two subject groups (HC/HD) was based entirely on the selection of a cut-off score applied on HPSQ–C, which may not reflect the true nature/presence of HD.

In our future studies, more granular FD α order search (step of 0.01 or even less) as well as investigation of other FD approximations (e. g. Capputo's approximation) will be analyzed. Finally, to investigate the power of FD-based features to not only discriminate HC/HD but also predict the presence/severity of HD in children population, advanced classification and regression models will be trained and evaluated.

REFERENCES

[1] A. P. Accardo, M. Genna, and M. Borean, "Development, maturation and learning influence on handwriting kinematics," *Human movement science*, vol. 32, no. 1, pp. 136–146, 2013.

[2] S. Rosenblum and G. Dror, "Identifying developmental dysgraphia characteristics utilizing handwriting classification methods," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 293–298, 2017.

[3] V. W. Berninger, K. B. Vaughan, R. D. Abbott, S. P. Abbott, L. W. Rogan, A. Brooks, E. Reed, and S. Graham, "Treatment of handwriting problems in beginning writers: Transfer from handwriting to composition." *Journal of Educational Psychology*, vol. 89, no. 4, pp. 652–666, 1997.

[4] S. Rosenblum, S. Parush, and P. Weiss, "The In Air Phenomenon: Temporal and Spatial Correlates of the Handwriting Process," *Percept Mot Skills.*, vol. 96, pp. 933–954, 2003.

[5] S. Rosenblum, P. L. Weiss, and S. Parush, "Handwriting evaluation for developmental dysgraphia: Process versus product," *Reading and Writing*, vol. 17, no. 5, pp. 433–458, Jul. 2004.

[6] S.-H. Chang and N.-Y. Yu, "Handwriting movement analyses comparing first and second graders with normal or dysgraphic characteristics," *Research in Developmental Disabilities*, vol. 34, no. 9, pp. 2433–2441, Sep. 2013.

[7] T. Asselborn, T. Gargot, ukasz Kidziski, W. Johal, D. A. Cohen, C. Jolly, and P. Dillenbourg, "Automated human-level diagnosis of dysgraphia using a consumer tablet," in *npj Digital Medicine*, 2018.

[8] K. P. Feder and A. Majnemer, "Handwriting development, competency, and intervention," *Developmental Medicine & Child Neurology*, vol. 49, no. 4, pp. 312–317, Apr. 2007.

[9] A. Richardson, S. Kraus, P. Weiss, and S. Rosenblum, "Coach - cumulative online algorithm for classification of handwriting deficiencies." vol. 3, 01 2008, pp. 1725–1730.

[10] C. Alonso-Martinez, M. Faundez-Zanuy, and J. Mekyska, "A comparative study of in-air trajectories at short and long distances in online handwriting," *Cognitive Computation*, vol. 9, no. 5, pp. 712–720, 2017.

[11] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and rating of developmental dysgraphia by handwriting analysis," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, Apr. 2017.

[12] V. Zvoncak, J. Mekyska, K. Safarova, Z. Smekal, and P. Brezany, "New approach of dysgraphic handwriting analysis based on the tunable q-factor wavelet transform," in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics*. IEEE, May 2019.

[13] V. Zvoncak, J. Mekyska, K. Safarova, Z. Galaz, J. Mucha, T. Kiska, Z. Smekal, B. Losenicka, B. Cechova, P. Francova, M. Faundez-Zanuy, and S. Rosenblum, "Effect of stroke-level intra-writer normalization on computerized assessment of developmental dysgraphia," in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*. IEEE, Nov. 2018.

[14] S. Rosenblum, "Inter-relationships between objective handwriting features and executive control among children with developmental dysgraphia," *PLOS ONE*, vol. 13, no. 4, p. e0196098, Apr. 2018.

[15] D. Morello, M. Ajcevic, I. Perrone, and A. Accardo, "Analysis of kinematic parameters relationships in normal and dysgraphic children," in *World Congress on Medical Physics and Biomedical Engineering 2018*. Springer, 2019, pp. 189–193.

[16] A. Kushki, H. Schwellnus, F. Ilyas, and T. Chau, "Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia," *Research in Developmental Disabilities*, vol. 32, no. 3, pp. 1058–1064, May 2011.

[17] K. Feder, A. Majnemer, and A. Synnes, "Handwriting: Current trends in occupational therapy practice," *Canadian Journal of Occupational Therapy*, vol. 67, no. 3, pp. 197–204, Apr. 2000.

[18] G. Silveri, F. De Dea, I. Perrone, and A. Accardo, *Influence of Dysgraphia on Kinematic Characteristics of Handwriting in Italian Primary School Children*. Springer Singapore, 05 2019, pp. 241–245.

[19] Z. WU, T. LIN, and M. LI, "Automated detection of children at risk of chinese handwriting difficulties using handwriting process information: An exploratory study," *IEICE Transactions on Information and Systems*, vol. E102.D, no. 1, pp. 147–155, Jan. 2019.

[20] J. Mucha, J. Mekyska, Z. Galaz, M. Faundez-Zanuy, K. Lopez-de Ipina, V. Zvoncak, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Identification and monitoring of parkinson's disease dysgraphia based on fractional-order derivatives of online handwriting," *Applied Sciences*, vol. 8, no. 12, p. 2566, 2018.

[21] J. Mucha, J. Mekyska, M. Faundez-Zanuy, K. Lopez-de Ipina, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, and I. Rektorova, "Advanced parkinson's disease dysgraphia analysis based on fractional derivatives of online handwriting." in *10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops.*, 2018.

[22] J. Mucha, V. Zvoncak, Z. Galaz, M. Faundez-Zanuy, J. Mekyska, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de Ipina, "Fractional derivatives of online handwriting: A new approach of parkinsonic dysgraphia analysis," in *2018 41st International Conference on Telecommunications and Signal Processing*. IEEE, 2018, pp. 214–217.

[23] S. Rosenblum and L. Gafni-Lachter, "Handwriting proficiency screening questionnaire for children (HPSQ–C): Development, reliability, and validity," *American Journal of Occupational Therapy*, vol. 69, no. 3, Apr. 2015.

[24] A. A. Alhusaini, G. R. Melam, and S. Buragadda, "Short-term sensorimotor-based intervention for handwriting performance in elementary school children," *Pediatrics International*, vol. 58, no. 11, pp. 1118–1123, Jul. 2016.

[25] I. Podlubny, *Fractional differential equations an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*. San Diego: Academic Press, 1999.

[26] M. Faundez-Zanuy and J. Mekyska, *User-Centric Privacy and Security in Biometrics*. Institution of Engineering and Technology, 2017, ch. Privacy of online handwriting biometrics related to biomedical analysis, pp. 17–39.

[27] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "A new modality for quantitative evaluation of Parkinson's disease: In-air movement," in *13th IEEE International Conference on BioInformatics and BioEngineering, Chania, Greece, November 10-13, 2013*. IEEE Computer Society, 2013, pp. 1–4.

[28] C. Bidet-Ildei, P. Pollak, S. Kandel, V. Fraix, and J.-P. Orliaguet, "Handwriting in patients with parkinson disease: Effect of l-dopa and stimulation of the sub-thalamic nucleus on motor anticipation," *Human Movement Science*, vol. 30, no. 4, pp. 783 – 791, 2011, special Issue: Progress in Graphonomics: A Perceptual Motor Skill Perspective.

[29] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[30] G. E. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 26, no. 2, pp. 211–243, 1964.

[31] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal.*, vol. 27, no. 8, pp. 1226–1238, 2005.

[32] M. M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Medical Journal*, vol. 24, no. 3, pp. 69–71, 2012.

[33] N. Nachar, "The mann-whitney u: A test for assessing whether two independent samples come from the same distribution," *Tutorials in quantitative Methods for Psychology*, vol. 4, no. 1, pp. 13–20, 2008.

[34] J. Mucha, M. Faundez-Zanuy, J. Mekyska, V. Zvoncak, Z. Galaz, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de Ipina, "Analysis of parkinson's disease dysgraphia based on optimized fractional order derivative features," in *2019 27th European Signal Processing Conference*. IEEE, 2019.

A.17 Advanced Parkinson's Disease Dysgraphia Analysis Based on Fractional Derivatives of Online Handwriting

Advanced Parkinson's Disease Dysgraphia Analysis Based on Fractional Derivatives of Online Handwriting

Jan Mucha*, Jiri Mekyska*, Marcos Faundez-Zanuy†, Karmele Lopez-de-Ipina‡, Vojtech Zvoncak*, Zoltan Galaz*§, Tomas Kiska*, Zdenek Smekal*, Lubos Brabenec§, Irena Rektorova§¶
*Department of Telecommunications and SIX Research Centre, Brno University of Technology
Technicka 10, 61600 Brno, Czech Republic
†Escola Superior Politecnica, Tecnocampus Avda. Ernest Lluch 32, 08302 Mataro, Barcelona, Spain
‡Department of Systems Engineering and Automation, University of the Basque Country UPV/EHU
Av de Tolosa 54, 20018 Donostia, Spain
§Applied Neuroscience Research Group, Central European Institute of Technology
Masaryk University, Kamenice 5, 62500 Brno, Czech Republic
¶First Department of Neurology, Masaryk University and St. Anne's University Hospital
Pekarska 53, 65691 Brno, Czech Republic

***Abstract*—Parkinson's disease (PD) is one of the most frequent neurodegenerative disorder with progressive decline in several motor and non-motor skills. Due to time-consuming and partially subjective conventional PD diagnosis, several more effective approaches based on signal processing and machine learning, e. g. online handwriting analysis, have been proposed. This paper introduces a new methodology of PD dysgraphia analysis based on fractional derivatives applied in PD handwriting quantification. The proposed methodology was evaluated on a database that consists 33 PD patients and 36 healthy controls who performed several handwriting tasks. Employing random forests classifier in combination with 5 kinematic features based on fractional-order derivatives we reached 90 % classification accuracy, 89 % sensitivity, and 91 % specificity. In comparison with the results of other related works dealing with the same database, the proposed approach brings improvements in PD dysgraphia diagnosis and confirms the impact of fractional derivatives in kinematic analysis.**

***Index Terms*—kinematic analysis; fractal calculus; fractional derivative; online handwriting; Parkinson's disease; Parkinson's disease dysgraphia;**

I. Introduction

Parkinson's disease (PD) affects millions of people all over the world as a second most frequent neurodegenerative disorder [1]. Prevalence rate of PD is estimated to approximately 1.5 % for people aged over 65 years [2], but the risk of being affected by this disease increases strongly with age [3]. The cardinal signs of PD include resting tremor, slowness of movement (bradykinesia), rigidity and postural instability [4]–[6]. Over the course of the disease a variety of non-motor symptoms may arise or can precede motor symptoms like depression, dementia, sleep disorders, anosmia, cognitive dysfunctions, psychosis etc. [4], [7], [8]. Even though, the precise pathophysiological cause of PD has not yet been discovered, the most significant biological finding is a rapid degeneration of dopaminergic cells in *substancia nigra pars compacta* [9].

Considering motor dysfunctions in people suffering from PD, in conjunction with complexity, proficiency and precision of handwriting performance, it is distinct that disrubted handwriting may be used as a significant biomarker for PD diagnosing [3], [6], [20]. With new technologies coming hand by hand with Health 4.0 systems we are able to acquire online handwriting signals, where a temporal information is added to x and y coordinates. Thus instead of quantifying PD micrographia by spatial features only, the use of digitalizing tablets gives us a new opportunity to quantify temporal, kinematic and dynamic manifestations of PD handwriting such as hesitations, pauses, and slow movement [5], which Letanneux et al. (2014) named PD dysgraphia [21].

The impact of many handwriting tasks in PD dysgraphia analysis has been explored, including simple (e. g. loops, circles, characters) as well as more complex ones (e. g. words, sentences, Archimedean spiral, figures) [6], [10], [11], [17], [22], [23]. Discrimination power of handwriting features is usually evaluated by correlation, classification and/or variance analysis. From the overview of related works (2015–now), which can be seen in Table I, it is obvious that kinematic features have irreplaceable place in PD dysgraphia analysis. Drotar et al. (2015, 2016) proved that combination of kinematic,

This work was supported by the grant of the Czech Ministry of Health 16-30805A (Effects of non-invasive brain stimulation on hypokinetic dysarthria, micrographia, and brain plasticity in patients with Parkinsons disease), grant of the Czech Science Foundation 18-16835S (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing) and the following projects: LO1401, FEDER and MEC, TEC2016-77791-C4-2-R, and TEC- 2016-77791-C4-4-R from the Ministry of Economic Affairs and Competitiveness of Spain. For the research, infrastructure of the SIX Center was used.

978-1-5386-9361-2/18/$31.00 ©2018 IEEE

TABLE I
OVERVIEW OF RELATED WORKS FOCUSED ON COMPUTERIZED ANALYSIS OF PD DYSGRAPHIA

First author	Year	PD/HC	Handwriting task	Analysis	Features	Conclusions
Drotar [10]	2015	37/38	letters, words, sentences	differential analysis (SVM)	kinematic, temporal, spatial, entropy, EMD, signal energy	The highest classification accuracy after feature selection approach was 88.1 %.
Drotar [11]	2015	37/38	letters, words, sentences	differential analysis (SVM)	kinematic, temporal, spatial, entropy, EMD, pressure	Classification performance was at its peak with on-surface features (89.09 %).
Heremans [12]	2015	34/10	up/down strokes at varying amplitudes	ANOVA	writing amplitude and velocity	Significant difference between groups was in writing amplitude ($F(2.41)=3.97; p=0.03$).
Pereira [13]	2015	37/18	Archimedean spiral	differential an. (SVM, NB, OPF)	mean relative tremor and spatial parameters	The best results were obtained by NB classifier, that provided around 79 % classification accuracy.
Drotar [6]	2016	37/38	letters, words, Archimedean spiral, sentences	differential an. (SVM, K-NN, AdaBoost)	kinematic, temporal, spatial, entropy, EMD, pressure	Combining all exercises, SVM proved to be the best classifier with 82.5 % accuracy.
Heremans [14]	2016	30/15	repetitive cursive loops	ANOVA, correlation an.	writing amplitude and velocity	Medical scale and writing amplitude had significant correlation ($r = -0.40$).
Pereira [15]	2016	14/21	Archimedean spiral meander	differential an. (CNN, OPF)	pen-based features	The best result was obtained by CNN with 87.14 % recognition rate using meander task.
Kotsavasil. [16]	2017	24/20	horizontal lines	differential analysis (NB)	normalized velocity variability	Average classification accuracy was 91 % for unlabelled PD and HC data.
Loconsole [17]	2017	4/7	sentence, repetitive loops	differential analysis (ANN)	execution time and average speed, density ratio, height ratio	Highest classification accuracy (96.81 %) was achieved using all the extracted features.
Taleb [18]	2017	16/16	letters, waves, words	differential analysis (SVM)	kinematic, stroke, pressure, entropy, energy, EMD	The highest classification accuracy was 96.88 % for 12 kinematic and pressure features.
Moetesum [19]	2018	37/38	letters, words, Archimedean spiral, sentence, loops	differential analysis (SVM)	CNN based features	Extraction of features using CNN applied on raw handwriting data resulted in 83 % classification accuracy.

SVM – support vector machine; EMD – empirical mode decomposition; r_s – Spearman's correlation coefficient; K-NN – K-nearest neighbours; ANOVA – analysis of variance; NB – naïve Bayes classifier; OPF – optimum path forest; ANN – artificial neural network; CNN – convolutional neural networks; F and p corresponds to variables of F distribution; articles are sorted by the year of release and then alphabetically.

pressure, energy or empirical mode decomposition (EMD) based features resulted in classification accuracy up to 89 % using several handwriting tasks [6], [10], [11]. Next, Kotsavasiloglou a et al. (2017) achieved an average prediction accuracy of 91 % using simple horizontal lines and features describing a variability of the pen tip's velocity, a deviation from the horizontal plane, and the trajectory's entropy [16]. Some other works report even higher classification accuracies results (approximately 97 %), e.g. Loconsole et al. (2017) who used computer vision and electromyography signal processing techniques but applied on a very small dataset (4 PD and 7 HC). Thus, the reliability of those results may be untrustworthy. Another promising approach was published by Moetesum et al. (2018) who reached 83 % classification accuracy by employing convolutional neural networks (CNN) that were used to extract discriminating visual features from raw handwriting data.

The main goal of this work is to introduce an advanced approach of kinematic features calculation based on fractional order derivation (FDE) as a new methodology of PD dysgraphia analysis. We aim to:

- proof the potential of FDE in PD dysgraphia quantification employing classification analysis,
- evaluate discrimination power of the newly designed features when comparing the results with a baseline,
- identify a handwriting task that (in combination with the newly designed parameters) provides best results in terms of PD dysgraphia classification accuracy.

The rest of this paper is organized as follows: section II describes cohort of patients and methodology, section III includes achieved results, discussion can be found in section IV and finally, the conclusions are drawn in section V.

II. DATASET & METHODS

A. Dataset

We used the Parkinson's disease handwriting database (PaHaW) [6] that consists 33 PD patients and 36 healthy controls (HC). Demographic and clinical data of the participants can be found in Table II. The participants were enrolled at the First Department of Neurology, St. Anne's University Hospital in Brno, Czech Republic. All participants reported Czech language as their native language and all participants were right-handed. The PD patients completed the tasks approximately 1 hour after their regular L-dopa medication. All participants signed an informed consent form approved by the local ethics committee.

B. Data Acquisition

PaHaW database [6] includes several handwriting tasks (see Fig. 1), namely: Archimedean spiral; repetitive loops; letter *l*; syllable *le*; Czech words *les*, *lektorka*, *porovnat*, and *nepopadnout*; Czech sentence *Tramvaj dnes už nepojede.* During all handwriting tasks the participants were rested and seated in a comfortable position with possibility to look at pre-filled template. A digitizing tablet (Wacom Intuos 4M) was overlaid with an empty paper template and participants were allowed to repeat a task in case of some mistakes. Online handwriting signals were recorded with $f_\mathrm{s} = 150$ Hz sampling rate. Following time sequences were acquired: x and y coordinates

TABLE II
DEMOGRAPHIC AND CLINICAL DATA OF PARTICIPANTS

Parkinson's disease patients					
Gender	N	Age [y]	PD dur [y]	UPDRS V	LED [mg]
Female	17	71.76 ± 7.93	9.88 ± 5.27	2.18 ± 0.86	1146.03 ± 543.89
Male	16	66.50 ± 13.44	7.44 ± 4.04	2.31 ± 0.75	1673.38 ± 616.66
All	33	69.21 ± 11.10	8.70 ± 4.82	2.24 ± 0.80	1401.72 ± 630.71
Healthy controls					
Gender	N	Age [y]			
Female	17	61.59 ± 10.17			
Male	19	63.32 ± 13.14			
All	36	62.50 ± 11.70			

N – number; y – years; PD dur – PD duartion; UPDRS V – Unified Parkinson's disease rating scale, part V: Modified Hoehn & Yahr staging score [24]; LED – L-dopa equivalent daily dose [25].

– $x[t]$, $y[t]$; time-stamp – t; in-air/on-surface status – $b[t]$; pressure – $p[t]$; azimuth $az[t]$; and tilt $al[t]$.

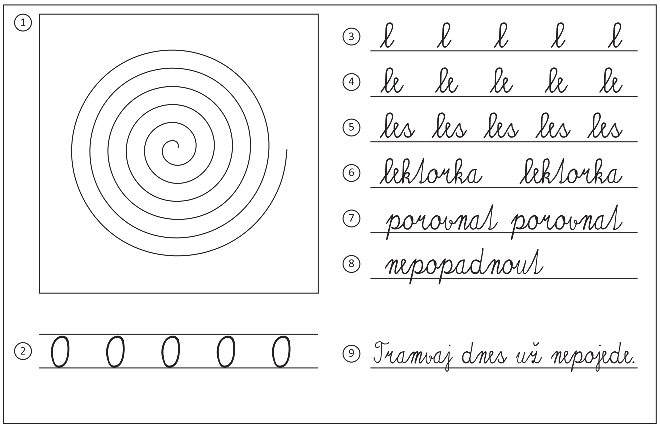


Fig. 1. Filled template of the PaHaW database.

C. Fractional Order Derivative

The idea of this study is to use the FDE as a substitution of the conventional differential derivative during calculation of the basic kinematic features. There are several definitions of FDE, namely, the Riemann-Liouville, Caputo, and Grünwald-Letnikov formulations [26]–[28]. For the purpose of this study we used the Jonathan Hadida's FDE implementation, which follows the Grünwald-Letnikov approximation [27], [29]. A direct definition of the FDE $D^\alpha y(t)$ is based on finite differences of an equidistant grid in $[0, \tau]$. Assume that the function $y(\tau)$ satisfies some smoothness conditions in every finite interval $(0, t), t \leq T$. Choosing the grid [27]

$$0 = \tau_0 < \tau_1 < ... < \tau_{n+1} = t = (n+1)h \tag{1}$$

with

$$\tau_{k+1} - \tau_k = h \tag{2}$$

and using the notation of the finite differences

$$\frac{1}{h^\alpha}\Delta_h^\alpha y(t) = \frac{1}{h^\alpha}\left(y(\tau_{n+1}) - \sum_{v=1}^{n+1} c_v^\alpha y(\tau_{n+1-v})\right), \tag{3}$$

where

$$c_v^\alpha = (-1)^{v-1}\binom{\alpha}{v}, \tag{4}$$

the Grünwald-Letnikov implementation is defined as:

$$D^\alpha y(t) = \lim_{h \to 0} \frac{1}{h^\alpha}\Delta_h^\alpha y(t), \tag{5}$$

where $D^\alpha y(t)$ means a derivative with order α of function $y(t)$, and h represents sampling lattice.

D. Handwriting Features

A wide range of handwriting features for analysis of PD dysgraphia is commonly used, but to demonstrate the impact of FDE, only basic on-surface kinematic features [6], [30], [31] extracted from all PaHaW tasks are considered. Feature set consists: *velocity* – rate at which a position of pen changes with time [mm/s]; *acceleration* – rate at which the velocity of pen changes with time [mm/s^2]; *jerk* – rate at which the acceleration of pen changes with time [mm/s^3]; and their horizontal and vertical variants. These features were extracted for different values of α going from 0.1 to 1.0 with 0.1 step. Consequently, statistical properties of the features were described using following statistics: mean, median, standard deviation (std), and maximum (max). Considering all combinations of tasks and features (with different FDE order), in total 5040 features were extracted.

E. Statistical Analysis

After feature extraction, univariate binary classification (PD/HC) model (stratified 7-fold cross-validation with 50 repetitions) based on random forests (RF) [32] was designed to evaluate a discrimination power of the features among all handwriting tasks. To eliminate non-significant features from results of univariate classification, Spearman's and Pearson's correlation analysis was performed (significance level of $p = 0.01$ was selected). Consequently, multivariate classification with the same classifier and the same cross-validation settings was performed in order to improve classification accuracy. To obtain the most appropriate combination of the features, the sequential floating forward selection (SFFS) algorithm was used [33]. Classification performance was evaluated by the Matthew's correlation coefficient [34], classification accuracy (ACC), sensitivity (SEN) and specificity (SPE).

III. RESULTS

Univariate and multivariate classification analysis results are summarized in Table III. In the upper part of table the results of univariate analysis sorted by ACC are reported. Only 10 best features that achieved condition of significance ($p < 0.01$) were chosen. The best ACC (74.96 %) was obtained using horizontal acceleration with $\alpha = 0.6$ extracted from repetitive loops task. Nevertheless, based on the results, the most useful task is sentence (8/10 features with ACC around 74 %). Fig. 2 displays dependence of ACC on FDE order for 3 most discriminative features extracted from this task. Dependence of average ACC for each α separately for XY, horizontal, vertical and altogether features extracted from

TABLE III
RESULTS OF UNIVARIATE AND MULTIVARIATE CLASSIFICATION ANALYSIS

Univariate classification analysis										
Feature	α	Task	ACC [%]	SEN [%]	SPE [%]	MCC	r_p	p_p	r_s	p_s
horizontal acceleration (mean)	0.6	repetitive loops	74.96	70.79	78.78	0.50	0.34	0.003821	0.52	0.000005
vertical velocity (max)	0.2	sentence	74.38	71.52	77.00	0.49	0.33	0.006085	0.44	0.000160
horizontal acceleration (mean)	0.5	repetitive loops	74.32	75.94	72.83	0.49	0.34	0.003785	0.50	0.000012
vertical jerk (max)	0.7	sentence	74.12	70.85	77.11	0.48	0.33	0.006086	0.44	0.000160
vertical velocity (max)	0.1	sentence	74.09	70.12	77.72	0.48	0.33	0.006086	0.44	0.000160
vertical acceleration (max)	0.6	sentence	74.00	70.24	77.44	0.48	0.33	0.006086	0.44	0.000160
vertical velocity (max)	0.8	sentence	74.00	70.36	77.33	0.48	0.33	0.006086	0.44	0.000160
vertical acceleration (max)	0.3	sentence	73.91	70.61	76.94	0.48	0.33	0.006086	0.44	0.000160
vertical jerk (max)	0.2	sentence	73.80	71.39	76.00	0.48	0.33	0.006086	0.44	0.000160
vertical jerk (max)	0.5	sentence	73.74	69.58	77.56	0.47	0.33	0.006086	0.44	0.000160

Multivariate classification analysis							
Feature set			Model information				
Feature	α	Task	Features quantity	ACC [%]	SEN [%]	SPE [%]	MCC
velocity (max)	0.1	repetitive character *l*	1	76,25	73.07	79.85	0.5325
horizontal velocity (median)	0.5	repetitive word *lektorka*	2	80.40	78.28	82.81	0.6112
vertical velocity (median)	0.9	word *les*	3	82.99	78.46	87.74	0.6699
acceleration (median)	0.8	syllables *le*	4	88.66	88.37	88.64	0.7785
velocity (median)	0.1	word *porovnat*	5	**89.81**	88.63	90.87	0.8039

α – order of FDE; ACC – accuracy; SEN – sensitivity; SPE – specificity; r_p – Pearson's correlation coefficient; MCC – Matthew's correlation coefficient; r_s – Spearman's correlation coefficient; p_p – significance level of correlation (r_p); p_s – significance level of correlation (r_s)

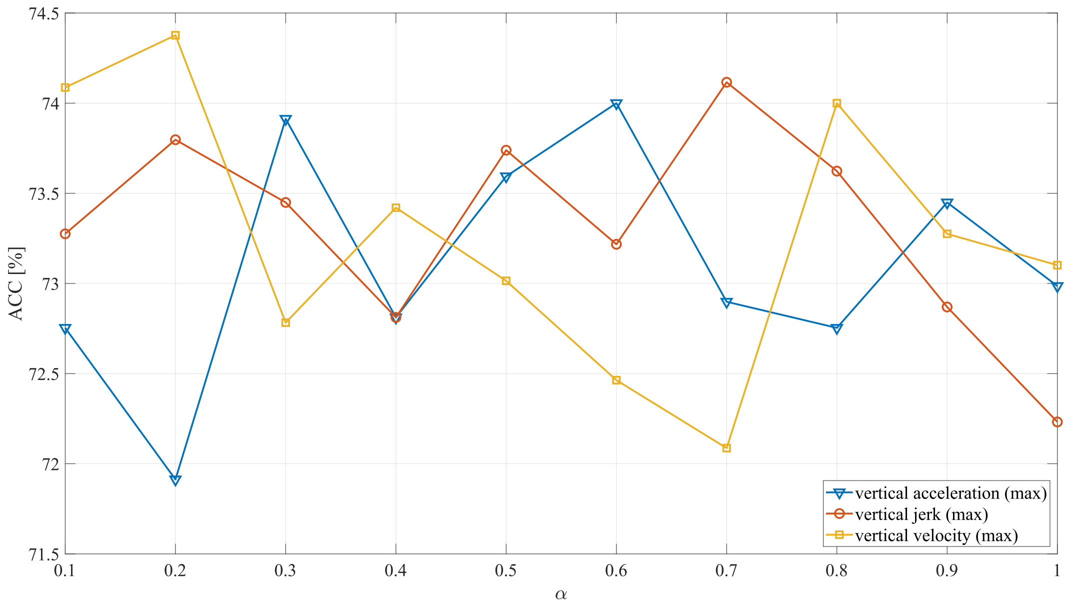


Fig. 2. Dependence of classification accuracy on FDE order for 3 most discriminative features extracted from the sentence task.

all tasks is visualized in Fig. 3. Regarding the multivariate classification analysis (bottom part of Table III), the best classification score (ACC = 89.81 %, SEN = 88.63 %, SPE = 90.87 %) was achieved using a combination of 5 kinematic features. The table contains information about RF performance as the features were gradually selected by the SFFS.

IV. Discussion

With respect to the results of univariate analysis, previous hypothesis that FDE utilization in PD dysgraphia analysis may improve classification performance can be confirmed. Following the α values reported in Table III, it is evident that features calculated by FDE fully substitute conventional kinematic parameters based on the differential derivative (full derivative; $\alpha = 1$). The sentence appears to be the most suitable handwriting task in univariate classification analysis, where 8 out of 10 most discriminative features (ACC around 74 %) are extracted from this task. This finding is in line with results reported by Drotar et al. ([6]). The sentence provides good discriminative power because the PD dysgraphia symp-

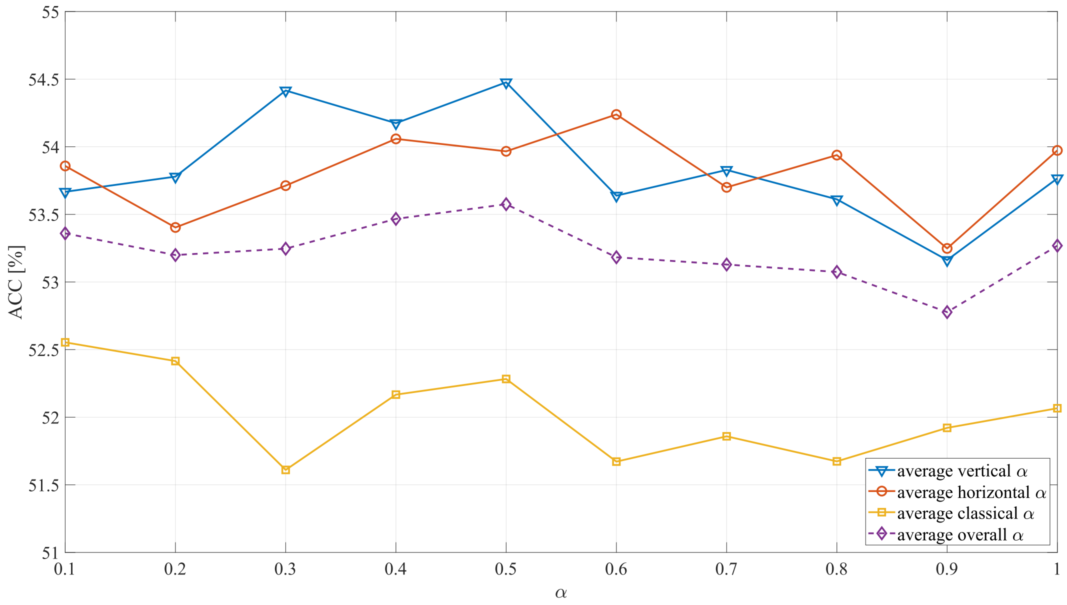


Fig. 3. Dependence of average classification accuracy on FDE order separately for XY, horizontal, vertical and overall features.

toms have more space to emerge in comparison with others tasks of the PaHaW database. I. e. the task contains more on-surface/in-air transitions, it can capture decreasing amplitude of letters (micrographia), variations in handwriting kinematics, etc. We can conclude that the univariate approach described in this paper brings remarkable improvements giving very similar classification accuracy using only basic kinematic features in comparison with the baseline published by Drotar et al. (2016), where the authors reported ACC = 76.5 % for the sentence task using combination of several kinematic and pressure features.

The effect of FDE order on the classification performance (as visualized by Fig. 2) has some local maxima for $\alpha \in < 0.2; 0.3 >$ and for $\alpha \in < 0.6; 0.8 >$. A decreasing character of ACC for α going from 0.8 towards the full derivation can be noticed. As can be seen in Fig. 3 horizontal and vertical features generally provide higher classification accuracies when compared to the XY features. This can be also confirmed by the nature of the most discriminative features whereas all of them are horizontal or vertical. Considering, that the average classification accuracy based on the XY features is lower than the overall average, we conclude that the importance of separate movement directions analysis is high.

Next, classification performance was improved by approximately 15 % using the multivariate classification analysis. The best classification model contains only 5 features (providing ACC = 89.81 %, SEN = 88.63 % and SPE = 90.87 %) extracted from different handwriting tasks, including cursive letter "l", syllable, words and repetitive word. This higher-dimensional feature space points to complexity of handwriting and directs to the need of considering various aspects of deficits in PD during PD dysgraphia analysis. Based on the values of α, which are different from 1, we can confirm full utilization of FDE in multivariate classification analysis too. The best classification accuracy reported in the frame of PaHaW database is ACC = 89 % employing combination of kinematic and pressure features [11]. We reached the same accuracy omitting the pressure ones.

The reached accuracy is interesting from a clinical point of view too. It is well known that L-dopa medication has a positive effect on upper limb in PD, which means that theoretically PD dysgraphia in patients who are in their ON state should not be manifested significantly. Nevertheless, we proved that using advanced kinematic analysis we are able to differentiate HC and patients 1 hour after their regular L-dopa medication with almost 90 % accuracy.

Several other research teams published PD dysgraphia classification accuracies in range between 91 % and 97 %, however, analysing different datasets (with significantly lower number of samples) and extracting advanced handwriting features [16]–[18]. A relevant comparison is thus not possible.

V. Conclusion

This pilot study proves that application of FDE in quantitative PD dysgraphia analysis brings new promising and enhancing methodology of PD diagnosis. Based on the results, we are able to identify PD dysgraphia with almost 90 % accuracy using only 5 basic kinematic features extracted from a few handwriting tasks. We hypothesise that combination of the newly designed features with spatial, temporal and dynamic ones could bring even better results. Some improvements could be made in machine learning too. For instance application of boosting algorithms such as XGBoost would be beneficial. Finally, a lot can be further explored in the case of FDE, i. e. finer selection of FDE order and individual tuning of α for horizontal and vertical movement.

The limitation of this study is the size of database. As already mentioned, this study has a pilot character. To be able to generalize the results, bigger and multilingual datasets should be analysed.

References

[1] M. C. de Rijk, L. J. Launer, K. Berger, M. M. Breteler, J. F. Dartigues, M. Baldereschi, L. Fratiglioni, J. Lobo, A. Martinez-Lage, C. Trenkwalder, and A. Hofman, "Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts," *Neurology*, vol. 54, pp. 21–23, 2000.

[2] S. Sapir, L. Ramig, and C. Fox, "Speech and swallowing disorders in Parkinson disease," *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 16, no. 3, pp. 205–10, 2008.

[3] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. S. di Freca, "Handwriting analysis to support neurodegenerative diseases diagnosis: A review," *Pattern Recognition Letters*, 2018.

[4] A. Elbaz, L. Carcaillon, S. Kab, and F. Moisan, "Epidemiology of parkinson's disease," *Revue Neurologique*, vol. 172, no. 1, pp. 14 – 26, 2016, neuroepidemiology. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0035378715009224

[5] J. L. Contreras-Vidal, P. Poluha, H.-L. Teulings, and G. E. Stelmach, "Neural dynamics of short and medium-term motor control effects of levodopa therapy in Parkinson's disease1supported in part by ninds grant ro1 ns33173, the flinn foundation and the bryng bryngelson fund.1," *Artificial Intelligence in Medicine*, vol. 13, no. 1, pp. 57 – 79, 1998.

[6] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," *Artificial Intelligence in Medicine*, vol. 67, pp. 39 – 46, 2016.

[7] L. Brabenec, J. Mekyska, Z. Galaz, and I. Rektorova, "Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation," *Journal of Neural Transmission*, pp. 1–32, 2017.

[8] J. Mucha, Z. Galaz, J. Mekyska, T. Kiska, V. Zvoncak, Z. Smekal, I. Eliasova, M. Mrackova, M. Kostalova, I. Rektorova *et al.*, "Identification of hypokinetic dysarthria using acoustic analysis of poem recitation," in *Telecommunications and Signal Processing (TSP), 2017 40th International Conference on*. IEEE, 2017, pp. 739–742.

[9] O. Hornykiewicz, "Biochemical aspects of Parkinson's disease," *Neurology*, vol. 51, no. 2 Suppl 2, pp. S2–S9, 1998.

[10] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. F. Zanuy, "Decision support framework for Parkinson's disease based on novel handwriting markers," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, pp. 1–1, 2015.

[11] P. Drotar, J. Mekyska, Z. Smekal, I. Rektorova, L. Masarova, and M. Faundez-Zanuy, "Contribution of different handwriting modalities to differential diagnosis of Parkinson's disease," in *Medical Measurements and Applications (MeMeA), 2015 IEEE International Symposium*, 2015, pp. 1–5.

[12] E. Heremans, E. Nackaerts, G. Vervoort, S. Vercruysse, S. Broeder, C. Strouwen, S. P. Swinnen, and A. Nieuwboer, "Amplitude manipulation evokes upper limb freezing during handwriting in patients with Parkinson's disease with freezing of gait," *PLOS ONE*, vol. 10, no. 11, p. e0142874, nov 2015.

[13] C. R. Pereira, D. R. Pereira, F. A. da Silva, C. Hook, S. A. Weber, L. A. Pereira, and J. P. Papa, "A step towards the automated diagnosis of parkinson's disease: Analyzing handwriting movements," in *Computer-Based Medical Systems (CBMS), 2015 IEEE 28th International Symposium on*. IEEE, 2015, pp. 171–176.

[14] E. Heremans, E. Nackaerts, S. Broeder, G. Vervoort, S. P. Swinnen, and A. Nieuwboer, "Handwriting impairments in people with Parkinson's disease and freezing of gait," *Neurorehabilitation and Neural Repair*, vol. 30, no. 10, pp. 911–919, jul 2016.

[15] C. R. Pereira, S. A. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided parkinson," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2016, pp. 340–346.

[16] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, and M. Arnaoutoglou, "Machine learning-based classification of simple drawing movements in parkinson's disease," *Biomedical Signal Processing and Control*, vol. 31, pp. 174–180, 2017.

[17] C. Loconsole, G. F. Trotta, A. Brunetti, J. Trotta, A. Schiavone, S. I. Tato, G. Losavio, and V. Bevilacqua, "Computer vision and EMG-based handwriting analysis for classification in Parkinson's disease," in *Intelligent Computing Theories and Application*. Springer International Publishing, 2017, pp. 493–503.

[18] C. Taleb, M. Khachab, C. Mokbel, and L. Likforman-Sulem, "Feature selection for an improved parkinson's disease identification based on handwriting," in *Arabic Script Analysis and Recognition (ASAR), 2017 1st International Workshop on*. IEEE, 2017, pp. 52–56.

[19] M. Moetesum, I. Siddiqi, N. Vincent, and F. Cloppet, "Assessing visual attributes of handwriting for prediction of neurological disordersa case study on parkinsons disease," *Pattern Recognition Letters*, 2018.

[20] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, and I. Schlesinger, "Handwriting as an objective tool for Parkinson's disease diagnosis," *Journal of neurology*, vol. 260, no. 9, pp. 2357–2361, 2013.

[21] A. Letanneux, J. Danna, J.-L. Velay, F. Viallet, and S. Pinto, "From micrographia to Parkinson's disease dysgraphia," *Movement Disorders*, vol. 29, no. 12, pp. 1467–1475, 2014.

[22] E. Nackaerts, S. Broeder, M. P. Pereira, S. P. Swinnen, W. Vandenberghe, A. Nieuwboer, and E. Heremans, "Handwriting training in Parkinson's disease: A trade-off between size, speed and fluency," *PLOS ONE*, vol. 12, no. 12, p. e0190223, dec 2017.

[23] J. Mucha, V. Zvoncak, Z. Galaz, M. Faundez-Zanuy, J. Mekyska, T. Kiska, Z. Smekal, L. Brabenec, I. Rektorova, and K. Lopez-de Ipina, "Fractional derivatives of online handwriting: A new approach of parkinsonic dysgraphia analysis," in *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2018, pp. 214–217.

[24] S. Fahn and R. L. Elton, *UPDRS Development Committee (1987) Unified Parkinson's Disease Rating Scale*. Recent developments in Parkinson's Disease. Macmillan, Florham Park, 1987.

[25] J. Y. Lee, J. W. Kim, W. Y. Lee, J. M. Kim, T. B. Ahn, H. J. Kim, J. Cho, and B. S. Jeon, "Daily dose of dopaminergic medications in Parkinson's disease: clinical correlates and a posteriori equation," *Neurol. Asia*, vol. 15, no. 2, pp. 137–143, 2010.

[26] J. Tenreiro Machado, *Fractional coins and fractional derivatives*, 2013, vol. 2013.

[27] I. Podlubny, *Fractional differential equations an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*. San Diego: Academic Press, 1999.

[28] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo, "Theory and applications of fractional differential equations," 2006.

[29] R. Scherer, S. L. Kalla, Y. Tang, and J. Huang, "The grünwald–letnikov method for fractional differential equations," *Computers & Mathematics with Applications*, vol. 62, no. 3, pp. 902–917, 2011.

[30] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "A new modality for quantitative evaluation of Parkinson's disease: In-air movement," in *13th IEEE International Conference on BioInformatics and BioEngineering, BIBE 2013, Chania, Greece, November 10-13, 2013*. IEEE Computer Society, 2013, pp. 1–4.

[31] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and rating of developmental dysgraphia by handwriting analysis," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, 2017.

[32] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[33] J. Pohjalainen, O. Rasanen, and S. Kadioglu, "Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits," *Comput. Speech Lang.*, 2014.

[34] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochim. Biophys. Acta (BBA)*, vol. 405, no. 2, pp. 442–51, 1975.

A.18 Effect of Stroke-level Intra-writer Normalization on Computerized Assessment of Developmental Dysgraphia

Effect of Stroke-level Intra-writer Normalization on Computerized Assessment of Developmental Dysgraphia

Vojtech Zvoncak*, Jiri Mekyska*, Katarina Safarova†, Zoltan Galaz*‡, Jan Mucha*, Tomas Kiska*, Zdenek Smekal*, Barbora Losenicka†, Barbora Cechova†, Pavlina Francova†, Marcos Faundez-Zanuy§, Sara Rosenblum¶

*Department of Telecommunications and SIX Research Centre, Brno University of Technology
Technicka 10, 61600 Brno, Czech Republic
†Department of Psychology, Faculty of Arts, Masaryk University
Uvoz 244/33, 60200 Brno, Czech Republic
‡Applied Neuroscience Research Group, Central European Institute of Technology
Masaryk University, Kamenice 5, 62500 Brno, Czech Republic
§Escola Superior Politecnica, Tecnocampus
Avda. Ernest Lluch 32, 08302 Mataro, Barcelona, Spain
¶Laboratory of Complex Human Activity and Participation, University of Haifa
Aba Khoushy Ave. 199, Mount Carmel, 3498838 Haifa, Israel

***Abstract*—Developmental dysgraphia (DD) can negatively influence writing speed, legibility of written product, and even planning and content generation. This could have a detrimental impact on self-expression, and communication in childhood. The goal of this study is to investigate new intra-writer and self-esteem normalization methods (IWN) in direction of improving computerized DD assessment based on the quantitative analysis of online handwriting. For this purpose we enrolled 97 children who wrote a paragraph using a digitizing tablet. Their handwriting proficiency was rated by the Handwriting Proficiency Screening Questionnaire (HPSQ). The handwriting was parametrized using a conventional set of features that were consequently normalized by four newly designed IWN. Based on the results we observed that a stroke-level ℓ_2 norm normalization decreased the computerized DD assessment error from 23 % to 18 %. This study proves that the stroke-level IWN has the significant potential in the field of computerized DD diagnosis and assessment.**

***Index Terms*—developmental dysgraphia; digitizer; intra-writer normalization; kinematic analysis; machine learning; online handwriting**

I. Introduction

To successfully participate in self-expression and communication in the academic environment in childhood, handwriting proficiency is required [1]. The handwriting is considered as proficient when a produced text is legible, performed with minimum effort and reasonable amount of time [2], [3]. According to Berninger et al. [4] the handwriting difficulties can be classified as lower and higher, where the lower difficulties correspond to actual mechanical forming of letters on the writing surface, and the higher difficulties are related to planning and content generation. These difficulties are usually linked with developmental dysgraphia (DD), which occurs in 10 – 30 % of school-aged children [5], and is associated with several symptoms: poor legibility [6], inadequate speed [7], slow performance time [8], and higher number of deletions and/or corrections [9].

The Handwriting Proficiency Screening Questionnaire (HPSQ) [14] proved to be reliable and valid tool for identifying and rating DD. The questionnaire consists of 10 items that describe three handwriting-related factors: legibility, performance time, and physical and emotional well-being. Each part is scored on a 5-point Likert scale. Severity of DD corresponds to higher values of the sum of all scores. Seven years later, authors Rosenblum et al. developed its modified version where the children assess themselves. This version is called The Handwriting Proficiency Screening Questionnaire for Children (HPSQ-C) [7].

Until now, there is no available objective (i. e. without an influence of human factor) method for diagnosing and rating DD in Czech Republic. Nowadays, the graphomotorical skills are assessed manually/visually based on shape of letters, spacing, and number of errors. This subjective approach is affected by rater's actual psychical state, visual abilities and experiences. Moreover, the raters are not able to accurately quantify information such as pressure of the tip of pen, fine motor tremor, in-air movement (movement of the pen when its tip is not touching a paper's surface), etc.

Digitizing tablets (sometimes called digitizers) enable us to overcome these limitations. The digitizers record various signals linked with the handwriting (see Fig. 1): x and y position of pen when it touches the surface of paper (on-surface move-

This work was supported by the grant of Czech Science Foundation 18-16835S (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing), project LO1401, and TEC2016-77791-C4-2-R from the Ministry of Economic Affairs and Competitiveness of Spain. For the research, infrastructure of the SIX Center was used.

978-1-5386-9361-2/18/$31.00 ©2018 IEEE

ment), x and y position of pen when it is up to 1.5 cm above the surface of paper (in-air movement), azimuth, altitude, and pressure. Because each sample of this information is associated with a time stamp, we generally call this kind of handwriting as online [15]. Previous research identified benefits and impact of digitizing tablets in quantitative DD analysis [2], [5], [6], [8], [11], [12], [16]–[18]. Nevertheless, just a few explored the possibility of automatic DD diagnosis and/or rating based on machine learning approaches. For example, Rosenblum et al. proved it is possible to discriminate between children's proficient and dysgraphic handwriting products with approximately 90 % accuracy using Support Vector Machine (SVM) classifier [1]. Similarly, in our recent study we proved that graphomotorical skills of children with DD (as assessed by HPSQ) could be automatically rated using classification and regression trees with error around 10 % [5]. We also proved that an intra-writer normalization can further decrease this error. Based on these results we hypothesise, that a more sophisticated intra-writer normalization working on a stroke level could introduce even better accuracy of computerized DD assessment.

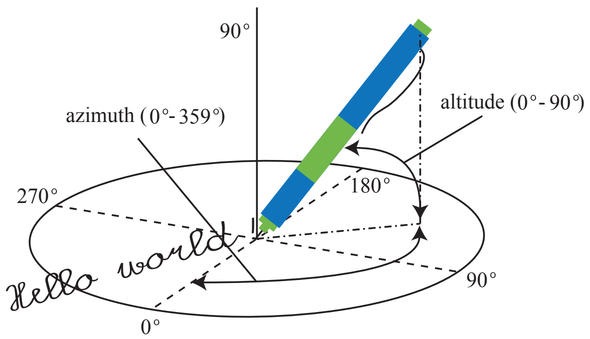


Fig. 1. Signals of online writing: position (on-surface/in-air), pressure, azimuth, altitude.

The objective of this paper is to introduce a new approach of intra-writer normalization (IWN). More specifically we aim to:

1) Introduce several techniques of IWN.
2) Evaluate these techniques and compare their performance with a baseline (i.e. without the use of IWN) in terms of DD assessment error.

The rest of this paper is organized as follows. A dataset of DD children, IWN techniques and a machine learning methodology is described in Section II. Results are reported in Section III and consequently discussed in Section IV. Conclusions are given in Section V.

II. Materials & Methods

A. Dataset

We enrolled altogether 97 children for this study. Almost all of them were right-handed writers (only 4 were left-handed). 35 children attended third and the rest fourth class of the elementary school. Based on the cut-off scores of HPSQ and HPSQ-C questionnaires, the children were divided into 4 groups (see Table I). Experimental group corresponds to children with higher values of HPSQ/HPSQ-C (i.e. children with DD). On the contrary, comparative group represents control children without DD. The study was approved by the local ethics committee, and parents of all the children signed an informed consent form.

TABLE I
Dataset structure

		HPSQ		
Gender	N	Age [y]	Grades [-]	Scores [-]
		Experimental group		
girls	20	9.40 ± 0.68	1.84 ± 0.47	21.70 ± 2.29
boys	9	9.00 ± 0.70	1.95 ± 0.45	23.22 ± 3.52
		Comparative group		
girls	25	9.12 ± 0.60	1.07 ± 0.22	5.12 ± 2.42
boys	10	9.10 ± 0.73	1.08 ± 0.23	3.8 ± 2.04
		HPSQ-C		
Gender	N	Age [y]	Grades [-]	Scores [-]
		Experimental group		
girls	17	9.17 ± 0.72	1.46 ± 0.37	15.23 ± 5.46
boys	16	9.19 ± 0.75	1.61 ± 0.47	20.87 ± 2.15
		Comparative group		
girls	18	9.05 ± 0.72	1.01 ± 0.05	4.72 ± 2.49
boys	14	9.21 ± 0.69	1.10 ± 0.21	6.57 ± 2.37

N – number; y – years; HPSQ/HPSQ-C – scores of HPSQ and HPSQC questionnaires

B. Data Acquisition

The children were asked to copy a paragraph on a lined A4 paper, which was lay down and fixed to a digitizing tablet. For this purpose we used the Wacom Intuos Pro L (PTH-850) digitizer and the Wacom inking pen, which enabled the children to have an immediate visual feedback and feeling like they were using a conventional inking pen. The online handwriting signals (i.e. x and y position, azimuth, altitude, and pressure) were sampled with $f_s = 150$ Hz. The paragraph was taken from a 3rd grade textbook and the children were asked to copy it from a printed template. It contains 63 words (310 characters, 371 characters including spaces). Example of one sentence written by a child with DD and a child without DD can be seen on Fig. 2.

C. Handwriting Features

Only the stroke-based handwriting features that can be further processed by the IWN techniques were considered in this study. More specifically we calculated a set of spatial (width, height, length of stroke), kinematic (velocity, acceleration, jerk), and temporal (duration) in-air/on-surface measures in a direction of stroke trajectory, as well as in its vertical and horizontal projection. For more information we refer to [5], [19]. Finally to transform a vector representation of the calculated features into scalar values we used these statistics: range, mean, median, variance, standard deviation.

D. Normalization Techniques

We introduced the four IWN techniques based on ℓ_1 and ℓ_2 norms [20], ℓ_∞ norm [21] and z-score. The normalization

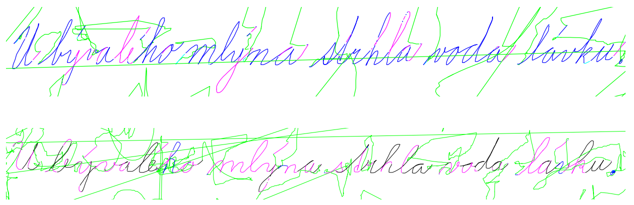


Fig. 2. Part of the paragraph written by a child without DD (HPSQ = 1, upper part of the picture) and with DD (HPSQ = 26, lower part of the picture). The color of letters corresponds to the tip pressure of the pen (cyan: 0 – 25 %, blue: 25 – 50 %, purple: 50 – 75 %, black: 75 – 100 %), where the green color represents the in-air movement.

process can be mathematically described by these formulas (in the previously mentioned order):

$$\mathbf{v}_1 = \frac{\mathbf{v}}{\sum_{i=1}^{N} |v_i|}, \quad (1)$$

$$\mathbf{v}_2 = \frac{\mathbf{v}}{\sqrt{\sum_{i=1}^{N} |v_i|^2}}, \quad (2)$$

$$\mathbf{v}_\infty = \frac{\mathbf{v}}{\max_i |v_i|}, \quad (3)$$

$$\mathbf{v}_\mathrm{z} = \frac{\mathbf{v} - \mathrm{mean}(\mathbf{v})}{\mathrm{std}(\mathbf{v})}, \quad (4)$$

where $\mathbf{v}$ is the input feature vector and N its length.

E. Statistical Analysis

First of all we employed an univariate analysis where we trained a Gradient Boosted Trees classifier [22], [23] (10-fold cross validation with 10 repetitions) fed by individual features, and evaluated their performance in HPSQ/HPSQ-C total score estimation by MAE (Mean Absolute Error), RMSE (Root Mean Squared Error) and EER (Estimation Error Rate):

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \quad (5)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \quad (6)$$

$$\mathrm{EER} = \frac{1}{n \cdot r} \sum_{i=1}^{n} |y_i - \hat{y}_i| \cdot 100\,[\%], \quad (7)$$

where n denotes the number of true/predicted values, y_i and $\hat{y}_i$ represents the true and predicted values of the HPSQ/HPSQ-C total scores, respectively, and r denotes the range of HPSQ/HPSQ-C total score values the dataset.

In the second step we performed a multivariate analysis where the HPSQ/HPSQ-C scores were modelled by a multidimensional feature space. To select the best combination of features we used the Sequntial Floating Forward Selection (SFFS) [24] in combination with the same classifier and the same cross validation settings. In both univariate and multivariate analysis the effect of IWN techniques was compared to a baseline (non-normalized features).

III. Results

The results of univariate regression analysis are reported in Table II. Due to a large number of features in each scenario (i.e. combination of on-surface/in-air movement and a particular normalization technique), we mention only features with the lowest MAE in each of these scenarios. In the case of HPSQ total score estimation, the ℓ_1 norm based normalization of median length of in-air stroke provided the lowest MAE = 6.22 (RMSE = 7.82, EER = 22.22 %). In the case of HPSQ-C, the lowest MAE = 5.76 (RMSE = 7.47, EER = 18.00 %) was achieved by ℓ_∞ norm normalization applied on the same feature, but acquired on-surface.

Table III reports the results of the multivariate analysis. In this case the HPSQ total score was estimated with minimum MAE = 4.99 (RMSE = 6.12, EER = 17.82 %) based on ℓ_2 norm normalization of eight in-air parameters. Regarding HPSQ-C, we did not observe any improvement in comparison to the baseline.

IV. Discussion

Although the results suggest that neglecting the effect of intra-writer variability plays some role in more accurate assessment of DD, the differences in comparison to the baseline are not so remarkable. This conclusion can have two explanations:

1) The stroke-level IWN is not so effective or the selected normalization algorithms particularly are not candidates suitable for this kind of normalization.
2) The intra-writer variability of children in primary school has not so significant effect on DD diagnosis. IWN is introduced to remove the variability of a specific task performed by the same person. This could have a positive effect, e.g. in biometric systems based on signature. Even though people are well trained to perform their signature, they usually perform it in a different way depending on their mood, actual psychical condition, etc. However, in the case of children that are still learning to write and whose handwriting is not proficient enough, the variability in a task performance is probably not so high.

Nevertheless, in some cases the normalization provided moderately better results than the baseline. In our recent work we reported a decrease in the HPSQ total score estimation error after a simple normalization based on a feature subtraction by approximately 3 % [5]. In this study, we proved that this error can be further decreased when employing more sophisticated stroke-level ℓ_2 norm normalization (decrease of EER by approximately 5 %), even applied on a smaller set of handwriting features. Some improvements in estimation error were reached by ℓ_1 norm (decrease of EER by approximately 3 %), and ℓ_∞ based normalization (decrease of EER by approximately 1 %) too.

Although some comparisons of the observed results with results reported by other scientific papers would be welcomed, this is a first study whose conclusions are based on a Czech dataset (containing Czech cursive handwriting) assessed by the

TABLE II
RESULTS OF UNIVARIATE REGRESSION ANALYSIS

		HPSQ				HPSQ-C			
Norm.	Movement	Feature Name	MAE	RMSE	EER	Feature Name	MAE	RMSE	EER
baseline	in-air	height of stroke (m)	7.28 ± 2.08	8.82 ± 2.25	25.99 ± 7.42	horizontal accel. (r)	6.37 ± 1.60	7.57 ± 1.65	19.90 ± 5.01
	on-surface	horizontal jerk (r)	7.97 ± 1.96	9.45 ± 2.03	28.48 ± 6.98	speed of writing (r)	6.20 ± 1.56	7.25 ± 1.69	19.35 ± 4.86
ℓ_∞	in-air	velocity (m)	6.39 ± 2.16	8.10 ± 2.48	22.84 ± 7.71	speed of writing (m)	6.39 ± 2.18	7.93 ± 2.53	19.98 ± 6.82
	on-surface	horizontal accel. (r)	7.94 ± 2.00	9.45 ± 1.97	28.35 ± 7.14	length of stroke (m)	5.76 ± 2.23	7.47 ± 2.98	18.00 ± 6.98
ℓ_1	in-air	length of stroke (m)	6.22 ± 1.89	7.82 ± 2.06	22.22 ± 6.74	vertical velocity (v)	6.35 ± 2.06	7.99 ± 2.48	19.86 ± 6.44
	on-surface	horizontal accel. (s)	6.97 ± 2.00	8.51 ± 2.15	24.88 ± 7.15	length of stroke (r)	6.24 ± 1.81	7.83 ± 2.25	19.51 ± 5.65
ℓ_2	in-air	velocity (m)	7.19 ± 2.43	8.83 ± 2.52	25.67 ± 8.67	dur. of stroke (m)	6.25 ± 1.83	7.76 ± 2.14	19.53 ± 5.72
	on-surface	speed of writing (m)	7.45 ± 1.94	8.95 ± 2.20	26.61 ± 6.95	length of stroke (v)	6.24 ± 1.73	7.46 ± 1.92	19.49 ± 5.42
z-score	in-air	jerk (m)	7.07 ± 2.18	8.73 ± 2.45	25.24 ± 7.78	velocity (r)	6.61 ± 1.58	7.70 ± 1.72	20.65 ± 4.94
	on-surface	speed of writing (m)	7.04 ± 1.92	8.79 ± 2.26	25.15 ± 6.84	height of stroke (mn)	6.26 ± 2.04	7.74 ± 2.50	19.58 ± 6.38

MAE – mean absolute error; RMSE – root mean square error; EER – equal error rate; HPSQ – Handwriting Proficiency Screening Questionnaire; HPSQ-C – Handwriting Proficiency Screening Questionnaire for Children; m – median; v – variance; r – range; mn – mean; s – std; accel. – acceleration; dur – duration.

TABLE III
MULTIVARIATE REGRESSION ANALYSIS

		HPSQ				HPSQ-C			
Normalization	Movement	NoF	MAE	RMSE	EER	NoF	MAE	RMSE	EER
baseline	in-air	4	7.61 ± 1.89	6.34 ± 2.04	22.64 ± 6.76	9	4.99 ± 1.52	6.26 ± 1.79	15.59 ± 4.75
	on-surface	14	6.62 ± 1.60	7.73 ± 1.66	23.63 ± 5.71	5	5.53 ± 1.61	6.71 ± 2.00	17.29 ± 5.02
ℓ_∞	in-air	3	5.72 ± 1.73	7.40 ± 1.91	20.45 ± 6.19	6	4.72 ± 1.44	5.79 ± 1.76	14.77 ± 4.51
	on-surface	9	6.72 ± 1.90	7.91 ± 1.97	23.99 ± 6.77	4	5.84 ± 1.55	7.18 ± 2.05	18.25 ± 4.85
ℓ_1	in-air	8	5.18 ± 1.69	6.32 ± 1.95	18.49 ± 6.03	6	5.58 ± 1.50	6.80 ± 1.85	17.45 ± 4.68
	on-surface	5	5.98 ± 1.62	7.06 ± 1.69	21.35 ± 5.77	5	5.55 ± 1.68	6.65 ± 1.94	17.36 ± 5.26
ℓ_2	in-air	8	4.99 ± 1.67	6.12 ± 1.99	17.82 ± 5.96	5	5.44 ± 1.61	6.70 ± 1.80	17.00 ± 5.03
	on-surface	7	5.74 ± 1.40	6.76 ± 1.51	20.49 ± 4.98	4	5.53 ± 1.65	6.73 ± 2.13	17.27 ± 5.14
z-score	in-air	6	6.05 ± 1.77	7.29 ± 1.98	21.62 ± 6.32	3	5.13 ± 1.66	6.17 ± 1.66	16.02 ± 5.20
	on-surface	3	6.17 ± 1.83	7.45 ± 2.05	22.03 ± 6.54	6	6.10 ± 1.57	7.67 ± 2.15	19.08 ± 4.91

NoF – number of features; MAE – mean absolute error; RMSE – root mean square error; EER – equal error rate; HPSQ – Handwriting Proficiency Screening Questionnaire; HPSQ-C – Handwriting Proficiency Screening Questionnaire for Children.

HPSQ/HPSQ-C scales. Therefore a relevant comparison is not possible. However, the goal of this paper was to explore the effect of IWN in computerized DD assessment. It has been proved that IWN has some impact in this field of science and that its further research could make computerized DD rating more accurate.

V. Conclusion

The aim of this study is to introduce new IWN techniques that will make computerized assessment of DD more accurate. To address this aim we developed a new stroke-level IWN methodology that is based on four normalization algorithms. Performance of these algorithms was compared to a baseline in automatic DD assessment scenarios utilizing a set of conventional online handwriting features. Based on the experimental analyses we proved that IWN has some benefits in computerized DD rating. Especially the ℓ_2 norm approach brought some promising results in comparison to the baseline (EER = 17.82 % vs. EER = 22.64 %).

To the best of our knowledge this is a first study dealing with the stroke-level IWN. Although the positive impact of this methodology in computerized DD assessment has been proved, this work has some limitations. The experimental results were conducted on a dataset that contains just a few dozens of samples. To generalize the results a larger dataset should be further analysed. Next, we proved that the idea of stroke-level IWN is promising, however, we used just a few simple normalization algorithms. We expect that more sophisticated normalization on the stroke-level could bring even better results. Taking into account the previously mentioned facts, this study should be considered as a pilot one.

References

[1] S. Rosenblum and G. Dror, “Identifying developmental dysgraphia characteristics utilizing handwriting classification methods,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 293–298, apr 2017. [Online]. Available: https://doi.org/10.1109/thms.2016.2628799

[2] S. Rosenblum, “Process versus product evaluation of poor handwriting among children with developmental dysgraphia and adhd,” 03 2018.

[3] A. W. V. Gemmert and H.-L. Teulings, “Advances in graphonomics: Studies on fine motor control, its development and disorders,” *Human Movement Science*, vol. 25, no. 4-5, pp. 447–453, oct 2006. [Online]. Available: https://doi.org/10.1016/j.humov.2006.07.002

[4] V. W. Berninger, K. B. Vaughan, R. D. Abbott, S. P. Abbott, L. W. Rogan, A. Brooks, E. Reed, and S. Graham, “Treatment of handwriting problems in beginning writers: Transfer from handwriting to composition.” *Journal of Educational Psychology*, vol. 89, no. 4, pp. 652–666, 1997. [Online]. Available: https://doi.org/10.1037/0022-0663.89.4.652

[5] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, “Identification and rating of developmental dysgraphia by handwriting analysis,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, apr 2017. [Online]. Available: https://doi.org/10.1109/thms.2016.2586605

[6] S. Rosenblum, S. Parush, and P. L. Weiss, "The in air phenomenon: Temporal and spatial correlates of the handwriting process," *Perceptual and Motor Skills*, vol. 96, no. 3, pp. 933–954, jun 2003. [Online]. Available: https://doi.org/10.2466/pms.2003.96.3.933

[7] S. Rosenblum and L. Gafni-Lachter, "Handwriting proficiency screening questionnaire for children (HPSQ–c): Development, reliability, and validity," *American Journal of Occupational Therapy*, vol. 69, no. 3, p. 6903220030p1, apr 2015. [Online]. Available: https://doi.org/10.5014/ajot.2015.014761

[8] S. Rosenblum, A. Y Dvorkin, and P. Weiss, "Automatic segmentation as a tool for examining the handwriting process of children with dysgraphic and proficient handwriting," vol. 25, pp. 608–21, 11 2006.

[9] R. P. Erhardt and V. Meade, "Improving handwriting without teaching handwriting: The consultative clinical reasoning process," *Australian Occupational Therapy Journal*, vol. 52, no. 3, pp. 199–210, 2005. [Online]. Available: http://dx.doi.org/10.1111/j.1440-1630.2005.00505.x

[10] V. Paz-Villagrán, J. Danna, and J.-L. Velay, "Lifts and stops in proficient and dysgraphic handwriting," *Human Movement Science*, vol. 33, pp. 381–394, feb 2014. [Online]. Available: https://doi.org/10.1016/j.humov.2013.11.005

[11] M. Prunty and A. L. Barnett, "Understanding handwriting difficulties: A comparison of children with and without motor impairment," *Cognitive Neuropsychology*, vol. 34, no. 3-4, pp. 205–218, may 2017. [Online]. Available: https://doi.org/10.1080/02643294.2017.1376630

[12] D. Morello, M. Ajcevic, I. Perrone, and A. Accardo, "Analysis of kinematic parameters relationships in normal and dysgraphic children," in *IFMBE Proceedings*. Springer Singapore, may 2018, pp. 189–193. [Online]. Available: https://doi.org/10.1007/978-981-10-9038-7_35

[13] H. L. McGlashan, C. C. Blanchard, N. J. Sycamore, R. Lee, B. French, and N. P. Holmes, "Improvement in children's fine motor skills following a computerized typing intervention," *Human Movement Science*, vol. 56, pp. 29–36, dec 2017. [Online]. Available: https://doi.org/10.1016/j.humov.2017.10.013

[14] S. Rosenblum, "Development, reliability, and validity of the handwriting proficiency screening questionnaire (hpsq)," vol. 62, pp. 298–307, 05 2008.

[15] C. Alonso-Martinez, M. Faundez-Zanuy, and J. Mekyska, "A comparative study of in-air trajectories at short and long distances in online handwriting," *Cognitive Computation*, vol. 9, no. 5, pp. 712–720, 2017.

[16] P. D'Antrassi, I. Perrone, A. Cuzzocrea, and A. Accardo, "A composite methodology for supporting early-detection of handwriting dysgraphia via big data analysis techniques," in *Intelligent Interactive Multimedia Systems and Services 2017*. Springer International Publishing, may 2017, pp. 241–253. [Online]. Available: https://doi.org/10.1007/978-3-319-59480-4_25

[17] A. Kushki, H. Schwellnus, F. Ilyas, and T. Chau, "Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia," *Research in Developmental Disabilities*, vol. 32, no. 3, pp. 1058–1064, may 2011. [Online]. Available: https://doi.org/10.1016/j.ridd.2011.01.026

[18] A. Richardson, S. Kraus, P. L. Weiss, and S. Rosenblum, "Coach-cumulative online algorithm for classification of handwriting deficiencies." in *AAAI*, 2008, pp. 1725–1730.

[19] M. Faundez-Zanuy and J. Mekyska, *User-Centric Privacy and Security in Biometrics*. Institution of Engineering and Technology, 2017, ch. Privacy of online handwriting biometrics related to biomedical analysis, pp. 17–39.

[20] G. Obozinski, B. Taskar, and M. Jordan, "Multi-task feature selection," *Statistics Department, UC Berkeley, Tech. Rep*, vol. 2, 2006.

[21] L. Wilkinson, A. Anand, and D. N. Tuan, "Linf: An l-infinity classifier."

[22] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.

[23] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, "Supervised feature learning for curvilinear structure segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*. Springer Berlin Heidelberg, 2013, pp. 526–533. [Online]. Available: https://doi.org/10.1007/978-3-642-40811-3_66

[24] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal.*, vol. 27, no. 8, pp. 1226–1238, 2005.

A.19 Fractional Derivatives of Online Handwriting: a New Approach of Parkinsonic Dysgraphia Analysis

Fractional Derivatives of Online Handwriting: a New Approach of Parkinsonic Dysgraphia Analysis

Jan Mucha*, Vojtech Zvoncak*, Zoltan Galaz*‡, Marcos Faundez-Zanuy†, Jiri Mekyska*, Tomas Kiska*, Zdenek Smekal*, Lubos Brabenec‡, Irena Rektorova‡§ and Karmele Lopez-de-Ipina¶
*Department of Telecommunications and SIX Research Centre, Brno University of Technology
Technicka 10, 61600 Brno, Czech Republic
†Escola Superior Politecnica, Tecnocampus
Avda. Ernest Lluch 32, 08302 Mataro, Barcelona, Spain
‡Applied Neuroscience Research Group, Central European Institute of Technology
Masaryk University, Kamenice 5, 62500 Brno, Czech Republic
§First Department of Neurology, Masaryk University and St. Anne's University Hospital
Pekarska 53, 65691 Brno, Czech Republic
¶Department of Systems Engineering and Automation
University of the Basque Country UPV/EHU, Av de Tolosa 54, 20018 Donostia, Spain

***Abstract*—Parkinson's disease (PD) is the second most frequent neurodegenerative disorder. One typical hallmark of PD is disruption in execution of practised skills such as handwriting. This paper introduces a new methodology of kinematic features calculation based on fractional derivatives applied on PD handwriting. Discrimination power of basic kinematic features (velocity, acceleration, jerk) was evaluated by classification analysis (using support vector machines and random forests). For this purpose, 30 PD patients and 36 healthy controls were enrolled. In comparison with results reported in other works, the newly designed features based on fractional derivatives increased classification accuracy by 8 % in univariate analysis and by 10 % when employing the multivariate one. This study reveals an impact of fractional derivatives based features in analysis of Parkinsonic dysgraphia.**

***Keywords*—Archimedean spiral; binary classification; fractal calculus; fractional derivative; online handwriting; Parkinson's disease;**

I. Introduction

Parkinsons disease (PD) is the second most frequent progressive neurodegenerative disorder in the world [1]. Its prevalence rate is estimated to approximately 1.5 % for people aged over 65 years [2]. Although, the exact pathophysiological cause of PD has not yet been discovered, a rapid degeneration of dopaminergic cells in *substancia nigra pars compacta* [3] emerged as the most significant biological finding associated with the disease. Tremor in rest, rigidity, bradykinesia, and loss of postural reflexes [4], [5] are considered as cardinal motor symptoms. PD also accompanies several non-motor symptoms such as sleep disorders, cognitive deficits, depression, dementia, etc. [6], [7].

Due to motor dysfunctions in people suffering from PD, some recent studies have suggested that quantitative analysis of handwriting can be used as a quick and accurate PD diagnosis method [8], [9]. Moreover, using digitizing tablets we are able to acquire online handwriting signals, where a temporal information is added to x and y coordinates. Therefore the analysis is not limited to spatial features quantifying mainly PD micrographia, but in addition, we are able to quantify temporal, kinematic and dynamic manifestations of PD (e. g. hesitations, pauses, and slow movement [10]), which are generally called PD dysgraphia [11].

For the purpose of PD handwriting analysis, several handwriting tasks were proposed (Archimedean spiral, repetitive loops, letters, words, sentences, etc.), but the most popular handwriting task for tremor assessment is currently the Archimedean spiral [5]. This task has been frequently used to evaluate motor performance in various movement disorders [12], [13], including PD. In view of these facts the Archimedean spiral was selected for the purposes of this study as well. Some related works (2014–now) focused on analysis of online handwriting in PD patients are summarized in Table I.

The aim of this paper is to introduce advanced kinematic features that replace the conventional ones by utilizing fractional derivative (FDE). The potential of FDE in PD dysgraphia quantification is demonstrated by classification analysis and a discrimination power of the newly designed features is compared with a baseline [5], [12], [9], [14].

This work was supported by the grant of the Czech Ministry of Health 16-30805A (Effects of non-invasive brain stimulation on hypokinetic dysarthria, micrographia, and brain plasticity in patients with Parkinsons disease), grant of the Czech Science Foundation 18-16835S (Research of advanced developmental dysgraphia diagnosis and rating methods based on quantitative analysis of online handwriting and drawing) and the following projects: LO1401, FEDER and MEC, TEC2016-77791-C4-2-R, and TEC- 2016-77791-C4-4-R from the Ministry of Economic Affairs and Competitiveness of Spain. For the research, infrastructure of the SIX Center was used.

TABLE I. OVERVIEW OF RELATED WORKS FOCUSED ON ANALYSIS OF PD DYSGRAPHIA

First author	Year	PD/HC	Handwriting task	Analysis	Features	Conclusions
Broeder [15]	2014	18/11	Repetitive loops	Correlation (Spearman)	Writing amplitude and velocity	The highest correlation for given medical scale and task was with velocity with r = 0.627.
Drotar [14]	2014	37/38	One sentence	Differential analysis (SVM)	Kinematic, temporal, spatial and its statistical representations	Using both in-air/on-surface features resulted in 85.61 % classification accuracy.
Drotar [16]	2015	37/38	Characters, words, sentences	Differential analysis (SVM)	Same as in [14], entropy, empirical mode decomposition, signal energy	The highest classification accuracy after feature selection approach was 88.1 %.
Drotar [12]	2015	37/38	Characters, words, sentences	Differential analysis (SVM)	Same as in [16], pressure	Classification performance was at its peak with on-surface features (89.09 %).
Drotar [5]	2016	37/38	Characters, words, Archimedean spiral, sentences	Differential an. (SVM, K-NN, AdaBoost)	Same as in [12], but in-air features were not computed	For classification of PD on all exercises SVM proved to be the best classifier with accuracy 82.5 %.
Heremans [17]	2015	34/10	Up/down strokes at varying amplitudes	ANOVA	Writing amplitude and velocity	Significant difference between groups was in writing amplitude (F(2.41) = 3.97; p = 0.03).
Heremans [18]	2016	30/15	Repetitive cursive loops	ANOVA, correlation	Writing amplitude and velocity	Medical scale and writing amplitude had significant correlation $r = -0.40$.
Loconsole [19]	2017	4/7	Sentence, repetitive loops	Differential analysis (ANN)	Execution time and average speed, density ratio, height ratio	Highest classification accuracy 96.81 % was achieved using all the extracted features.
Masarova [20]	2014	40/40	Characters, words, Archimedean spiral, sentences	Correlation (Spearman)	Velocity, acceleration, jerk, statistical representations of each one	The most significant relative difference between groups was 19.5 % for mean velocity of writing extracted from long sentence.
Nackaerts [21]	2017	38/0	Repetitive loops, eight-like figure	Correlation (Spearman)	Stroke duration, writing velocity, normalized jerk	Amplitude training has as negative effect on fluency and stroke duration.
Smits [13]	2014	10/10	Circle, spiral, line characters, sentence	t-test	Kinematic, temporal, spatial and its statistical representations	Time per repetition, velocity, and acceleration have the highest discriminative power.

SVM – support vector machine; r_s – Spearman's correlation coefficient; K-NN – K-nearest neighbours; ANOVA – analysis of variance; ANN – artificial neural network; F and p corresponds to variables of F distribution; articles are sorted alphabetically and then by year of release.

II. MATERIALS AND METHODS

A. Dataset

The dataset consisted of 66 participants: 36 healthy controls (HC) with (mean $\pm$ std) age: 62.50 $\pm$ 11.70 years, and 30 PD patients with (mean $\pm$ std) age: 68.37 $\pm$ 11.08 years, PD duration: 8.67 $\pm$ 4.49 years, UPDRS V (Unified Parkinson's disease rating scale, part V: Modified Hoehn & Yahr staging score) [22]: 2.23 $\pm$ 0.83 and LED (L-dopa equivalent daily dose) [23]: 1474.67 $\pm$ 614.81 mg. The participants were enrolled at the First Department of Neurology, St. Anne's University Hospital in Brno, Czech Republic. All participants reported Czech language as their native language and all participants were right-handed. The PD patients completed the tasks approximately 1 hour after their regular L-dopa medication. All participants signed an informed consent form approved by the local ethics committee.

B. Data Acquisition

The Archimedean spiral task is a part of the PaHaW database [5]. During this task, a template was shown to a subject for visual guidance. Participants drew the spiral from its center, but were not asked to draw it within particular boundaries or to follow a pre-drawn line. Online handwriting signals were recorded using the Intuos 4M (Wacom technology) digitizing tablet, with sampling rate $f_{\mathrm{s}} = 100$ Hz. The tablet was overlaid with an empty paper template. The following features were acquired (time sequences): x and y coordinates – $x[t]$, $y[t]$; time-stamp – t; in-air/on-surface status – $b[t]$; pressure – $p[t]$; azimuth $az[t]$; and tilt $al[t]$.

C. Fractional Derivative

Several approaches of fractional derivative calculation exist [24]. In this paper, the implementation of FDE by Jonathan Hadida, which follows the Grünwald-Letnikov approximation [25], was used. A direct definition of the FDE $D^{\alpha}y(t)$ is based on finite differences of an equidistant grid in $[0, \tau]$. Assume that the function $y(\tau)$ satisfies some smoothness conditions in every finite interval $(0, t), t \leq T$. Choosing the grid

$$0 = \tau_0 < \tau_1 < ... < \tau_{n+1} = t = (n+1)h \tag{1}$$

with

$$\tau_{k+1} - \tau_k = h \tag{2}$$

and using the notation of the finite differences

$$\frac{1}{h^{\alpha}}\Delta_h^{\alpha} y(t) = \frac{1}{h^{\alpha}}\left(y(\tau_{n+1}) - \sum_{v=1}^{n+1} c_v^{\alpha} y(\tau_{n+1-v})\right), \tag{3}$$

where

$$c_v^{\alpha} = (-1)^{v-1}\binom{\alpha}{v}, \tag{4}$$

the Grünwald-Letnikov implementation is defined as [24]:

$$D^{\alpha}y(t) = \lim_{h\to 0}\frac{1}{h^{\alpha}}\Delta_h^{\alpha}y(t), \tag{5}$$

where $D^{\alpha}y(t)$ means a derivative with order α of function $y(t)$, and h represents sampling lattice.

In our case, the FDE substitutes the conventional differential derivative during calculation of the kinematic features. A detailed description of the FDE can be found at [24], [25].

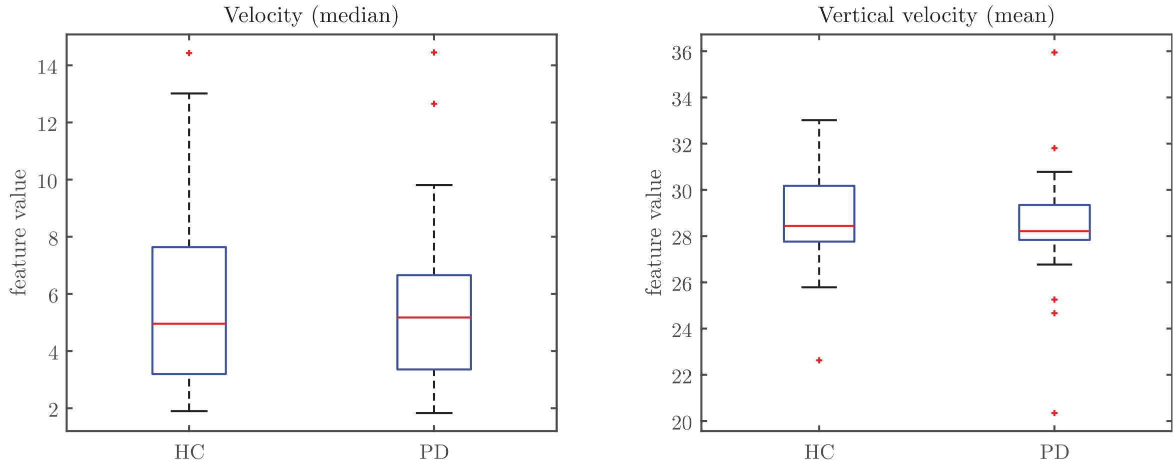


Fig. 1. Box plots of two features with the highest MCC (univariate models).

D. Handwriting Features

To demonstrate the impact of FDE in analysis of PD dysgraphia, we extracted only basic on-surface kinematic parameters [5], [9]: velocity – rate at which a position of pen changes with time [mm/s]; acceleration – rate at which the velocity of pen changes with time [mm/s^2]; jerk – rate at which the acceleration of pen changes with time [mm/s^3]; and their horizontal and vertical variants. These features were calculated for different orders α of the FDE in range from 0.1 to 1.0 with 0.1 steps. Consequently, statistical properties of the features were described using following statistics: mean, median, standard deviation (std), and maximum (max) [5], [12], [9], [14]. In total 360 features were extracted.

E. Statistical Analysis

To evaluate a discrimination power of the features, univariate binary classification (PD/HC) models (stratified 7-fold cross-validation with 50 repetitions) based on random forests (RF) [26] and support vector machines (SVM) [27] with radial basis function (RBF) were employed. Next, some improvements in classification accuracy were done by multivariate approach with the same classifiers and the same cross-validation settings. In this case, the sequential floating forward selection (SFFS) algorithm was used [28] in order to select the most appropriate combination of the features. Classification performance was evaluated by the Matthew's correlation coefficient [29], classification accuracy (ACC), sensitivity (SEN) and specificity (SPE).

III. Results

Results of the univariate and multivariate analysis are summarized in Table II. Regarding the univariate classification, only the best features (in terms of the MCC values) are reported. The best feature of the univariate classification is median of velocity with $\alpha = 0.1$ (ACC $= 70.55\,\%$ classified by SVM). Box plots of two features with the highest MCC are visualized in Figure 1. Regarding the multivariate classification analysis, ACC of $72.39\,\%$ (MCC $= 0.44$) was achieved using combination of 10 features classified by RF. The set of these features as gradually selected by SFFS can be found in Table III.

TABLE II. Results of the univariate and multivariate classification analysis

Univariate analysis						
Classifier	Feature	α	MCC	ACC [%]	SEN [%]	SPE [%]
SVM	velocity (median)	0.1	0.40	70.55	62.00	77.67
SVM	vertical velocity (mean)	0.6	0.40	70.24	52.40	85.11
RF	jerk (max)	1.0	0.33	67.06	59.53	73.34
RF	vertical jerk (median)	0.1	0.29	65.34	54.40	74.45
Multivariate analysis						
Classifier	Number of features		MCC	ACC [%]	SEN [%]	SPE [%]
RF	10		0.44	72.39	65.52	77.87
SVM	11		0.39	67.55	57.42	79.96

α – order of fractional derivative; RF – random forests; SVM – support vector machine; MCC – Matthew's correlation coefficient; ACC – accuracy; SEN – sensitivity; SPE – specificity.

TABLE III. The best combination of features in the multivariate classification (employing RF) selected by SFFS

Feature	α	ACC [%]
vertical jerk (median)	0.1	64.89
acceleration (mean)	0.3	65.97
horzintal velocity (median)	0.1	68.17
vertical jerk (mean)	0.8	70.34
horzintal velocity (median)	0.4	71.52
vertical acceleration (median)	0.5	71.15
horzintal acceleration (median)	1.0	72.36
vertical acceleration (median)	0.6	71.77
velocity (median)	0.8	72.01
horzintal jerk (median)	1.0	72.39

IV. Discussion

According to the reported results we can confirm our previous hypothesis that application of the FDE in calculation of kinematic features brings promising potential in automatic diagnosis of PD dysgraphia. Considering that only the basic kinematic features such as velocity, acceleration, and jerk were extracted, the results of discrimination analysis are promising, especially when compared with previous related papers (baseline) [5], [12], [9], [14], where the Archimedean spiral task was eliminated from the final classification models due to low ACC (62–65 %). In the case of univariate analysis we can claim that the ACC was improved by 3–8 %. Based on the results summarized in Figure 1, we can confirm reduced movement abilities in PD cohort, which is caused mainly by rigidity and bradykinesia. The best result of multivariate analysis (ACC $= 72.38\,\%$, MCC $= 0.44$) was achieved by the RF classifier in combination with 10 features selected by SFFS. In comparison to the baseline, this result means improvement by 10 %. Moreover, from the feature set description (see Table III), it is evident that most of the parameters were based on $\alpha \neq 1$, which confirms full utilization of the FDE.

V. Conclusion

With respect to the results we can conclude that using the FDE in kinematic analysis brings new improvements in quantitative PD dysgraphia processing and add-on to the existing conventional techniques. This study is considered as a pilot one and its conclusions should be confirmed and extended by further research. For instance, it would be interesting to combine the newly developed parameters with other features such as temporal, spatial or dynamic ones. Moreover, the other tasks (e.g. overlapped circles, words, drawings) could be quantified. Another implementation of the FDE should be evaluated as well. Finally, a bigger dataset must be used to be able to generalize the conclusions.

References

[1] M. C. de Rijk, L. J. Launer, K. Berger, M. M. Breteler, J. F. Dartigues, M. Baldereschi, L. Fratiglioni, J. Lobo, A. Martinez-Lage, C. Trenkwalder, and A. Hofman, “Prevalence of Parkinson’s disease in Europe: A collaborative study of population-based cohorts,” *Neurology*, vol. 54, pp. 21–23, 2000.

[2] S. Sapir, L. Ramig, and C. Fox, “Speech and swallowing disorders in Parkinson disease,” *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 16, no. 3, pp. 205–10, 2008.

[3] O. Hornykiewicz, “Biochemical aspects of Parkinson’s disease,” *Neurology*, vol. 51, no. 2 Suppl 2, pp. S2–S9, 1998.

[4] J. Mekyska, Z. Smekal, M. Kostalova, M. Mrackova, S. Skutilova, and I. Rektorova, “Motor aspects of speech impairment in Parkinson’s disease and their assessment,” *Cesk Slov Neurol N*, vol. 74, no. 6, pp. 662–668, 2011.

[5] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, “Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson’s disease,” *Artificial Intelligence in Medicine*, vol. 67, pp. 39 – 46, 2016.

[6] L. Brabenec, J. Mekyska, Z. Galaz, and I. Rektorova, “Speech disorders in Parkinson’s disease: early diagnostics and effects of medication and brain stimulation,” *Journal of Neural Transmission*, pp. 1–32, 2017.

[7] J. Mucha, Z. Galaz, J. Mekyska, T. Kiska, V. Zvoncak, Z. Smekal, I. Eliasova, M. Mrackova, M. Kostalova, I. Rektorova *et al.*, “Identification of hypokinetic dysarthria using acoustic analysis of poem recitation,” in *Telecommunications and Signal Processing (TSP), 2017 40th International Conference on*. IEEE, 2017, pp. 739–742.

[8] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, and I. Schlesinger, “Handwriting as an objective tool for Parkinsons disease diagnosis,” *Journal of neurology*, vol. 260, no. 9, pp. 2357–2361, 2013.

[9] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, “A new modality for quantitative evaluation of Parkinson’s disease: In-air movement,” in *13th IEEE International Conference on BioInformatics and BioEngineering, BIBE 2013, Chania, Greece, November 10-13, 2013*. IEEE Computer Society, 2013, pp. 1–4.

[10] J. L. Contreras-Vidal, P. Poluha, H.-L. Teulings, and G. E. Stelmach, “Neural dynamics of short and medium-term motor control effects of levodopa therapy in Parkinson’s disease1supported in part by ninds grant ro1 ns33173, the flinn foundation and the bryng bryngelson fund.1,” *Artificial Intelligence in Medicine*, vol. 13, no. 1, pp. 57 – 79, 1998.

[11] A. Letanneux, J. Danna, J.-L. Velay, F. Viallet, and S. Pinto, “From micrographia to Parkinson’s disease dysgraphia,” *Movement Disorders*, vol. 29, no. 12, pp. 1467–1475, 2014.

[12] P. Drotar, J. Mekyska, Z. Smekal, I. Rektorova, L. Masarova, and M. Faundez-Zanuy, “Contribution of different handwriting modalities to differential diagnosis of Parkinson’s disease,” in *Medical Measurements and Applications (MeMeA), 2015 IEEE International Symposium*, 2015, pp. 1–5.

[13] E. J. Smits, A. J. Tolonen, L. Cluitmans, M. van Gils, B. A. Conway, R. C. Zietsma, K. L. Leenders, and N. M. Maurits, “Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinson’s disease,” *PloS one*, vol. 9, no. 5, p. e97614, 2014.

[14] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, “Analysis of in-air movement in handwriting: A novel marker for Parkinson’s disease,” *Computer Methods and Programs in Biomedicine*, vol. 117, no. 3, pp. 405–411, 2014.

[15] S. Broeder, E. Nackaerts, A. Nieuwboer, B. Smits-Engelsman, S. Swinnen, and E. Heremans, “The effects of dual tasking on handwriting in patients with Parkinson’s disease,” *Neuroscience*, vol. 263, pp. 193–202, mar 2014.

[16] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. F. Zanuy, “Decision support framework for Parkinson’s disease based on novel handwriting markers,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, pp. 1–1, 2015.

[17] E. Heremans, E. Nackaerts, G. Vervoort, S. Vercruysse, S. Broeder, C. Strouwen, S. P. Swinnen, and A. Nieuwboer, “Amplitude manipulation evokes upper limb freezing during handwriting in patients with Parkinson’s disease with freezing of gait,” *PLOS ONE*, vol. 10, no. 11, p. e0142874, nov 2015.

[18] E. Heremans, E. Nackaerts, S. Broeder, G. Vervoort, S. P. Swinnen, and A. Nieuwboer, “Handwriting impairments in people with Parkinson’s disease and freezing of gait,” *Neurorehabilitation and Neural Repair*, vol. 30, no. 10, pp. 911–919, jul 2016.

[19] C. Loconsole, G. F. Trotta, A. Brunetti, J. Trotta, A. Schiavone, S. I. Tato, G. Losavio, and V. Bevilacqua, “Computer vision and EMG-based handwriting analysis for classification in Parkinson’s disease,” in *Intelligent Computing Theories and Application*. Springer International Publishing, 2017, pp. 493–503.

[20] L. Masarova, P. Drotar, J. Mekyska, , Z. Smekal, and I. Rektorova, “Assessing handwriting in patients with Parkinson’s disease,” *The Czech and Slovak Neurology and Neurosurgery*, pp. 456–462, 2014.

[21] E. Nackaerts, S. Broeder, M. P. Pereira, S. P. Swinnen, W. Vandenberghe, A. Nieuwboer, and E. Heremans, “Handwriting training in Parkinson’s disease: A trade-off between size, speed and fluency,” *PLOS ONE*, vol. 12, no. 12, p. e0190223, dec 2017.

[22] S. Fahn and R. L. Elton, *UPDRS Development Committee (1987) Unified Parkinson’s Disease Rating Scale*. Recent developments in Parkinson’s Disease. Macmillan, Florham Park, 1987.

[23] J. Y. Lee, J. W. Kim, W. Y. Lee, J. M. Kim, T. B. Ahn, H. J. Kim, J. Cho, and B. S. Jeon, “Daily dose of dopaminergic medications in Parkinson’s disease: clinical correlates and a posteriori equation,” *Neurol. Asia*, vol. 15, no. 2, pp. 137–143, 2010.

[24] I. Podlubny, *Fractional differential equations an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*. San Diego: Academic Press, 1999.

[25] R. Scherer, S. L. Kalla, Y. Tang, and J. Huang, “The grünwald–letnikov method for fractional differential equations,” *Computers & Mathematics with Applications*, vol. 62, no. 3, pp. 902–917, 2011.

[26] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[27] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 1995.

[28] J. Pohjalainen, O. Rasanen, and S. Kadioglu, “Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits,” *Comput. Speech Lang.*, 2014.

[29] B. W. Matthews, “Comparison of the predicted and observed secondary structure of T4 phage lysozyme,” *Biochim. Biophys. Acta (BBA)*, vol. 405, no. 2, pp. 442–51, 1975.

A.20 A Comparative Study of In-Air Trajectories at Short and Long Distances in Online Handwriting

Cogn Comput (2017) 9:712–720
DOI 10.1007/s12559-017-9501-5

A Comparative Study of In-Air Trajectories at Short and Long Distances in Online Handwriting

Carlos Alonso-Martinez [1] · Marcos Faundez-Zanuy [1] · Jiri Mekyska [2]

Received: 17 February 2017 / Accepted: 18 July 2017 / Published online: 27 July 2017

Abstract Existing literature about online handwriting analysis to support pathology diagnosis has taken advantage of in-air trajectories. A similar situation occurred in biometric security applications where the goal is to identify or verify an individual using his signature or handwriting. These studies do not consider the distance of the pen tip to the writing surface. This is due to the fact that current acquisition devices do not provide height formation. However, it is quite straightforward to differentiate movements at two different heights (a) short distance: height lower or equal to 1 cm above a surface of digitizer, the digitizer provides x and y coordinates; (b) long distance: height exceeding 1 cm, the only information available is a time stamp that indicates the time that a specific stroke has spent at long distance. Although short distance has been used in several papers, long distances have been ignored and will be investigated in this paper. In this paper, we will analyze a large set of databases (BIOSECUR-ID, EMOTHAW, PaHaW, OXYGEN-THERAPY, and SALT), which contain a total amount of 663 users and 17,951 files. We have specifically studied (a) the percentage of time spent on-surface, in-air at short distance, and in-air at long distance for different user profiles (pathological and healthy users) and different tasks; (b) the potential use of these signals to improve classification rates. Our experimental results reveal that long distance movements represent a very small portion of the total execution time (0.5% in the case of signatures and 10.4% for uppercase words of BIOSECUR-ID, which is the largest database). In addition, significant differences have been found in the comparison of pathological versus control group for letter "l" in PaHaW database ($p = 0.0157$) and crossed pentagons in SALT database ($p = 0.0122$).

Keywords Handwriting · Biometrics · In-air trajectories

✉ Carlos Alonso-Martinez
calonso@tecnocampus.cat

Marcos Faundez-Zanuy
faundez@tecnocampus.cat

Jiri Mekyska
mekyska@feec.vutbr.cz

1 ESUP Tecnocampus (Pompeu Fabra University), Av. Ernest Lluch 32, 08302 Mataró, Spain

2 Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technicka 10, 616 00 Brno, Czech Republic

Introduction

Speech and handwriting are probably the most difficult tasks performed by human beings, because they differentiate us from animals. Handwriting requires very fine motor skills, probably more so than speech, because some animals can imitate human sounds but no animal can write. In addition, we learn to speak first and then we learn how to read and write, when the brain is more mature.

Handwriting analysis is a good way to study the human brain in a non-invasive way. This knowledge, once acquired, can be applied to artificial systems that emulate the human brain. We consider that handwriting movements are more complex by far than what has been analyzed in the past. In fact, some parts of the movements have been neglected. With this paper, we will analyze this kind of movements, which will be defined in posterior sections as in-air at long distance. This kind of movements can be used to improve artificial intelligence for biometric applications such as health and security [1–4].

Fig. 1 Intuos Pro L digitizing tablet and pen

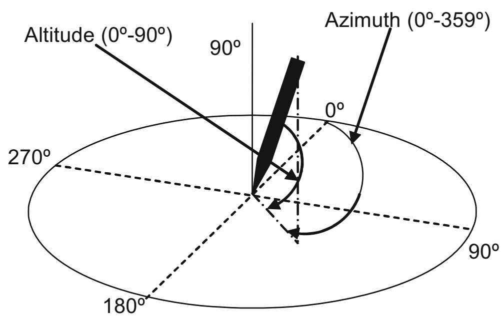


Fig. 2 Azimuth and altitude angles of the pen with respect to the plane of the writing surface

In the past, the analysis of handwriting had to be performed in an offline manner. Only the writing itself (strokes on a piece of paper) were available for analysis. Nowadays, modern-capturing devices like digitizing tablets and pens or online whiteboards can gather data without losing its temporal dimension. When spatiotemporal information is available, its analysis is referred to as online. A typical modern-digitizing tablet (Fig. 1) not only gathers the x-y coordinates that describe the movement of the writing device as it changes its position, but it can also collect other data, mainly the pressure exerted by the writing device on the writing surface, the azimuth (the angle of the pen in the horizontal plane), and the altitude (the angle of the pen with respect to the vertical axis) (see (Fig. 2)). From now own, x-y coordinates, pressure, azimuth, and altitude will be referred to as *features of the handwriting*.

A very interesting aspect of the modern online analysis of handwriting is that it can consider information gathered when the writing device was not exerting pressure on the writing surface. Thus, the movements performed by the hand while writing a text can be split into two classes:

1. On-surface trajectories (pen-downs), corresponding to the movements executed while the writing device is touching the writing surface. Each of these trajectories produces a

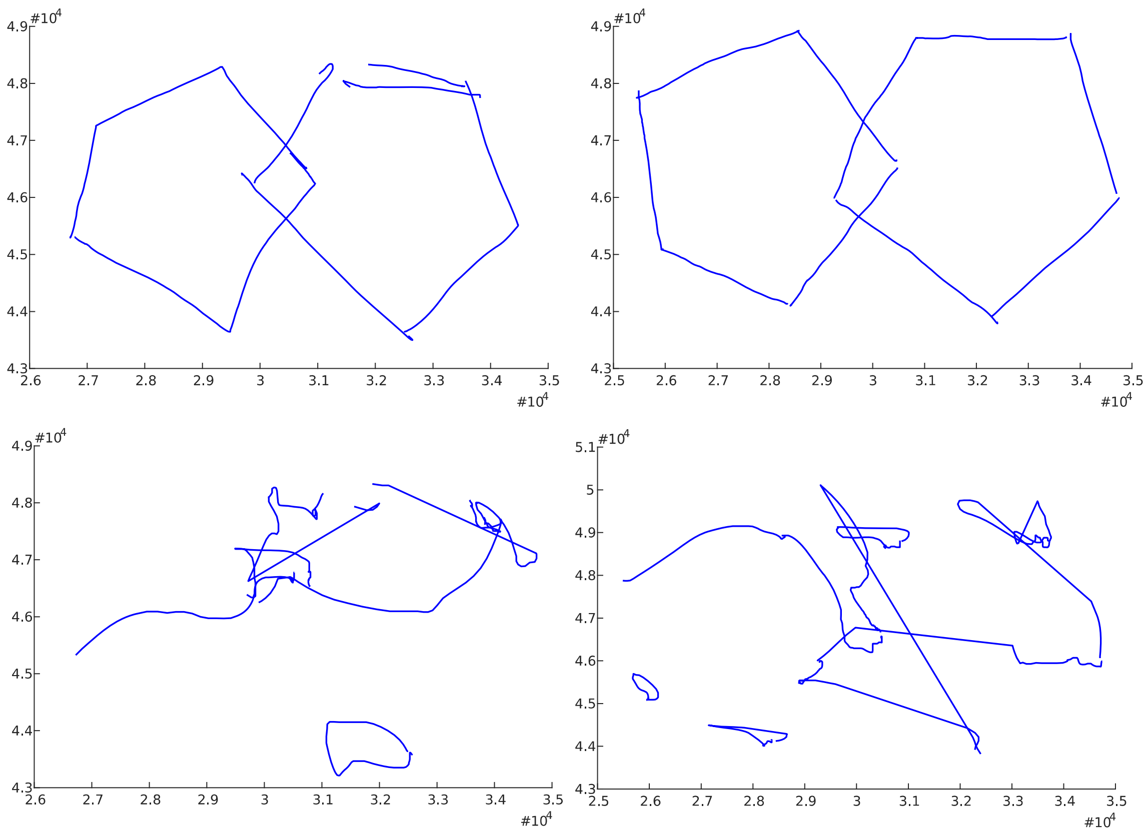


Fig. 3 On-surface (top) and in-air (bottom) trajectories from two executions of two crossed pentagons

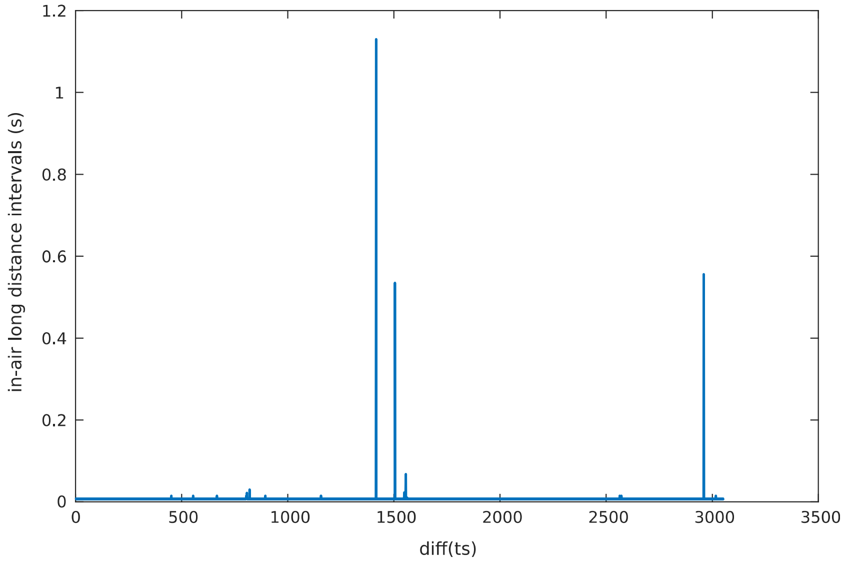


Fig. 4 Time stamp difference of consecutive samples for an example of accepted file from PaHaW database task write *lektorka* word twice

visible stroke. We will call this kind of movement on-surface.

2. In-air trajectories (pen-ups), corresponding to the movements performed by the hand while transitioning from one stroke to the next one. During these movements, the writing device exerts no pressure on the surface. This class can be split into two subsets:

    a. In-air at short distances (in-air$_S$), when the distance from the tip of the pen to the writing surface is lower or equal to 1 cm. In this case, the digitizing device can track the (x, y) coordinates during the pen movement.
    b. In-air at long distances (in-air$_L$), when distances from the tip of the pen to the writing surface are higher than 1 cm. In this case, the digitizing device is not able to track the movements and we only know the time spent at high distance.

In our previous research, we have focused on on-surface and in-air$_S$ movements discarding in-air$_L$ movements because they do not provide the same amount of data as the previous ones. In fact, the unique parameters are just the number of strokes at long distance and time spent at long distance. For instance, in [5], we applied information theory to demonstrate that on-surface and in-air$_S$ contain almost the same amount of information and they are not redundant. This was an important milestone because in-air trajectories had received almost no attention at all, even in online approaches where spatiotemporal information is available.

Figure 3 shows two examples of on-surface and in-air$_S$ trajectories taken from two executions of the pentagon test performed by two different writers from the Emothaw database.

In-air$_L$ can be detected looking at the time stamp provided by the digitizing tablet. During in-air$_L$ time, the tablet is

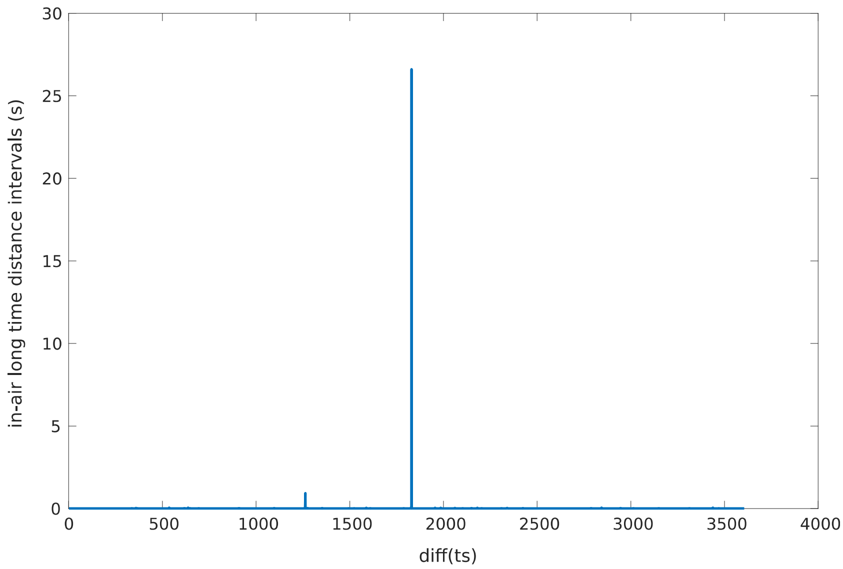


Fig. 5 Time stamp difference of consecutive samples for an example of discarded file from PaHaW database task write *lektorka* word twice

Table 1 BIOSECUR-ID database. Time in absolute units and relative time in parenthesis

	Time			Strokes		
Task	On-surface	In-air$_S$	In-air$_L$	On-surface	In-air$_S$	In-air$_L$
Genuine signature	2857.6 (79.6%)	715.4 (19.9%)	17.5 (0.5%)	6.62	5.94	0.32
Skilled forgeries	5447.9 (68.5%)	2373.4 (29.9%)	128.5 (1.6%)	6.58	6.21	0.63
Lower case words	110,445.1 (55.9%)	76,454 (38.7%)	10,644.4 (5.4%)	335.01	367.16	33.16
Numbers	3677.3 (53.6%)	3071.1 (44.7%)	117.0 (1.7%)	11.66	11.46	0.79
Uppercase words	73,608.8 (54.4%)	47,756.2 (35.3%)	14,073.4 (10.4%)	313.49	343.29	30.81

unable to track the tip of the pen and no samples are acquired. Nevertheless, time stamp is increasing and the next time that the pen touches the surface, the samples are stored again in the file and the time jump can be detected. Figure 4

Table 2 EMOTHAW database. Time in absolute units and relative time in parenthesis

	Time			Strokes		
Task	On-surface	In-air$_S$	In-air$_L$	On-surface	In-air$_S$	In-air$_L$
a. Depression						
Two-pentagon	11394.0 (55.5%)	7755.8 (37.7%)	1393.3 (6.8%)	9.26	13.15	9.47
House	18765.4 (53.6%)	13933.1 (39.8%)	2329.4 (6.7%)	23.74	33.00	20.97
Capital letters	15789.4 (51.0%)	13112.3 (42.4%)	2050.1 (6.6%)	59.79	65.91	12.15
Loops with left hand	10183.9 (97.7%)	215.8 (2.1%)	21.3 (0.2%)	1.26	0.41	0.21
Loops with right hand	8542.7 (98.9%)	58.6 (0.7%)	39.3 (0.4%)	1.18	0.21	0.06
Clock	14228.8 (45.0%)	14905.2 (47.2%)	2468.7 (7.8%)	27.35	36.91	21.44
Sentence	15288.8 (60.4%)	8052.4 (31.8%)	1958.5 (7.8%)	41.24	47.41	11.09
b. Stress						
Two-pentagon	11283.0 (55.0%)	7768.6 (37.9%)	1444.1 (7.1%)	9.41	13.89	11.39
House	18868.4 (52.5%)	14378.6 (40.0%)	2685.6 (7.5%)	25.45	35.14	21.32
Capital letters	15732.3 (50.1%)	13555.7 (43.1%)	2135.2 (6.8%)	60.80	67.09	12.04
Loops with left hand	10648.5 (97.3%)	233.5 (2.1%)	66.6 (0.6%)	1.59	0.77	0.95
Loops with right hand	9264.1 (99.3%)	40.0 (0.4%)	23.9 (0.3%)	1.13	0.14	0.04
Clock	14481.5 (44.8%)	14934.1 (46.2%)	2896.2 (9.0%)	27.63	37.80	21.41
Sentence	15756.6 (59.4%)	8539.8 (32.2%)	2215.8 (8.4%)	42.55	48.95	10.84
c. Anxiety						
Two-pentagon	11474.7 (57.3%)	7135.2 (35.6%)	1420.7 (7.1%)	8.70	12.75	10.16
House	18871.9 (53.6%)	13683.5 (38.8%)	2672.7 (7.6%)	23.77	32.89	18.95
Capital letters	16010.0 (50.9%)	13356.9 (42.5%)	2082.7 (6.6%)	60.48	66.39	10.96
Loops with left hand	10248.4 (96.9%)	224.3 (2.1%)	103.0 (1.0%)	1.57	0.79	0.96
Loops with right hand	8793.2 (99.3%)	35.6 (0.4%)	23.9 (0.3%)	1.11	0.13	0.04
Clock	14175.3 (46.5%)	13487.9 (44.3%)	2811.3 (9.2%)	26.27	35.48	19.54
Sentence	15676.5 (59.9%)	8402.2 (32.1%)	2107.5 (8.0%)	41.96	48.14	10.38
d. Control						
Two-pentagon	10256.0 (49.7%)	8670.5 (42.1%)	1684.9 (8.2%)	10.13	15.27	12.91
House	17468.1 (49.0%)	15150.2 (42.5%)	3044.5 (8.5%)	26.63	36.23	22.27
Capital letters	15699.2 (48.9%)	13721.2 (42.8%)	2677.1 (8.3%)	61.46	67.84	11.68
Loops with left hand	9737.1 (98.5%)	133.3 (1.3%)	17.7 (0.2%)	1.18	0.30	0.41
Loops with right hand	8992.1 (98.4%)	123.2 (1.3%)	23.9 (0.3%)	1.07	0.09	0.04
Clock	12365.6 (38.9%)	16180.8 (50.9%)	3229.5 (10.2%)	27.13	37.25	22.63
Sentence	15660.0 (53.6%)	9539.6 (32.6%)	4024.3 (13.8%)	42.41	49.43	11.43

Table 3 PAHAW database. Time in absolute units and relative time in parenthesis

Task	Time			Strokes		
	On-surface	In-air$_S$	In-air$_L$	On-surface	In-air$_S$	In-air$_L$
a. Control						
Spiral	18,665.8 (98.6%)	171.5 (0.9%)	103.2 (0.5%)	1.40	1.97	1.94
Letter *l*	8077.8 (57.6%)	3868.3 (27.6%)	2069.6 (14.8%)	5.21	18.16	15.50
Bigram *le*	10,545.9 (71.2%)	2998.4 (20.2%)	1274.3 (8.6%)	5.13	14.03	11.00
Word *les*	12,309.1 (69.2%)	3513.0 (19.7%)	1977.7 (11.1%)	5.29	15.11	11.82
Word *lektorka*	14,931.2 (73.0%)	3238.1 (15.9%)	2279.8 (11.1%)	7.00	16.97	12.00
Word *porovnat*	13,071.5 (74.4%)	3356.7 (19.1%)	1139.4 (6.5%)	8.08	18.08	11.82
Word *nepopadnout*	8757.5 (83.8%)	1512.5 (14.5%)	179.0 (1.7%)	5.29	8.47	4.50
Sentence	14,481.3 (58.4%)	7457.9 (30.1%)	2844.6 (11.5%)	15.24	31.87	19.34
b. Parkinson patients						
Spiral	24,057.4 (95.4%)	618.3 (2.4%)	536.6 (2.2%)	2.03	6.78	7.31
Letter *l*	8928.1 (63.8%)	4132.5 (29.5%)	939.1 (6.7%)	5.51	16.08	12.59
Bigram *le*	12,143.2 (69.1%)	4094.1 (23.3%)	1330.4 (7.6%)	5.57	17.08	13.76
Word *les*	14,702.7 (69.6%)	4093.1 (19.4%)	2330.9 (11.0%)	5.76	19.22	15.54
Word *lektorka*	17,716.2 (76.3%)	36,045.0 (15.5%)	1890.1 (8.2%)	7.22	17.97	12.49
Word *porovnat*	14,690.6 (75.8%)	3808.9 (19.6%)	891.1 (4.6%)	8.86	18.11	11.00
Word *nepopadnout*	9784.0 (79.8%)	2115.7 (17.2%)	365.6 (3.0%)	6.76	11.30	5.86
Sentence	16,176.5 (58.2%)	8252.3 (29.9%)	3300.1 (11.9%)	16.57	36.81	23.62

shows the difference of consecutive time stamps for an example file. For most of the samples (on-surface and in-air$_S$), this value is small (typically two units). However, there are some peaks, which correspond to in-air$_L$ movements. Figure 4 reveals 11 strokes of the type in-air$_L$. Sometimes, this time is abnormally long. This is probably due to some acquisition problem, where the user started to speak with the database acquisition supervisor for minutes. We will label these cases and will not include them in the average computation of time spent at in-air$_L$. We consider these cases when time in-air$_L$ is greater than 70% of the total time. In particular, we have found this phenomenon in 5 files from the analyzed databases (total amount of analyzed files is 17,951 files) (e.g. see Fig. 5).

Experimental Databases

In this paper, we have analyzed a set of different databases that contain different tasks and user profiles. The databases share the existence of handwritten tasks. In this section, we will summarize the main characteristics of the analyzed databases.

BIOSECUR-ID

This database is a multimodal biometric one and includes eight biometric traits: speech, iris, face (still images and videos), handwritten signature and handwritten text, fingerprints, hand, and keystroking. This database acquired inside the Biosecur-ID project was developed by a consortium of six

Table 4 OXIGEN-THERAPY database. Time in absolute units and relative time in parenthesis

Task	Time			Strokes		
	On-surface	In-air$_S$	In-air$_L$	On-surface	In-air$_S$	In-air$_L$
a. Before O_2						
House	32,699.0 (49.6%)	22,184.8 (33.7%)	11,033.8 (16.7%)	28.88	131.13	141.29
Clock	20,144.0 (40.2%)	21,824.0 (43.6%)	8104.3 (16.2%)	27.25	94.13	79.00
b. After O2						
House	26,572.1 (53.6%)	18,429.1 (37.1%)	4606.5 (9.3%)	27.70	74.57	51.96
Clock	16,619.8 (46.4%)	16,007.8 (44.7%)	3197.6 (8.9%)	25.21	57.21	37.29

Table 5 SALT database. Time in absolute units and relative time in parenthesis

Task	Time			Strokes		
	On-surface	In-air$_S$	In-air$_L$	On-surface	In-air$_S$	In-air$_L$
a. DCLI						
Crossed pentagons	18,292.8 (60.2%)	8497.3 (27.9%)	3612.6 (11.9%)	10.00	20.33	27.50
Spiral	8219.3 (99.0%)	26.75 (0.3%)	60.8 (0.7%)	1.42	1.75	2.25
3D house	33,503.83 (52.0%)	19,388.6 (30.1%)	11,534.3 (17.9%)	29.50	49.17	50.42
Clock	18,931.9 (31.2%)	24,807.2 (40.9%)	16,917.0 (27.9%)	26.67	52.17	70.50
Spontaneous sentence	16,500.3 (48.8%)	14,322.9 (42.4%)	2966.5 (8.8%)	40.67	47.75	15.33
Sentence copied	26,535.4 (49.3%)	21,918.3 (40.7%)	5404.9 (10.0%)	57.58	69.08	29.58
Sentence dictation	20,710.7 (59.1%)	11,717.8 (33.4%)	2633.0 (7.5%)	43.25	50.08	16.33
b. Alzheimer						
Crossed pentagons	21,535.4 (48.4%)	15,430.1 (34.6%)	7555.4 (17.0%)	14.05	28.00	48.14
Spiral	11,312.2 (88.7%)	1108.8 (8.7%)	327.2 (2.6%)	1.71	1.67	2.52
3D house	40,341.6 (47.3%)	30,465.8 (35.8%)	14,386.2 (16.9%)	31.55	55.23	75.77
Clock	24,524.7 (36.1%)	33,060.4 (48.6%)	10,420.8 (15.3%)	29.36	48.41	50.95
Spontaneous sentence	19,555.9 (48.6%)	17,090.1 (42.4%)	3606.1 (9.0%)	37.23	44.05	17.09
Sentence copied	34,023.8 (45.1%)	33,451.3 (44.4%)	7951.2 (10.5%)	54.32	69.50	35.95
Sentence dictation	26,640.6 (52.7%)	20,723.6 (41.0%)	3189.6 (6.3%)	44.27	54.86	20.45
c. Control						
Crossed pentagons	17,077.7 (50.1%)	13,085.8 (38.4%)	3918.6 (11.5%)	11.88	36.47	36.65
Spiral	6198.3 (91.0%)	426.6 (5.4%)	251.3 (3.6%)	1.63	3.06	2.94
3D house	29,170.5 (43.3%)	26,094.5 (38.7%)	12,152.4 (18.0%)	30.82	72.24	68.12
Clock	18,986.1 (30.2%)	31,299.1 (49.8%)	12,547.1 (20.0%)	29.94	71.38	71.06
Spontaneous sentence	14,990.5 (43.8%)	14,648.8 (42.8%)	4566.4 (13.4%)	35.41	56.88	31.12
Sentence copied	24,684.2 (45.5%)	23,968.8 (44.1%)	5654.8 (10.4%)	53.59	78.00	37.53
Sentence dictation	19,531.1 (56.9%)	13,131.1 (38.2%)	1676.5 (4.9%)	38.71	50.24	16.76

Spanish Universities, more details can be found in [6]. With respect to handwriting and signatures, this database defines five different tasks: a Spanish text in lower-case, ten digits written separately, 16 Spanish words in upper-case, four genuine signatures, and one forgery of the three precedent subjects.

EMOTHAW

As described in [7], this database includes samples of 129 participants who are classified on the basis of their emotional states: anxiety, depression, and stress or health. This classification is assessed by the Depression–Anxiety–Stress Scales (DASS) questionnaire. Seven tasks are recorded through a digitizing tablet: pentagons and house drawing, words in capital letters copied in handprint, circles with left and right hand, clock drawing, and one sentence copied in cursive writing.

PAHAW

The Parkinson's Disease Handwriting Database (PaHaW) consists of multiple handwriting samples from 37 Parkinson's disease patients, and 38 gender and age matched controls. Eight different tasks were recorded through a digitizing tablet: spiral drawing, letters, words, and a sentence. The details about this database can be found in [8].

OXIGEN-THERAPY

This database described in [9] includes eight patients with hypoxemia who performed two tasks: house and clock drawing, before and after breathing 30 min with O_2 with the aim of evaluating changes in psychomotor functions.

SALT

As described in [10], the database includes samples of 52 participants: 23 with Alzheimer's disease, 12 with mild cognitive impairment (MCI), and 17 healthy controls. Seven tasks were recorded: crossed pentagons, spiral, 3D house, clock drawings, spontaneous, copied, and dictated handwriting.

Table 6 EMOTHA (Mann-Whitney U test)

Task	p T_S	p T_{AS}	p T_{AL}	p $strokes_S$	p $strokes_{AS}$	p $strokes_{AL}$
a. Depression/control						
Two-pentagon	0.4316	0.3082	0.0589	0.1374	0.0731	0.0561
House	0.7329	0.0495	0.5002	0.0315	0.0774	0.4217
Capital letters	0.5771	0.5771	0.8744	0.0904	0.2317	0.5994
Loops with left hand	0.7613	0.2380	0.1292	0.2954	0.2742	0.1542
Loops with right hand	0.6592	0.5316	0.7322	0.5067	0.5005	0.7322
Clock	0.1267	0.6293	0.2196	0.6641	0.7739	0.9688
Sentence	0.8992	0.3273	0.1849	0.3794	0.2870	0.5764
b. Anxiety/control						
Two-pentagon	0.2429	0.1020	0.1678	0.1546	0.1010	0.1505
House	0.4564	0.0417	0.4086	0.0652	0.1777	0.2550
Capital letters	0.3770	0.6374	0.3503	0.1731	0.1888	0.7751
Loops with left hand	0.7711	0.1575	0.1723	0.1560	0.1429	0.2017
Loops with right hand	0.9374	1	1	0.9822	0.9762	1
Clock	0.0414	0.1540	0.2410	0.4462	0.5801	0.8294
Sentence	0.7234	0.4259	0.1296	0.5392	0.3971	0.2913
c. Stress/control						
Two-pentagon	0.5665	0.4886	0.3429	0.6173	0.4131	0.3678
House	0.5221	0.2565	0.4705	0.5562	0.7621	0.9188
Capital letters	0.4741	0.9907	0.7934	0.2769	0.4367	0.4662
Loops with left hand	0.3859	0.1625	0.2801	0.1466	0.1498	0.3173
Loops with right hand	0.4795	0.7184	1	0.6875	0.6875	1
Clock	0.0241	0.7401	0.4670	0.6199	0.6496	0.6623
Sentence	0.6819	0.5753	0.1011	0.7034	0.5335	0.4764

T_S time on-surface, T_{AS} time in-air_S, T_{AL} time in-air_L, $Strokes_S$ strokes on-surface, $Strokes_{AS}$ strokes in-air_s, $Strokes_{AL}$ strokes in-air_L

Experimental Results

The first experiments consisted of analyzing the three kinds of time in absolute and relative values as well as the number of strokes in all the scenarios. Tables 1, 2, 3, 4, and 5 summarize the results for the analyzed databases. It is worth remarking that different databases contain different tasks described in the previous section.

For a given user, the number of strokes is an integer number. However, the table shows the average number of strokes

Table 7 PaHaW (Mann-Whitney U test)

Task	p T_S	p T_{AS}	p T_{AL}	p $strokes_S$	p $strokes_{AS}$	p $strokes_{AL}$
a. Parkinson/control						
Spiral	0.3947	0.5621	0.0939	0.2857	0.0919	0.0949
Letter *l*	0.4614	0.5529	0.0157	0.2390	0.3611	0.2718
Bigram *le*	0.3015	0.0403	0.5671	0.0090	0.1173	0.1710
Word *les*	0.3015	0.3166	0.6601	0.2941	0.2453	0.4385
Word *lektorka*	0.5166	0.9440	0.3019	0.8111	0.6928	0.4744
Word *porovnat*	0.3878	0.7226	0.4025	0.3778	0.9239	0.7963
Word *nepopadnout*	0.5780	0.1776	0.2836	0.0630	0.1287	0.2538
Sentence	0.2000	0.5850	0.9612	0.3229	0.2720	0.5773

T_S time on-surface, T_{AS} time in-air_S, T_{AL} time in-air_L, $Strokes_S$ strokes on-surface, $Strokes_{AS}$ strokes in-air_s, $Strokes_{AL}$ strokes in-air_L

Table 8 OXYGEN THERAPY (Mann-Whitney *U* test)

Task	*p* T_S	*p* T_{AS}	*p* T_{AL}	*p* $strokes_S$	*p* $strokes_{AS}$	*p* $strokes_{AL}$
a. Pre/post O_2						
House	0.8968	0.8764	0.9174	0.9174	0.8968	0.8968
Clock	0.9218	0.8936	0.9077	0.8665	0.8795	0.8795

T_S time on-surface, T_{AS} time in-air$_S$, T_{AL} time in-air$_L$, $Strokes_S$ strokes on-surface, $Strokes_{AS}$ strokes in-air$_s$, $Strokes_{AL}$ strokes in-air$_L$

for a specific database and task (in addition to the number of strokes done by the whole set of users split by the number of users). This number is not integer anymore.

Experimental results of BIOSECUR-ID database, which is the largest one according to the number of users and files, reveal that in-air$_L$ is almost negligible in the case of signatures, but interestingly, it is three times larger for skilled forgeries than for genuine signatures. For uppercase words, the time in-air$_L$ is larger than for the other tasks but still quite modest (10.4%). Thus, this kind of movement is less important than the other two and can probably be ignored without sacrificing a lot of information. For the other databases, a statistical test will be performed after presenting the experimental results.

From all the databases related to diseases, we computed the Mann-Whitney *U* test between study and control groups to determine the existence of statistically significant difference ($p < 0.05$) in the studied features (time and strokes). The results are shown in Table 6.

We can observe in Table 6 (a. Depression/control) how in crossed pentagon task, the values are very close to the threshold for long distance time and strokes. In house draw, the near time and on-surface strokes show statistical significance. In Table 6 (b. Anxiety/control), house draw shows again that near-distance time is statistically signficant. Finally, in Table 6 (c. Stress/control), we obtain $p < 0.05$ for on-surface time in clock draw only.

As is shown in Table 7, for PaHaW database we obtain statistically significant results in letter *l* long distance time and in bigram *le* for near-distance time and on-surface strokes.

In OXYGEN THERAPY database, the times and number of strokes do not show statistical significance and do not seem to offer a valid classification pattern between pre- and post-O_2 results (Table 8).

In Table 9 (SALT, a. Alzheimer/control), we can observe how on crossed pentagons draw, statistical significance can be found in on-surface time and long distance time. Also, on-surface time presents significance on the sentence copied. No results with $p < 0.05$ were obtained for mild cognitive impairment (MCI)/control (Table 9, b).

Discussion

Although most of the results in previous tables are not significant, even for on-surface and in-air$_S$ information, we should

Table 9 SALT (Mann-Whitney *U* test)

Task	*p* T_S	*p* T_{AS}	*p* T_{AL}	*p* $strokes_S$	*p* $strokes_{AS}$	*p* $strokes_{AL}$
a. Alzheimer/control						
Crossed pentagons	0.0303	0.1609	0.0122	0.3941	0.6604	0.0891
Spiral	0.0063	0.5132	0.1995	0.9185	0.1338	0.1869
3D house	0.0677	0.1370	0.1297	0.3493	0.7533	0.0720
Clock	0.1071	0.1984	0.1785	0.5526	0.2033	0.6256
Spontaneous sentence	0.1878	0.0524	0.9210	0.3875	0.8316	0.8761
Sentence copied	0.0096	0.1080	0.1096	0.5612	0.3954	0.2629
Sentence dictation	0.0132	0.0721	0.0920	0.2510	0.3953	0.1604
b. MCI/control						
Crossed pentagons	0.1915	0.4925	0.1915	0.2758	0.1688	0.5643
Spiral	0.0968	0.5358	0.0865	0.4290	0.0889	0.0973
3D house	0.2069	0.5500	0.3879	0.6729	0.7734	0.5206
Clock	0.4437	0.9445	0.0738	1	0.8892	0.2854
Spontaneous sentence	0.5500	0.8421	0.9119	0.4124	0.5496	0.6094
Sentence copied	0.1501	0.4384	0.7735	0.2777	0.6260	0.8075
Sentence dictation	0.3640	0.7398	0.2878	0.3407	0.4784	0.5203

T_S time on-surface, T_{AS} time in-air$_S$, T_{AL} time in-air$_L$, $Strokes_S$ strokes on-surface, $Strokes_{AS}$ strokes in-air$_s$, $Strokes_{AL}$ strokes in-air$_L$

point out that this kind of measurements offers a large set of features that can be extracted, such as speed and acceleration of trajectories and complexity measurements extracted from coordinates x, y. In fact, a classifier would not be based on a single measurement. It will take advantage of a set of measurements. Thus, high p values for on-surface and in-air$_S$ do not imply the impossibility to perform a classification. These values are provided just for comparison purpose with in-air$_L$ values. In-air$_L$ extracted features are limited to time and number of strokes. Thus, the analysis of relevance of this information is simpler.

Nevertheless, this paper points out the tasks and pathologies where more potential improvements can be achieved, because in some tasks, $p < 0.05$ has been obtained.

Looking at the experimental results of pathologies, we can establish that in-air$_L$ movements are not relevant but there are some exceptions: crossed pentagon task for depression patients in EMOTHAW, which is near significance ($p = 0.0589$ for time and $p = 0.0561$ for strokes), letter l task for PaHaW database ($p = 0.0157$ for time), and crossed pentagons task for Alzheimer/control comparison ($p = 0.0122$ for time). We consider these results especially interesting because crossed pentagons are a very useful measurement in pathological analysis, in fact, it is the only drawing that subjects must perform in the well-established mini-mental state examination, also known as the Folstein test [11].

Conclusions

One of the main goals of this paper was to study if in-air$_L$ information can be discarded in handwritten tasks analysis. Looking at the experimental results, we can conclude that little time is spent by healthy writers at long distance so most of the information is contained on-surface and in-air$_S$ distances. This implies that the development of a new acquisition device able to track x and y coordinates and long distances will probably not be very useful, because few samples will be acquired in this condition. However, experimental results reveal that time spent at long distance is more than three times higher for skilled forgeries than for genuine signatures. This opens a possible research line in security biometrics. A similar consideration can be established for the number of strokes, which is doubled in the case of skilled forgeries with respect to short distance in-air movements. Thus, the existence of long distance movements can be indicative of a signature forgery.

On the other hand, when looking at pathologies, we have found statistically significant differences in the pentagon tasks for Alzheimer/control comparison. This result opens the possibility of investigating in-air at long distance movements further.

Acknowledgements This work has been supported by FEDER and MEC, TEC2016-77791-C4-2-R, 16-30805A, SIX (CZ.1.05/2.1.00/03.0072), and LO1401.

Compliance with Ethical Standards All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this type of study formal consent is not required.

Conflict of Interest The authors declare that they have no conflict of interest.

Statement of Human and Animal Rights This chapter does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

1. Faundez-Zanuy M, et al. Biometric applications related to human beings: there is life beyond security. Cogn Comput. 2013;5(1):136–51.
2. Lopez-de-Ipiña K, et al. On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature. Cogn Comput. 2015;7(1):44–55.
3. Sesa-Nogueras E, Faundez-Zanuy M, Roure-Alcobé J. Gender classification by means of online uppercase handwriting: a text-dependent allographic approach. Cogn Comput. 2016;8(1):15–29.
4. Rosenblum S, Luria G. Applying a handwriting measurement model for capturing cognitive load implications through complex figure drawing. Cogn Comput. 2016;8(1):69–77.
5. Sesa-Nogueras E, Faundez-Zanuy M, Mekyska J. An information analysis of in-air and on-surface trajectories in online handwriting. Cogn Comput. 2012;4(2):195–205.
6. Fierrez J, et al. BiosecurID: a multimodal biometric database. Pattern Anal Applic. 2010;13(2):235–46.
7. Likforman-Sulem L, Esposito A, Faundez-Zanuy M, Clémençon S, Cordasco G. EMOTHAW: "A Novel Database for Emotional State Recognition from Handwriting and Drawing". IEEE Trans Hum Mach Syst. 2017;PP(99):1–12. https://doi.org/10.1109/THMS.2016.2635441.
8. Drotár P, Mekyska J, Rektorová I, Masarová L, Smékal Z, Faundez-Zanuy M. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. Artif Intell Med. 2016;67:39–46.
9. Fiz JA, Faundez-Zanuy M, Monte-Moreno E, Alcobé JR, Andreo F, Gomez R, et al. Short term oxygen therapy effects in hypoxemic patients measured by drawing analysis. Comput Methods Prog Biomed. 2015;118(3):330–6.
10. Garre-Olmo, J., et al.. Kinematic and pressure features of handwriting and drawing: preliminary results between patients with mild cognitive impairment, Alzheimer disease and healthy controls. Curr Alzheimer Res. 2017.
11. Folstein MF, Folstein SE, McHugh PR. "Mini-mental status". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3):189–98. https://doi.org/10.1016/0022-3956(75)90026-6.

A.21 Privacy of online handwriting biometrics related to biomedical analysis

Chapter 2
Privacy of online handwriting biometrics related to biomedical analysis

Marcos Faundez-Zanuy[1] *and Jiri Mekyska*[2]

Online handwritten signals analysis for biomedical applications has received lesser attention from the international scientific community than other biometric signals such as electroencephalogram (EEG), electrocardiogram (ECG), magnetic resonance imaging signals (MRI), speech, etc. However, handwritten signals are useful for biometric security applications, especially in the case of signature, but to support pathology diagnose/monitoring as well. Obviously, while utilising handwriting in one field, there are implications in the other one and privacy concerns can arise. A good example is a biometric security system that stores the whole biometric template. It is desirable to reduce the template to the relevant information required for security, removing those characteristics that can permit the identification of pathologies.

In this paper, we summarize the main aspects of handwritten signals with special emphasis on medical applications (Alzheimer's disease, Parkinson's disease, mild cognitive impairment, essential tremor, depression, dysgraphia, etc.) and security. In addition, it is important to remark that health and security issues cannot be easily isolated, and an application in one field should take care of the other.

2.1 Introduction

Online handwritten biometrics belongs to behavioural biometrics because it is based on an action performed by a user. This is opposed to morphological biometrics, which is based on direct measurements of physical traits of the human body. From human behaviour and health condition point of view, it appears more appealing than other hard biometrics such as fingerprint or iris. Although health applications based on online handwriting today have not been deeply explored, there is a nice set of possibilities that will probably grow in the future, such as diagnosis/monitoring of depression, neurological diseases, drug abuse, etc. It can be noted that nowadays,

[1]Pompeu Fabra University, Spain
[2]Department of Telecommunications, Brno University of Technology, Brno, Czech Republic

most of the published research in biometric signal processing is based on image and speech, reasons for which can be that these signals are easier to acquire and cheaper than online handwriting tasks. The price of a webcam or a microphone has been low since the past century, while digitizing devices for online handwritten tasks was by far more expensive. Fortunately, in the recent years, tactile screens have become more popular and online handwritten signals are more present in the society than a few years ago. This has permitted a reduction in the cost of acquiring devices. Thus, nowadays, the price of the acquisition device is not a drawback anymore. We can forecast a growing in applications in this field, and we should take care of privacy issues. In this chapter, we will present an introduction to online handwritten signals and discuss several applications of them in the medical field, which we consider relevant for the biometric community.

This chapter is written for signal-processing engineers devoted to security biometric applications. Even if readers have a background in speech and/or image but are not familiar with online handwritten signals, they will find an explanation including fundamentals of the acquisition process as a starting point. However, and even more challenging, this part of the book is also written for people outside the biometric community, including the audience of medical doctors, willing to enter into this topic and collaborate with engineers. Today, it seems hard to establish collaborations between engineers and medical doctors. Quite often, we do not understand each other due to our different background. Thus, we tried to write the chapter in an easy-to-read way. Breaking innovations are hardly produced in the core of a knowledge area, and the main contribution is seen rather in terms of focussing on the borders between different areas.

The structure of this chapter is as follows: Section 2.2 introduces to the properties and characteristics of the acquisition devices as well as the online handwritten signal. Section 2.3 is devoted to examples of implications between both fields, security and health, with special emphasis on those situations where the privacy of the user can be compromised, and the authentication task is performed under pressure or without consciousness of the users (e.g. suffering a severe disease). Section 2.4 summarizes the chapter.

2.2 Online handwritten signals – an introduction

Online handwritten signals acquisition consists of dynamic acquisition of various properties of the moving pen during the writing process in real time, whereas the digital representation of the signals is typically given by time-stamped sequences of measurement points/tupels. For instance, using a digitizing tablet, smartphone, etc., which typically acquires information listed in Table 2.1.

Using this set of dynamic data, further information can be inferred by analytical computation, which is usually more suitable for certain applications (e.g. handwriting velocity, duration, width, height). This results in what is usually called feature sets, being similar to the use of body mass index for overweight classification. Body mass

Table 2.1 Information acquired from a digitizing tablet

Abbreviation	**Description**
x	Position of pen tip in x axis
y	Position of pen tip in y axis
s/a	On-surface/in-air pen position information
p	Pressure applied by the pen tip
az	Azimuth angle of the pen with respect to the tablet's surface (see Figure 2.1)
al	Altitude angle (sometimes called tilt) of the pen with respect to the tablet's surface (see Figure 2.1)
t	Timestamp

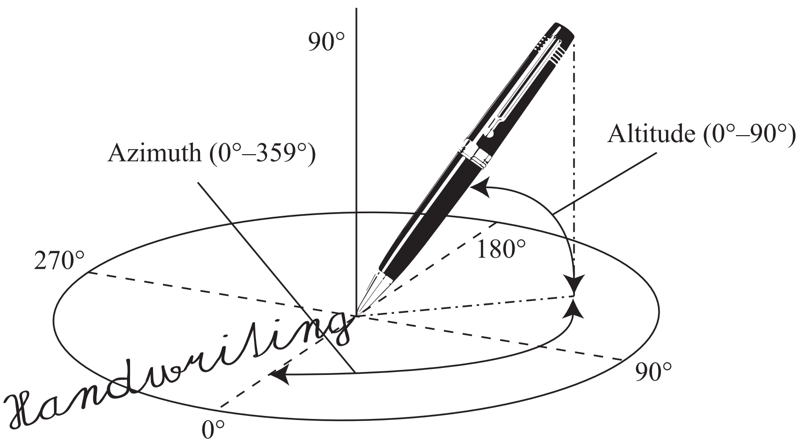


Figure 2.1 Handwriting online information acquired in typical cases (x and y position, pressure, azimuth, altitude)

index is not a direct measure. In fact, it is based on weight and height but it is more useful than body/weight alone.

2.2.1 In-air and on-surface movements

Some digitizing devices, such as Intuos Wacom Tablet™, Samsung Galaxy Note™, etc., are able to track the pen-tip movement even when it is not touching the surface. Thus, it is possible to record the x and y coordinates of in-air movements when pressure is equal to zero. Unfortunately, this is only possible when the distance between the tip of the pen and the surface is less or equal to approximately 1 cm, otherwise the tracking is lost. Nevertheless, the time spent in air is still known because the acquisition device provides a timestamp of each sample. By looking at the difference between consecutive samples, it is possible to know the exact amount of time spent in-air,

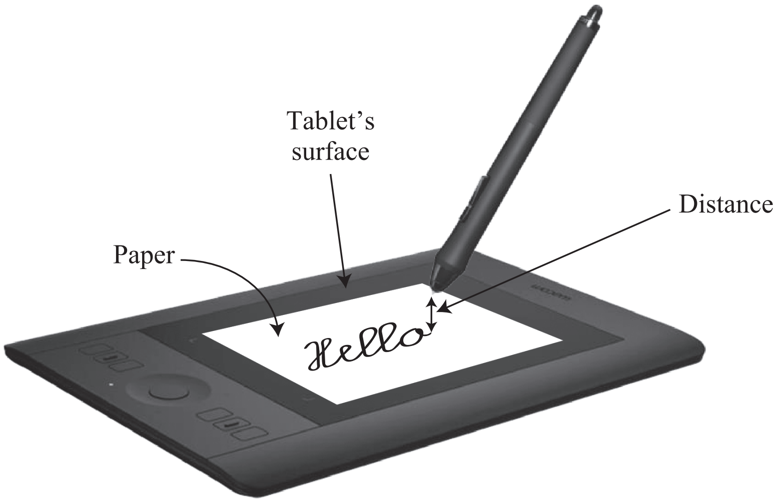


Figure 2.2 Illustration of the distance from pen tip to surface

although the x and y coordinates are only known when the height is smaller or equal to 1 cm (see Figure 2.2).

While some devices can be operated with a sheet of paper and a special ink pen, others do not permit this kind of pen, and the handwriting must be directly done on the tablet's surface using plastic pen without an immediate visual feedback.

Thus, we know three kinds of data:

1. Movement on-surface: typically provides the five features described in the previous section (x, y, pressure, azimuth, altitude).
2. Movement in-air at short distance to surface: provides x and y position, azimuth and altitude.
3. Movement in-air at long distances to surface: when distance is higher than approximately 1 cm, we only know the time spent in-air, as no samples are acquired.

Figure 2.3 shows the aspect of raw samples acquired by a digitizer. For each sampling instance, a set of features is acquired: x coordinate; y coordinate; timestamp t provided by the machine; surface/air bit s/a, which is equal to zero when there is no contact between tip of pen and surface, and one where there is contact; pressure value p; azimuth az and altitude al. In this example, we may observe some samples in-air at short distance plus some time in-air (between $t = 11{,}253{,}657$ and 11,253,827), with a subsequent measurement at long distance. This can be observed because the jump in timestamp between $t = 11{,}253{,}827$ and 11,253,843 is higher than the usual sampling rate for on-surface samples. For the later, the time-stamp progress in t is 10 units, while for the last sample in-air at short distance, it is 16 time units. Time in-air at long distance can appear after in-air at short distance before touching again the surface. For most of the users and tasks, this time is negligible, because movements between strokes tend to be short.

Looking at Figure 2.3, we observe that raw data provided by digitizing tablet is really simple in structure and thus can be processed in a straightforward way, even

x	*y*	*t*	*s/a*	*al*	*az*	*p*	
10,609	8,915	11,253,637	1	1,680	630	574	On-surface samples
10,774	8,907	11,253,647	1	1,700	630	171	
10,774	8,907	11,253,657	0	1,720	630	0	
10,801	8,906	11,253,667	0	1,740	630	0	
10,706	8,924	11,253,677	0	1,760	630	0	
10,585	8,955	11,253,687	0	1,760	620	0	
10,466	8,993	11,253,697	0	1,760	620	0	
10,368	9,025	11,253,707	0	1,760	620	0	
10,295	9,049	11,253,717	0	1,740	620	0	
10,242	9,060	11,253,727	0	1,730	620	0	
10,206	9,060	11,253,737	0	1,680	630	0	In-air samples
10,191	9,054	11,253,747	0	1,660	620	0	
10,186	9,044	11,253,757	0	1,650	620	0	
10,186	9,033	11,253,767	0	1,630	620	0	
10,186	9,023	11,253,777	0	1,610	620	0	
10,186	9,013	11,253,787	0	1,590	610	0	
10,186	9,006	11,253,797	0	1,590	610	0	
10,186	9,001	11,253,807	0	1,570	610	0	
10,191	8,992	11,253,817	0	1,590	610	0	
10,205	8,981	11,253,827	0	1,590	610	0	
10,205	8,981	11,253,843	1	1,610	620	173	On-surface samples
10,205	8,981	11,253,853	1	1,610	620	153	

Figure 2.3 Example of digital representation of samples acquired with digitizer in two scenarios: on-surface, in-air. x – x position, y – y position, t – timestamp, s/a – on-surface/in-air pen position information, p – pressure, az – azimuth, al – altitude

by people without programming skills. For instance, it can be easily imported in any standard spreadsheet software and processed there to extract simple and useful statistics such as mean time on-surface/in-air, variation in pressure, etc.

Although most of the many existing works related to handwritten signals in biometrics and handwriting recognition have been based on surface movements (see e.g. [1]), there are evidences of the importance of in-air movements as well. Sesa-Nogueras *et al.* [2] presented an analysis of in-air and on-surface signals from an information theory point of view. They performed the entropy analysis of handwriting samples acquired in a group of 100 people (see the BiosercurID database for more information [3]) and observed that both types of movements contain approximately the same amount of information. Moreover, based on the values of mutual information, these movements appear to be notably non-redundant. This property has been advantageously used in several fields of science. For instance, Drotar *et al.* [4,5] proved that in-air movement increases the accuracy of Parkinsonic dysgraphia identification. Specifically, when classifying the Parkinsonic dysgraphia by support vector machine (SVM) in combination with the in-air features, they reached 84% accuracy

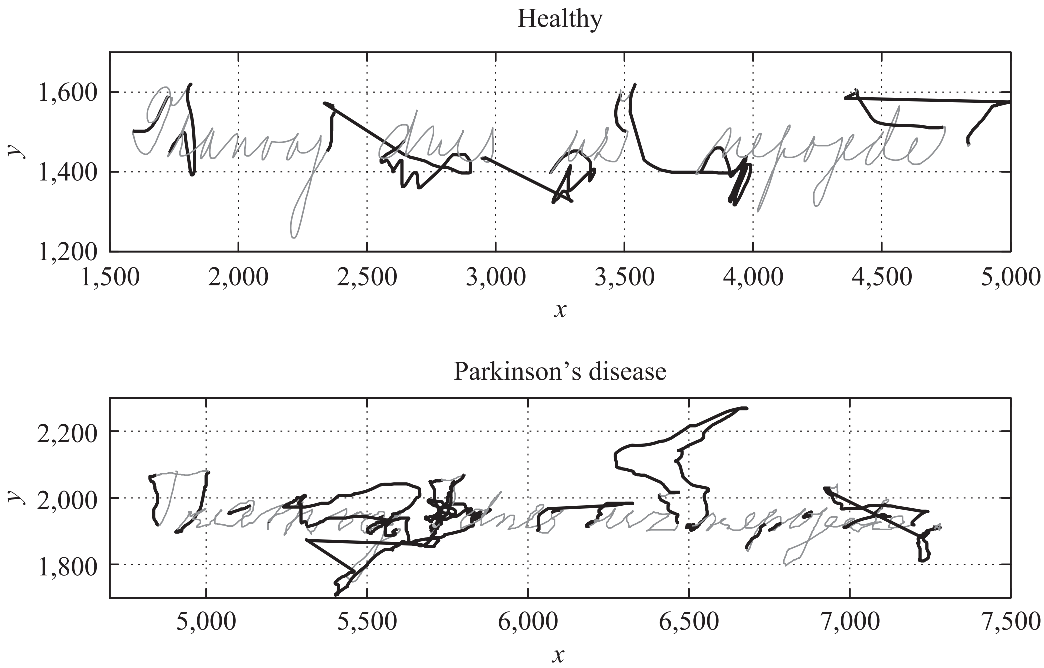


Figure 2.4 Example of on-surface (grey line) and in-air (black line) movement. Czech sentence written by a healthy writer and patient with PD (samples from the PaHaW database, Drotar et al. [11])

which is by 6% higher in comparison to classification based on the on-surface features only. When combining both feature sets, they observed 86% classification accuracy. Faundez-Zanuy *et al.* [6] reported that the in-air movement supports diagnosis of Alzheimer's disease (AD). They observed that patients with AD spend seven times longer in-air when comparing to a control group. In the case of on-surface movement, it is only three times longer. Similarly, Rosenblum *et al.* [7] found out that the in-air duration can be a good measure for performance analysis of children with high-functioning autism spectrum disorder. The in-air movement has also been used for identification and quantitative analysis of developmental dysgraphia in children population [8–10]. Mekyska *et al.* [8] proved that kinematic features derived from this kind of movement (especially jerk, which is rate at which the acceleration of a pen changes with time) provide good discrimination power between children with dysgraphia and control group.

Figure 2.4 contains an example of Czech sentence written by a healthy writer and writer with Parkinson's disease (PD). As can be seen, the in-air movement (transition between strokes plotted in black bold) is in the case of PD writer very unsmooth and irregular. We can see that the writer spent a lot of time in-air before he initiated the writing of next word. This is tightly related to cognitive functions, the writer has to think about the next movement, and sometimes, he forgets what to write. We wouldn't be able to objectively describe these cognitive processes without the in-air movement.

2.3 Handwriting signals from biometrics to medical applications

The analysis of handwriting in security applications, i.e. for the automated identification or verification of subjects by means of biometric methods, today appears to be a well-studied domain. We thus in this section discuss this part very briefly, with reviews of some relevant works. Further, we expand the views to metadata analysis (also referred to as Soft Biometrics) with a brief review of selected works published. Finally, we bridge the gap towards the analysis of handwriting signals for medical purposes, for example to support diagnostics of some diseases. These aspects will be the main focus of discussions in the following subsections.

2.3.1 Biometric security applications

Biometric security applications based on handwritten tasks are mainly based on signatures. Several international competitions summarize the state of the art achieved by dozens of teams, such as Houmani *et al.* [12], signature verification competition (SVC) [13] and SigWiComp (competitions on signature verification and writer identification for on- and offline skilled forgeries) [14]. Although less known, there are also some works where biometric recognition is based on handwritten text, either text-dependent or independent.

The individuality of handwriting has been demonstrated by several authors. Srihari *et al.* [15] assessed the individuality of handwriting in the off-line case. They collected a database of 1,500 writers selected to be representative of the US population and conducted experiments on identification and verification. Regarding identification, they reached accuracy of about 83% at the word level (88% at the paragraph-level and 98% at the document-level). These results allowed the authors to conclude that the individuality hypothesis, with respect to the target population, was true with a 95% confidence level. Zhang and Srihari [16] complemented the previous work of [15]. They analysed the individuality of four handwritten words (*been*, *Cohen*, *Medical* and *referred*) taken from 1,027 US individuals, who wrote each word three times. The combination of the four words yielded an identification accuracy of about 83% and a verification accuracy of about 91%.

With regard to the online case, some authors have addressed the issue of individuality of single words and short sentences. Hook *et al.* [17] showed that single words (the German words *auch*, *oder*, *bitte* and *weit*) and the short sentence *Guten Morgen* exhibit both considerable reproducibility and uniqueness (i.e. equal items written by the same person match well while equal items written by different people match far less well). They used a small database consisting of 15 writers that produced, in a single session, ten repetitions of each item captured by a prototype of a digitizing pen. Chapran [18] used the English words *February*, *January*, *November*, *October* and *September* (25 repetitions of each word donated by 45 writers). The identification rate reached 95%. In Sesa and Faundez-Zanuy [19], a writer identification rate of 92.38% and a minimum of detection cost function [20] of 0.046 (4.6%) was achieved with 370 users using just one word written in capital letters. Results were improved up to 96.46% and 0.033 (3.3%) when combining two words.

2.3.2 Metadata applications

Behavioural biometrics, in addition to security and health applications, can provide a set of additional information, known as metadata. Sometimes also referred to as Soft Biometrics, it can be based on system hardware specifics (technical metadata) and on the other side on personal attributes (non-technical metadata) [21,22]. System-related metadata represent physical characteristics of biometric sensors and are essential for ensuring comparable quality of the biometric raw signals. Previous work in personal related metadata has shown that it is possible to estimate some metadata like script language, dialect, origin, gender and age by statistically analysing human handwriting. In this section, we will summarize some non-technical metadata applications.

Gender recognition attempts to classify the writer as a male or a female. In [23] using only four repetitions of a single uppercase word, the average rate of well-classified writers is 68%; with 16 words, the rate rises to an average of 72.6%. Statistical analysis reveals that the aforementioned rates are highly significant. In order to explore the classification potential of the in-air strokes, these are also considered. Although in this case, results are not conclusive, and an outstanding average of 74% of well-classified writers is obtained when information from in-air strokes is combined with information from on-surface ones. This rate is slightly better than the one achieved by calligraphic experts. However, we should keep in mind that this is a two-class problem and even by pure chance (for instance, flipping a coin) we would get 50% accuracy.

Bandi *et al.* [24] proposed a system that classifies handwritings into demographic categories using measurements such as pen pressure, writing movement, stroke formation and word proportion. The authors reported classification accuracies of 77.5%, 86.6% and 74.4% for gender, age and handedness classification, respectively. In this study, all the writers produced the same letter. Liwicki *et al.* [25] also addressed the classification of gender and handedness in the on-line mode. The authors used a set of 29 features extracted from both on-line information and its off-line representation and applied support vector machines and Gaussian mixture models to perform the classification. The authors reported an accuracy of 67.06% for gender classification and 84.66% for handedness classification. In [26], the authors separately reported the performance of the offline mode, the on-line mode and their combination. The accuracy reported for the off-line mode was 55.39%.

Emotional states, such as anxiety, depression and stress, can be assessed by the depression anxiety stress scales (DASS) questionnaire. Likforman-Sulem *et al.* [27] presents a new database that relates emotional states to handwriting and drawing tasks acquired with a digitizing tablet. Experimental results show that anxiety and stress recognition perform better than depression recognition. This database includes samples of 129 participants whose emotional states are assessed by the DASS questionnaire and is freely distributed for those interested in researching in this line.

2.3.3 Biometric health applications

As to be seen from the example on emotional states and the reasons for emotional changes, the transition from metadata to medical analysis is somewhat fluent. In this

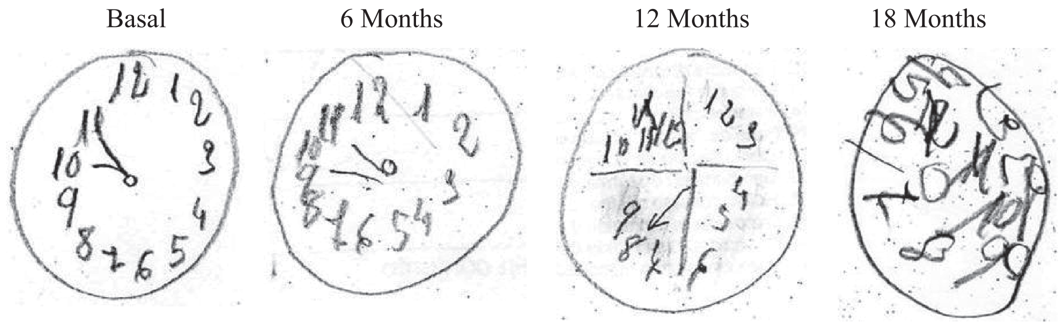


Figure 2.5 Clock drawing test of ACE-R for a person with AD, showing initial baseline on the left, and then from left to right, samples from the same person after 6, 12 and 18 months

section, we focus on selected analysis for the latter case, with regards to handwriting modality. While signature and handwritten script samples are also useful for health issues, we focus on a set of probably more interesting tasks such as drawings or sketches. These kinds of signals can also be used for biometric recognition, although they are not as usual in real life as handwriting or signature (some examples can be found in [28]).

One important unsolved problem is how the dementia syndrome is associated with diseases such as Parkinson's and Alzheimer's, etc. In the case of Alzheimer's, it is estimated that the cost per year for a single patient is 35,000 USD in the USA. One in ten patients is below 60 years old. The incidence of Alzheimer's is doubled for every 5 years after 65, and beyond 85 years old the incidence is between one-third and half of the amount of population. If a solution is not found, this problem will be unbearable for society. Consequently, a relevant issue related to dementia is its diagnostic procedure. For example, AD is the most common type of dementia, and it has been pointed out that early detection and diagnosis may confer several benefits. However, intensive research efforts to develop a valid and reliable biomarker with enough accuracy to detect AD in the very mild stages or even in pre-symptomatic stages of the disease have not been conclusive. Nowadays, the diagnostic procedure includes the assessment of cognitive functions by using psychometric instruments such as general or specific tests that assess several cognitive functions. A typical test for AD is the clock drawing test (CDT) [29] that consists of drawing a circle and distributing the 12 hours inside. An example of this is shown in Figure 2.5. The initial result produced by a person (baseline) is shown on the left, and on the right, several samples of the same person after 6, 12 and 18 months of being damaged are also shown. This same test has also been used for detecting drug abuse, depression, etc. Figure 2.6 shows a similar situation when copying two interlinking pentagons, which is one of the tasks of the mini-mental state examination (MMSE) [30]. The MMSE or Folstein test is a brief 30-point questionnaire test that is used to screen for cognitive impairment. It is also used to estimate the severity of cognitive impairment at a specific time and to follow the course of cognitive changes in an individual over time, thus making it an effective way to document an individual's response to treatment.

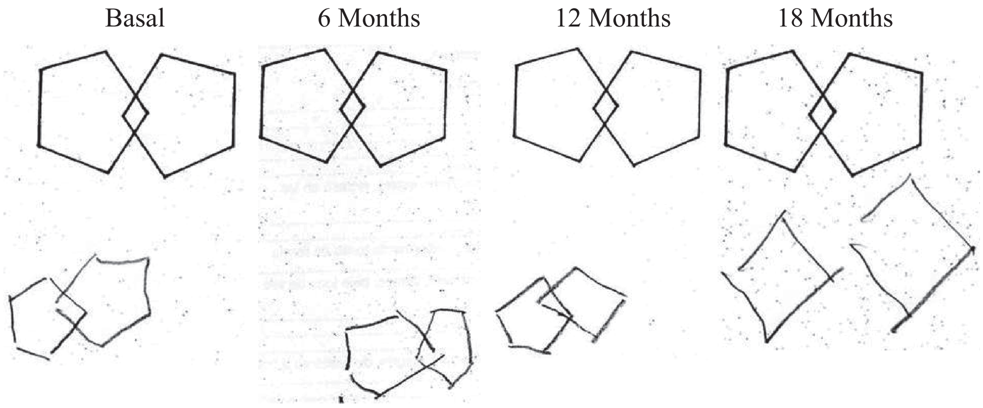


Figure 2.6 Pentagons of MMSE for a person with AD, showing initial baseline on the left, and then from left to right, samples from the same person after 6, 12 and 18 months. A pentagon template copied by the patients appears in the top

Figure 2.7 presents a house drawing that includes perspective notions (3D aspect [31]). The first two rows are performed by individuals with AD of different clinical severity. The visual inspection of the on-surface image suggests a progressive degree of impairment, where drawing becomes more disorganized and the three dimensional effect is only achieved in the second row (mild case). The visual information provided by the in-air drawing between AD individuals also indicates a progressive impairment and disorganization when the individuals try to plan the drawing. It is also important to note that the comparison of the on-surface drawing between the mild case of AD and the control (third and fourth rows) also shows important differences. Even in the case when the drawing is performed with the non-dominant hand. Besides the increased time in-air, there is an increased number of hand movements before writers decide to put the pen on the surface to drawn. We consider that these graphomotor measures applied to the analysis of drawing and writing functions may be a useful alternative to study the precise nature and progression of the drawing and writing disorders associated with several neurodegenerative diseases [6,31]. Figure 2.7 illustrates the potential of in-air information, which is neglected when medical doctors use the classical ink pen system in off-line mode.

Generally, in the area of diagnostics in medical context, drawings are widely used. In summary, some common drawings and their potential usage in medical field (included the cases described already above) are

1. *Pentagon test* – used in the MMSE to assess cognitive impairment [30] (see Figure 2.6). A template provided to the patient appears in the top row. The second row is the result produced by the patient when trying to copy the template, which is always the same.
2. *CDT* – can be utilized as a precursory measure to indicate the likelihood of further/future cognitive deficits. It is used in the Addenbrooke's cognitive examination-revised (ACE-R) test [32] (see Figure 2.5). As in the previous case,

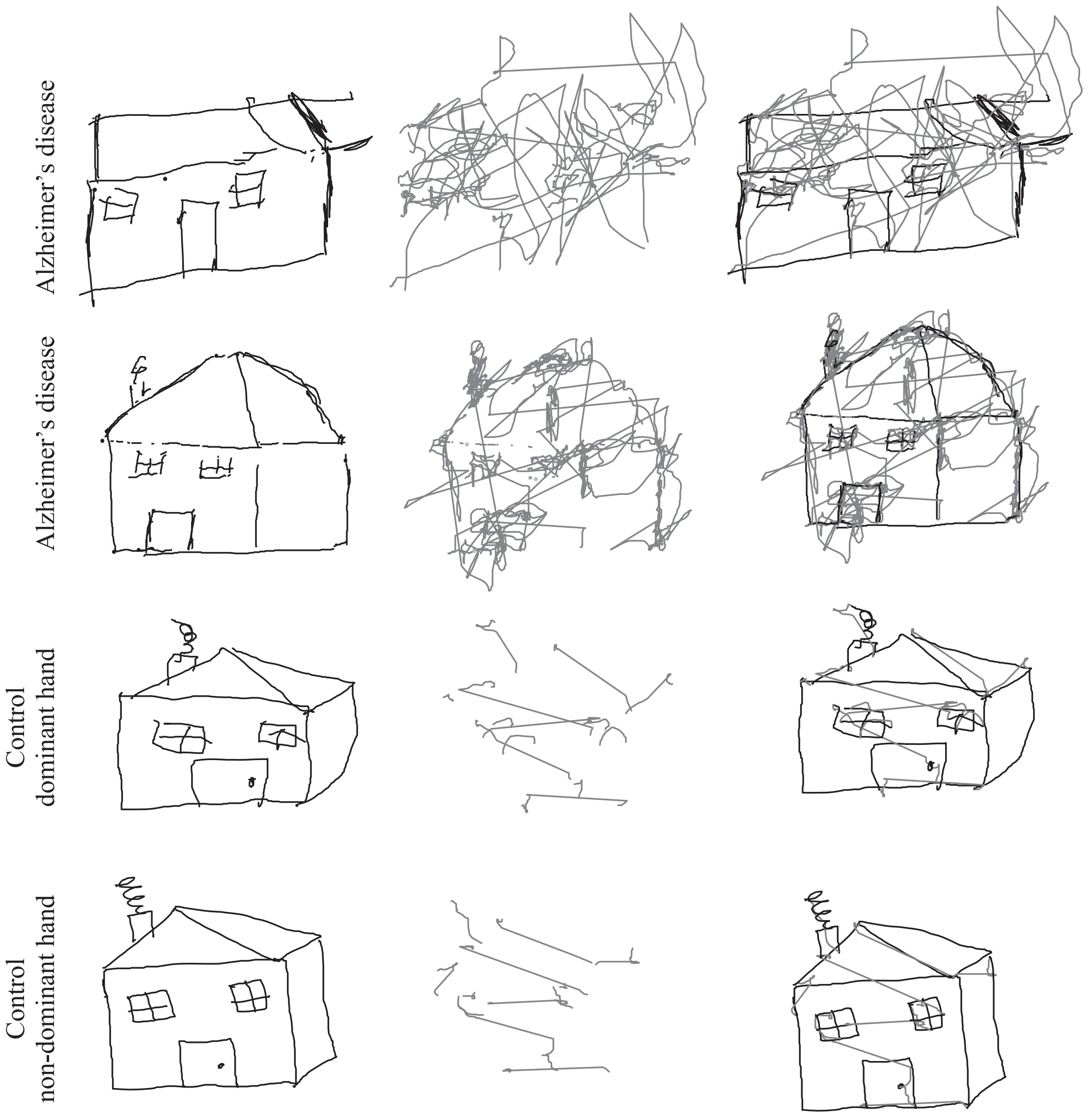


Figure 2.7 House drawing performed by four individuals. First and second rows correspond to individuals with Alzheimer's disease (one per row). Third and fourth rows correspond to a healthy user when drawing with dominant and non-dominant hand. First column – on-surface movement, second column – in-air movement, third column – combination of both. Extracted and adapted with permission from Reference [33]

the different clocks (from left to right) are produced by the same patient passing 6 months.

3. *House drawing copy* – used for identification of AD [6,33] (see Figure 2.7). Patients have to copy a shown image of a house sketch.
4. *Archimedes spiral and straight line (drawing between points)* – useful to discriminate between PD and essential tremor; diagnose mild cognitive impairment, AD, dysgraphia, etc. [11,34–38] (see Figure 2.8). In the case of the Archimedes spiral

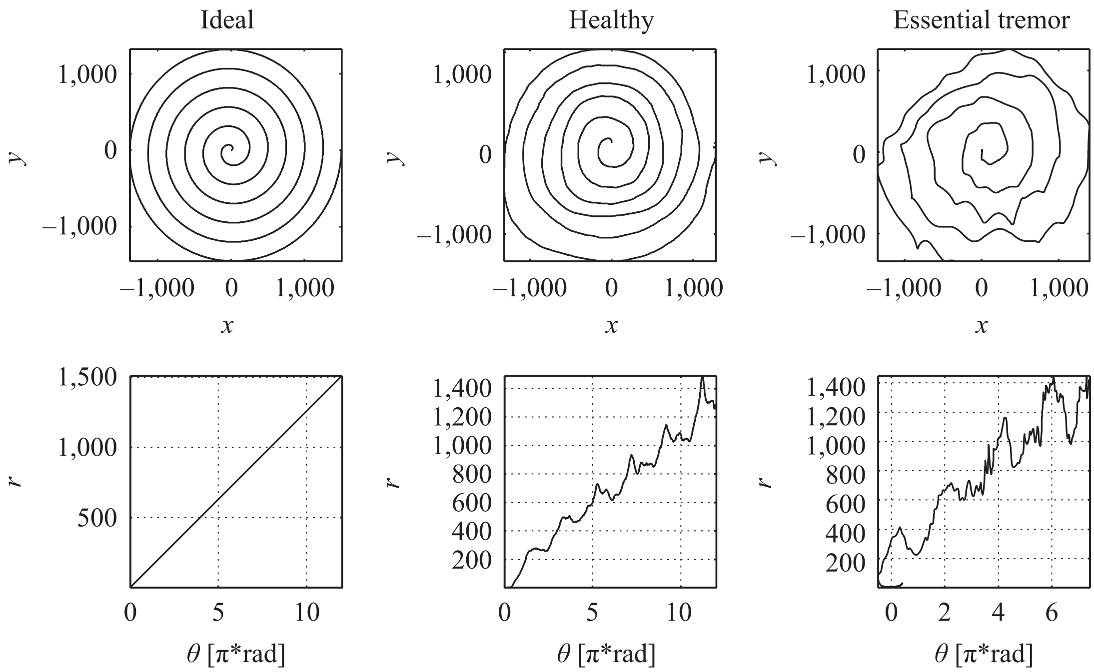


Figure 2.8 Examples of Archimedean spiral: ideal, spiral written by a healthy writer, spiral written by a patient with essential tremor. The second row of picture represents an unwrapped version, where x axis contains angle of each sample on spiral with respect to its centre, and y axis a distance from this centre. This representation is usually used for discrimination among healthy controls and patients with essential tremor or PD [43–45]

acquisition and straight lines, the participants can have a printed spiral on a sheet of paper and a couple of dots to be connected, and they are asked to trace it by a pen without touching the spiral neither the bars (see Figure 2.9). Or, the spiral is shown to them on a template, and they are asked to replicate it on a blank sheet of paper. Similarly, the straight lines can be acquired. In addition, the participants can be asked to connect printed points.

5. *Overlapped circles (ellipses)* – can be used for quantitative analysis of schizophrenia or PD [39–41]. See Figures 2.10 and 2.11, which represents some simple kinematic features that can be used for an effective diagnosis.
6. *Rey–Osterrieth complex figure test* – developed in 1941 and further consists of copying a complex drawing [42]. It is frequently used to further explain any secondary effect of brain injury in neurological patients, to test for the presence of dementia or to study the degree of cognitive development in children. In this task, patients have to memorize an image, and later, they have to replicate it without looking at the example.

Changes in handwriting are usually among the first manifestations of the second most common neurodegenerative disorder – PD [46]. PD patients are usually associated with several motor features, such as tremor in rest, rigidity (resistance to passive

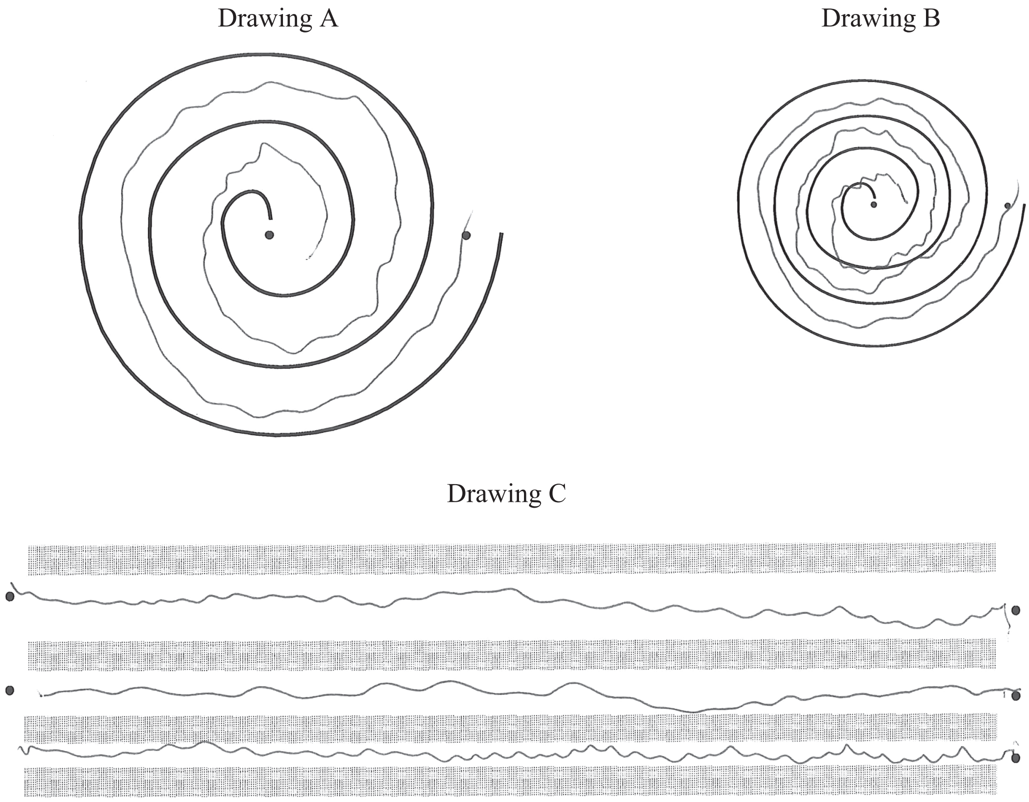


Figure 2.9 Spiral and straight lines test. In the spiral test, the user has to trace a spiral without touching the walls of the traced spiral. In the line test, he has to connect the dots with a straight line without touching the upper and lower bars

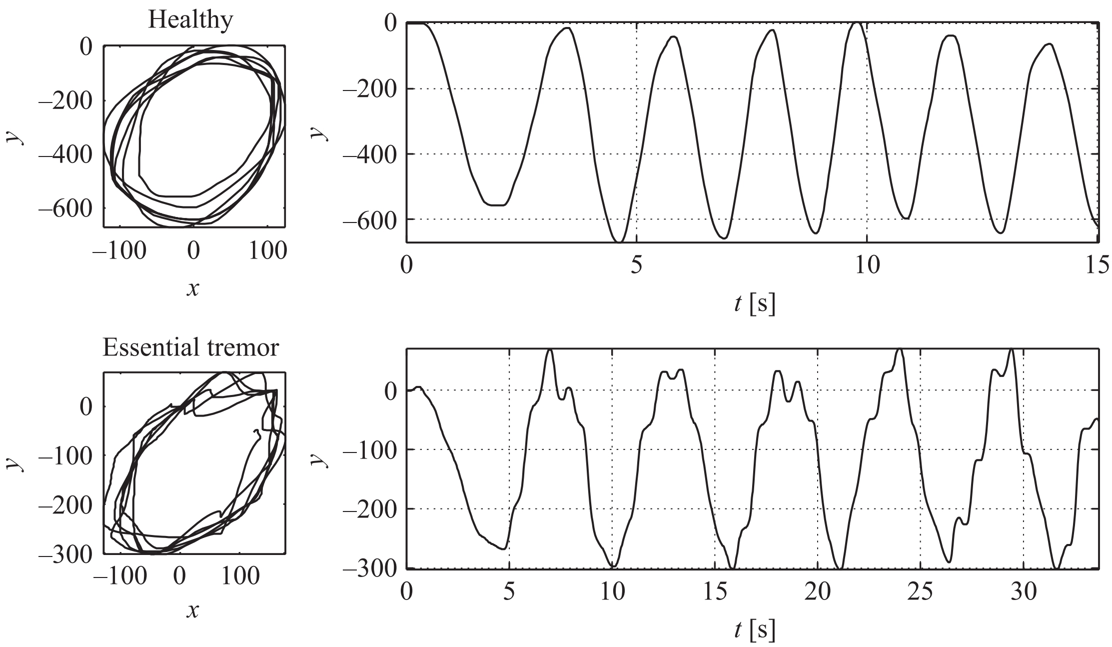


Figure 2.10 Examples of overlapped circles (ellipses): healthy subject and patient with essential tremor. The right part of figure represents vertical movement in time

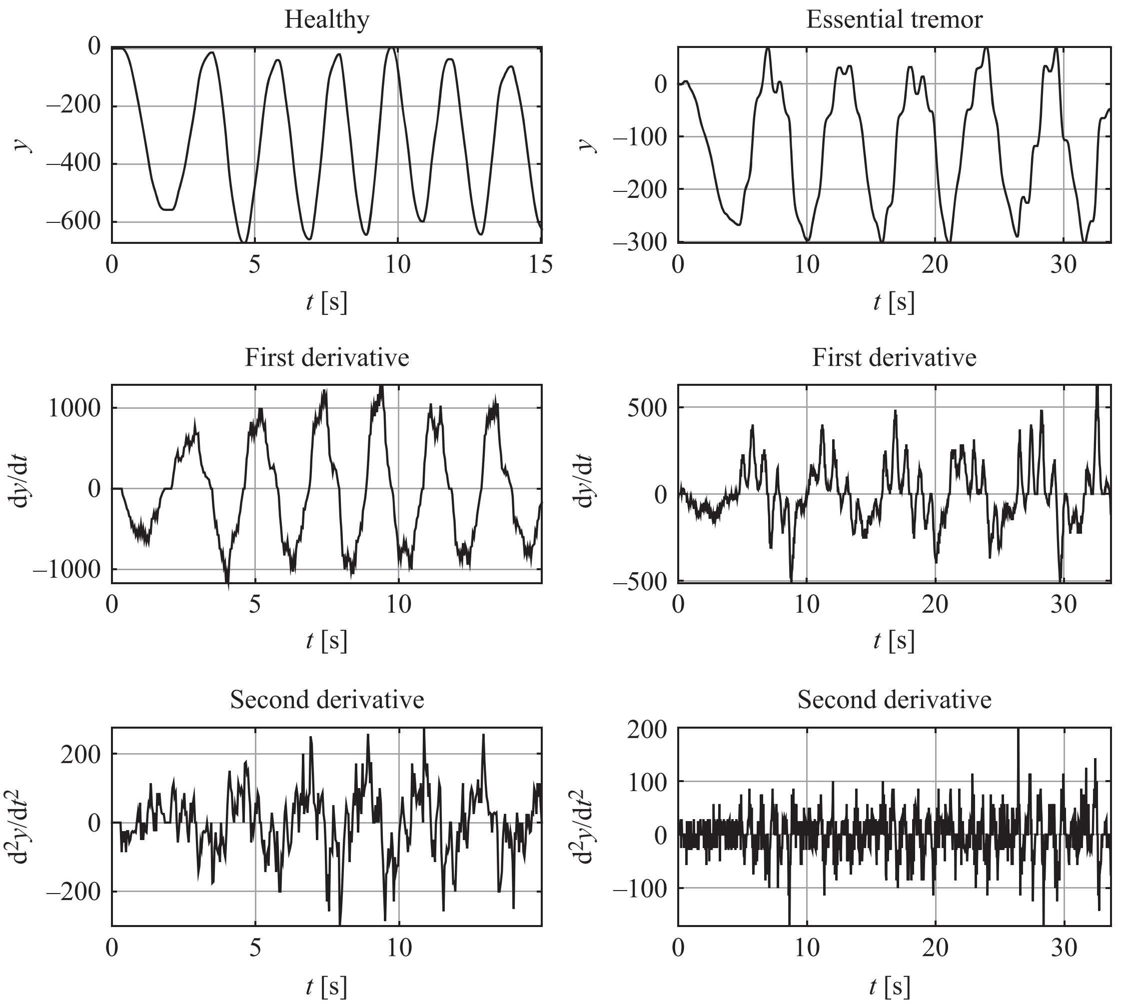


Figure 2.11 Examples of signals from Figure 2.10 (vertical movement during overlapped ellipses writing) after application of first (speed) and second (acceleration) derivative

movement) and bradykinesia (slowness of movement). These motor features affect handwriting as well. Usually, we can observe micrographia (characterized by abnormal reduction in writing size) in patients with PD [46,47]. However, several recent studies showed that micrographia is not the most characteristic feature of PD handwriting [48]. The availability of graphical tablets allowed investigating the PD handwriting in great detail, which resulted in the definition of the term PD dysgraphia. PD dysgraphia encompasses all deficits characteristic of Parkinsonian handwriting [46,48], e.g. deficits in geometry, kinematic variables, pressure patterns, in-air movement. All these deficits can be effectively and objectively quantified using a protocol of handwriting/drawing tasks [4,5,11,49].

Beside AD and PD, mild cognitive impairment, essential tremor, dysgraphia and schizophrenia, the online handwriting analysis found its place in many other health applications, e.g. analysis of depression [50], autism [51] or obsessive–compulsive disorder [52].

2.4 Security–health implications and concerns

Security and health applications have been described as isolated fields to each other in previous sections. They indeed are separated for most of the scientific community. However, they are not separated in real life.

Engineers can develop a technological solution for security or health issues. In general, they tend to be experts in just one of the two fields, and they do not take care of the other field. This solution can be based on biometrics but hardly can we consider that security is not related to health and the opposite. In some cases, both should be considered jointly. In other ones, we cannot isolate one from the other one. In the next subsections, we will describe several examples.

2.4.1 Security where health aspects influence biometric security

Most biometric security applications only try to determine the identity of a user or to verify if he is who claims to be. However, in the common knowledge that automated biometric systems are subject to erroneous classifications, it is important to extract some additional information in context of the acquisition of biometric signals. In the following, we summarise three of such possible scenarios, motivated by the questions as follows:

1. *Is the user under stress?* It is not the same to put a finger in a sensor, confirm the identity, and open a door if, for example, his heart is beating at 70 beats per minute (bpm) than if it is beating at 120 bpm. If his heart is much more accelerated than expected, some suspicious activity can be happening. To solve this, some fingerprint systems have a mechanism called duress finger, which is a way to notify security guards or the police about the threatening situation without letting the threatening person notice it. To do that the user enrols at least two fingers. Both will open the door but one of them will activate a silent alarm. This concept is known as duress detection. See for instance [53]. Some examples are the commercialized fingerprint products by Fingertec[1] and Suprema[2]. This is a simple example to illustrate the idea, but obviously, it is quite simple for the threatening person to force the user to use the specific finger that does not activate the alarm. This knowledge can be obtained just looking how the user interacts with the sensor in previous days. Similarly, the user can enrol a couple of different signatures, one for duress recognition and the other one for normal operation system. Again, it would be possible for a third party to be familiar with the genuine signature that does not activate any silent alarm and force the user to use that signature.

   A robust biometric security system should be able to detect the stress situation based on characteristics that cannot be easily controlled by the user. Detection of user stress from signature or handwriting is a challenging research topic that can indeed improve security systems.

[1] http://www.fingertec.com/ver2/english/vd-duressfingerprint.html
[2] https://www.supremainc.com/es/node/613

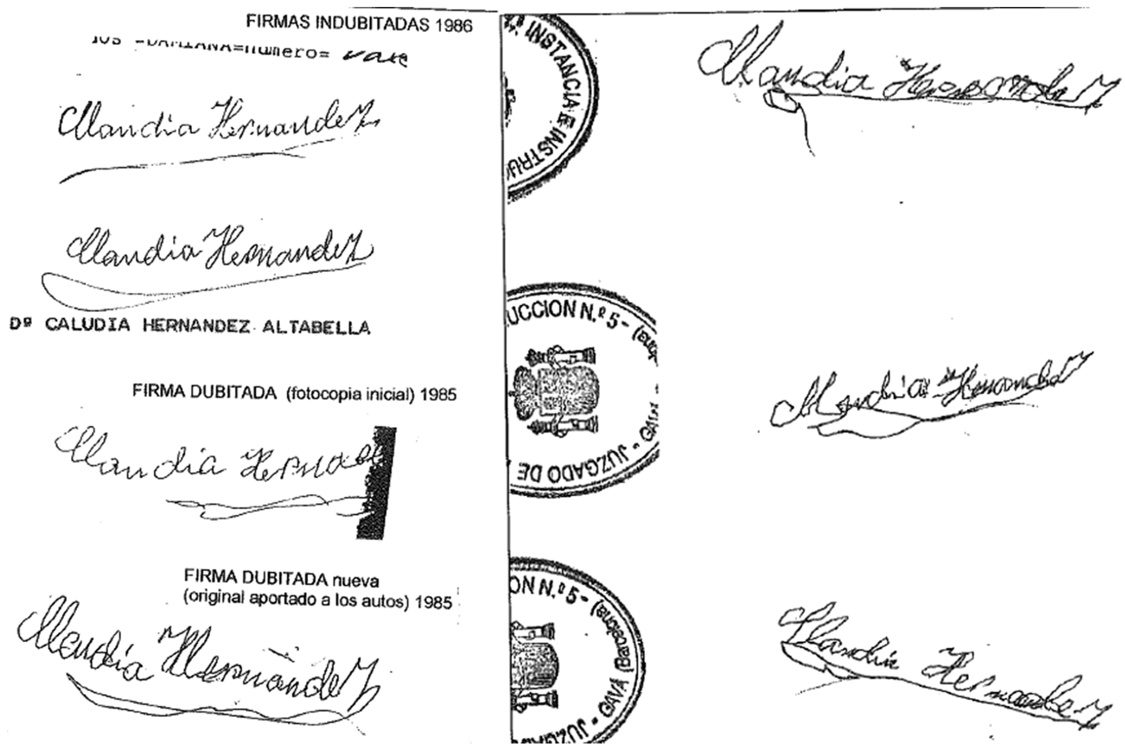


Figure 2.12 On the left, documents signed in 1985 (hesitated) and 1986. On the right, some signatures on blank sheets, when the elder woman was suffering dementia. Extracted with permission from Reference [33]

2. *Is he suffering any disease that makes him unable to understand the real implication on his acts?* In [33], we presented the case of a woman affected by AD. This case was presented to us by Viñals and described in his book [54]. In this case, several women made an elder woman sign her name on blank sheets of paper (see Figure 2.12). Theoretically, due to some issues related to medicines. When the elder person died, the women took advantage of the signed sheets in order to write a rental agreement. The declared date of this agreement was 1985 (Figure 2.12 on the bottom left), but several documents signed in 1986 (Figure 2.12 on the top left) showed better control of calligraphic movements. In fact, the hesitantly written signature document signed in 1985 was closer in appearance to the blank sheets signed when the elder woman had dementia than to the 1986 document. Thus, it was demonstrated that in fact the rental document was not signed in 1985. It was signed later.

   Another possibility is to be affected by depression. Heinik *et al.* [55] used drawings for analysing depressive disorders in older people.

   These two examples indicate that while even if in the context of biometric signature verification one can conclude that the signature is genuine, this may be not enough. One should in addition take into account aspects such as the health state of the user. Considering both aspects (identity, i.e. degree of signature matching and health), one can conclude in doubt of a good health condition of the subject the signature, from a legal point of view, may not be valid. In these cases, the biometric authentication of the individual does not solve the problem, and some additional considerations should be taken. This is not just related to health. Another similar situation where a genuine biometric sample is used in a fraudulent way is a replay attack. In a replay attack, the biometric signal is usually genuine, but it was acquired/recorded in the past and presented again and should be considered as a fake attempt.

3. *Is he temporarily affected by drug substances abuse?* References [56,57] found changes in handwriting due to alcohol. Tucha *et al.* [58] detected the effects of caffeine on handwriting. Foley and Miller [59] performed similar experiments about the effects of marijuana and alcohol. While this consumption could be hardly detected by fingerprint analysis, for instance, this is not the case with biometric signals such as handwriting/signature and speech.

2.4.2 Situations where health information can be extracted from security applications

One of the main concerns of biometrics applied to security is about privacy issues [33]. Technological advances let to store, gather and compare a wide range of information on people. Using identifiers such as name, address, passport or social security number, institutions can search databases for individuals' information. This information can be related to salary, employment, sexual preferences, religion, consumption habits, medical history, etc. This information can be collected with the consent of the user, but in some cases it could also be extracted from biometric samples without the knowledge of the user. Thus, the user could ignore that some additional and private information can be extracted from his biometric samples.

Though in most of the scenarios there should be no problem, there is a potential risk. Let us think, for instance, in sharing medical information. Obviously, in case of emergency, this sharing between hospitals would be beneficial. On the contrary, if this information is transferred to a personal insurance company or a prospective employer, the insurance or the job application can be denied. The situation is especially dramatic when biometric data collection is intended for security biometric recognition to grant access to a facility or information but a third party tries to infer the health condition of the subject. For instance, in the case of retina and iris recognition, an expert can determine that a patient suffers from diabetes, arteriosclerosis, hypertension, etc.

For any biometric identifier, there is a portion of population for which it is possible to extract relevant information about their health, with similar implications to the ones described in the previous paragraph. This is not a specific problem of handwritten signals. Some other biometric signals exhibit the same potential problems. For example, speech disorders, hair or skin colour problems, etc. An important question is what exactly is disclosed when biometric scanning is used. In some cases, additional information not related to identification might be obtained. One possible scenario could be a company where an attendance sheet must be signed each day. The main purpose of this task could be to check if the worker is at his workplace all the labouring days. However, once the handwriting is provided, the company could decide to analyse the signature to detect some pathologies or drugs abuse and to fire out those workers who do not show a good health. And last but not the least, once we provide our biometric samples, they can last in a database for dozens of years, and due to technological advances, they can be used in a simple way in the future to extract additional information that was not intended during acquisition. For this reason, we should think about technical solutions to preserve privacy and legal regulations to avoid that.

2.4.3 Situations where the identity information must be removed

Sometimes, the situation is just opposite to that one mentioned in the previous section. With the growth of eHealth and telemedicine fields, scientists started to develop automatic handwriting analysis systems that can be used for disease diagnosis, rating or monitoring. However, to introduce a robust analysis system, it is necessary to develop it using a large database consisting of hundreds or thousands of subjects. This could be problematic, especially when acquiring patients with rare diseases or patients with cognitive deficits (e.g. patients with dementia). In these cases, it is difficult to find enough samples and explain the handwriting tasks, respectively.

One possibility to overcome the lack of data is to fuse databases acquired by different research or clinical teams around the world, i.e. make the data publicly available (or at least for research purposes). But this is usually not allowed by the local ethics committee or by the participants themselves. Just a few people would make their health data available when containing identity information. Therefore, during last few years, scientists started to develop de-identification methods, that would remove this information, but that would still keep the information about pathology (see the next paragraph). Usually, this is done using a sophisticated parameterization process. For example, in future datasets used for analysis of handwriting in patients with PD, it would be enough to keep and disseminate kinematic, in-air, tremor and pressure characteristics.

In this field, there was a European Cooperation in Science and Technology action devoted to de-identification for privacy protection in multimedia content. De-identification in multimedia content can be defined as the process of concealing the identities of individuals captured in a given set of data (images, video, audio, text), for the purpose of protecting their privacy. This will provide an effective means for supporting the EU's Data Protection Directive (95/46/EC), which is concerned with the introduction of appropriate measures for the protection of personal data. The fact that a person can be identified by such features as face, voice, silhouette and gait, indicates the de-identification process as an interdisciplinary challenge, involving such scientific areas as image processing, speech analysis, video tracking and biometrics. This action aims to facilitate coordinated interdisciplinary efforts (related to scientific, legal, ethical and societal aspects) in the introduction of person de-identification and reversible de-identification in multimedia content by networking relevant European experts and organisations.

2.5 Summary and conclusions

In this chapter, we have described the main characteristics of online handwritten signals as well as their applications on biometric recognition and health. We have emphasized the importance of taking into account that security and health should not be isolated to each other. Care must be taken to protect privacy in health applications (in some studies, the identity should not be revealed), and vice versa, it is important to preserve health state privacy in security ones.

To sum up, some background of both fields is desirable although we are working only in the field of security or health. In addition, privacy aspects must be carefully considered in our technological solutions.

References

[1] Plamondon R., Shirari S. 'On-line and off-line handwriting recognition: a comprehensive survey'. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000;22(1):63–84.

[2] Sesa-Nogueras E., Faundez-Zanuy M., Mekyska J. 'An information analysis of in-air and on-surface trajectories in online handwriting'. Cognitive Computation. 2012;4(2):195–205.

[3] Fierrez-Aguilar J., Galbally J., Ortega-Garcia J., *et al.* 'BiosecurID: a multimodal biometric database'. Pattern Analysis and Applications. 2010;13(2):235–246.

[4] Drotar P., Mekyska J., Rektorova I., Masarova L., Smekal Z., Faundez-Zanuy M. 'A new modality for quantitative evaluation of Parkinson's disease: In-air movement'. Proceedings of the IEEE 13th International Conference on Bioinformatics and Bioengineering; Chania, GR, Nov. 2013. Chania: IEEE; 2013, pp. 1–4.

[5] Drotar P., Mekyska J., Rektorova I., Masarova L., Smekal Z., Faundez-Zanuy M. 'Analysis of in-air movement in handwriting: a novel marker for Parkinson's disease'. Computer Methods and Programs in Biomedicine. 2014;117(3): 405–411.

[6] Faundez-Zanuy M., Sesa-Nogueras E., Roure-Alcobé J., Garré-Olmo J., Lopez-de-Ipiña K., Solé-Casals J. 'Online drawings for dementia diagnose: in-air and pressure information analysis'. Proceedings of the 8th Mediterranean Conference on Medical and Biological Engineering and Computing; Seville, ES, Sep. 2013. New York: Springer; 2013, pp. 567–570.

[7] Rosenblum S., Simhon H. A. B., Gal E. 'Unique handwriting performance characteristics of children with high-functioning autism spectrum disorder'. Research in Autism Spectrum Disorders. 2016;23:235–244.

[8] Mekyska J., Faundez-Zanuy M., Mzourek Z., Galaz Z., Smekal Z., Rosenblum S. 'Identification and rating of developmental dysgrahia by handwriting analysis'. IEEE Transactions on Human-Machine Systems. 2017;47(2), pp. 235–248. doi: 10.1109/THMS.2016.2586605.

[9] Rosenblum S., Parush S., Weiss P. L. 'Computerized temporal handwriting characteristics of proficient and non-proficient handwriters'. American Journal of Occupational Therapy. 2003;57(2):129–138.

[10] Rosenblum S., Dvorkin A. Y., Weiss P. L. 'Automatic segmentation as a tool for examining the handwriting process of children with dysgraphic and proficient handwriting'. Human Movement Science. 2006;25(45):608–621.

[11] Drotar P., Mekyska J., Rektorova I., Masarova L., Smekal Z., Faundez-Zanuy M. 'Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease'. Artificial Intelligence in Medicine. 2016;67:39–46.

[12] Houmani N., Mayoue A., Garcia-Salicetti S., *et al.* 'BioSecure Signature Evaluation Campaign (BSEC'2009): evaluating online signature algorithms depending on the quality of signatures'. Pattern Recognition. 2012;45:993–1003.

[13] Yeung D. Y., Chang H., Xiong Y., *et al.* 'SVC2004: first international signature verification competition'. Proc. of Intl. Conf. on Biometric Authentication, ICBA, LNCS-3072. New York: Springer, 2004, pp. 16–22.

[14] Malik M. I., Ahmed S., Marcelli A., *et al.* 'ICDAR2015 competition on signature verification and writer identification for on- and off-line skilled forgeries (SigWIcomp2015)'. IEEE International Conference on Document Analysis and Recognition; Nancy, France, 23–26 Aug. 2015.

[15] Srihari S., Sung-Hyuk C., Sangjik L. 'Establishing handwriting individuality using pattern recognition techniques'. Proceedings of the Sixth International Conference on Document Analysis and Recognition; 2001, pp. 1195–1204.

[16] Zhang B., Srihari S. 'Analysis of handwriting individuality using word features'. Proceedings of the Seventh International Conference on Document Analysis and Recognition; 2003, pp. 1142–1146.

[17] Hook C., Kempf J., Scharfenberg G. 'A novel digitizing pen for the analysis of pen pressure and inclination in handwriting biometrics'. Biometric Authentication Workshop, Prague 2004, Lecture Notes in Computer Science. 2004;3087:283–294.

[18] Chapran J. 'Biometric writer identification: feature analysis and classification'. International Journal of Pattern Recognition & Artificial Intelligence. 2006;20:483–503.

[19] Sesa-Nogueras E., Faundez-Zanuy M. 'Biometric recognition using online uppercase handwritten text'. Pattern Recognition. 2012;45:128–144.

[20] Martin A. A. F., Doddington G., Kamm T., Ordowski M., Przybocki M., 'The DET curve in assessment of detection task performance'. Proc. Eurospeech '97; Rhodes, Greece, Sept. 1997;4:1899–1903.

[21] Vielhauer C., Basu T., Dittmann J., Dutta P. K. 'Finding meta data in speech and handwriting biometrics'. Proceedings of the SPIE. 2005;5681: 504–515.

[22] Scheidat T., Wolf F., Vielhauer C. 'Analyzing handwriting biometrics in metadata context'. Proceedings of the SPIE. 2006;6072:60720H1–60720H12.

[23] Sesa-Nogueras E., Faundez-Zanuy M., Roure-Alcobe J. 'Gender classification by means of online uppercase handwriting: a text-dependent allographic approach'. Cognitive Computation. 2016;8(1):15–29.

[24] Bandi K. R., Srihari S. N. 'Writer demographic identification using bagging and boosting'. Proceedings of the 12th Conference of the International Graphonomics Society (IGS); Salerno, IT, Jun. 2005. Salerno: IGS; 2005, pp. 133–137.

[25] Liwicki M., Schlapbach A., Loretan P., Bunke H. 'Automatic detection of gender and handedness from on-line handwriting'. Proceedings of the 13th Conference of the International Graphonomics Society (IGS); Melbourne, AU, Jun 2007. Melbourne: IGS; 2007, pp. 179–183.

[26] Liwicki M., Schlapbach A., Loretan P., Bunke H. 'Automatic gender detection using on-line and off-line information'. Pattern Analysis and Applications. 2011;14:87–92.

[27] Likforman-Sulem L., Esposito A., Faundez-Zanuy M., Clemençon S., Cordasco G. 'EMOTHAW: a novel database for emotional state recognition from handwriting and drawing'. IEEE Transactions on Human-Machine Systems. 2017;47(2):273–284. doi: 10.1109/THMS.2016.2635441.

[28] Vielhauer C. 'Processing of handwriting and sketching dynamics' in Thiran J. P., Marqués F., Bourlard H. (ed.). Multimodal Signal Processing: Theory and Applications for Human–Computer Interaction. Oxford: Academic Press; 2009. pp. 119–142.

[29] Sunderland T., Hill J. L., Mellow A. M., *et al*., 'Clock drawing in Alzheimer's disease: a novel measure of dementia severity'. Journal of the American Geriatrics Society. 1989;37:725–729.

[30] Folstein M. F., Folstein S. E., McHugh P. R. 'Mini-mental state: a practical method for grading the cognitive state of patients for the clinician'. Journal of Psychiatric Research. 1975;12(3):189–198.

[31] Faundez-Zanuy M., Sesa-Nogueras E., Roure-Alcobe J., *et al.* 'A preliminary study of online drawings and dementia diagnose'. Neural Nets and Surroundings. Volume 19 of the series Smart Innovation, Systems and Technologies; 2013, pp. 367–374, Springer.

[32] Larner A. J. 'Addenbrooke's cognitive examination-revised (ACE-R) in day-to-day clinical practice'. Age Ageing. 2007;36(6):685–686.

[33] Faundez-Zanuy M., Hussain A., Mekyska J., *et al.* 'Biometric applications related to human beings: there is life beyond security'. Cognitive Computation. 2013;5(1):136–151.

[34] Albert D., Opwis K., Regard M. 'Effect of drawing hand and age on figural fluency: a graphomotor study with the five-point test in children'. Child Neuropsychology. 2009;16(1):32–41.

[35] Louis E. D., Yu Q., Floyd A. G., Moskowitz C., Pullman S. L. 'Axis is a feature of handwritten spirals in essential tremor'. Movement Disorders. 2006;21(8):1294–1295.

[36] Popovic M. B., Dzoljic E., Kostic V. 'A method to assess hand motor blocks in Parkinson's disease with digitizing tablet'. Tohoku Journal of Experimental Medicine. 2008;216(4):317–324.

[37] Pullman S. L. 'Spiral analysis: a new technique for measuring tremor with a digitizing tablet'. Movement Disorders. 1998;13(Suppl. 3):85–89.

[38] Thenganatt M. A., Louis E. D. 'Distinguishing essential tremor from Parkinson's disease: bedside tests and laboratory evaluations'. Expert Review of Neurotherapeutics. 2012;12(6):687–696.

[39] Yan J. H., Rountree S., Massman P., Doody R. S., Li H. 'Alzheimer's disease and mild cognitive impairment deteriorate fine movement control'. Journal of Psychiatric Research. 2008;42(14):1203–1212.

[40] Caligiuri M. P., Teulings H. L., Dean C. E., Lohr J. B. 'A quantitative measure of handwriting dysfluency for assessing tardive dyskinesia'. Journal of Clinical Psychopharmacology. 2015;35(2):168–174.

[41] Eichhorn T. E., Gasser T., Mai N., *et al.* 'Computational analysis of open loop handwriting movements in Parkinson's disease: a rapid method to detect dopamimetic effects'. Movement Disorders. 1996;11(3):289–297.

[42] Lepelley M. C., Thullier F., Bolmont B., Lestienne F. G. 'Age-related differences in sensorimotor representation of space in drawing by hand'. Clinical Neurophysiology. 2010;121(11):1890–1897.

[43] Rey A. 'The psychological examination in cases of traumatic encepholopathy. Problems'. Archives de Psychologie. 1941;28:215–285.

[44] Saunders-Pullman R., Derby C., Stanley K., *et al.* 'Validity of spiral analysis in early Parkinson's disease'. Movement Disorders. 2008;23(4):531–537.

[45] Wang H., Yu Q., Kurtis M. M., Floyd A. G., Smith W. A., Pullman S. L. 'Spiral analysis-improved clinical utility with center detection'. Journal of Neuroscience Methods. 2008;171(2):264–270.

[46] Pinto S., Velay J. L. 'Handwriting as a marker for PD progression: a shift in paradigm'. Neurodegenerative Disease Management. 2015;5(5):367–369.

[47] Rosenblum S., Samuel M., Zlotnik S., Erikh I., Schlesinger I. 'Handwriting as an objective tool for Parkinson's disease diagnosis'. Journal of Neurology. 2013;260(9):2357–2361.

[48] Letanneux A., Danna J., Velay J. L., Viallet F., Pinto S. 'From micrographia to Parkinson's disease dysgraphia'. Movement Disorders. 2014;29(12):1467–1475.

[49] Drotar P., Mekyska J., Rektorova I., Masarova L., Smekal Z., Faundez-Zanuy M. 'Decision support framework for Parkinson's disease based on novel handwriting markers'. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2015;23(3):508–516.

[50] Schröter A., Mergl R., Bürger K., Hampel H., Möller H. J., Hegerl U. 'Kinematic analysis of handwriting movements in patients with Alzheimer's disease, mild cognitive impairment, depression and healthy subjects'. Dementia and Geriatric Cognitive Disorders. 2003;15(3):132–142.

[51] Johnson B. P., Phillips J. G., Papadopoulos N., Fielding J., Tonge B., Rinehart N. J. 'Understanding macrographia in children with autism spectrum disorders'. Research in Developmental Disabilities. 2013;34(9):2917–2926.

[52] Mavrogiorgou P., Mergl R., Tigges P., *et al.* 'Kinematic analysis of handwriting movements in patients with obsessive–compulsive disorder'. Journal of Neurology, Neurosurgery & Psychiatry. 2001;70(5):605–612.

[53] Martin C., Oh E., Addy K., Eskildsen K. 'Biometric verification and duress detection system and method'. Patent US 20070198850 A1, 2007.

[54] Viñals F., Puente M. L. 'Alteraciones neurológicas y biológicas' in Herder (ed.). Grafología criminal. 2009. Chapter 3.

[55] Heinik J., Werner P., Dekel T., Gurevitz I., Rosenblum S. 'Computerized kinematic analysis of the clock drawing task in elderly people with mild major depressive disorder: an exploratory study'. International Psychogeriatrics. 2010;22(3):479–488.

[56] Faruk A., Turan N. 'Handwritten changes under the effect of alcohol'. Forensic Science International. 2003;132(3):201–210.

[57] Phillips J. G., Ogeil R. P., Muller F. 'Alcohol consumption and handwriting: a kinematic analysis'. Human Movement Science. 2009;28:619–632.

[58] Tucha O., Mecklinger L., Walitza S., Lange K. W. 'The effect of caffeine on handwriting movements in skilled writers'. Human Movement Science. 2006;25:523–535.

[59] Foley R. G., Miller L. 'The effects of marijuana and alcohol usage on handwriting'. Forensic Science International. 1979;14(3):159–164.

A.22 Identification and Rating of Developmental Dysgraphia by Handwriting Analysis

Identification and Rating of Developmental Dysgraphia by Handwriting Analysis

Jiri Mekyska, Marcos Faundez-Zanuy, Zdenek Mzourek, Zoltan Galaz, Zdenek Smekal, *Senior Member, IEEE*, and Sara Rosenblum

***Abstract*—Developmental dysgraphia, being observed among 10–30% of school-aged children, is a disturbance or difficulty in the production of written language that has to do with the mechanics of writing. The objective of this study is to propose a method that can be used for automated diagnosis of this disorder, as well as for estimation of difficulty level as determined by the handwriting proficiency screening questionnaire. We used a digitizing tablet to acquire handwriting and consequently employed a complex parameterization in order to quantify its kinematic aspects and hidden complexities. We also introduced a simple intrawriter normalization that increased dysgraphia discrimination and HPSQ estimation accuracies. Using a random forest classifier, we reached 96% sensitivity and specificity, while in the case of automated rating by the HPSQ total score, we reached 10% estimation error. This study proves that digital parameterization of pressure and altitude/tilt patterns in children with dysgraphia can be used for preliminary diagnosis of this writing disorder.**

***Index Terms*—Dysgraphia, handwriting analysis, handwriting proficiency screening questionnaire (HPSQ), intrawriter normalization, rating.**

I. Introduction

WRITING is a complex form of language production ranging from the idea of conceptualization to motor execution by hand, meaning handwriting [1]. Handwriting is a complex human activity, considered to be an "overlearned" skill involving particularly rapid sequencing of movements in time, which reflects the relationship between planning and product generation [2]. Handwriting is one of the functional daily activities required of school-aged children for their adequate participation in the academic process [3]. In fact, 50% of a child's school day is spent performing handwriting tasks [4], [5]. It is proposed that the mastery of lower level transcription skills such as handwriting and spelling is required for idea conceptualization and production of high-level content text [6]. For example, handwriting speed was found to be important for note-taking-recording important information [7], and handwriting automaticity was correlated with children's composition performance variance [8]. Despite the widespread use of computers, handwriting still serves as a medium of communication and is a necessary life skill [3].

Most children are capable of coping with their handwriting requirements and become proficient writers. Their handwriting is legible, and they invest little effort in the handwriting process [9], [10]. The process of adopting the letters' form begins at the first grade (age 6 in Israel). Following two school years of handwriting experience leads to unify an automatic proficient manner of letters production, which occurs around age eight [11]. Thus, children at that age who do not succeed in developing proficient handwriting face developmental dysgraphia [12], [13], and it is not just a matter of "tim" or "maturatio." Their functional limitations are manifested in inadequate speed and/or product legibility [9], [14].

Dysgraphia is found among children of at least average intelligence and who have not been identified as having any obvious neurological or perceptual-motor problems. The prevalence of handwriting difficulties or developmental dysgraphia among school-aged children varies between 10% and 30% [13], whereas children with neurodevelopmental disabilities were found at high risk for handwriting difficulties [15].

Researchers suggested that handwriting difficulties might have serious consequences for the student's overall academic success, emotional well-being, attitude, and behavior [3], [16]. These findings reinforce the importance of identifying handwriting difficulties as early as possible, both as a preventive and as a corrective aid [17].

Previous studies indicated the benefits of several methods for both identifying and detecting unique handwriting characteristics of children with dysgraphia. Two short and practical questionnaires enable dysgraphia identification: one is designated for teachers or parents [handwriting proficiency screening questionnaire (HPSQ)] [18] and one for the child's self-report [handwriting proficiency screening questionnaire for child (HPSQ-C)] [11]. Those two scales cover three domains of dysgraphic writing production: legibility, performance time, and physical and emotional well-being.

Further studies indicated the benefits of a computerized system for detection of the handwriting process and further evaluation of the written product for detection of handwriting

Manuscript received February 10, 2016; revised May 17, 2016; accepted June 28, 2016. Date of publication August 3, 2016; date of current version March 13, 2017. Research described in this paper was financed by the National Sustainability Program under grant LO1401. For the research, infrastructure of the SIX Center was used. This paper was recommended by Guest Editor E. Anquetil. *(Corresponding Author: Jiri Mekyska.)*

J. Mekyska, Z. Mzourek, Z. Galaz, and Z. Smekal are with the Department of Telecommunications, Brno University of Technology, 61600 Brno, Czech Republic (e-mail: mekyska@feec.vutbr.cz; z.mzourek@phd.feec.vutbr.cz; xgalaz00@stud.feec.vutbr.cz; smekal@feec.vutbr.cz).

M. Faundez-Zanuy is with the Escola Superior Politecnica, Tecnocampus, 08302 Barcelona, Spain (e-mail: faundez@tecnocampus.cat).

S. Rosenblum is with the Laboratory of Complex Human Activity and Participation, Department of Occupational Therapy, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa 3498838, Israel (e-mail: rosens@univ.haifa.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2016.2586605

features of children with dysgraphia [13], [19]–[21]. However, knowledge from the field of machine learning may shed more light on the performance of features and may serve for improving both diagnosis and detection of handwriting deficits among children with dysgraphia. Previous literature indicated the benefits of sequential motor task, which requires consideration of quantity, time, space, and pressure for diagnosis of deficient performance [19], [22]–[24]. Thus, the assumption of the current study was that unique characteristics taken from machine learning discipline of sequential task as repetitions of the same letter may contribute to identification of children with dysgraphia. In fact, such task performance reflects the child's ability to control quantitative, temporal, spatial, and pressure characteristics of handwriting production simultaneously. Therefore, the aim of this work is to:

1) introduce an automated and complex parameterization approach that can be used for quantification of different dysgraphia domains;
2) propose a simple intrawriter normalization method that can increase discrimination/automated HPSQ total score estimation accuracies;
3) propose a system of dysgraphia discrimination;
4) propose a system of dysgraphia rating and evaluate its performance in terms of HPSQ total score estimation error.

The estimation error is defined here as a relative difference between HPSQ total score given by clinician and score calculated automatically by the system.

The rest of this paper is organized as follows. Section II describes the dataset and methodology. Section III provides some preliminary insight into extracted features, as well as the results of dysgraphia discrimination and HPSQ total score estimation using feature selection and intrawriter normalization. A discussion of the results can be found in Section IV, and, finally, conclusion is given in Section V.

II. Patients and Methods

A. Concept of Automatic Dysgraphia Classification and Rating

Because a reader of this work can be outside the field of signal processing and machine learning, first, we decided to describe the idea of automatic developmental dysgraphia rating using a simple picture (see Fig. 1). This section is suitable more for clinicians and researchers from the field of occupational therapy or human–machine interaction. Those who are more interested in the theory of handwriting signal parameterization, mathematical model construction, and evaluation can move to Sections II-E and II-F.

The activation of system (the way how the system works in office of clinician during examination of children) is described in the lower part of the picture. First, consider that the system is a black box that has a digitizer as an input and that shows the estimated HPSQ score on an output. The child writes a sequence of letters on a paper using an ink pen. The paper is placed on a surface of digitizer that is recording the whole process of writing. After the child finishes, the system performs automatic handwriting analysis and estimates child's HPSQ score. This can be used as a preliminary diagnosis and potential recommendation of seeing a therapist who can do a deeper analysis and a final diagnosis.

In our case, the black box has four main parts (see the Fig. 1).

1) Sequential writing acquisition—a digitizer used for the acquisition and specific sequential handwriting task consisted of repeated cursive Hebrew letter HET (for more information, see Section II-D).
2) Feature extraction—in this part of system, the handwriting is quantified, i.e., we describe kinematic aspects (velocity, jerk, acceleration, duration, etc.), handwriting geometry (width, height, orientation, etc.), handwriting fluency (tremor, irregularities in velocity, etc.), in-air movement (movement of pen when its tip is not touching the surface of paper), and other characteristics (see Section II-E).
3) Intrawriter normalization—some of the previously extracted features are very writer dependent. For example, consider that velocity is very individual and it affects the value of all other features. In order to suppress the individuality in handwriting and emphasize dysgraphic features, it is necessary to introduce some kind of intrawriter normalization. In our work, it is performed by a simple subtraction (see Section II-E5).
4) Dysgraphia rating—in this part, we send the values of previously normalized features to a mathematical model (equation) whose output is the estimated HPSQ score. In our work, we employed a model based on trees whose nodes are the values of the features. During the estimation of the HPSQ, we begin from the root and ask a question: "Is the value of the feature greater or smaller than a specific threshold?" If the value is smaller, we move to the next node on the right, otherwise on the left. We repeat the questions with the different features until we reach the last node. This last node holds a specific value of HPSQ that is used as an estimate (for deeper description of classification and regression trees (CART), see Section II-F4).

We described the process of activation. However, to activate this system, we must first train it, which means that we must identify the features that are significant for dysgraphia rating, we must find the feature that is suitable for intrawriter normalization, and, finally, we must construct the tree (identify its structure, features on nodes, and thresholds). Generally, we call it a training phase. Training is performed during development of system, and it is described by the upper part of Fig. 1.

5) Acquisition of the training database—in order to create a robust model with good sensitivity and specificity, we must train the system using a complex database that contains handwriting of children (healthy and with dysgraphia) with different HPSQ scores (given by clinicians). For more information about our database, see Section II-B.
6) Extraction of all the possible features—at the beginning of training, we do not know what features are significant for dysgraphia rating. Therefore, we extract all the possible features (see Section II-E), and then, we try to identify those that are significant.
7) Finding optimal feature for intrawriter normalization—in the next step of system training, we must identify a feature that is suitable for intrawriter normalization. In

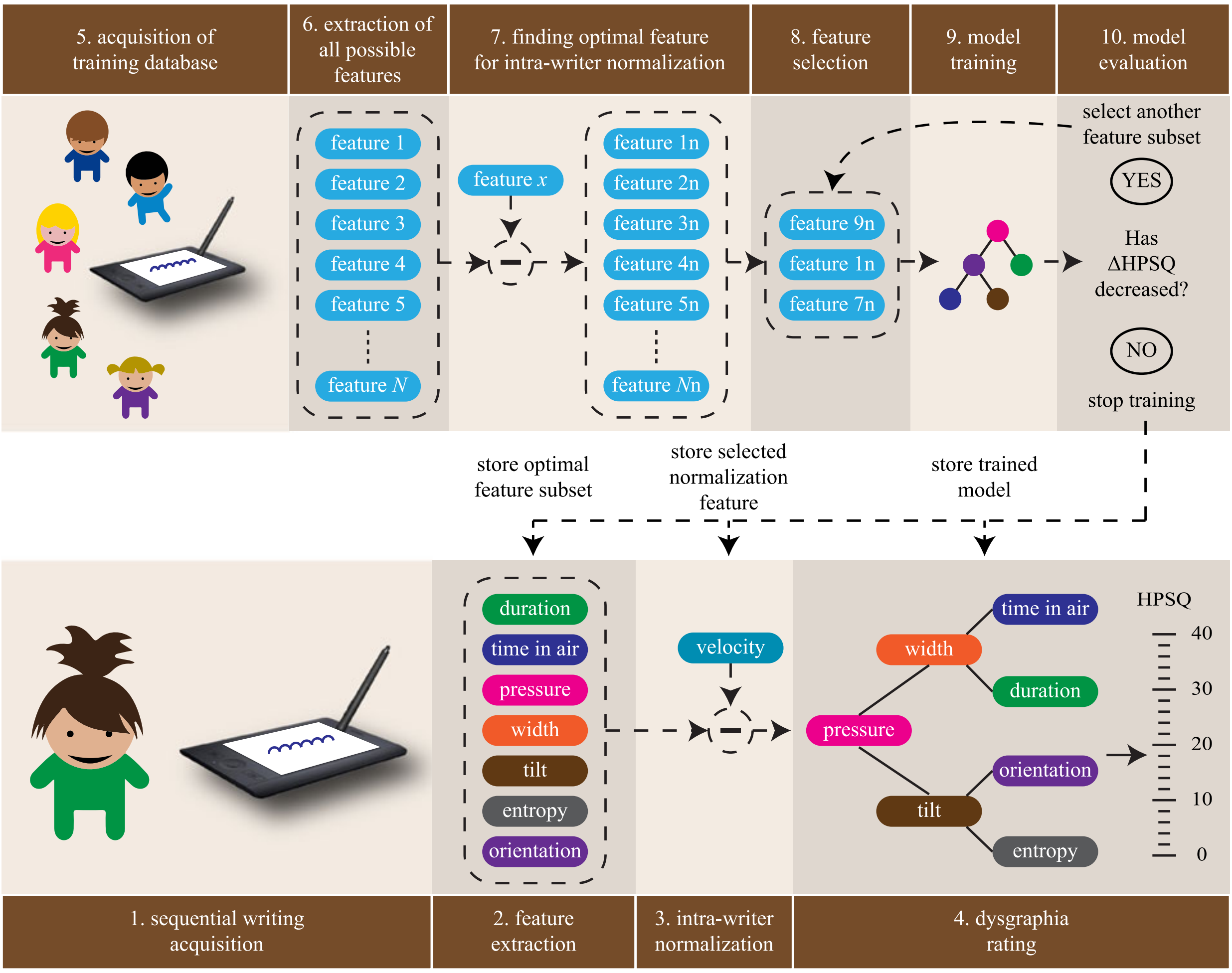


Fig. 1. Process of training and activation of the automatic dysgraphia classification and rating system.

our work, we tested all the possible combinations (see Section II-E5).

8) Feature selection—as was already mentioned, not all features are significant for dysgraphia rating. Some can be redundant, some irrelevant, etc. We try to find the smallest possible subset of features that will provide the best estimation power. For this purpose, we used a combination of two techniques; see Section II-F2.
9) Model training—in this step, we find the best structure of tree and optimal thresholds (see Section II-F4).
10) Model evaluation—in step 8, we try to find the best feature subset. For each subset, we train one model (tree) and evaluate it (see Section II-F1). It means we use some children from the database for training and some for testing (see Section II-F5). In the case of testing, we know the real HPSQ score given by clinician and the HPSQ score estimated by the system. We can calculate difference of these two values and monitor its trend. If the value of this difference decreased, it means that we can still improve rating accuracy and we are trying to find better subset of features. If the difference reached its minimum, we stop training and store information about optimal feature subset, normalization feature, and model parameters. At this point, we can distribute the system and use it for rating.

B. *Study Participants*

Two groups of handwriters (proficient and dysgraphic), each consisting of 27 third-grade male and female pupils, aged 8 and 9, were included in the study. Dysgraphic handwriters were identified via the standardized and validated ten-item questionnaire for handwriting proficiency (HPSQ) [18]. All participants were born in Israel, used the Hebrew language as their primary means of verbal and written communication, and were right-hand dominant. The proficient handwriters were matched to the participants in the poor handwriting group on the basis of gender, age, school, and class. There were no significant differences between the two groups with respect to their age (8.38 ± 0.22 years for the proficient handwriters and 8.32 ± 0.30 years for the dysgraphic handwriters) and gender ratio (13 girls and 14 boys in each group). Children with known psychiatric/emotional disorders, autistic tendencies, physical disabilities, or neurological or systemic disease were excluded from the study.

C. *Handwriting Proficiency Screening Questionnaire*

The HPSQ is a ten-item questionnaire developed to spot school-aged children with handwriting difficulties based on their teacher's observation. These ten items cover the most important indicators of handwriting deficiencies, i.e., legibility, performance time, and physical and emotional well-being were scored

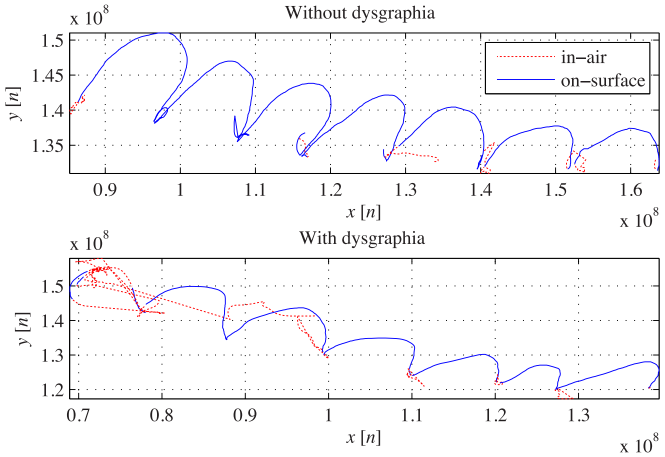


Fig. 2. Representative sequential writing samples selected from our database.

Fig. 3. Template that was copied by children.

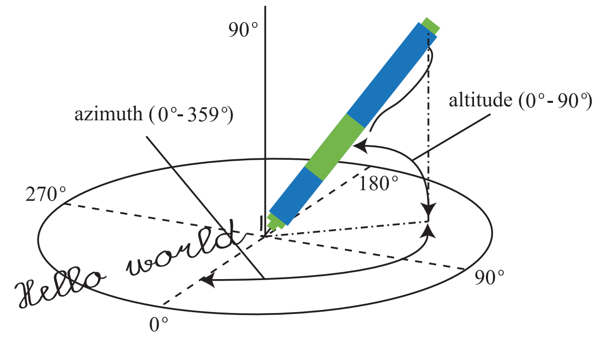


Fig. 4. Altitude/tilt and azimuth information.

on a five-point Likert scale (from 0—never to 4—always) and a final score is summed. The examples of the items in each category are: *Legibility:* Is the child's handwriting readable? *Performance time:* Does the time given in class for copying tasks from the blackboard is enough for the child? *Physical and emotional well-being:* Does the child complain about pain while writing? (for more details, see [18]). The questionnaire's reliability and validity were established, and a sum of 14 was determined as the cut of score for handwriting deficiency among school-aged children in Israel [18]. In the present study, the internal reliability for the entire scale was found to be $\alpha = 0.95$. The HPSQ is now in the process of adaptation in more than ten countries worldwide (e.g., Spain).

D. Handwriting Acquisition

We used Computerized Penmanship Evaluation Tool (ComPET, previously named POET) [25], which is standardized and validated handwriting assessment utilizing a digitizing tablet and online data collection and analysis software. It was developed to collect objective measures of the handwriting process (for more details, see [25]). In the present study, children were asked to write a sequence of seven semi-HET (letters (as presented to them in Fig. 3). This pattern has been used in previous studies, where authors analyzed nature of control of pen stroke size while writing [26], [27]. The pattern was written on A4 size lined paper affixed to the surface of a WACOM Intuos II xy digitizing tablet (404 $\times$ 306 $\times$ 10 mm) using a wireless electronic pen with a pressure-sensitive tip (Model GP-110). Displacement, pressure, and pen tip angle were sampled at 100 Hz via a 1300-MHz Pentium (R) M laptop computer.

Fig. 2 represents sequential semi-HET letter writing of children without (C, controls) and with dysgraphia (D) from our database. As can be seen, while a child without dysgraphia performs the writing task with almost no "in-air" motions, in a sequential way, the child with dysgraphia needs more time, and motion in-air between the written strokes and his performance is less sequential (especially at the beginning of writing, see the left part of the second figure). Such findings are similar to previous, while children with dysgraphia stayed in-air more time than proficient writers, and their writing was less automatic [19], [25].

The ComPET system enables dynamic handwriting evaluation while analyzing temporal, spatial, and pressure measures for each writing stroke. In the present study, a stroke is defined as the sequential written line from the point at which the pen touches the paper (applying pressure of more than 50 nonscale units) until the point it leaves the paper [28].

E. Handwriting Features

Our first objective was to introduce a complex parameterization approach that can be used for quantification of different manifestations of dysgraphia. In this work, we considered a large and complex set of the features that quantify handwriting geometry, dynamics, tremor, pressure, and altitude. We used the conventional features [19], [21], [25], [29] as well as new and more complicated parameters proposed in our recent works for analysis of micrographia and generally Parkinsonic (PD) dysgraphia [30], [31]. The digitizing tablet is able to capture several signals. These include position of the pen tip in terms of x and y coordinates $(x[n], y[n])$, time stamp $(t[n])$, and a binary variable $(b[n])$, being 0 for pen-up state (in-air movement) and 1 for pen-down state (on-surface movement). This means that besides the on-surface movement, the tablet is able to track $x[n]$ and $y[n]$ information when the pen is not touching the tablet surface too. It has been already proven that in-air movement brings the additional and valuable information to the overall handwriting analysis [32]. Moreover, it increases PD dysgraphia identification accuracy [33], [34]. Therefore, we included in-air movement analysis as well. In addition to the previously mentioned signals, we processed pressure exert on the tablet surface during writing $(p[n])$, information of pen altitude/tilt $(a[n])$, and azimuth $(az[n])$ (see Fig. 4).

The on-surface or in-air movement can be further divided into horizontal (in direction of $x[n]$) and vertical (in the direction of $y[n]$) one. A sample of trajectory is then defined as $\mathrm{tr}[m] = \sqrt{(x[n+1] - x[n])^2 + (y[n+1] - y[n])^2}$ for $m = 0, \ldots, n-1$. Notation of handwriting feature has format *INF:*

DIR-FN (HL), where *INF* indicates processed information (ON for on-surface, AIR for in-air, and P for pressure), *DIR* denotes direction (H for horizontal and V for vertical), *FN* holds feature name, and *HL* statistic used for transformation to scalar value (see Section II-E4). For example, AIR: H-VEL (med) means median of horizontal velocity during in-air movement.

We divided the feature set into three groups: kinematic features, nonlinear dynamic features, and other features. A deeper description of these features is beyond the scope if this paper; however, we always refer to the publications that contain a sufficient definition.

1) Kinematic Measures: It has been shown that various kinematic aspects are affected in children with dysgraphia [13], [19], [21], [25], and several measures were proposed for quantification of these disruptions. We can consider these measures as conventional ones, because they are used in a wide range of handwriting analysis applications. In our work, we used these kinematic and spatiotemporal features (to analyze the whole handwriting and particular strokes as well): speed (SPEED, trajectory divided by duration), velocity (VEL, rate at which the position of a pen changes with time), acceleration (ACC, rate at which the velocity of a pen changes with time), jerk (JERK, rate at which the acceleration of a pen changes with time, reflects the smoothness of the movement, and can provide the information about the coordination of finger and wrist movements), normalized jerk (N_JERK, normalized by movement duration), width (WIDTH), height (HEIGHT), orientation (ORIENT, describe slope of handwriting), duration (DUR), and length (LEN, length of trajectory) [29], [33], [35], [36].

2) Nonlinear Dynamic Features: From the signal processing point of view, handwriting is time series that is the result of several interacting physiological mechanisms. This kind of signal contains complex fluctuations, which could provide the information related to underlying processes and states of the physiological system. Disfluent movement and irregular muscle contractions introduce randomness to the handwriting and increase its complexity (e.g., add tremor, more handwriting interruptions, sudden changes in velocity, etc.). We hypothesize that D children with deficient fine motor skills, poor dexterity, poor muscle tone, or unspecified motor clumsiness manifest higher complexity of handwriting. However, this complexity randomness is difficult to analyze using only kinematic measures. To uncover hidden complexities, several features were proposed by the research community, and some of them were first used for analysis of dysgraphia in this work.

The first of them is correlation dimension (CD), which statistically measures attractor geometry in the phase space. CD is related to a number of independent variables necessary for generating the attractor [37]. Another dimension measure fractal dimension is based on a number of basic building blocks that form a pattern [38]. To quantify the regularity embedded in a time series, we used Ziv-Lempel complexity (ZLC) [39]. Possible long-term dependencies in the analyzed signal were described by Hurst exponent [40].

Next, we used a set of features based on entropies. The entropy is a measure of uncertainty, and it can be used to quantify the complexity of a system. In our work, we focused on Shannon entropy (SHE) [30], first-order Shannon entropy (SHE1) [41], second-order Rényi entropy (RE) [30], correlation entropy (CE) [42], first-order Rényi block entropy (RBE1) [43], second-order Rényi block entropy (RBE2) [43], permutation entropy (PE) [44], normalized PE, approximate entropy (AE) [45], sample entropy (SE) [45], and normalized recurrence time probability density entropy [46]. REs quantify the loss of information in time in a dynamic system, CE and first-order SHE give an indication of the predictability of the nonlinear time series, and PE takes into account temporal information in the time series. The only difference between AE and SE is that SE does not evaluate a comparison of embedding vectors with themselves. Approximate and sample entropies are usually used with Heaviside kernel (HEAV); however, in this work, we extended these features by the other kernels like Gaussian (GAUSS), exponential (EXP), Laplacian (LAPL), circular (CIRC), spherical (SPHE), Cauchy (CAUCH), and triangular (TRIAN) proposed in [47].

Another representative of nonlinear dynamic features is first minimum of mutual information [43]. To include also a measure of sensitivity to an initial condition, the largest Lyapunov exponent [48] was employed (its absolute value LLE and its prediction error PE_LLE). Finally, we used detrended fluctuation analysis (DFA) to characterize the self-similarity of the graph of a signal from a stochastic process [49]. In this field, normalized scaling exponent (NSE_DFA) and fluctuation amplitudes (FA_DFA) were evaluated.

3) Other Features: The ability to make handwriting fluent with minimum of unnecessary interruptions (interruption is considered as a change from on-surface to in-air movement and it is closely related to a number of strokes in handwriting) can be quantified by a number of interruptions (N_INT) and its normalized version, relative number of interruptions (RN_INT, normalized by duration).

The last measure that quantifies tremulous/noisy parts of handwriting (and indirectly its complexity) is median of power spectral density. This parameter takes into account a distribution of energy in power spectrum, and its increased value indicates a significant presence of high-frequency parts [50].

A feature that is based on energy of an analyzed signal is Teager–Kaiser energy operator (TKEO). The advantage of this feature is that it takes into account also signal frequency [51]. This feature found its place in the field of handwriting analysis; usually, it is used for description of pressure profile (PRESS) [52]. It quantifies the total pressure over a small period of time.

Some of the signals recorded by digitizer were directly used as the features and were forwarded to a postprocessing step described in Section II-E4. Specifically, we used pressure (PRESS), azimuth (AZIM), and altitude (ALT).

4) Handwriting Features Postprocessing: The feature extraction stage produces the parameters represented either by scalar values or by vectors. An example of vector representation is pressure profile or velocity calculated for each stroke. However, to be able to carry out the next processing like statistical analysis, classification, and regression, the vector representation must be transformed to a scalar value. This is usually done by an extraction of some kind of statistics (we call these statistics high-level features). We extracted these 62 high-level features:

1) maximum (max), minimum (min), position of max (pmax), position of min (pmin), relative position of max (rpmax), relative position of min (rpmin);
2) range, relative range (rr), interquartile range (iqr), relative interquartile range (riqr), interdecile range (idr), relative interdecile range (ridr), interpercentile range (ipr), relative interpercentile range (ripr), studentized range (sr);
3) mean, geometric mean (gmean), harmonic mean (hmean), mean excluding 10%, 20%, 30%, 40% and 50% of outliers (mean x%), median (med), mode;
4) variation (var), standard deviation (std), mean absolute deviation (mad), median absolute deviation (mead), geometric standard deviation (gsd), coefficient of variation (cv), index of dispersion (id);
5) 3rd, 4th, 5th and 6th moment (xm), kurtosis (kurt), skewness (skew), Pearson's first skewness coefficient (skew1), Pearson's second skewness coefficient (skew2);
6) 1st, 5th, 10th, 20th, 30th, 40th, 60th, 70th, 80th, 90th, 95th and 99th percentile (xp), 1st and 3rd quartile (xq);
7) slope, offset (off) and error (err) of linear regression;
8) modulation (mod), Shannon entropy (ent), first-order entropy (1ent), second-order RE (2ren), first correlation coefficient (1cc).

5) Intrawriter Normalization: The second objective of this study is to propose a simple intrawriter normalization method that can increase discrimination/automated HPSQ total score estimation accuracies. In pattern recognition field, the convenience of score normalization before classification is well known. There is variability in tasks performed by human beings. This variability is evident between different people (interwriter) but also when looking at different realizations of a same task performed by a specific writer. One example could be a signature. The signature of different people is different. On the other hand, different realizations of a specific signature are not exactly the same due to human variability. This means that a direct comparison between measurements could provide wrong conclusions. A classic way to improve the results is trying to remove this variability. It can be done using the samples of the same user or using the samples of different users, usually known in the technical literature as cohorts or universal background model (UBM).

For instance, in speaker verification, the UBM is a reference speaker model to which the target speaker is compared during the classification process to produce a log-likelihood ratio (LLR) [53]. During classification, the LLR can be calculated from the target speaker model and background model. In essence, this configuration can be viewed as the UBM normalizing for the characteristics of the impostor population that has potential to affect the classification score.

Another successful technique is called cohort normalization [54], which tries to normalize classification scores in the same manner as the UBM. Rather than using a world model, an impostor person model with similar characteristics to the target person is dynamically selected from a small cohort of similar impostor people. This cohort is selected based on a distance metric between models.

This implies that a predefined classification threshold is setup to decide if a sample belongs to an impostor or a genuine user. Thus, rather than having an adaptive threshold, the score is adapted itself using a reference (UBM or cohort).

In this paper, we propose a normalization scheme similar to the cohort strategy, but rather than using other users, we use other handwriting features of the same user. Mathematically, for a specific user, we extract a set of M features $F = f_1, f_2, \ldots, f_M$. For instance, the feature f_1 could be the mean speed of handwriting, f_2 the mean acceleration, and so on. Considering that a normalized feature is called f_{norm}, the following normalization can be applied for a specific user n:

$$f_{\mathrm{norm}} = f_a(n) - f_b(n) \tag{1}$$

where f_a and f_b are different features, and $a, b \in 1, \ldots, M$. We tested all combinations of features' normalization.

F. Statistical Analysis and Classification

1) Preliminary Statistical Analysis: To obtain some preliminary insight into statistical properties of the features, we followed [49] and calculated nonparametric Spearman's rank sum correlation [55] coefficient and mutual information (MI) between the features and associated clinical diagnosis. MI is a measure of the amount of the information shared by two random variables. The larger the value of MI, the stronger statistical association between the feature and the response can be observed. MI is defined as

$$I(X;Y) = \int_X \int_Y f(x,y) \log_2 \left(\frac{f(x,y)}{f_X(x) f_Y(y)} \right) \tag{2}$$

where X and Y are random variables with associated joint probability density function $f(x,y)$ and marginal density functions $f_X(x)$ and $f_Y(y)$, respectively. We calculated MI using marginal entropies $H(X)$ and $H(Y)$ and joint entropy $H(X,Y)$, defined as

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \tag{3}$$

We also performed the Mann–Whitney U test to compare the handwriting features between C and D children. The Mann-Whitney U test is a nonparametric statistical test used to assess whether two independent groups are significantly different from each other. Additionally, every feature was used separately as an input to the linear discriminant analysis (LDA) and random forest (RF) classifiers to evaluate its discrimination power. The classifiers were evaluated using classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and tradeoff between sensitivity and specificity (TSS). Definition of these metrics can be found in Section II-F2. For each feature, we also calculated its median and standard deviation in D or C group. Finally, we mentioned relative difference of both medians D $>$ C defined as (median D $-$ median C)/median C.

We evaluated statistical properties of the features using Spearman's rank sum correlation, MI, Mann–Whitney U test, TSS (LDA), and TSS (RF) separately. For each case, we selected five most significant features that were included into the final overview table. In addition, we plotted density estimations (computed using kernel density estimation with Gaussian kernels) of top three features with the highest discrimination power according to the Mann–Whitney U test.

In the next step, we were interested in ability of the features to rate the dysgraphia and estimate HPSQ total score. In addition, in this step, we employed Spearman's rank sum correlation and Mann–Whitney U test; however, in these cases, we tested an association between the features and HPSQ. Finally, for each parameter individually, we trained CART and evaluated its estimation power by mean absolute error (MAE) and two kinds of estimation errors [56]

$$\mathrm{EE1} = \frac{\mathrm{MAE}}{\mathrm{range(HPSQ)}} \tag{4}$$

$$\mathrm{EE2} = \frac{\mathrm{MAE}}{\max(\mathrm{HPSQ})} \tag{5}$$

where the function range(HPSQ) calculates the range from HPSQ data available during the analysis, while the function max(HPSQ) returns the maximal score that can be theoretically reached in the specific scale. Theoretically, EE1 is more relevant because it takes into account only HPSQ values that were inside the database. To have some visual insight into the associations between the features and HPSQ, we selected three parameters with the highest Spearman's rank sum correlation coefficients and plotted correlation graphs.

2) Feature Selection: A general problem in data analysis is the curse of dimensionality that can be summarized as follows: The presence of a very large number of the features (parameters) inhibits the detection of the useful patterns underlying the data. The curse of dimensionality often obstructs the subsequent classification process. In order to reduce the dimensionality of the input feature set and also remove the nonrelevant features before the classification, we employed a feature selection process in which the high-dimensional feature space is analyzed in direction of obtaining a compact subset of features holding the maximum clinically relevant information without a loss of predictive information. In addition, the feature selection process does considerably reduce a risk of overfitting and also associated computational performance requirements. Moreover, it has been proven in many research articles that reduction of the feature space before the classification can significantly improve the model's predictive power [57].

Many different feature selection methods exist. They are generally divided into the following categories: filters, wrappers, and embedded methods [58]. The filter methods select the feature subsets from entire feature set independently of chosen learning algorithm. Wrapper methods search for best feature subset for a given classifier; however, wrapper methods are often computationally very expensive. Last, embedded methods select feature subset using the information obtained from a classifier. In this work, we used a two-step feature selection process consisting of the preprocessing step using a filter-based algorithm and the processing step using a wrapper approach. The rationale behind this particular approach is the fact that the filter methods are faster and provide better generalization than wrapper or embedded methods, and therefore, its usage as the preprocessing step can mitigate the requirements of the subsequent wrapper-based methods [59]. In contrast, the wrapper methods do utilize the learning algorithm of interest to score the subsets of features according to their predictive power.

In the preprocessing step, the minimum redundancy maximum relevance (mRMR) algorithm was considered. We used the implementation of Tsanas, Little, and McSharry [60] and computed mRMR using Spearman's correlation coefficient as a criterion to quantify the statistical relationships between the features and the response in mRMR. The features that did pass the preprocessing step (for the purpose of this paper, we used 20 features as a reasonable tradeoff between a size of the feature set and the associated computational requirements) were selected as the candidates for further processing and classification. Consequently, we performed the sequential floating forward selection (SFFS) algorithm to determine the best possible subsets of the features. Furthermore, we have recently introduced a novel classification performance criterion for determining the best available feature subset by taking into account the relationship between classification sensitivity and specificity, named TSS [61]. The value of this parameter was used for stop criteria in SFFS. Classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and TSS are defined as

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{6}$$

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{7}$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{8}$$

$$\mathrm{TSS} = 2^{\sin\left(\frac{\pi \mathrm{SEN}}{2}\right)\sin\left(\frac{\pi \mathrm{SPE}}{2}\right)} \tag{9}$$

where TP (true positive) and FP (false positive) represents the number of correctly identified D subjects and a number of subjects identified as D, but being without dysgraphia. Similarly, TN (true negative) and FN (false negative) represent the total number of correctly identified controls (C), and D children identified as controls.

3) Classification: The third objective of this work is to propose a system of dysgraphia discrimination. Term *discrimination* in this work is considered as binary classification from the machine-learning point of view. During preliminary statistical analysis, we used the LDA and RF classifiers to evaluate the features' discrimination power. LDA is a statistical method used for separation of two or more classes. For the prediction, LDA utilizes the conditional probability density functions f_k, where k represents the particular class. For the purpose of this study, f_0 and f_1 will represent the conditional probability functions of C and D children, respectively. LDA assumes that the f_k are normally distributed, and they share a common variance matrix. LDA estimates the parameters of f_k (mean and covariance) using the data and uses them to construct a linear function called the decision criterion used for the prediction. In practice, the assumptions of LDA never hold; however, LDA still gives reasonably good results comparable with logistic regression [62].

However, we observed that RF provided better sensitivity and specificity; therefore, we decided to use this classifier in consequent classification based on feature selection and intrawriter normalization. RF is an ensemble learning algorithm that operates by constructing a multiple of base learners (weighted decision trees) [63]. Typically, the RF classifies the

TABLE I
PRELIMINARY FEATURE ANALYSIS IN TERMS OF DISCRIMINATION POWER

Feature	**r**	**p (SC)**	MI	p (MW)	D > C	median (C)	std (C)	median (D)	stdxbrk (D)	ACC (LDA)	SEN (LDA)	SPE (LDA)	TSS (LDA)	ACC (RF)	SEN (RF)	SPE (RF)	TSS (RF)
ALT (cv)	**0.5074**	**0.0001**	0.4204	**0.0002**	46.97	0.1823	0.302	0.3437	0.6865	64.29	40.74	86.21	1.50	50.00	51.85	48.28	1.4144
ALT (id)	**0.4787**	**0.0002**	0.2841	**0.0004**	62.62	8.8984E+05	2.3154E+06	2.3808E+06	9.1177E+06	66.07	37.04	93.10	1.46	67.86	70.37	65.52	1.7002
PRESS (60p)	**0.4202**	**0.0013**	0.4926	**0.0019**	19.63	172	43.21	214	35.85	71.43	74.07	68.97	1.75	67.86	66.67	68.97	1.6995
ON: H-N_JERK (cv)	**0.4168**	**0.0014**	0.4164	**0.0021**	13.61	0.9566	0.2595	1.1073	0.2407	66.07	62.96	68.97	1.67	50.00	55.56	44.83	1.4102
PRESS (med)	**0.4024**	**0.0021**	0.5622	**0.0029**	19.81	166	43.66	207	38.58	73.21	70.37	75.86	1.78	57.14	59.26	55.17	1.5277
Feature	r	p (SC)	**MI**	p (MW)	D > C	median (C)	std (C)	median (D)	std (D)	ACC (LDA)	SEN (LDA)	SPE (LDA)	TSS (LDA)	ACC (RF)	SEN (RF)	SPE (RF)	TSS (RF)
ON: V-FA_DFA (mean 10%)	0.0851	0.5328	**0.6442**	0.5332	12.39	9.6679E+07	6.5636E+07	1.1036E+08	5.9985E+07	55.36	48.15	62.07	1.48	58.93	51.85	65.52	1.5403
ALTxbrk (5p)	−0.2057	0.1282	**0.6354**	0.1291	−70.00	1.5938E+07	1.9028E+07	9.3750E+06	1.1344E+07	55.36	40.74	68.97	1.44	76.79	74.07	79.31	1.8279
ON: V-PE_LLE (1q)	0.2686	0.0453	**0.6237**	0.0472	23.92	0.1307	0.0569	0.1718	0.0562	67.86	70.37	65.52	1.70	55.36	48.15	62.07	1.4825
PRESS (off)	0.2134	0.1144	**0.6132**	0.1155	18.36	155.61	47.08	190.61	50.28	60.71	59.26	62.07	1.58	57.14	55.56	58.62	1.5261
ON: V-FA_DFA (1q)	−0.147	0.2795	**0.6085**	0.2792	−24.67	1.4036E +06	6.7119E +05	1.1259E +06	6.5486E +05	66.07	74.07	58.62	1.66	66.07	55.56	75.86	1.6377
Feature	r	p (SC)	MI	p (MW)	D > C	median (C)	std (C)	median (D)	std (D)	ACC (LDA)	**SEN (LDA)**	**SPE (LDA)**	**TSS (LDA)**	ACC (RF)	SEN (RF)	SPE (RF)	TSS (RF)
ALT (idr)	0.369	0.0051	0.3351	0.0064	30.00	1.3125E +07	5.3131E +06	1.8750E +07	9.1506E +06	75.00	**74.07**	**75.86**	**1.81**	58.93	66.67	51.72	1.5462
PRESS (3q)	0.3936	0.0027	0.4317	0.0036	17.81	180	41	219	33.68	73.21	**77.78**	**68.97**	**1.78**	55.36	62.96	48.28	1.4892
PRESS (80p)	0.3926	0.0028	0.4703	0.0037	17.65	182	39.72	221	32.65	73.21	**77.78**	**68.97**	**1.78**	67.86	66.67	68.97	1.6995
PRESS (70p)	0.3925	0.0028	0.3902	0.0037	17.89	179	41.94	218	34.4	73.21	**77.78**	**68.97**	**1.78**	64.29	62.96	65.52	1.6425
PRESS (95p)	0.3529	0.0076	0.4426	0.0091	11.45	201	31.23	227	28.51	73.21	**77.78**	**68.97**	**1.78**	55.36	66.67	44.83	1.4749
Feature	r	p (SC)	MI	p (MW)	D > C	median (C)	std (C)	median (D)	std (D)	ACC (LDA)	SEN (LDA)	SPE (LDA)	TSS (LDA)	ACC (RF)	**SEN (RF)**	**SPE (RF)**	**TSS (RF)**
ALT (60p)	−0.1339	0.3251	0.3702	0.3247	−14.29	2.2500E +07	1.8863E +07	1.9688E +07	8.9280E +06	44.64	59.26	31.03	1.30	82.14	**81.48**	**82.76**	**1.8961**
ALT (1p)	−0.1791	0.1865	0.5793	0.1867	−116.67	1.2188E +07	1.9513E +07	5.6250E +06	1.1679E +07	50.00	44.44	55.17	1.40	82.14	**77.78**	**86.21**	**1.8891**
ON: H-ZLC	−0.0785	0.5653	0.5206	0.566	−0.64	0.3604	0.0735	0.3581	0.0738	50.00	59.26	41.38	1.40	78.57	**74.07**	**82.76**	**1.8464**
P: ZLC	−0.0785	0.5653	0.5206	0.566	−0.64	0.3604	0.0735	0.3581	0.0738	50.00	59.26	41.38	1.40	78.57	**74.07**	**82.76**	**1.8464**
ON: V-ZLC	−0.0785	0.5653	0.5206	0.566	−0.64	0.3604	0.0735	0.3581	0.0738	50.00	59.26	41.38	1.40	78.57	**74.07**	**82.76**	**1.8464**

[1] D—children with dysgraphia, C—children without dysgraphia, r—Spearman's rank correlation coefficient, p (SC)—significance level of Spearman's rank correlation, MI—mutual information, p (MW)—significance level of Mann-Whitney U test, D > C—relative difference of medians of D and C group [%], ACC—accuracy [%], SEN—sensitivity [%], SPE—specificity [%], TSS—trade-off between sensitivity and specificity, LDA—linear discriminant analysis, RF random forests, ON: *—on surface movement, P: *—pressure information, *: H-*—horizontal movement, *: V-*—vertical movement, ALT—altitude, PRESS—pressure, N_JERK—normalized jerk, FA_DFA—fluctuation amplitudes of detrended fluctuation analysis, PE_LLE—prediction error of largest Lyapunov exponent, ZLC—Ziv–Lempel complexity, cv—coefficient of variation, id—index of dispersion, xp—xth percentile, med—median, mean x%—mean excluding x% of outliers, xq—xth quartile, off—offset of linear regression, mead—median absolute deviation, idr—interdecile range.

new sample by the majority of votes from the base learners. RF has got two tuning parameters: the number of the features over which to search to construct each branch of each tree and the number of trees in the classifier. We used the settings, where the number of the features over which to search to construct each branch is equal to the square root of the number of input features and the number of trees is decided to be equal to 500. We chose this particular algorithm for our classification setup considering its ability to deal reasonably well with high-dimensional and highly correlated data with complex interactions and its ability to rank the importance of used variables.

4) Classification and Regression: The fourth and last objective of this study is to propose a system of dysgraphia rating based on estimation of HPSQ total score. To estimate this score, we employed the CART. This is a nonparametric learning algorithm used either for classification of regression. The output of this algorithm is a single decision tree, which is basically a collection of rules based on the training data. The rules are designed to get the best split to model the dependent variable. A recursive procedure, which continues in constructing new rules for each split, is used to acquire a tree-like structure of rules. When the algorithm detects no further gain by splitting, then the splitting process is stopped. After this step, the technique called pruning is applied to reduce the size and complexity of the decision tree and, hence, improves the predictive capability of the final model.

5) Validation: Cross validation is a technique used for model validation. Its primary purpose is assessing how the results of an analysis will generalize to a previously unseen data. A lot of different cross-validation setups exist. In this work, we decided to use leave-one-out cross validation. In this approach, the dataset is divided into two groups. The first one is used as a training dataset, and its size is $N-1$, where N is the number of samples contained in the whole dataset. The other group is used as a testing dataset, and its size is 1.

III. Results

The results of the preliminary feature analysis in terms of discrimination power can be found in Table I. We considered five evaluation scenarios, while in each scenario, we identified first five most significant features (scenarios 1 and 2 are fused in Table I, because the same significant features were selected). Regarding the Spearman's rank correlation, the most significant feature is ALT (cv) ($r = 0.5074, p = 0.0001$). Generally, the features based on altitude, pressure, and jerk (on-surface) correlated well with binary state (C/D). Next, similarly to the correlation measure, the Mann–Whitney U test identified ALT (cv) ($p = 0.0002$) as the feature with the highest discrimination power. Density estimation plots of top three features with the highest discrimination power according to the Mann–Whitney U test can be found in Fig. 5. The most significant feature according to the MI is ON: V-FA_DFA (mean 10%), where we observed MI $= 0.6442$.

The first scenario that evaluated the features according to performance of classifier employed classification by LDA, where the best results were observed in the case of altitude feature ALT (idr) (ACC $=$ 75.00%, SEN $=$ 74.07%, SPE $=$ 75.86%, and TSS $=$ 1.81). The last and probably the most interesting scenario included classification by RFs, where, using just 60th percentile of altitude, we are able to identify dysgraphia with 81.48% sensitivity (SPE $=$ 82.76%, TSS $=$ 1.8961).

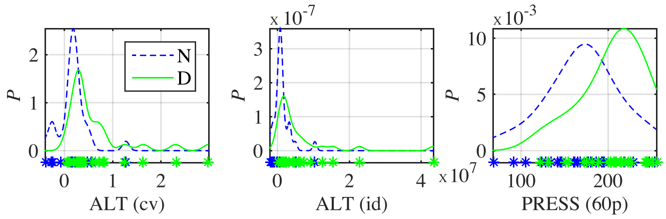


Fig. 5. Density estimation plots (computed using kernel density estimation with Gaussian kernels) of top three features with the highest discrimination power according to the Mann–Whitney U test (C—children without dysgraphia, D—children with dysgraphia).

TABLE II
DYSGRAPHIA DISCRIMINATION BASED ON FEATURE SELECTION AND INTRAWRITER NORMALIZATION

Normalized by	ACC	SEN	SPE	TSS	No.
ON: V-JERK (10p)	96.43	96.30	96.55	1.9956	7
ON: H-JERK (rpmin)	94.64	96.30	93.10	1.9896	6
ON: V-JERK (1p)	94.64	92.59	96.55	1.9886	5
no normalization	92.86	92.59	93.10	1.9826	4
AIR: V-JERK (10p)	92.86	92.59	93.10	1.9826	5
ON: JERK (1q)	92.86	88.89	96.55	1.9771	5
AIR: JERK (1q)	91.07	92.59	89.66	1.9727	6
AIR: H-JERK (10p)	91.07	88.89	93.10	1.9711	5
AIR: H-JERK (1q)	91.07	88.89	93.10	1.9711	7
AIR: V-JERK (1q)	91.07	88.89	93.10	1.9711	5

[1] ACC—accuracy [%], SEN—sensitivity [%], SPE—specificity [%], TSS—trade-off between sensitivity and specificity, No.—number of selected features, ON: *—on surface movement, AIR: *—in-air movement, *: H-*—horizontal movement, *: V-*—vertical movement, xp—xth percentile, xq—xth quartile, rpmin—relative position of minimum.

Although, using just a single-feature classification, we are able to exceed 80% limit in terms of sensitivity and specificity, we expected better results when employing feature selection and simple intrawriter normalization. These results can be found in Table II (they are related to the second and third objective of this work: intrawriter normalization and dysgraphia discrimination). As can be seen, normalization by on-surface features derived from jerk can increase ACC, SEN, and SPE by approximately 4%. When considering classification without normalization, we reached ACC $=$ 92.86%, SEN $=$ 92.59%, SPE $=$ 93.10%, and TSS $=$ 1.9826 (four features selected by SFFS), while in the case of normalization by ON: V-JERK (10p), we reached ACC $=$ 96.43%, SEN $=$ 96.30%, SPE $=$ 96.55%, and TSS $=$ 1.9956. In this case, the SFFS selected seven features based on altitude, azimuth, velocity, acceleration, jerk, and fluctuation analysis: ALT (1q), ON: H-FA_DFA (skew2), AIR: V-VEL (pmin), AZIM (4m), ON: H-ACC (80p), ALT (99p), and ON: H-JERK (kurt).

In the next step, we investigated features' ability to rate dysgraphia and estimate HPSQ total score (objective 4). The results of preliminary analysis can be found in Table III. Performance was evaluated in terms of correlation and MI with the HPSQ, MAE, and estimation errors evaluating CART. The most significant correlation ($r = -0.4546, p = 0.0004$) with the HPSQ total score was identified in feature derived from pressure

TABLE III
PRELIMINARY FEATURE ANALYSIS IN TERMS OF HPSQ TOTAL SCORE ESTIMATION

Feature	r	p	MI	D > C	median (C)	std (C)	median (D)	std (D)	MAE	EE1	EE2
P: TKEO (off)	**−0.4546**	**0.0004**	2.0030	−76.54	−2.0912	24.1664	−8.9153	12.6800	8.26	26.64	20.65
ON: H-JERK (mean)	**−0.4107**	**0.0017**	2.6343	2577.31	1.4256E-10	3.9512E-10	−5.7546E-12	3.5646E-10	10.20	32.92	25.51
ALT (cv)	**0.4046**	**0.0020**	1.9957	46.97	0.1823	0.3020	0.3437	0.6865	9.78	31.55	24.45
ON: H-JERK (skew1)	**−0.3848**	**0.0034**	2.8705	1789.82	0.0012	0.0036	−0.0001	0.0027	8.18	26.4	20.46
ON: H-JERK (skew2)	**−0.3848**	**0.0034**	2.8560	1789.82	0.0012	0.0036	−0.0001	0.0028	8.27	26.69	20.69
Feature	r	p	**MI**	D > C	median (C)	std (C)	median (D)	std (D)	MAE	EE1	EE2
ON: V-PE_LLE (3q)	0.2098	0.1207	**3.2349**	29.81	0.1503	0.0635	0.2141	0.0674	10.02	32.33	25.06
ON: V-PE_LLE (80p)	0.2012	0.137	**3.2057**	29.57	0.1509	0.0650	0.2143	0.0685	10.43	33.64	26.07
ON: V-PE_LLE (1q)	0.1612	0.2351	**3.1875**	23.92	0.1307	0.0569	0.1718	0.0562	9.04	29.15	22.59
ON: V-PE_LLE (mean 50%)	0.2130	0.1150	**3.1857**	25.69	0.1444	0.0594	0.1943	0.0608	10.22	32.98	25.56
ON: V-PE_LLE (mean 40%)	0.2099	0.1204	**3.1857**	25.89	0.1442	0.0595	0.1946	0.0606	10.22	32.98	25.56
Feature	r	p	MI	D > C	median (C)	std (C)	median (D)	std (D)	**MAE**	**EE1**	**EE2**
ALT (1p)	−0.1291	0.3429	2.8151	−116.67	1.2188E+07	1.9513E+07	5.6250E+06	1.1679E+07	**6.24**	**20.14**	**15.61**
AIR: V-JERK (pmax)	−0.0756	0.5799	1.3721	24.22	1.2200E+02	2.7765E+02	1.6100E+02	6.1075E+02	**6.62**	**21.35**	**16.55**
ON: V-TKEO (1ent)	−0.0071	0.9586	2.6118	−0.24	2.4878	0.4260	2.4818	0.5476	**6.93**	**22.36**	**17.33**
ALT (5p)	−0.1629	0.2303	2.8643	−70.00	1.5938E+07	1.9028E+07	9.3750E+06	1.1344E+07	**7.02**	**22.66**	**17.56**
P: TKEO (95p)	−0.0951	0.4857	2.2813	−13.19	2.8070E+02	1.5074E+02	2.4800E+02	8.0905E+01	**7.04**	**22.72**	**17.61**

[1] D—children with dysgraphia, C—children without dysgraphia, r—Spearman's rank correlation coefficient, p—significance level of Spearman's rank correlation, MI—mutual information, D > C—relative difference of medians of D and C group [%], MAE—mean absolute error, EE1—equal error rate of type 1 [%], EE2—equal error rate of type 2 [%], ON: *—on surface movement, AIR: *—in-air movement, P: *—pressure information, *: H-*—horizontal movement, *: V-*—vertical movement, ALT—altitude, TKEO—Teager-Kaiser energy operator, PE_LLE—prediction error of largest Lyapunov exponent, xp—xth percentile, xq—xth quartile, off—offset of linear regression, cv—coefficient of variation, skew1—Pearson's first skewness coefficient, skew2—Pearson's second skewness coefficient, mean x%—mean excluding x% of outliers, pmax—position of maximum, 1ent—first-order Shannon entropy.

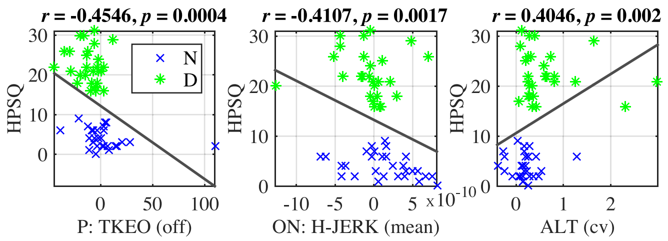


Fig. 6. Correlation graphs of top three features with resulting lowest significance level according to Spearman's rank correlation (C—children without dysgraphia, D—children with dysgraphia).

P: TKEO (off). The resulting correlation graph of this parameter and other significant features can be found in Fig. 6.

The highest MI was observed between HPSQ and features based on largest Lyapunov exponent (on-surface). Finally, using a single-feature approach (again feature derived from altitude), we are able to estimate the HPSQ total score with MAE = 6.24, EE1 = 20.14, and EE2 = 15.61. However, it is obvious that these results are poor, and the estimation can be improved by feature selection and intrawriter normalization (see Table IV, objective 2 and 4: intrawriter normalization and dysgraphia rating). In the best case [normalization by ON: V-JERK (rpmin)], we estimated the HPSQ with EE1 = 10.05% (MAE = 3.12), which is by 3.48% better than standard (without normalization) approach and by 10.09% better than the single-feature approach. In this case, the SFFS selected 13 features based on pressure, altitude, velocity, acceleration, jerk, fluctuation analysis, and largest Lyapunov exponent: AIR: ACC (2ren), ON: H-PE_LLE (skew2), ALT (40p), AIR: VEL (40p), AIR: V-VEL (5m), ON: V-N_JERK (3m), AIR: H-VEL (range), AIR: JERK (ripr), ON: H-FA_DFA (gsd), AIR: V-N_JERK (max), AIR: V-N_JERK (rr), ON: H-N_JERK (1q), and P: TKEO (60p).

TABLE IV
HPSQ TOTAL SCORE ESTIMATION BASED ON FEATURE SELECTION AND INTRAWRITER NORMALIZATION

Normalized by	MAE	EE1	EE2	No.
ON: V-JERK (rpmin)	3.12	10.05	7.79	13
PRESS (rr)	3.30	10.65	8.25	13
AIR: H-VEL (3q)	3.95	12.73	9.87	11
ON: JERK (mad)	4.08	13.17	10.21	13
ON: H-JERK (2ren)	4.09	13.19	10.22	7
no normalization	4.20	13.53	10.49	10
AIR: V-VEL (idr)	4.43	14.28	11.07	4
ON: JERK (1cc)	4.74	15.28	11.84	7
P: FA_DFA (40p)	5.15	16.62	12.88	5
PRESS (80p)	5.28	17.03	13.20	4

[1] MAE—mean absolute error, EE1—equal error rate of type 1 [%], EE2—equal error rate of type 2 [%], No.—number of selected features, ON: *—on surface movement, AIR: *—in-air movement, P: *—pressure information, *: H-*—horizontal movement, *: V-*—vertical movement, PRESS—pressure, VEL—velocity, FA_DFA—fluctuation amplitudes of detrended fluctuation analysis, xp—xth percentile, xq—xth quartile, rpmin—relative position of minimum, rr—relative range, mad—median absolute deviation, 2ren—second-order Rényi entropy, idr—interdecile range, 1cc—first correlation coefficient.

IV. DISCUSSION

The previous findings indicated lower values of sensitivity and specificity of methods designated for handwriting deficiency evaluation. For example, although the developmental test of visual-motor integration (VMI) [64] was not designated for this purpose, it was identified as the most commonly used tool to assess handwriting difficulties by occupational therapists in Canada [65]. When the VMI's sensitivity was analyzed, it was found that the VMI correctly identified only a small number of the children with handwriting dysfunction (34% sensitivity) [65]. When focusing on handwriting evaluation scales, the values of 71% sensitivity and 75% specificity were reported for the evaluation tool of children handwriting cursive (ETCH-C) scale among children aged five to six years [66]. Similar values of discrimination were found in further studies focusing on validating handwriting evaluation scales [67], [68]. When applying a computerized system for differentiation between healthy people and patients with varied diseases, as for example depression, based on their handwriting, sensitivity values raised to 82% [69]. In light of such previous findings, achieving 96% of sensitivity and specificity in the present study is novelty.

We observed that most of the features useful for diagnosis are based on pressure (pressure was by 17% higher in the case of children with dysgraphia; median of pressure is 166 ± 43.66 and 207 ± 38.58 for C and D children, respectively; $p = 0.0032$ identified by the Mann–Whitney U test). Pressure was previously found as a sensitive measure of the individual's performance and served for individual's identification and verification [70], [71]. More specifically, this finding is in line with the previous studies, where pressure measure differentiated between children with developmental coordination disorders (DCD) who have handwriting difficulties and controls with typical development [72]. As investigated by Kao, Mak, and Lam, relationship between pressure and complexity of task exists [73]; therefore, we hypothesize that more complex task would further increase the difference between both groups.

Other interesting finding is that coefficient of variation of altitude has one of the best discrimination power among features (this parameter was by 47% higher in the case of children with dysgraphia; std of altitude/tilt is 0.1823 ± 0.3020 and 0.3437 ± 0.6865 for C and D children, respectively; and $p = 0.0002$ identified by the Mann–Whitney U test). This probably means that children with dysgraphia are not able to hold pen in a stable position, and therefore, the pen tilt is changing a lot in time. In fact, the tilt measure reflects the ability of controlling a tool (pen) in order to produce a brain–hand action. Such a control requires a proficient hand movement while taking into account the required form, space, time, and amount of pressure, and thus, it is sensitive to the individual's writing proficiency.

As single features, those based on pressure and tilt correlate with the HPSQ better than the features based on in-air movement (see Table III). However, as shown in the last paragraph of Section III, all these three categories bring specific information about dysgraphia, and their fusion into multidimensional model provides the best estimation of HPSQ. Therefore, it is more advantageous not to consider these features separately, but rather in combination.

Besides pressure and tilt-based features, we also observed that nonlinear dynamic ones provide good discrimination or HPSQ estimation power. Specifically, this is the case of ZLC. This measure has been widely used in biomedical signal analysis as a metric to estimate the complexity of discrete-time physiological signals. Aboy, Hornero, Abasolo, and Alvarez showed that ZLC is useful as a scalar metric to estimate the bandwidth of random processes, quasiperiodic or periodic signals [39]. In other words, ZLC can be used to quantify regularity in handwriting. Theoretically, fluent and well-coordinated movement results in lower complexity of handwriting; therefore, D children with deficient fine motor skills, poor dexterity, poor muscle tone, or unspecified motor clumsiness manifest higher values of ZLC.

Developing methods for intraindividual variability detection with objective measures of real activity performance goes in line with the up-to-date trend to find the unique performance features of each individual. The framework presented by the World Health Organization extended the concept of health, while the way individual performs actual daily activities (as handwriting) and participate in life domains is relevant for his health/illness status [74], [75]. Furthermore, the worldwide movement toward personalized medicine [76] leading to the need to get insight about the individual's performance features for adapting personally the appropriate intervention method. For example, individual differences were found among children with DCD, attention deficit hyperactive disorders, and autism who struggle with handwriting difficulties [77], [78]. Application of the method described in the present study may shed light on each child's performance features and adapted the most appropriate intervention method for his specific needs.

We proposed a system that is able to rate developmental dysgraphia with approximately 10% HPSQ estimation error, and we believe it is possible to further decrease it, e.g., using some additional handwriting features (e.g., the features based on empirical mode decomposition [30], [31] or another machine learning methods (e.g., quantile regression [79]). However, the HPSQ is given subjectively according to examination of trained person; therefore, it can happen that two trained experts rate one writer with inconsistent results. It would be interesting to make some tests, where ten experts rate one writer and compare interexpert differences with the estimation error of automatic rating.

V. Conclusion

This paper has dealt with an automatic rating of developmental dysgraphia in children population using the state-of-the-art handwriting parameterization techniques and a simple intrawriter normalization approach. The work is innovative from several points of view. First of all, it employs complex parameterization based on ten kinematic measures, 34 nonlinear dynamic, and other seven features. Moreover, we analyzed both, on-surface and in-air movements. Second, it is the first work that introduces intrawriter normalization applied in the field of dysgraphia analysis. Finally, to our best knowledge, it is the first work that introduces automatic rating of dysgraphia based on HPSQ total score estimation.

In this paper, we analyzed sequential writing of 27 children with dysgraphia and 27 age-matched controls who had an experience of two years of writing in school. We achieved all goals of this work.

1) We introduced parameterization based on 51 features and observed that those based on altitude/tilt and pressure discriminate well D and C children.
2) We proposed a simple intrawriter normalization method based on subtraction and proved that it can increase discrimination accuracy by 4% and decrease HPSQ score estimation error by 3.48%.
3) We proposed a system of automatic dysgraphia discrimination with 96% sensitivity and specificity.

4) We proposed a system that is able to rate developmental dysgraphia and estimate HPSQ total score with 10% estimation error.

Our findings may be used clinically for discriminating dysgraphia among school-aged children who at least have two years of writing experience. However, our dataset consists of 54 participants; therefore, further research works should address the same topic and verify the proposed concept of dysgraphia rating. Another possible limitation of this work is related to the analysis of in-air movement. The vertical dimension (z) is not captured in our protocol. The digitizer only acquires orthogonal projection of the pen coordinates when the tip of the pen is no more than approximately 1 cm above the tablet's surface. Therefore, it can happen that when the writer elevates pen more than this limit, the time series of coordinates can lose continuity between two specific samples. Some features, especially kinematic ones, can be distorted by this limitation. In order to suppress an impact of this limitation on our system, during parameter postprocessing, we used some statistics less sensitive to outliers. Moreover, the nonrelevant parameters were consequently filtered during a two-step feature selection based on mRMR and SFFS.

References

[1] V. Connelly, J. E. Dockrell, and J. Barnett, "The slow handwriting of undergraduate students constrains overall performance in exam essays," *Educ. Psychol.*, vol. 25, no. 1, pp. 99–107, 2005.

[2] A. W. V. Gemmert and H.-L. Teulings, "Advances in graphonomics: Studies on fine motor control, its development and disorders," *Human Movement Sci.*, vol. 25, no. 45, pp. 447–453, 2006.

[3] K. P. Feder and A. Majnemer, "Handwriting development, competency, and intervention," *Develop. Med. Child Neurol.*, vol. 49, no. 4, pp. 312–317, 2007.

[4] K. McHale and S. A. Cermak, "Fine motor activities in elementary school: Preliminary findings and provisional implications for children with fine motor problems," *Amer. J. Occupat. Therapy*, vol. 46, no. 10, pp. 898–903, 1992.

[5] M. H. Tseng and S. M. K. Chow, "Perceptual-motor function of school-age children with slow handwriting speed," *Amer. J. Occupat. Therapy*, vol. 54, no. 1, pp. 83–88, 2000.

[6] S. Graham, "Want to improve children's writing? Don't neglect their handwriting," *Amer. Educ.*, vol. 33, no. 4, pp. 20–40, 2010.

[7] S. T. Peverly, J. K. Garner, and P. C. Vekaria, "Both handwriting speed and selective attention are important to lecture note-taking," *Reading Writing*, vol. 27, no. 1, pp. 1–30, 2014.

[8] J. Medwell and D. Wray, "Handwriting automaticity: The search for performance thresholds," *Lang. Educ.*, vol. 28, no. 1, pp. 34–51, 2014.

[9] S. Rosenblum, P. Weiss, and S. Parush, "Product and process evaluation of handwriting difficulties," *Educ. Psychol. Rev.*, vol. 15, no. 1, pp. 41–81, 2003.

[10] R. P. Erhardt and V. Meade, "Improving handwriting without teaching handwriting: The consultative clinical reasoning process," *Aust. Occupat. Therapy J.*, vol. 52, no. 3, pp. 199–210, 2005.

[11] S. Rosenblum and L. Gafni-Lachter, "Handwriting proficiency screening questionnaire for children (HPSQ-C): Development, reliability, and validity," *Amer. J. Occupat. Therapy*, vol. 69, no. 3, pp. 6903220030, 2015.

[12] A. O'Hare, "Hands up for handwriting," *Develop. Med. Child Neurol.*, vol. 46, no. 10, pp. 651–651, 2004.

[13] A. Kushki, H. Schwellnus, F. Ilyas, and T. Chau, "Changes in kinetics and kinematics of handwriting during a prolonged writing task in children with and without dysgraphia," *Res. Develop. Disabilities*, vol. 32, no. 3, pp. 1058–1064, 2011.

[14] S. S. Lam, R. K. Au, H. W. Leung, and C. W. Li-Tsang, "Chinese handwriting performance of primary school children with dyslexia," *Res. Develop. Disabilities*, vol. 32, no. 5, pp. 1745–1756, 2011.

[15] C. T. Fuentes, S. H. Mostofsky, and A. J. Bastian, "Perceptual reasoning predicts handwriting impairments in adolescents with autism," *Neurology*, vol. 75, no. 20, pp. 1825–1829, 2010.

[16] S. T. Peverly, P. C. Vekaria, L. A. Reddington, J. F. Sumowski, K. R. Johnson, and C. M. Ramsay, "The relationship of handwriting speed, working memory, language comprehension and outlines to lecture note-taking and test-taking among college students," *Appl. Cognitive Psychol.*, vol. 27, no. 1, pp. 115–126, 2013.

[17] M. R. I. Martins, J. A. Bastos, A. T. Cecato, M. de Lourdes Souza Araujo, R. R. Magro, and V. Alaminos, "Screening for motor dysgraphia in public schools," *J. Pediatria*, vol. 89, no. 1, pp. 70–74, 2013.

[18] S. Rosenblum, "Development, reliability, and validity of the handwriting proficiency screening questionnaire (HPSQ)," *Amer. J. Occupat. Therapy*, vol. 62, no. 3, pp. 298–307, 2008.

[19] S. Rosenblum, A. Y. Dvorkin, and P. L. Weiss, "Automatic segmentation as a tool for examining the handwriting process of children with dysgraphic and proficient handwriting," *Human Movement Sci.*, vol. 25, no. 45, pp. 608–621, 2006.

[20] S. Rosenblum, S. Goldstand, and S. Parush, "Relationships among biomechanical ergonomic factors, handwriting product quality, handwriting efficiency, and computerized handwriting process measures in children with and without handwriting difficulties," *Amer. J. Occupat. Therapy*, vol. 60, no. 1, pp. 28–39, 2006.

[21] S. Rosenblum, S. Parush, and P. Weiss, "The in-air phenomenon: Temporal and spatial correlates of the handwriting process," *Perceptual Motor Skills*, vol. 96, pp. 933–954, 2003.

[22] S. Rosenblum and N. Regev, "Timing abilities among children with developmental coordination disorders (DCD) in comparison to children with typical development," *Res. Dev. Disabil.*, vol. 34, no. 1, pp. 218–227, 2013.

[23] S. Rosenblum, "Do motor ability and handwriting kinematic measures predict organizational ability among children with Developmental Coordination Disorders?" *Human Movement Sci.*, vol. 43, pp. 201–215, 2015.

[24] B. J. Rhodes, D. Bullock, W. B. Verwey, B. B. Averbeck, and M. P. Page, "Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives," *Human Movement Sci.*, vol. 23, no. 5, pp. 699–746, 2004.

[25] S. Rosenblum, S. Parush, and P. L. Weiss, "Computerized temporal handwriting characteristics of proficient and non-proficient handwriters," *Amer. J. Occupat. Therapy*, vol. 57, no. 2, pp. 129–138, 2003.

[26] H. Teulings and L. Schomaker, "Invariant properties between stroke features in handwriting," *Acta Psychol.*, vol. 82, no. 1–3, pp. 69–88, 1993.

[27] A. Lelivelt, R. Meulenbroek, and A. Thomassen, *Mapping Abstract Main Axes in Handwriting on Hand and Finger Joints*. Amsterdam, The Netherlands: IOS Press, 1996, pp. 29–40.

[28] R. Plamondon and F. Maarse, "An evaluation of motor models of handwriting," *IEEE Trans. Syst., Man Cybern.*, vol. 19, no. 5, pp. 1060–1072, Sep./Oct. 1989.

[29] G. Luria, A. Kahana, and S. Rosenblum, "Detection of deception via handwriting behaviors using a computerized tool: Toward an evaluation of malingering," *Cognitive Comput.*, vol. 6, no. 4, pp. 849–855, 2014.

[30] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "Decision support framework for Parkinson's disease based on novel handwriting markers," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 3, pp. 508–516, May 2015.

[31] P. Drotar, J. Mekyska, Z. Smekal, I. Rektorova, L. Masarova, and M. Faundez-Zanuy, "Contribution of different handwriting modalities to differential diagnosis of Parkinson's disease," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, 2015, pp. 344–348.

[32] E. Sesa-Nogueras, M. Faundez-Zanuy, and J. Mekyska, "An information analysis of in-air and on-surface trajectories in online handwriting," *Cognitive Comput.*, vol. 4, no. 2, pp. 1–11, 2012.

[33] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy, "A new modality for quantitative evaluation of Parkinson's disease: In-air movement," in *Proc. IEEE 13th Int. Conf. Bioinformat. Bioeng*, 2013, pp. 1–4.

[34] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faúndez-Zanuy, "Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease," *Comput. Methods Programs Biomed.*, vol. 117, no. 3, pp. 405–411, 2014.

[35] P. Drotar, J. Mekyska, Z. Smekal, I. Rektorova, L. Masarova, and M. Faundez-Zanuy, "Prediction potential of different handwriting tasks for diagnosis of Parkinson's," in *Proc. E-Health Bioeng. Conf.*, 2013, pp. 1–4.

[36] J. H. Yan, S. Rountree, P. Massman, R. S. Doody, and H. Li, "Alzheimer's disease and mild cognitive impairment deteriorate fine movement control," *J. Psychiatric Res.*, vol. 42, no. 14, pp. 1203–1212, 2008.

[37] J. Shao, J. K. Maccallum, Y. Zhang, A. Sprecher, and J. J. Jiang, "Acoustic analysis of the tremulous voice: Assessing the utility of the correlation dimension and perturbation parameters," *J. Commun. Disorders*, vol. 43, pp. 35–44, 2010.
[38] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, "A comparison of waveform fractal dimension algorithms," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 48, no. 2, pp. 177–183, Feb. 2001.
[39] M. Aboy, R. Hornero, D. Abasolo, and D. Alvarez, "Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2282–2288, Nov. 2006.
[40] J. R. O. Arroyave *et al.*, "Automatic selection of acoustic and non-linear dynamic features in voice signals for hypernasality detection," in *Proc. 12th Annu. Conf. Int. Speech Commun. Assoc.*, 2011, pp. 529–532.
[41] M. Faundez-Zanuy, "N-dimensional nonlinear prediction with MLP," in *Proc. XI Eur. Signal Process. Conf.*, 2002, pp. 1–4.
[42] A. W. Jayawardena, P. Xu, and W. K. Li, "Modified correlation entropy estimation for a noisy chaotic time series," *Chaos*, vol. 20, no. 2, pp. 1–11, 2010.
[43] P. Henriquez, J. Alonso, M. Ferrer, C. Travieso, J. Godino-Llorente, and F. Diaz-de Maria, "Characterization of healthy and pathological voice through measures based on nonlinear dynamics," *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, no. 6, pp. 1186–1195, Aug. 2009.
[44] M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, "Permutation entropy and its main biomedical and econophysics applications: A review," *Entropy*, vol. 14, no. 8, pp. 1553–1577, 2012.
[45] J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou, "The appropriate use of approximate entropy and sample entropy with short data sets," *Ann. Biomed. Eng.*, vol. 41, no. 2, pp. 349–365, 2013.
[46] M. A. Little, P. E. Mcsharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomed. Eng. Online*, vol. 6, p. 23, 2007.
[47] J. Mekyska *et al.*, "Robust and complex approach of pathological speech signal analysis," *Neurocomputing*, vol. 167, pp. 94–111, 2015.
[48] G. Vaziri, F. Almasganj, and R. Behroozmand, "Pathological assessment of patients' speech signals using nonlinear dynamical analysis," *Comput. Biol. Med.*, vol. 40, no. 1, pp. 54–63, 2010.
[49] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *J. R. Soc. Interface*, vol. 8, no. 59, pp. 842–855, 2010.
[50] M. Gonzalez-Izal *et al.*, "sEMG wavelet-based indices predicts muscle power loss during dynamic contractions," *J. Electromyogr. Kinesiol.*, vol. 20, no. 6, pp. 1097–1106, 2010.
[51] D. Dimitriadis, A. Potamianos, and P. Maragos, "A comparison of the squared energy and Teager-Kaiser operators for short-term energy estimation in additive noise," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2569–2581, Jul. 2009.
[52] J. A. Fiz *et al.*, "Short term oxygen therapy effects in hypoxemic patients measured by drawing analysis," *Comput. Methods Programs Biomed.*, vol. 118, no. 3, pp. 330–336, 2015.
[53] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," in *Proc. Digital Signal Process.*, vol. 10, pp. 19–41, 2000.
[54] A. E. Rosenberg, J. DeLong, C. H. Lee, B. H. Juang, and F. K. Soong, "The use of cohort normalized scores for speaker verification," in *Proc. 2nd Int. Conf. Spoken Language Process.*, 1992, pp. 13–16.
[55] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 5th ed. London, U.K.: Chapman & Hall/CRC Press, 2011.
[56] J. Mekyska *et al.*, "Assessing progress of Parkinson's disease using acoustic analysis of phonation," in *Proc. 4th Int. Work Conf. Bioinspired Intell.*, 2015, pp. 111–118.
[57] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction: Foundations and Applications*, I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Eds. Berlin, Germany: Springer-Verlag, 2006.
[58] H. Deng and G. Runger, "Feature selection via regularized trees," in *Proc. Int. Joint Conf. Neural Netw.*, 2012, pp. 1–8.
[59] N. Sanchez-Marono, A. Alonso-Betanzos, and M. Tombilla-Sanroman, "Filter methods for feature selection—A comparative study," in *Proc. 8th Int. Conf. Intell. Data Eng. Autom. Learn.*, 2007, pp. 178–187.
[60] A. Tsanas, M. Little, and P. McSharry, "A methodology for the analysis of medical data," in *Handbook of Systems and Complexity in Health*, J. P. Sturmberg and C. M. Martin, Eds. New York, NY, USA: Springer, 2013, pp. 113–125.
[61] Z. Smekal, J. Mekyska, Z. Galaz, Z. Mzourek, I. Rektorova, and M. Faundez-Zanuy, "Analysis of phonation in patients with Parkinson's disease using empirical mode decomposition," in *Proc. Int. Symp. Signals, Circuits Syst.*, 2015, pp. 1–4.
[62] T. Hastie, R. Tibshirani, and J. Friedman, *Feature Extraction: Foundations and Applications*. New York, NY, USA: Springer, 2009.
[63] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[64] K. E. Beery, *The BeeryBuktenica Developmental Test of Visualmotor Integration*, 4th ed. Parsippany, NJ, USA: Modern Curriculum Press, 1997.
[65] T.-A. Goyen and S. Duff, "Discriminant validity of the developmental test of visual-motor integration in relation to children with handwriting dysfunction," *Aust. Occupat. Therapy J.*, vol. 52, no. 2, pp. 109–115, 2005.
[66] S. Duff and T.-A. Goyen, "Reliability and validity of the evaluation tool of children's handwriting cursive (ETCHC) using the general scoring criteria," *Amer. J. Occupat. Therapy*, vol. 64, no. 1, pp. 37–46, 2010.
[67] M. Brossard-Racine, B. Mazer, M. Julien, and A. Majnemer, "Validating the use of the evaluation tool of children's handwriting-manuscript to identify handwriting difficulties and detect change in school-age children," *Amer. J. Occupat. Therapy*, vol. 66, no. 4, pp. 414–421, 2012.
[68] S. M. Koziatek and N. J. Powell, "A validity study of the evaluation tool of children's handwriting-cursive," *Amer. J. Occupat. Therapy*, vol. 56, no. 4, pp. 446–453, 2002.
[69] S. Rosenblum, P. Werner, T. Dekel, I. Gurevitz, and J. Heinik, "Handwriting process variables among elderly people with mild major depressive disorder: a preliminary study," *Aging Clin. Exp. Res.*, vol. 22, no. 2, pp. 141–147, 2010.
[70] C. Hook, J. Kempf, and G. Scharfenberg, "A novel digitizing pen for the analysis of pen pressure and inclination in handwriting biometrics," in *Proc. Int. Workshop Biometric Authentication*, Prague, Czech Republic, May 15, 2004, pp. 283–294.
[71] J. Wann and I. Nimmo-Smith, "The control of pen pressure in handwriting: A subtle point," *Human Movement Sci.*, vol. 10, nos. 2/3, pp. 223–246, 1991.
[72] S. Rosenblum and M. Livneh-Zirinski, "Handwriting process and product characteristics of children diagnosed with developmental coordination disorder," *Human Movement Sci.*, vol. 27, no. 2, pp. 200–214, 2008.
[73] H. Kao, P. Mak, and P. Lam, "Handwriting pressure: Effects of task complexity, control mode and orthographic differences," in *Graphonomics: Contemporary Research in Handwriting*. Amsterdam, The Netherlands: North Holland, 1986, pp. 47–66.
[74] *International Classification of Functioning, Disability and Health (ICF)*. World Health Org., Geneva, Switzerland, 2001.
[75] G. M. Ronen and P. L. Rosenbaum, *Life Quality Outcomes in Children and Young People with Neurological and Developmental Conditions: Concepts, Evidence and Practice*. London, U.K.: Mac Keith Press, 2013.
[76] C. T. Scott, T. Caulfield, E. Borgelt, and J. Illes, "Personal medicine—The new banking crisis," *Nature Biotechnol.*, vol. 30, pp. 141–147, 2012.
[77] R. Blank, B. Smits-Engelsman, H. Polatajko, and P. Wilson, "European academy for childhood disability (EACD): Recommendations on the definition, diagnosis and intervention of developmental coordination disorder (long version)," *Develop. Med. Child Neurol.*, vol. 54, no. 1, pp. 54–93, 2012.
[78] P. C. M. Molenaar, H. M. Huizenga, and J. R. Nesselroade, "The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory," in *Understanding Human Development*, U. M. Staudinger and U. Lindenberger, Eds. New York, NY, USA: Springer, 2003, pp. 339–360.
[79] R. Koenker, *Quantile Regression*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

Jiri Mekyska was born in Holesov, Czech Republic. He received the Ph.D. degree in telecommunications from the Brno University of Technology (BUT), Brno, Czech Republic, in 2014.

He is currently a Researcher and Head of the Brain Diseases Analysis Laboratory, and the Head of the Human–Machine Interaction Group of the Signal Processing Laboratory, Department of Telecommunications, BUT. His research interests include noninvasive neurodegenerative disorders analysis based on speech and handwriting processing. He is also involved with thermal image processing and information theory.

Marcos Faundez-Zanuy was born in Barcelona, Spain. He received the Ph.D. degree in telecommunications from the Polytechnic University of Catalunya, Barcelona, Spain in 1998.

He is currently a Full Professor with ESUP Tecnocampus Mataro and heads the Signal Processing Group there. Since 2009, he has been the Dean of the Escola Superior Politècnica (Polytechnic University of Catalunya until 2015, Pompeu Fabra University thereafter), Barcelona, where he has been the Head of Research at the Tecnocampus since 2010. He is the author of more than 50 papers indexed in the ISI Journal Citation Report, more than 100 conference papers, about ten books, and has been the Responsible of 10 national and European research projects. His research interests include biometrics applied to security and health.

Prof. Faundez-Zanuy is a Spanish liaison of EURASIP.

Zdenek Mzourek was born in Brno, Czech Republic. He received the M.Sc. degree in telecommunications from the Brno University of Technology (BUT), Brno, in 2014, where he is currently working toward the Ph.D. degree.

He is a Junior Researcher and a Member of the Brain Diseases Analysis Laboratory, Department of Telecommunications, BUT. His research interests include signal processing, machine learning, and data analysis.

Zoltan Galaz was born in Galanta, Slovak Republic. He received the M.Sc. degree in telecommunications from the Brno University of Technology (BUT), Brno, Czech Republic, in 2014, where he is currently working toward the Ph.D. degree.

He is a Junior Researcher and a Member of the Brain Disease Analysis Laboratory, Department of Telecommunications, BUT. His research interests include pathological speech signal and handwritten text analysis and its application in biomedicine.

Zdene Smekal (SM'06) received the doctoral degree from the Military Academy Brno, Brno, Czech Republic, in 1978.

Currently, he is a full-time Professor with the Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic. His research interests include discrete and digital processing of one- and multidimensional signals with application in digital signal processors, emotional speech and image analysis and recognition, and also biometrics system identification. He is an author of three monographs about digital signal processing and digital signal processors.

Prof. Smekal has been a Member of the IEEE Signal Processing Society and the IEEE Communications Society since 1998.

Sara Rosenblum received the Ph.D. degree in occupational therapy from the Hebrew University of Jerusalem, Jerusalem, Israel, in 2002.

Currently, she is an Associate Professor in occupational therapy and Head of the Laboratory for Complex Human Activity and Participation, University of Haifa, Haifa, Israel, with special interest in the characteristics of human daily function. She aims to gain better insight into interactions between varied body functions (e. g., cognitive, motor, sensory), activity performance, and participation abilities of people faced with functional deficits in everyday life. A main focus is placed on trying to understand the relationships between brain mechanisms and actual daily functions among varied populations along the life cycle.

A.23 Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease

Artificial Intelligence in Medicine 67 (2016) 39–46

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease

Peter Drotár [a], Jiří Mekyska [a], Irena Rektorová [b,*], Lucia Masarová [b], Zdeněk Smékal [a], Marcos Faundez-Zanuy [c]

[a] Department of Telecommunications, Brno University of Technology, Technická 12, 61200 Brno, Czech Republic
[b] First Department of Neurology, Faculty of Medicine, St. Anns University Hospital, Pekarska 664, 66591 Brno, Czech Republic
[c] Signal Processing Group, Tecnocampus, Escola Universitaria Politecnica de Mataro, Avda. Ernest Llunch 32, 08302 Mataro, Spain

ARTICLE INFO

Article history:
Received 17 February 2015
Received in revised form
30 December 2015
Accepted 13 January 2016

Keywords:
Decision support system
Support vector machine classifier
Handwriting database
Handwriting pressure
Parkinson's disease
PD dysgraphia

ABSTRACT

Objective: We present the PaHaW Parkinson's disease handwriting database, consisting of handwriting samples from Parkinson's disease (PD) patients and healthy controls. Our goal is to show that kinematic features and pressure features in handwriting can be used for the differential diagnosis of PD.
Methods and material: The database contains records from 37 PD patients and 38 healthy controls performing eight different handwriting tasks. The tasks include drawing an Archimedean spiral, repetitively writing orthographically simple syllables and words, and writing of a sentence. In addition to the conventional kinematic features related to the dynamics of handwriting, we investigated new pressure features based on the pressure exerted on the writing surface. To discriminate between PD patients and healthy subjects, three different classifiers were compared: K-nearest neighbors (K-NN), ensemble AdaBoost classifier, and support vector machines (SVM).
Results: For predicting PD based on kinematic and pressure features of handwriting, the best performing model was SVM with classification accuracy of $P_{acc} = 81.3\%$ (sensitivity $P_{sen} = 87.4\%$ and specificity of $P_{spe} = 80.9\%$). When evaluated separately, pressure features proved to be relevant for PD diagnosis, yielding $P_{acc} = 82.5\%$ compared to $P_{acc} = 75.4\%$ using kinematic features.
Conclusion: Experimental results showed that an analysis of kinematic and pressure features during handwriting can help assess subtle characteristics of handwriting and discriminate between PD patients and healthy controls.

1. Introduction

Parkinson's disease is a complex neurodegenerative disease that affects a large portion of the worldwide population [1]. With current prevalence rates, ranging from 10 to 800 people per 100,000, PD is one of the most common neurodegenerative disorders [2]. PD is a movement disorder characterized by resting tremor, rigidity, slowness of movement (bradykinesia), and loss of postural reflexes. The disturbances of motor control in PD involve processing of motor planning, motor programming, motor sequencing, movement initiation and movement execution [3].

There is currently no objective method for diagnosing PD. It can take months to get a reliable PD diagnosis, and symptoms need to be carefully monitored. Even then the probability of an inaccurate diagnosis is approximately 25% [4]. The diagnosis can be confirmed only by a pathological analysis at autopsy; this further highlights the complexity of the diagnosis. Decision support tools for accurate diagnosis would be beneficial for early diagnosis and for the development of treatment strategies for PD patients [5,6]. Identifying biomarkers is an important goal of the research on neurodegenerative diseases [7].

One typical hallmark of PD is disruption in the execution of practiced skills such as handwriting [8,9]. People with PD frequently have severe difficulties in coordinating of the components of a motor sequence movement. They tend to perform sequential movements in a more segmented fashion. Hesitations and pauses are often observed between the components of the sequence [10,11]. Continuous handwriting and similar motor tasks occur more slowly than in a healthy person. Some recent studies have suggested that handwriting can be used as a biomarker for diagnosing PD [12,13]. The reasoning behind this suggestion is that handwriting is no longer an automated process for PD patients and their handwriting depends on a visual closed loop [14].

* Corresponding author. Tel.: +420 543182639.
E-mail address: rektorova@fnusa.cz (I. Rektorová).

http://dx.doi.org/10.1016/j.artmed.2016.01.004

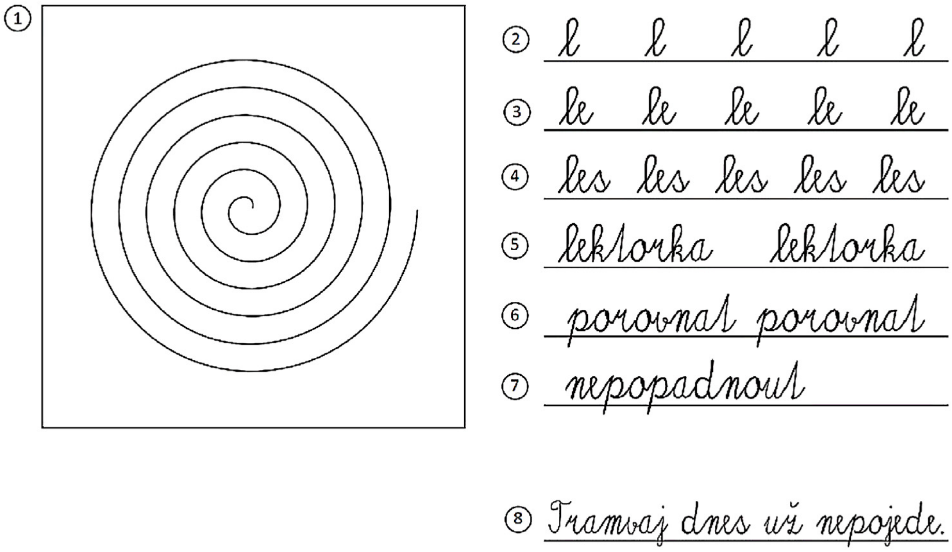


Fig. 1. Illustration of filled template (not actual handwriting samples).

Several handwriting tasks were proposed for use in analyzing the handwriting of PD patients and for obtaining insight into the motor disruption caused by PD. Probably the most popular handwriting exercise for tremor assessment is currently the Archimedean spiral. Spiral drawing has been frequently used for evaluating of the motor performance in various movement disorders, including PD [4,8,15–17]. Words containing one or more repetitions of the cursive letter "l" are the second-most common exercises in handwriting assessment [10,18]. In addition to these established tasks, we proposed new ones consisting of writing simple words and short sentences. The words used in these handwriting tasks were selected for their simple orthography and easy syntax.

It has been shown that the absolute positioning of the pen during handwriting is relevant for PD diagnosis, as are pen movements above the writing surface (when the pen does not leave the trajectory) [12,13,19]. Pressure exerted on the surface during handwriting plays a significant role too [13].

In this paper, we extend our previous work [12,20] by providing a more detailed analysis of the pressure modality of handwriting and by introducing novel pressure features. Moreover, we introduce the Parkinson's disease handwriting (*PaHaW*) database, which can be used for developing predictive models for PD diagnosis. The *PaHaW* database contains recorded in-air/on-surface trajectories and pressure, i.e. modalities that have been shown to be significant for PD classification. The results confirmed that handwriting is relevant in diagnosing and monitoring PD. We also compared three frequently used classifiers on the *PaHaW* database: SVM, Adaboost and K-NN.

We believe that the *PaHaW* database can encourage further research and provide additional information to other available databases related to PD such as Parkinson's disease speech datasets [21,22].

In the next section, the database of handwriting samples is introduced and described in detail. Section 3 presents our methods and obtained results. We provide a discussion and conclusions in the last section.

2. Parkinson's disease handwriting (PaHaW) database

We created a handwriting database of 37 PD patients (19 men) and 38 sex- and age-matched healthy controls (20 men). The database was acquired in cooperation with the First Department of Neurology, Masaryk University and St. Anne's University Hospital in Brno, Czech Republic.

Subjects were rested and seated in front of the table in comfortable position. Each subject was asked to complete a handwriting task according to the prepared pre-filled template at a comfortable speed. Subjects were allowed to repeat the task in case of some error or mistake during handwriting. The pre-filled template is depicted in Fig. 1 [23]. The pre-filled template was shown to the subjects; no restrictions about the number of repetitions of syllables/words in tasks or their height were given.

A tablet was overlaid with an empty paper template (containing only printed lines and square box specifying area for Archimedean spiral), and a special ink pen was held in a normal fashion, allowing for immediate full visual feedback. The signals were recorded using the Intuos 4M (Wacom technology) digitizing tablet.

Digitized signals were acquired during the movements executed while exerting pressure on the writing surface (on-surface movement) and during the movement above the writing surface (in-air movement). The perpendicular pressure exerted on the tablet surface was also recorded. The recordings started when the pen touched the surface of the digitizer and finished when the task was completed. The tablet captured the following dynamic features (time-sequences): x-coordinate, $x[t]$; y-coordinate, $y[t]$; time stamp, $s[t]$; button status, $b[t]$; pressure, $p[t]$; and discrete time t. Button status is a binary variable, being 0 for in-air movement and 1 for on-surface movement.

The tablet sampling rate was 100 samples per second; the acquisition software was developed by the research team. Subsequent analysis was performed using Matlab and Python programming language.

2.1. Subjects

Altogether, 75 subjects (37 PD patients and 38 healthy controls) participated in the study. The participants were enrolled in the First Department of Neurology, St. Anne's University Hospital in Brno. A complete list of all participants is provided in Appendix A, with information about sex, age, disease duration, UPDRS-part V score[1]

[1] Unified Parkinson's Disease Rating Scale (UPDRS) is a rating scale used to follow the longitudinal course of PD [24].

Table 1
Parkinson's handwriting dataset. Characteristics of healthy controls (H) and Parkinson's disease (PD) group.

	Age		UPDRS (part V)		Years since diag.		LED		Male/female
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	
PD	69.3	10.9	2.27	0.84	8.37	4.8	1373.4	714	19/18
H	62.4	11.3	–	–	–	–	–	–	20/18

and levodopa equivalent daily dose.[2] Mean and standard deviation of age, UPDRS-Part V – Modified Hoehn and Yahr staging score [24] and disease duration are summarized in Table 1. No significant differences related to gender or age were found between the PD and healthy control groups.

All the subjects had completed at least ten years of education and reported Czech as their native language. All subjects used their dominant right hand. None of the subjects had a history or presence of any psychiatric symptoms or any disease affecting the central nervous system (other than PD in the PD cohort). All PD patients completed the tasks under L-DOPA medication.

Prior to handwriting acquisition, each patient was evaluated by a clinical neurologist. The healthy controls were examined by a clinician in order to make sure that there was no movement disorder or injury present that could significantly affect handwriting. We removed subjects whose handwriting was apparently affected by another diseases or who were not in the suitable physical condition. Basic information and instructions regarding the upcoming task were provided for each subject, and they were allowed to practice the task before the recording.

2.2. Handwriting tasks

The template consisted of eight different handwriting tasks. Based on a survey of the literature, we included drawing of an Archimedean spiral and repetitively writing cursive a letter "l", or a syllable "le", respectively.

The Archimedean spiral is an established task used in assessing of akinesia in PD and essential tremor [16,18]. During this task, the template was shown to the subject for visual guidance. Subjects drew the spiral from inside to out, but were not asked to draw spiral within particular boundaries or to follow a pre-drawn line.

In tasks 2, 3, and 4 participants wrote the cursive letter l, bigram le or the trigram les. Similar tasks (writing the letter l – or its variations) are frequently used for handwriting analysis [10,17].

Tasks 5, 6, and 7 were to write words lektorka – female teacher, porovnat – to compare, and nepopadnout – to not catch (written in Czech –, the native language of all participants). These words are characterized by simple orthography and quite easy syntax. The common characteristic is that they can be written continuously, without lifting the pen above the surface, i.e. they can be written in one continuous movement.

Task 8 was to write a longer sentence: Tramvaj dnes už nepojede (The tram won't go today). Use of the whole sentence allowed us to acquire also movements above the writing surface, i.e. in-air movements, during transitions between individual words in the sentence.

3. Methods and results

3.1. Feature extraction

The handwriting features were computed from on-surface movements (in the form of Cartesian coordinates) and pressure.

Table 2
Overview of kinematic handwriting features.

Feature	Description
Stroke speed	Stroke length divided by stroke duration in mm/s
Speed	Trajectory during handwriting divided by handwriting duration in mm/s
Velocity	Rate at which the position of a pen changes with time in mm/s
Acceleration	Rate at which the velocity of a pen changes with time in mm/s^2
Jerk	Rate at which the acceleration of a pen changes with time in mm/s^3
Horizontal velocity/acceleration/jerk	Velocity/acceleration/jerk in horizontal direction
Vertical velocity/acceleration/jerk	Velocity/acceleration/jerk in vertical direction
Number of changes in velocity direction (NCV)	The mean number of local extrema of velocity [25]
Number of changes in acceleration direction (NCA)	The mean number of local extrema of acceleration [25]
Relative NCV	NCV relative to writing duration
Relative NCA	NCA relative to writing duration
On-surface time	Time spent on-surface during writing
Normalized on-surface time	Time spent on-surface during writing normalized by whole writing duration

The kinematic features used in this study are listed in Table 2 [12]. The term stroke represents single connected continuous component of trace, i.e. on-surface movement between two successive pen lifts. According to this definition spiral or letter l are usually drawn as one stroke. Strokes were used only to calculate stroke speed.

Novel pressure handwriting features were computed to take advantage of all tablet functionalities. The typical pressure profile during writing is depicted in Fig. 2 for Archimedean spiral and Fig. 3 for tasks 2, 4 and 6.

The fundamental pressure features are the value of *pressure* as captured by the tablet during the particular task and the rate at which *pressure changes with respect to time*. Similarly to the concept of the number of changes in velocity [25] we proposed a *number of changes in pressure* (NCP) and *relative NCP*. Since there can be rapid changes in pressure that lead to incorrect NCP, we smoothed the data using a local regression using weighted linear least squares and a first degree polynomial model[3]. Relative NCP is NCP normalized by the whole length of writing. We introduced correlation coefficients to capture the relationship between pressure and kinematic features. In particular, we computed *correlation between pressure and (horizontal/vertical) velocity/acceleration*. Altogether, six correlation coefficients $\rho_{(horizontal/vertical),vel/acc}$ were computed.

Figs. 2 and 3 show that the main trend of the typical pressure trajectory starts with a rising edge, continues with a slowly increasing main movement, and finishes with a falling edge. The progress of the main trend is similar for all tasks. On the other hand, there is visible difference in smoothness of the main part of the signal for different tasks. As can be seen from Fig. 3, the main part is relatively smooth for the second task, however, as the performed task become longer the main part is more rough. The exterted pressure varies

[2] Levodopa equivalent dose (LED) of a drug that produces the same antiparkinsonian effect as 100 mg of immediate-release levodopa.

[3] We used Matlab's built-in function *smooth*.

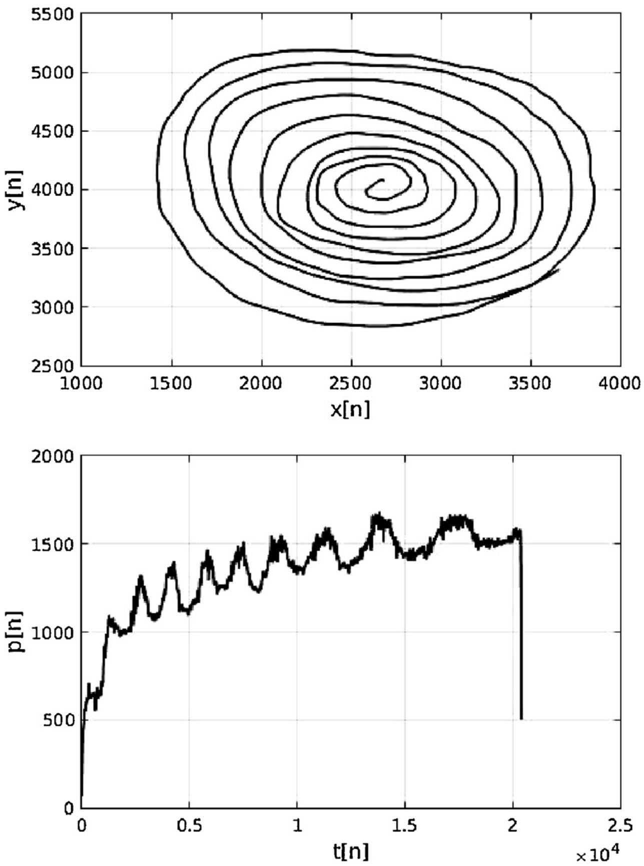


Fig. 2. Example of handwriting from task 1 (Archimedean spiral). On surface trajectory recorded by tablet (top figure) and related pressure exerted on tablet surface (bottom figure).

with the length of the word and complexity of the drawn letter. This may indicate why are some tasks more useful for classification than others.

The drawing of the spiral is different from the other tasks used in our study. Handwriting features are generally more stable for writing of the words, whereas, when drawing the spiral features change from beginning to the end of the task. This can be seen in Fig. 2 where the pressure trajectory is continuously increasing. In this study, we focus on the analysis of the handwriting and we did not introduce any spiral specific features. Therefore we extracted the same features from signals from all tasks.

We expected that there would be indicators for particular parts of the pressure signal, therefore we calculated the above mentioned *pressure features* separately for rising edge, main movement, and falling edge. The features corresponding to the different parts of the signal are denoted by superscript *rise*, *main* or *fall*. The boundary between the edges and the main movement is given by a median of signal pressure. Additionally, the range (maximal value–minimal value) between the rising edge and falling edge time duration ($R_{time}^{fall/rise}$) and the range of the rising edge and falling edge pressure ($R_{press}^{fall/rise}$) were included in the analysis. The final included feature was the *pressure overshoot*, providing the distance between the pressure maximum and pressure median.

Additionally, six basic statistical functionals (mean, median, standard deviation, 1st percentile, 99th percentile, 99th percentile − 1st percentile) were computed. Features were normalized before classification on a per-feature basis to have a zero mean and a standard deviation of one.

3.2. Statistical classification

Our aim was to build a discriminative model to differentiate between people with PD and healthy subjects. It is a binary classification task that can be resolved by statistical machine learning algorithms.

We compared three frequently used machine learning techniques: SVM [26], AdaBoost classifier with a decision tree base estimator [27] and *K*-NN algorithm. We used *Python* implementation of *scikit-learn* library [28].

The underlying idea of SVM classifiers is to calculate a maximal margin hyperplane separating two classes of the data. To learn non-linearly separable functions, the data are implicitly mapped to a higher dimensional space by means of a kernel function, where a separating hyperplane is found. New samples are classified according to the side of the hyperplane they belong to. We used radial basis function (RBF) kernel [26]. The RBF kernel is defined as

$$K(x, x_i) = e^{\frac{-\|x-x_i\|^2}{2\gamma^2}} \tag{1}$$

where γ controls the width of RBF function.

The parameters kernel gamma γ and penalty parameter C were optimized using grid search of possible values. Specifically, we searched over the grid (C, γ) defined by the product of the sets $C = [2^{-8}, 2^{-5}, \ldots, 2^{7}, 2^{8}]$, $\gamma = [2^{-9}, 2^{-4}, \ldots, 2^{8}, 2^{9}]$.

AdaBoost is one of the important *ensemble methods* known as *boosting*. The key idea behind boosting techniques is to use ensemble methods to combine weak classifiers in order to build a strong learner. AdaBoost is an iterative boosting algorithm constructing a strong classifier as a linear combination of weak classifiers, each performing at least above chance level (50% correct classification). We used decision trees classifiers as weak classifiers [29]. The maximum number of estimators at which boosting is terminated was set to 500. Settings used for decision trees were as follows. The number of features to consider when looking for the best split was the square root of the number of features and the maximum depth of the tree was set to 3.

In the *K*-NN algorithm, *k*-the nearest samples in a reference set are found, by taking a majority vote among the classes of these *k* samples. The goal is to determine the true class of an undefined test pattern by finding the nearest neighbors within a hyper-sphere of predefined radius. For the *K*-NN classifier, the best results were obtained with $k = 3$.

3.3. Numerical results

Classifier validation was conducted using stratified 10-fold cross-validation. The data was divided into ten mutually exclusive and exhaustive equal-sized subsets. For each subset, the union of all other subsets was considered as training data and the error rate was determined. Errors over different subsets were averaged to obtain the classification error. The process was repeated a total of ten times; the original dataset was randomly permuted in each repetition prior to splitting into training and testing subsets. The results were averaged over all ten runs.

The classification test performance was determined by the classification accuracy P_{acc}, sensitivity P_{sen} and specificity P_{spe} [30].

From all computed features we kept only those that passed the Mann–Whitney *U* test, i.e. those that showed a statistically significant ($p < 0.05$) difference between the PD and control groups. Table 3 shows 20 most relevant features and median of their values for PD and healthy control group. Features are sorted according Spearman's correlation coefficient ρ.

At first, we evaluated the prediction potential of different handwriting tasks considering both conventional kinematic features and pressure features. The classification accuracies for all tasks are

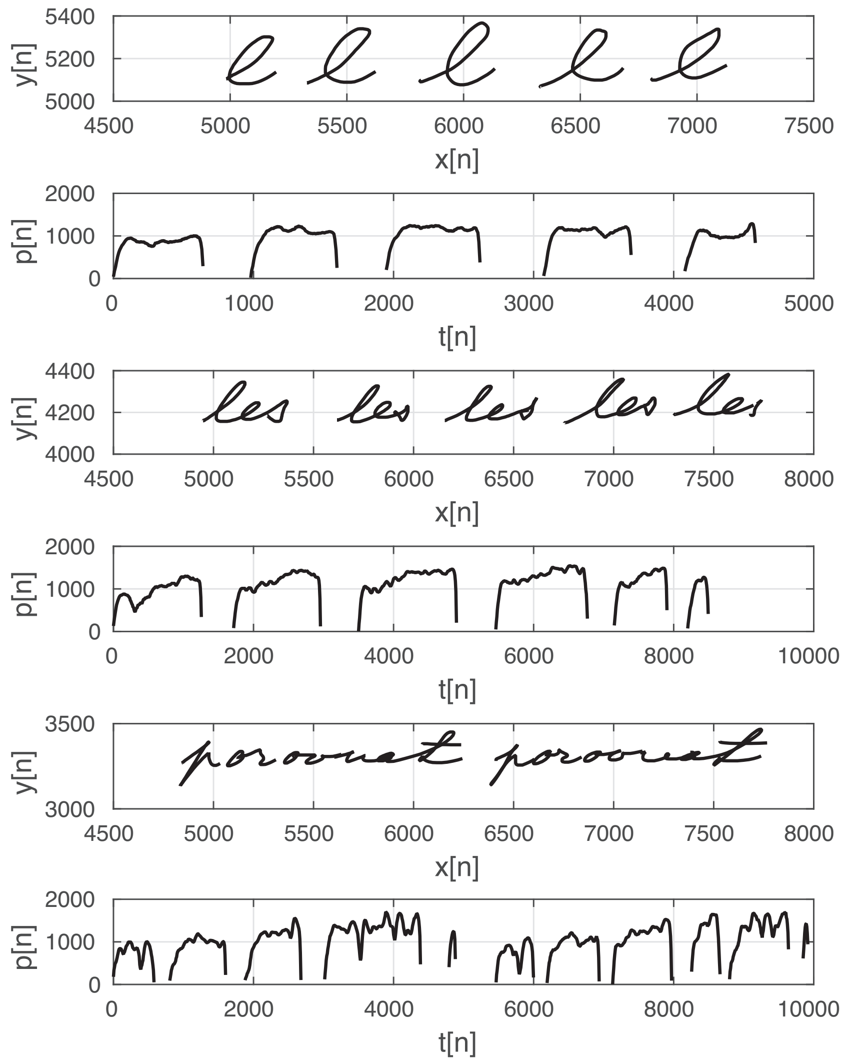


Fig. 3. Example of handwriting from tasks 2, 4 and 6. On surface trajectory recorded by tablet and related pressure exerted on tablet surface.

Table 3
Twenty kinematic and pressure features with largest relevance to class label sorted according to the Spearman's correlation coefficient $|\rho|$. Displayed are medians and standard deviations (std) for healthy controls (H) and Parkinson's disease group.

Feature, stat. functional, task number	$|\rho|$	PD median (std)	H median (std)
Stroke speed, std, task 8	0.39	0.45 (0.88)	−0.47 (0.97)
Relative NCP, task 8	0.37	−0.22 (0.63)	−0.06 (1.16)
Horizontal velocity, std, task 8	0.35	0.20 (0.99)	−0.33 (0.89)
ρ_{vel}, 99th percentile, task 2	0.35	−0.57 (0.86)	0.37(1.02)
Horizontal velocity, 99th percentile, task 8	0.33	0.4 (0.95)	−0.44 (0.95)
R^{fall}_{time}, median, task 8	0.33	−0.42 (0.57)	0.11(1.22)
$\rho^{rise}_{horizontal,acc}$, 1st percentile, task 8	0.33	−0.17 (0.83)	−0.37 (1.06)
Relative NCP, –, task 6	0.33	−0.40 (0.56)	−0.1 (1.23)
R^{rise}_{press}, 99th percentile − 1st percentile, task 3	0.33	−0.33(0.61)	0.1 (1.2)
$\rho^{main}_{vertical,vel}$, 99th percentile – 1st percentile, task 2	0.32	−0.47 (0.75)	−0.11(1.12)
R^{rise}_{press}, std, task 3	0.32	−0.29(0.6)	0.06 (1.21)
Horizontal jerk, std, task 8	0.32	0.23 (1.0)	−0.41 (0.9)
Horizontal velocity, 99th percentile 99 – 1st percentile, task 8	0.32	0.22 (1.01)	−0.48 (0.89)
Horizontal jerk, 99th percentile, task 8	0.32	0.29 (0.99)	−0.25 (0.91)
R^{fall}_{time}, median, task 3	0.31	−0.17(0.65)	−0.1 (1.19)
Velocity, median, task 8	0.31	0.2(0.9)	−0.35 (1.0)
Horizontal velocity (rising edge), std, task 8		0.26 (0.98)	−0.44(0.94)
Horizontal jerk, percentile 99th – percentile 1st, task 8	0.31	0.26 (0.97)	−0.33 (0.95)
$\rho_{vertical,vel}$, std, task 3	0.3	−0.6 (0.93)	0.23 (0.99)
Velocity (rising edge), mean, task 8	0.3	0.16 (0.98)	−0.41 (0.93)

Table 4
Classification accuracies of different handwriting tasks for kinematic and pressure features (support vector machines classifier).

Evaluated task	P_{acc} pressure features	P_{acc} kinematic features	P_{acc} kinematic and pressure
1 (Archimedean spiral)	62.8	–	62.8
2 (letter l)	72.1	69.2	72.3
3 (bigram le)	71.5	72.5	71.0
4 (word les)	66.4	–	66.4
5 (word lektorka)	66.9	65.1	65.2
6 (word porovnat)	74.2	64.9	73.3
7 (word nepopadnout)	66.8	66.4	67.6
8 (sentence)	73.2	74.9	76.5
overall	82.5	75.4	81.3

Table 5
Comparison of different classifiers for diagnosis of Parkinson's disease from handwriting. Kinematic and pressure features obtained from tasks 2 to 8 were used.

Classifier	P_{acc} [%]	P_{spe} [%]	P_{sen}[%]
SVM	81.3	80.9	87.4
AdaBoost	78.9	79.2	82.4
K-NN	71.7	70.8	78.5

depicted in Table 4. We did not find any statistically significant kinematic features for the tasks 1 and 4; therefore, for these two tasks we considered only pressure features. The highest classification accuracy using pressure features P_{acc} = 74.2% was obtained for data from the task 6. The slightly lower classification accuracy of 73.2% was achieved for the task 8. The best results for kinematic features were provided by data from the task 8. Similarly, task 8 was the most discriminative for merged kinematic and pressure features. Both modalities, i.e. pressure and kinematic features showed relatively similar classification accuracies for all tasks. Interestingly, the fusion of pressure and kinematic features[3] did not result in any improvement in terms of classification accuracy. The only exception was the most predictive task 8.

Merging all of the tasks together noticeably improved the classification accuracy, probably due to the different nature of the handwriting tasks. The results presented in Table 4 were obtained using an SVM classifier. To obtain more confidence in our results and to compare different classifiers we also employed the AdaBoost classifier and *K*-NN classifier. Overall classification accuracy (P_{acc}), sensitivity (P_{sen}), and specificity (P_{spe}) for all tasks and merged kinematic and pressure features are provided in Table 5. Comparing all three classifiers, it is clear that the best results in terms of accuracy, specificity, and sensitivity were obtained using the SVM classifier.

4. Discussion

PD is a very complex disease with different symptoms that can vary from patient to patient. The handwriting process is a complex motor activity requiring the coordination of several muscles. Both these aspects make it very difficult to explain or exactly link any handwriting characteristics or features to particular symptoms of PD. The results of our study show that pressure or kinematic features can be used to support a differential diagnosis of PD; however, the exact relationship between PD symptoms and particular features is not known. From a clinical point of view, kinematic features reflect complex cognitive processes and are influenced by a wide range of clinical aspects such as tremor, muscle stiffness, rigidity, and variance in movement speed. On the other hand, the pressure features can provide additional information that is not captured in kinematic features.

As indicated in Table 4, not all handwriting tasks provide the same level of discrimination power. After evaluating our results, it is evident that some tasks are more useful for diagnosis than others. Task 8 appeared to be the most promising task. This is probably because the task involves writing a whole sentence, and some representations of PD appear only when the task has some temporal extension. This is similar to typical symptom of PD – micrography, where the letter size is reduced as the subject spends more time writing a sentence line. As in task 8, task 2 (writing the letter l) provided good predictive performance from kinematic and pressure features. There was a gap in the predictive performance derived from pressure features and kinematic features for the task 6 (writing the word nepopadnout). Tasks 1, 4, 5, and 7 did not contribute to overall predictive performance significantly, as they reached only 62–66% accuracy. These findings indicate that special attention should be paid by researchers and clinicians when designing handwriting templates or even handwriting standardized tests since the task selection strongly influences the results or the potential of acquired data. In this study we focus mainly on handwriting analysis and we did not utilize any spiral specific features. This may explain why the task 1 does not have a significant impact on classification. Therefore in our future work we plan to perform deeper spiral analysis and evaluate spiral specific features.

Decision support tools are gaining significant research interest due to their potential to improve health-care provision [5,6]. Among many possible approaches, those that provide noninvasive monitoring and diagnosis of diseases are of increased interest to clinicians and biomedical engineers. We contribute to this area with the publication of our *PaHaW* database[3], containing eight different handwriting samples from 75 healthy and PD subjects. To prove the relevance of the database, we proposed a methodology to build a predictive model of PD from kinematic and pressure handwriting measures. It was shown that using of basic kinematic and pressure features allowed for a classification accuracy of 82%. The proposed approach is not intended to replace the clinician but rather to provide assistance for a more accurate and objective diagnosis. When employed with other approaches such as speech processing [22,31], even better results can be achieved in terms of accuracy of prediction. We showed that both kinematic and pressure features contribute in discriminating between PD and healthy subjects.

In this study, we almost 200 features. To follow and analyze such a high number of features can be very difficult for clinicians. It would be more convenient to specify a smaller representative group of features that would make it possible to map features to

[3] Only tasks 2–8 were merged. Task 1 contained data from only 69 subjects and did not show any significant discrimination potential, therefore we did not include this task.

[3] The database can be downloaded from BDALab webpage (http://bdalab.utko.feec.vutbr.cz/) or UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/).

standard metrics and provide quantification of a PD severity. However this again requires a relatively high number of subjects with different levels of disease symptom severity.

The results presented in this study indicate that different aspects of handwriting can be with advantage used in diagnosis of PD. However, several limitations have to be recognized when interpreting the results. Firstly, in this study we decided to focus only on PD and healthy control group. Other diseases have been analyzed in other papers [32,33]. Inclusion of the cognitively well characterized PD patients at different motor stages of PD as well as inclusion of other relevant patient groups that suffer from both micrographia and cognitive impairment such as progressive supranuclear palsy or Huntington's disease are warranted in order to investigate whether the proposed technique can be used to discriminate between PD and other diseases. Classification of different diseases may be possible if they alter handwriting in diverse way, i.e. there are different patterns across the variables. However, this further underlines the importance of deeper handwriting analysis that include all modalities, since discriminative features may be hidden in handwriting signal and simple evaluation of conventional kinematic features may not be sufficient. Secondly, all patients with PD performed the handwriting tasks under medication. It suggests that proposed methodology may be sensitive enough to identify PD even if the symptoms are attenuated by the medication. On the other hand, medication can have side effects impacting the movements of the patients that can influence classification process. Before implementation of the proposed approach in the clinical settings a future study on patients without medication should be performed to investigate how would classifier perform under this condition. Thirdly, we have shown that handwriting can be used as biomarker for PD, however this should be considered only as a first step in further investigation. Longitudinal study is required to investigate subjects with a high risk of PD development, and confirm whether proposed methodology successfully identified participants that actually developed PD. This can reveal whether it is possible to use handwriting markers for early diagnosis of PD. Similarly, repeated measurements from the same subjects can be obtained to increase test-retest reliability or to investigate how can be proposed approach used for symptoms severity monitoring. Additionally, it is not possible to control for all intentional and unintentional alterations of handwriting. There may be many factors influencing handwriting that might impact classification decision. Therefore further studies are needed to provide confirmation of conclusions drawn in this paper. Standardized data collection and testing of subjects on multiple occasions is necessary and it is aim of our future research.

Acknowledgements

This work was supported by the projects COST IC 1206, NT13499 (Speech, its impairment and cognitive performance in Parkinsons disease), LO1401, CZ.1.07/2.3.00/20.0094, project CEITEC, Central European Institute of Technology: (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund and by FEDER and Ministerio de Economa y Competitividad TEC2012-38630-C04-03. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation. Peter Drotár was supported by the CZ.1.07/2.3.00/30.0005 project of Brno University of Technology.

Appendix A. Subject data

Table A.6.

Table A.6
Detailed clinical and demographic information about participants.

ID	Sex	Diagnosis	Age [years]	LED	UPDRS (part V)	Years since diag.
01	F	PD	68	1115	2	6
02	F	PD	78	2110	2	8
03	F	PD	69	1557	2	7
04	F	PD	79	1691	2	12
05	F	PD	69	600	2	2
06	F	PD	57	1272	2	9
07	F	PD	78	666	3	19
08	F	PD	58	397	1	5
09	M	PD	78	2066	1	3
10	M	PD	74	1480	2.5	3
13	M	PD	65	990	1	2
14	M	PD	64	1253	3	8
15	F	PD	69	990	2.5	17
16	M	PD	67	1188	2	4
17	F	PD	75	1370	5	18
18	F	PD	76	1250	2.5	17
19	F	PD	86	750	2	6
20	F	PD	79	2227	2	8
22	F	PD	67	645	2	14
23	F	PD	73	1235	2	9
24	M	PD	70	1317	4	7
25	M	PD	60	1143	3	10
26	F	Healthy	57	–	–	–
27	M	Healthy	92	–	–	–
28	F	Healthy	52	–	–	–
29	F	Healthy	58	–	–	–
30	M	Healthy	69	–	–	–
31	M	Healthy	76	–	–	–
32	F	Healthy	59	–	–	–
33	F	PD	62	750	2	4
34	M	PD	61	2547	2	5
36	M	PD	90	750	2	3
39	M	Healthy	65	–	–	–
40	M	Healthy	53	–	–	–
41	M	Healthy	78	–	–	–
43	M	PD	48	1080	1	4
44	F	PD	62	397	1	5
48	M	PD	87	1450	4	12
49	M	Healthy	58	–	–	–
51	F	Healthy	48	–	–	–
52	F	Healthy	44	–	–	–
53	M	PD	84	1942	2	2
54	M	PD	69	2546	2	10
55	M	PD	63	1930	2.5	14
57	M	Healthy	80	–	–	–
60	M	Healthy	65	–	–	–
61	F	Healthy	59	–	–	–
62	F	Healthy	65	–	–	–
66	F	Healthy	69	–	–	–
67	M	Healthy	59	–	–	–
69	F	Healthy	74	–	–	–
70	F	Healthy	47	–	–	–
71	M	Healthy	52	–	–	–
72	M	Healthy	45	–	–	–
73	F	Healthy	64	–	–	–
74	M	PD	53	2387	2.5	9
75	M	PD	73	2010	2.5	12
76	M	Healthy	56	–	–	–
77	M	PD	74	2337	3	1
78	M	PD	36	800	2	2
80	M	PD	67	3544	3	5
82	M	Healthy	45	–	–	–
83	F	Healthy	74	–	–	–
84	F	Healthy	62	–	–	–
85	F	Healthy	75	–	–	–
87	M	Healthy	57	–	–	–
89	M	Healthy	63	–	–	–
90	M	Healthy	71	–	–	–
91	F	Healthy	64	–	–	–
92	F	Healthy	58	–	–	–
94	M	Healthy	64	–	–	–
95	M	Healthy	74	–	–	–
96	F	Healthy	77	–	–	–
97	M	Healthy	44	–	–	–
98	F	PD	77	1210	2	6

References

[1] de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. Lancet Neurol 2006;5(6):525–35.
[2] von Campenhausen S, Bornschein B, Wick R, Botzel K, Sampaio C, Poewe W, et al. Prevalence and incidence of Parkinson's disease in Europe. Eur Neuropsychopharm 2005;15(4):473–90.
[3] Contreras-Vidal JL, Stelmach GE. Effects of parkinsonism on motor control. Life Sci 1995;58(3):165–76.
[4] Hughes AJ, Daniel SE, Ben-Shlomo Y, Lees AJ. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain 2002;125(4):861–70, http://dx.doi.org/10.1093/brain/awf080.
[5] Ammenwerth E, Nykanen P, Rigby M, de Keizer N. Clinical decision support systems: need for evidence, need for evaluation. Artif Intell Med 2013;59(1): 1–3.
[6] Dreiseitl S, Binder M. Do physicians value decision support? a look at the effect of decision support systems on physician opinion. Artif Intell Med 2005;33(1):25–30.
[7] Mattison HA, Stewart T, Zhang J. Applying bioinformatics to proteomics: is machine learning the answer to biomarker discovery for PD and MSA? Mov Disord 2012;27(13):1595–7.
[8] Saunders-Pullman R, Derby C, Stanley K, Floyd A, Bressman S, Lipton RB, et al. Validity of spiral analysis in early Parkinson's disease. Mov Disord 2008;23(4):531–7.
[9] Broderick M, Van Gemmert A, Shill H, Stelmach G. Hypometria and bradykinesia during drawing movements in individuals with Parkinsons disease. Exp Brain Res 2009;197(3):223–33.
[10] Bidet-Ildei C, Pollak P, Kandel S, Fraix V, Orliaguet J-P. Handwriting in patients with Parkinson's disease: effect of L-DOPA and stimulation of the sub-thalamic nucleus on motor anticipation. Hum Mov Sci 2011;30(4):783–91.
[11] Contreras-Vidal JL, Poluha P, Teulings H-L, Stelmach GE. Neural dynamics of short and medium-term motor control effects of levodopa therapy in Parkinson's disease. Artif Intell Med 1998;13(1–2):57–79.
[12] Drotar P, Mekyska J, Rektorova I, Masarova L, Smekal Z, Faundez-Zanuy M. A new modality for quantitative evaluation of Parkinson's disease: in-air movement. In: 2013 IEEE 13th international conference on Bioinformatics and bioengineering (BIBE). 2013. p. 1–4.
[13] Rosenblum S, Samuel M, Zlotnik S, Erikh I, Schlesinger I. Handwriting as an objective tool for parkinsons disease diagnosis. J Neurol 2013;260(9):2357–61.
[14] Unlu A, Brause R, Krakow K. Handwriting analysis for diagnosis and prognosis of parkinsons disease. In: Maglaveras N, Chouvarda I, Koutkias V, Brause R, editors. Biological and medical data analysis, Vol. 4345 of lecture notes in computer science. Berlin, Heidelberg: Springer; 2006. p. 441–50.
[15] Kraus PH, Hoffmann A. Spiralometry: computerized assessment of tremor amplitude on the basis of spiral drawing. Mov Disord 2010;25(13):2164–70, http://dx.doi.org/10.1002/mds.23193.
[16] Stanley K, Hagenah J, Bruggemann N, Reetz K, Severt L, Klein C, et al. Digitized spiral analysis is a promising early motor marker for Parkinson's disease. Parkinsonism Relat Disord 2010;16(3):233–4.
[17] Smits EJ, Tolonen AJ, Cluitmans L, van Gils M, Conway BA, Zietsma RC, et al. Standardized handwriting to assess bradykinesia, micrographia and tremor in parkinson's disease. PLOS ONE 2014;9:1–8.
[18] Teulings H-L, Contreras-Vidal JL, Stelmach GE, Adler CH. Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control. Exp Neurol 1997;146(1):159–70.
[19] Sesa-Nogueras E, Faundez-Zanuy M, Mekyska J. An information analysis of in-air and on-surface trajectories in online handwriting. Cogn Comput 2012;4(2):195–205, http://dx.doi.org/10.1007/s12559-011-9119-y.
[20] Drotar P, Mekyska J, Rektorova I, Masarova L, Smekal Z, Faundez-Zanuy M. Decision support framework for Parkinson's disease based on novel handwriting markers. IEEE Trans Neural Syst Rehab Eng 2015;23(3):508–16, http://dx.doi.org/10.1109/TNSRE.2014.2359997.
[21] Sakar B, Isenkul M, Sakar C, Sertbas A, Gurgen F, Delil S, et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Biomed Health Inform 2013;17(4):828–34.
[22] Tsanas A, Little M, Fox C, Ramig L. Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease. IEEE Trans Neural Syst Rehab Eng 2014;22(1):181–90.
[23] Drotar P, Mekyska J, Smekal Z, Rektorova I, Masarova L, Faundez-Zanuy M. Prediction potential of different handwriting tasks for diagnosis of Parkinson's. In: E-Health and bioengineering conference (EHB). 2013. p. 1–4.
[24] Fahn S, Elston R. Unified parkinsons disease rating scale. In: Fahn S, Marsden C, Calne D, Goldstein M, editors. Recent developments in Parkinson's disease, vol. 2. Florham Park, NJ: Macmillan; 1987. p. 153–63.
[25] Eichhorn T, Gasser T, Mai N, Marquardt C, Arnold G, Schwarz J, Oertel W. Computational analysis of open loop handwriting movements in Parkinson's disease: a rapid method to detect dopamimetic effects. Mov Disord 1996;11(3):289–97.
[26] Vapnik V. Statistical learning theory. 1st Ed. London: John Willey & Sons; 1998.
[27] Zhu J, Zou H, Rosset S, Hastie T. Multi-class adaboost. Stat Interface 2009;2(3):349–60.
[28] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[29] Breiman L, Friedman J, Olshen R, Stone C. Classification and regression trees. Belmont, CA: Wadsworth; 1984.
[30] Cherkassky V, Mulier F. Learning from data: concepts, theory, methods. Hoboken, NJ: John Wiley & Sons; 2007.
[31] Mekyska J, Janousova E, Gomez-Vilda P, Smekal Z, Rektorova I, Eliasova I, et al. Robust and complex approach of pathological speech signal analysis. Neurocomputing 2015;167:94–111.
[32] Lopez-de Ipina K, Bergareche A, de la Riva P, Faundez-Zanuy M, Calvo P, Zelarain F. Analysis of non-invasive writing signals, applied to essential tremor: a non-linear approach. In: International work conference on bio-inspired intelligence (IWOBI). 2014. p. 56–9, http://dx.doi.org/10.1109/IWOBI.2014.6913938.
[33] Faundez-Zanuy M, Sesa-Nogueras E, Roure-Alcobe J, Garre-Olmo J, Mekyska J, Lopez-de Ipina K, et al. A preliminary study of online drawings and dementia diagnose. In: Apolloni B, Bassis S, Esposito A, Morabito FC, editors. Neural nets and surroundings, Vol. 19 of smart innovation, systems and technologies. Berlin, Germany: Springer; 2013. p. 367–74.

A.24 Contribution of different handwriting modalities to differential diagnosis of Parkinson's Disease

This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication.

Contribution of Different Handwriting Modalities to Differential Diagnosis of Parkinson's Disease

Peter Drotár, Jiří Mekyska, Zdeněk Smékal
Faculty of Electrical Engineering and Communications
Brno University of Technology
Brno, Czech Republic
Email: drotarp@feec.vutbr.cz

Irena Rektorová, Lucia Masarová
First Department of Neurology
CEITEC MU
Masaryk University
Brno, Czech Republic

Marcos Faundez-Zanuy
Tecnocampus
Escola Universitaria Politecnica
de Mataro
Mataro, Spain

***Abstract*—In this paper, we evaluate the contribution of different handwriting modalities to the diagnosis of Parkinson's disease. We analyse on-surface movement, in-air movement and pressure exerted on the tablet surface. Especially in-air movement and pressure-based features have been rarely taken into account in previous studies. We show that pressure and in-air movement also possess information that is relevant for the diagnosis of Parkinson's Disease (PD) from handwriting. In addition to the conventional kinematic and spatio-temporal features, we present a group of the novel features based on entropy and empirical mode decomposition of the handwriting signal. The presented results indicate that handwriting can be used as biomarker for PD providing classification performance around 89% area under the ROC curve (AUC) for PD classification.**

I. Introduction

According to the recent estimates, more than seven million people are affected by Parkinson's Disease (PD) worldwide [1]. The high number of affected people makes PD the second most common neurodegenerative disorder. Moreover, with the aging population, it is expected that the prevalence rates will further increase and impose social and economic burden for healthcare. Despite intensive research effort, the causes of PD are still not known and the reliable easily applicable diagnostic test is not available yet [1].

A diagnosis of PD is currently based mainly on clinical symptoms such as bradykinesia, rigidity, tremor or postural imbalance. Several alternative solutions and decision support systems have been proposed to improve diagnosis of PD. The neuroimaging methods show significant potential but require expensive equipment [2]. Other approaches include attempts to detect PD from breath [3] or voice [4], [5], [6]. Especially speech processing for diagnosis of PD gained significant attention and offered very promising results. Recent studies indicate that also handwriting can be with advance used for differential diagnosis of PD [7], [8], [9]. This is due to the alterations in parkinsonian handwriting represented by micrographia and PD dysgraphia [10]. Micrographia was reported for the first time by Pick [11] as an abnormal reduction in handwriting size associated with PD. On the other hand, PD dysgraphia is a new term proposed by Letanneaux et al. [10] that "*encompasses the whole spectrum of disorders that affect the writing of PD patients*" including tremor, rigidity, bradykinesia, akinesia, freezing of the upper limb etc.

Micrographia was studied in [12] or [13], however analyzing only micrographia alone is not enough, since micrographia occurs only in 30 % to 60 % of patients with PD [14],[15]. This was motivation for several authors to investigate also kinematic aspects of movement including speed, acceleration or stroke duration [16], [17]. Even though these studies provided the significant insight into the handwriting in PD, they did not assess tremor. Recently, two studies have been published that provide quantitative measures to assess multiple motor symptoms of the PD handwriting allowing for the clinically acceptable differential diagnosis of PD [9],[8]. The current technologies allow to exploit new modalities, such as in-air movement and pressure exerted on the surface, not only conventional handwriting trace on surface [18]. Some initial studies employing the in-air movement [19], [7] or pressure [7] indicate that also these modalities can be useful for diagnosis of PD.

In this study, we employ an approach presented in our previous work [8] and use advanced handwriting markers based on entropy, energy and intrinsic measures of handwriting. Here, we apply these measures also to in-air movement and pressure to exploit full potential of the handwriting for classification of PD. To achieve this goal we make use of our Parkinsonian Handwriting database (PaHaw) which consists of seven different handwriting tasks. The task template contains the exercises found in similar studies that are used to assess PD. In addition to conventional handwriting tasks we added the new tasks: simple words and one sentence. The orthography of these task is intentionally simple to minimize cognitive effort during the writing task.

The paper is organized as follows. After the introductory section, the description of used database is given. Next, methodology of extracting the features from handwriting signal is given, followed by a brief overview of used classifier. Finally, the numerical results and conclusions are provided.

II. Data

The Parkinsons handwriting dataset consists of multiple handwriting samples from 37 parkinsonian patients (19 men/18 women) and 38 gender and age matched controls (20 men/18 women). Mean age was 69.3 ± 10.9 for parkinsonian patients and 62.4 ± 11.3 for control subjects, respectively. All subjects were right-handed, had completed at least 10 years of education, and reported Czech as their first language. For parkinsonian patients the mean value of Unified Parkinson's Disease Rating Scale-Part V was 2.27 ± 0.84 and all patients

978-1-4799-6477-2/15/$31.00 ©2015 IEEE

Fig. 1. Filled task sheet used as template for handwriting tasks.

completed the tasks under medication L-DOPA. A more detailed description of dataset is shown in [8].

Each subject was asked to complete handwriting task according to the prepared template. The template consists of 7 tasks that were a part of a more detailed test battery. The tasks were designed to have simple orthography and all but last one can be written as one long stroke. The filled task sheet is depicted in Fig.1. The template was shown to the patients and they were free to write the words without no need to follow the exact pattern.

III. Feature extraction

Signals were acquired using Wacom Intuos 4M pen tablet. The tablet itself does not provide visual feedback, therefore, it was overlaid with the paper so the pen can be held in a normal fashion and allows full visual feedback during writing. The inked pen is commercially available from the tablet producer, it allows a user to draw on the paper and at the same time the signal is captured by the tablet.

The tablet is able to capture several signals related to the handwriting. These include position of the pen tip in the form of x and y coordinate ($x[n]$, $y[n]$), binary variable ($b[n]$), being 0 for pen-up state (in-air movement) and 1 for pen-down state (on-surface movement). This means that the tablet is able to capture $x[n]$ and $y[n]$ also in case when pen is not touching the tablet surface while moving in the air. Additionally, pressure exerted on the tablet surface during writing ($p[n]$) and time stamp ($t[n]$) are recorded. An example of handwriting sample from the 7th task is depicted in Fig. 2. The figure shows the example of recorded on-surface (blue solid line) and in-air (red dashed line) movement. Even though it illustrates how the actual handwriting sample is realized it does not provide the insight into the shape of the signals that were used to extract handwriting features. Signals corresponding to the sample depicted in Fig. 2 can be seen in Fig. 3. Fig. 3 shows pressure and coordinates of in-air/on-surface movement as a function of time.

A. Spatio temporal and kinematic features

Using Cartesian coordinates ($x[n]$,$y[n]$) and time stamp it is possible to determine several kinematic and spatio-temporal features. These include handwriting velocities and derived measures such as acceleration and jerk: speed, a number of changes in velocity(NCV)/acceleration(NCA), relative NCV/NCA, stroke speed, velocity, acceleration, jerk, horizontal velocity/acceleration/jerk, vertical velocity/acceleration/jerk. Considered spatio-temporal features are stroke height/width, writing duration, writing length and in-air to on-surface ratio. The definition of these features is provided in [19].

B. Pressure features

To make use of recorded pressure signal $p[n]$ several pressure features were extracted. Similarly to kinematic features, we computed rate at which pressure changes with respect to time, a number of changes in velocity pressure (NCP) and relative NCP. Relative NCP is NCP normalized by the whole writing length. Additionally, six correlation coefficient were introduced: correlation between pressure and (horizontal/vertical) velocity/acceleration. Fig. 3 indicates that typical pressure stroke starts with rising edge, continues with slowly varying main part and ends with falling edge. We derived pressure features for each part of the stroke (edge, main part of signal and falling edge) separately. The boundary between edges and main part is given by median of signal pressure. Finally, a range of rising and falling edge in terms of pressure and time was included in the analysis.

C. Nonstandard handwriting features

In order to uncover also hidden complexities of handwriting we also employ nonstandard features proposed in [8]. Out of these, entropy based features have potential to capture randomness of the movement during handwriting. We calculated the Shannon entropy H_S and Rényi entropy of the second $H_{R,2}$ and the third $H_{R,3}$ order [20].

Similarly, the energy based features express amount of noise in handwriting in relation to useful handwriting signal. Note that what we refer to as noise is not unwanted interference from an external source, but the irregular movement resulting from muscle contractions and irregularities. Therefore, we computed estimated noise variance $\mathcal{N}(s[n])$ of the signal and energy of the signal $\Theta(s[n])$. The operator Θ represents conventional energy operator $CE(s[n]) = 1/N \sum_{n=0}^{N-1} s[n]^2$ or Teager-Kaiser energy operator $TKE_r(s[n]) = 1/N \sum_{n=0}^{N-r-1} s[n]^2 - s[n+r] \cdot s[n-r]$. Signal to noise ratios are given by $SNR_{\Theta} = \Theta(s[n])/\mathcal{N}(s[n])$. The signal $s[n]$ is the signal under evaluation i.e. $x[n]$, $y[n]$ (obtained from both on-surface and in-air movement) or $p[n]$.

To obtain intrinsic features we apply empirical mode decomposition (EMD) to decompose signal into its intrinsic mode functions (IMF). EMD is an intuitive data-dependent decomposition of a time series that allows decomposition of non-linear and non-stationary data into IMFs. EMD is conducted by the iterative extraction based on the local representation of the signal as the sum of a local oscillating component and a local trend [21]. Given a time series $s[n]$, combining all the IMFs gives the original signal $s[n] = \sum_{j=1}^{N} IMF_j[n] + r_N[n]$.

We obtained intrinsic entropies and energies by applying the above mentioned methods on first and second intrinsic function. Similarly, intrinsic signal to noise rations were derived as the ratio of energy of first two intrinsic functions to energy of the rest of intrinsic functions.

IV. Machine Learning

Six basic statistical functionals (mean, median, standard deviation, 1st percentile, 99th percentile, 99th percentile - 1st percentile) were computed from the extracted features. This produces more than 5000 features. The features were normalized before classification on a per-feature basis to have zero mean and a standard deviation of one. As a preprocessing step we applied Mann-Whitney U test for significant differences identification to remove features that did not show statistical significance to class label. Application of Mann-Whitney U test reduced the number of features to less than 700. The distribution of features for different tasks and modalities is depicted in Tab. I.

Support Vector Machine (SVM) was used as a classifier to predict class labels. The underlying idea of SVM classifiers is to calculate a maximal margin hyperplane separating two classes of the data. To learn non-linearly separable functions, the data are implicitly mapped to a higher dimensional space by means of a kernel function. The new samples are classified according to the side of the hyperplane they belong to. We used Radial Basis Function (RBF) kernel [22]. The RBF kernel is defined as

$$K(x, x_i) = \mathrm{e}^{\frac{-\|x - x_i\|^2}{2\gamma^2}} \tag{1}$$

where γ controls the width of RBF function.

The parameters kernel gamma γ and penalty parameter C were optimized using grid search of possible values. Specifically, we searched over the grid (C, γ) defined by the product of the sets $C = [2^{-10}, 2^{-9}, \ldots, 2^{6}, 2^{7}]$, $\gamma = [2^{-7}, 2^{-6}, \ldots, 2^{6}, 2^{7}]$. We used scikit-learn implementation of SVM [23].

V. Numerical Results and Discussion

The prediction performance was evaluated in terms of the area under ROC curve (AUC). In order to obtain prediction potential of each modality we assess the AUC performance for every modality (in-air, on-surface and pressure) individually. Additionally, prediction performance of seven different handwriting tasks was also considered separately. The numerical results achieved by SVM classifier with 10 fold cross validation are provided in Table II.

As it can be seen from Table II, the highest $AUC = 89.09\%$ was achieved when the features extracted from on surface movement were used. This provides significantly higher AUC than other two modalities. The promising results are also yielded from the pressure features that have not been used for purpose of PD classification before, giving rise to more than 83% prediction accuracy. Within this scenario in-air movement does not appear to be significantly contributing to differential diagnosis of PD from handwriting. However, we should note that entropy, energy and intrinsic features were originally designed for on-surface movement and as such may not explore full potential of in-air movement.

TABLE II. AUC of PD classification based on different modalities (in-air, on-surface and pressure).

task/modality	on-surface	in-air	pressure
1	72.39	67.58	72.5
2	70.16	66.75	76
3	70.86	66.75	72.16
4	66.08	65.25	64.25
5	62.75	67.33	69.66
6	65.66	-	71.66
7	83.83	73	72.58
all	89.09	74.16	83.83

When comparing the contribution of a different handwriting task to classification of PD, it is clear that most of the prediction performance comes from the seventh task. This is true especially for the on-surface and in-air modality. In case of pressure, different handwriting tasks contribute more equally. The seventh task (*Tramvaj dnes už nepojede)* is the longest task and in contrast to other tasks this task cannot be written as one long stroke. Writing longer sentence probably requires higher cognitive effort and escalates effect of disease on handwriting. These results are in agreement with the previous findings where the last task also appears to be the most predictive one [24].

We tried to identify a smaller subset of the features that provide the strongest discriminative power. However, reducing number of the features resulted in decline in classification performance.

VI. Conclusion

We have evaluated a prediction performance of the different handwriting modalities for differential classification of PD. It was shown that not only on-surface movement, but also pressure and in-air movement contribute to the classification and can be with advantage used for diagnosis of PD from handwriting. By using 7 different handwriting tasks, standard kinematic and also novel intrinsic and entropy features we showed that handwriting is a promising tool for diagnosis of PD achieving almost 90% prediction performance.

Acknowledgment

This work was supported by the projects COST-IC1206,NT13499 (Speech, its impairment and cognitive performance in Parkinsons disease), CZ.1.07/2.3.00/20.0094, project CEITEC, Central European Institute of Technology: (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund and by FEDER and Ministerio de Economa y Competitividad TEC2012-38630-C04-03. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation. Peter Drotár was supported by the project CZ.1.07/2.3.00/30.0005 of Brno University of Technology.

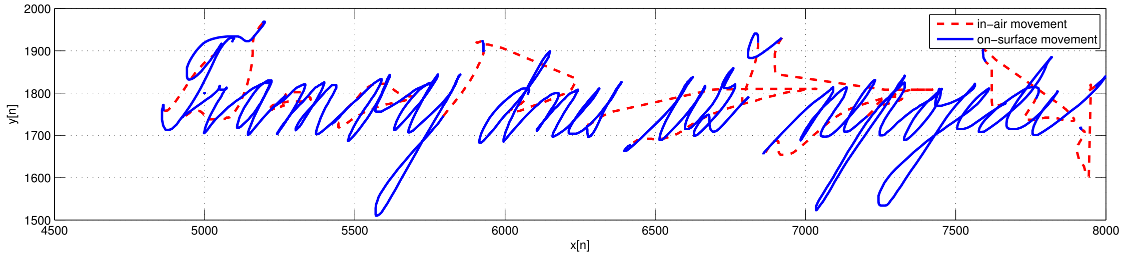


Fig. 2. Handwriting sample from seventh task of the template. In-air and on-surface movement.

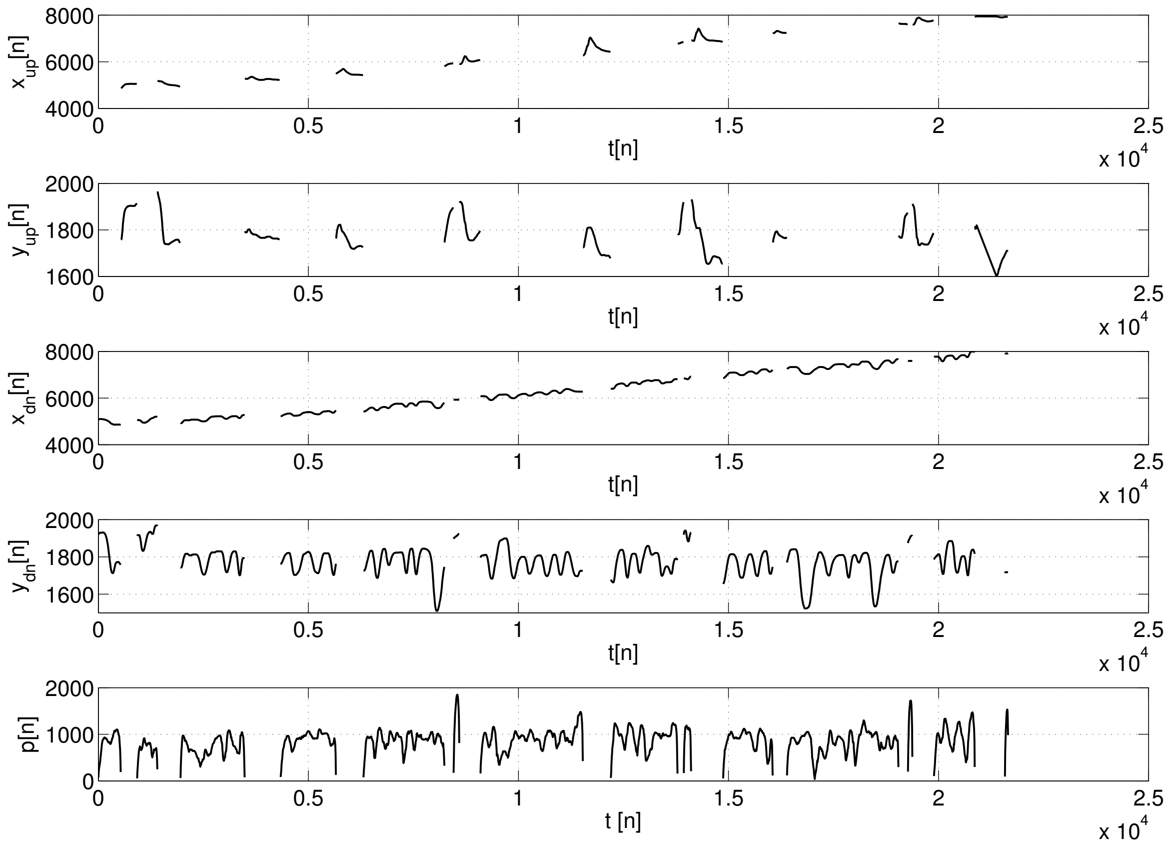


Fig. 3. Recorded signals during the execution of the seventh task. The figure displays x and y coordinates of in-air (up) and on-surface (dn) movement as a function of time. Additionally bottom figure illustrates pressure exerted on tablet surface as an function of time.

TABLE I. NUMBER OF EXTRACTED FEATURES AND NUMBER OF FEATURES THAT PASSED THE MANN-WHITNEY U TEST

	on-surface (task 1 - task 7)	in-air (task 1 - task 7)	pressure (task 1 - task 7)
all	268 per task	290 per task	188 per task
afer Mann-Whitney U test	11, 61, 12, 10, 10, 35, 138	58, 81, 2, 7, 8, 0, 80	32, 39, 11, 10, 24, 29, 34

REFERENCES

[1] L. M. de Lau and M. M. Breteler, "Epidemiology of Parkinson's disease," *Lancet Neurol.*, vol. 5, no. 6, pp. 525 – 535, 2006.

[2] Y. Wu, W. Le, and J. Jankovic, "Preclinical biomarkers of parkinson disease," *Archiv. Neurol.*, vol. 68, no. 1, pp. 22–30, 2011.

[3] U. Tisch, I. Schlesinger, R. Ionescu, M. Nassar, N. Axelrod, D. Robertman, Y. Tessler, F. Azar, A. Marmur, J. Aharon-Peretz, and H. Haick, "Detection of alzheimers and parkinsons disease from exhaled breath using nanomaterial-based sensors," *Nanomed.*, vol. 8, no. 1, pp. 43–56, 2013.

[4] A. Tsanas, M. Little, C. Fox, and L. Ramig, "Objective automatic assessment of rehabilitative speech treatment in parkinsons disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, no. 22, pp. 181–190, 2014.

[5] J. Rusz, R. mejla, H. Rikov, J. Klemp, V. Majerov, J. Picmausov, J. Roth, and E. Rika, "Acoustic assessment of voice and speech disorders in parkinson's disease through quick vocal test," *Mov. Disor.*, vol. 26, no. 10, pp. 1951–1952, 2011.

[6] J. Mekyska, Z. Smekal, M. Kostalova, M. Mrackova, S. Skutilova, and I. Rektorova, "Motor aspect of speech impairment in parkinsons disease and their assesment," *Czech and Slovak Neurology and Neurosurgery*, vol. 74, no. 6, pp. 662–668, 2011.

[7] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, and I. Schlesinger, "Handwriting as an objective tool for parkinsons disease diagnosis," *J. Neurol.*, pp. 1–5, 2013.

[8] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Decision support framework for parkinson's disease based on novel handwriting markers," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 1, no. 1, 2015, to appear.

[9] E. J. Smits, A. J. Tolonen, L. Cluitmans, M. van Gils, B. A. Conway, R. C. Zietsma, K. L. Leenders, and N. M. Maurits, "Standardized handwriting to assess bradykinesia, micrographia and tremor in parkinson's disease," *Plos One*, vol. 9, pp. 1 – 8, 2014.

[10] A. Letanneux, J. Danna, J.-L. Velay, F. Viallet, and S. Pinto, "From micrographia to parkinson's disease dysgraphia," *Movement Disorders*, vol. 29, no. 12, pp. 1467–1475, 2014. [Online]. Available: http://dx.doi.org/10.1002/mds.25990

[11] A. Pick, "Ueber eine eigenthuemliche schriebstueung, mikrographie. folge cerebraler erkrankung," *Prag Med Wochenschr*, vol. 28, pp. 1 – 4, 1903.

[12] A. W. A. Van Gemmert, C. H. Adler, and G. E. Stelmach, "Parkinsons disease patients undershoot target size in handwriting and similar tasks," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 74, no. 11, pp. 1502–1508, 2003.

[13] M. Broderick, A. Van Gemmert, H. Shill, and G. Stelmach, "Hypometria and bradykinesia during drawing movements in individuals with parkinsons disease," *Experimental Brain Res.*, vol. 197, no. 3, pp. 223–233, 2009.

[14] E.-J. Kim, B. H. Lee, K. C. Park, W. Y. Lee, and D. L. Na, "Micrographia on free writing versus copying tasks in idiopathic parkinson's disease," *Parkinsonism & Related Disorders*, vol. 11, no. 1, pp. 57 – 63, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1353802004001312

[15] L. Contreras-Vidal, J, L. Teulings, H, and E. Stelmach, G, "Micrographia in parkinsons disease," *Neuroreport*, vol. 6, pp. 2089–2092, 1995.

[16] H.-L. Teulings and G. E. Stelmach, "Control of stroke size, peak acceleration, and stroke duration in parkinsonian handwriting," *Human Movement Science*, vol. 10, no. 23, pp. 315 – 334, 1991. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016794579190010U

[17] A. Unlu, R. Brause, and K. Krakow, "Handwriting analysis for diagnosis and prognosis of parkinsons disease," in *Biological and Medical Data Analysis*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2006, vol. 4345, pp. 441–450.

[18] E. Sesa-Nogueras, M. Faundez-Zanuy, and J. Mekyska, "An information analysis of in-air and on-surface trajectories in online handwriting," *Cognitive Comput*, vol. 4, no. 2, pp. 195–205, 2012.

[19] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "A new modality for quantitative evaluation of Parkinsons disease:in-air movement," in *13th IEEE International Conference on BioInformatics and BioEngineering (BIBE)*, Chania, Greece, November 2013.

[20] T. M. Cover, *The elements of Information Theory*. Wiley, 1991.

[21] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. A*, vol. 454, no. 1971, pp. 903–995, 1998.

[22] V. Vapnik, *Statistical Learning Theory*, 1st ed. London: John Willey & Sons, 1998.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[24] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Prediction potential of different handwriting tasks for diagnosis of Parkinsons," in *IEEEE International Conference on e-Health and Bioengineering (EHB 2013)*, Iasi, Romania, November 2013.

A.25 Decision support framework for Parkinsons disease based on novel handwriting markers

Decision Support Framework for Parkinson's Disease Based on Novel Handwriting Markers

Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, *Senior Member, IEEE*, and Marcos Faundez-Zanuy, *Member, IEEE*

***Abstract*—Parkinson's disease (PD) is a neurodegenerative disorder which impairs motor skills, speech, and other functions such as behavior, mood, and cognitive processes. One of the most typical clinical hallmarks of PD is handwriting deterioration, usually the first manifestation of PD. The aim of this study is twofold: (a) to find a subset of handwriting features suitable for identifying subjects with PD and (b) to build a predictive model to efficiently diagnose PD. We collected handwriting samples from 37 medicated PD patients and 38 age- and sex-matched controls. The handwriting samples were collected during seven tasks such as writing a syllable, word, or sentence. Every sample was used to extract the handwriting measures. In addition to conventional kinematic and spatio-temporal handwriting measures, we also computed novel handwriting measures based on entropy, signal energy, and empirical mode decomposition of the handwriting signals. The selected features were fed to the support vector machine classifier with radial Gaussian kernel for automated diagnosis. The accuracy of the classification of PD was as high as 88.13%, with the highest values of sensitivity and specificity equal to 89.47% and 91.89%, respectively. Handwriting may be a valuable marker as a diagnostic and screening tool.**

***Index Terms*—Biomarker, decision support system, handwriting, Parkinson's disease (PD), tablet.**

I. Introduction

PARKINSON'S DISEASE (PD) is a progressive neurodegenerative disorder characterized by tremor, rigidity, bradykinesia, and loss of postural reflexes. With current prevalence rates, ranging from 10 to 800 people per 100 000, PD is one of the most common neurodegenerative disorders [1]. PD usually affects people at an average age of 60 [2], and prevalence rates are expected to increase further as the population ages. There is no objective quantitative method of clinical diagnosis. It is thought that PD can only be definitively diagnosed at postmortem, which further highlights the complexities of diagnosis. Clinicopathological studies show that up to 25% of the patients with PD are diagnosed incorrectly in the final stage of their disease, even by specialists in movement disorders [3]. Therefore, there is intensive effort to develop expert systems for the analysis and diagnosis of PD.

There have been several attempts to approach the assessment of PD through technology and to develop decision support tools for accurately diagnosing PD. This usually includes signal acquisition through wearable sensors during the Timed Up and Go clinical test [4], [5] or monitoring during free movement [6]. Another approach that attracted the attention of the speech processing community is based on recent findings that one frequent symptom of PD is significant vocal impairment [7], [8]. Research of automatic PD detection with machine learning tools using acoustic voice impairment measurement achieved promising results [9], [10]; the latest reported results showed as high as 98% overall classification accuracy [11].

It has been well documented that handwriting is affected in Parkinson's disease [12], [13], and some preliminary data suggest that handwriting might serve as a diagnostic marker for PD [14], [15]. Today there are wide range of high precision tablets and touch screen devices that make the acquisition and evaluation of handwriting signals very feasible. Moreover, there is no requirement for any special sensors and no need to solve the typical problems of acoustic signal acquisition and processing.

Handwriting is a highly skilled and complex coordinated motor activity requiring the dynamic interaction of the lower arm, wrist, and finger muscles. The writing process involves accurate sequencing and online scaling of automated movements and planning of subsequent strokes [16]. In healthy subjects, these movements are automated and do not require additional attentional resources. The handwriting of patients with PD is often characterized by decreased letter size, changes in kinematic aspects of movement including decreased velocities and acceleration, an increased number of changes in velocity, and increased movement time [17].

A number of studies have been performed assessing the kinematic aspects of movement execution during handwriting and showing the potential of digitized handwriting analysis to be more sensitive than the UPDRS [17], [18]. However, a complete picture of the extent to which any one measurement

Manuscript received November 08, 2013; revised May 20, 2014 and September 09, 2014; accepted September 15, 2014. Date of publication September 25, 2014; date of current version May 06, 2015. This work was supported by Projects GAP102/12/1104, COSTIC1206, NT13499 (Speech, its impairment and cognitive performance in Parkinson's disease), CZ.1.07/2.3.00/20.0094, Project CEITEC, Central European Institute of Technology (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund and by FEDER and Ministerio de Econom y Competitividad TEC2012-38630-C04-03. This work was performed in laboratories supported by the SIX Project registration CZ.1.05/2.1.00/03.0072, operational program Research and Development for Innovation. The work of P. Drotár was supported by Project CZ.1.07/2.3.00/30.0005 of Brno University of Technology.

P. Drotár, J. Mekyska, and Z. Smékal are with the Department of Telecommunications, Brno University of Technology, 61200 Brno, Czech Republic.

I. Rektorová and L. Masarová are with the First Department of Neurology, Faculty of Medicine, Applied Neuroscience Research Group, CEITEC MU, Masaryk University, 625 00 Brno, Czech Republic.

M. Faundez-Zanuy is with Escola Universitaria Politecnica de Mataro, Tecnocampus, 08302, Mataro, Spain.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNSRE.2014.2359997

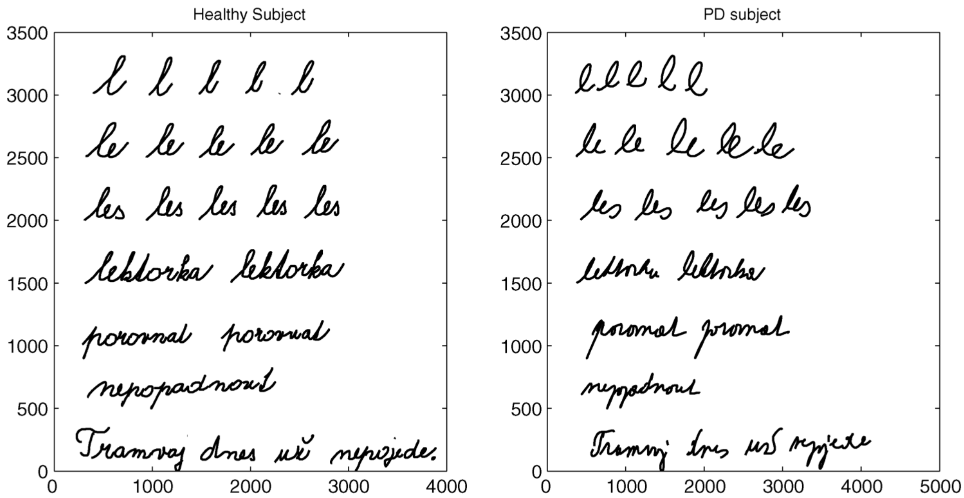


Fig. 1. Handwriting sample of healthy and PD subject.

or set of measurements is useful in discriminating PD has not yet been given. The most commonly measured parameters relate to the kinematics and dynamics of writing and drawing. To provide more insight and better understanding of the data, nonlinear, local structure-based measures can be used with advantage. Many measures based on time series analysis, different modeling techniques in frequency domain, and empirical mode decomposition were shown to be effective in various domains of biomedical research [9], [19]. We were therefore motivated to introduce several novel features describing the kinematic aspects of handwriting while also taking into account tremor, randomness, and hidden irregularities. These new features are based on the measures of entropy, energy, and empirical mode decomposition of the handwriting data.

A template for acquiring handwriting signals was proposed and used for diagnosis of PD. The support vector machines (SVM) algorithm was used to build a predictive model and identify subjects with PD. The reported results show that the proposed predictive model achieves medically relevant results in identifying subjects with PD.

The remainder of this paper is organized as follows. Sections II and III present the dataset and the methods used, respectively. Section IV provides a statistical analysis of handwriting features and selects a subset of the most significant features. Finally, Section V presents the experimental results, in Section VI is a discussion, and conclusions are given in the Section VII.

II. Data

Altogether, 37 PD (19 men/18 women) and 38 (20 men/18 women) age and gender matched healthy controls were enrolled at the First Department of Neurology, St. Annes University Hospital, Brno, Czech Republic. All subjects were right-handed, had completed at least ten years of education, and reported Czech as their first language. PD patients were examined only in their ON-state while on dopaminergic medication, i.e., 1–2 hours after taking their regular dose of dopaminergic medication. All patients were taking L-dopa COMT (catechol-o-methyl transferase) inhibitor and/or a dopamine agonist.

TABLE I
Parkinson's Handwriting Dataset Characteristics

	Age		UPDRS (part V)		Years since diag.		LED	
	mean	*std*	*mean*	*std*	*mean*	*std*	*mean*	*std*
PD	69.3	10.9	2.27	0.84	8.37	4.8	1373.4	714
H	62.4	11.3	-	-	-	-	-	-

The mean daily levodopa-equivalent dose (LED) [20] was 1373.4 ± 714 mg. Mean and standard deviation of age, Unified Parkinson's Disease Rating Scale-Part V Modified Hoehn and Yahr Staging score [21] and disease duration are summarized in Table I. The PD patients were diagnosed by an experienced neurologist following the clinical criteria of PD according to [3] and nondemented based on the clinicians judgment, caregivers interview, and the MMSE [22] score (>27 points).

Each subject was asked to complete a handwriting task according to the prepared template. The template was shown to patients and they were free to write the words without needing to follow the exact pattern. A completed task sheet is depicted in Fig. 1. The template consists of seven handwriting tasks (part of a more detailed test battery). From the first to the third task, participants wrote cursive letters or bi/tri-grams of letters. Similar tasks (the letter l-or repetitions of it) are commonly used for handwriting analysis [23]. The next three tasks involved words that can be written in one long stroke, i.e., the writing device is in continuous contact with the writing surface while writing these words. The words were written in Czech (the native language of participants) with the following translation to English: lektorka—teacher (female), porovnat—to compare, nepopadnout—to not catch. The final task involved a longer sentence that also allowed the capture of the effect of fatigue during writing (Tramvaj dnes už nepojede—The tram won't go today).

Handwritten signals were acquired using the digitizing tablet Intuos 4M (Wacom technology) in the x-y plane and in the pressure axis. The tablet itself does not provide visual feedback; therefore, it was overlaid with paper so that the ink pen could be held in a normal fashion and allow for full visual feedback during writing.

III. Methodology

The recording starts when the pen touches the surface of digitizer and finishes when the task is completed. The digitizing tablet captures the following dynamic features (time-sequences): x-coordinate ($x[n]$); y-coordinate ($y[n]$); time stamp ($t[n]$); and button status ($b[n]$). The x and y components are segmented into on-surface and in-air strokes and analyzed in terms of handwriting measures. The feature calculation stage involves the application of traditional and nonstandard measurement methods to process handwriting signals. Each method produces either a single value or a vector of numbers for each of 75 signals. If there is a resulting vector, we further compute six basic functionals (mean, median, standard deviation, 1st percentile, 99th percentile and 99th percentile—1st percentile).

A. Kinematic Measures

It has been shown that various kinematic aspects are affected in PD and several measures have been established to capture disruptions during handwriting [13]. Based on the literature survey, several kinematic and spatio-temporal parameters were considered for each task. These include: speed, number of changes in velocity(NCV)/acceleration(NCA), relative NCV/NCA, writing duration and length, stroke speed, velocity, acceleration, jerk, horizontal velocity/acceleration/jerk, vertical velocity/acceleration/jerk, and stroke height/width. Definition of these features can be found in [14] or in references therein.

B. Measures of Entropy and Energy

The digital representation of handwriting is a physiologically based time series that is the result of several interacting physiological mechanisms. Such signals contain complex fluctuations which could provide information related to underlying processes and states of the physiological system. Entropy features have the potential to uncover hidden complexities in the handwriting process. For example, tremor and irregular muscle contractions introduce randomness to the movements during handwriting; however, this randomness is difficult to analyze using only kinematic measures. In order to uncover hidden complexities, two types of features were extracted separately from the x-coordinate ($x[n]$) and y-coordinate ($y[n]$) signals: signal entropy and signal energy based features.

Entropy is a numerical measure of the randomness or uncertainty of a signal. The widely established and well known Shannon entropy of random variable X is defined as

$$H_S(X) = -\sum_{x \in X} p(x) \log_2 p(x) \tag{1}$$

where $p(x)$ is probability density function computed using kernel density estimation with a Gaussian kernel. The generalization of Shannon entropy, describing the diversity and randomness of the system, is Rényi entropy. The Rényi entropy is defined as follows:

$$H_{R,r}(X) = \frac{1}{1-r} \log \left(\sum_{i=1}^{n} p_i^r \right) \tag{2}$$

where $r \geq 0$ is order of Rényi entropy. In summary, Shannon entropy H_S and second and third order Rényi entropy $H_{R,2}, H_{R,3}$ were calculated for both $x[n]$ and $y[n]$.

In order to obtain signal to noise ratios of handwriting, conventional energy (CE) and Teager-Kaiser (TKE) energy operators were computed as follows. The conventional energy operator of signal $s[n]$ is defined as

$$\mathrm{CE}\,(s[n]) = \sum_{n=0}^{N-1} s[n]^2 \tag{3}$$

where N is the length of series $s[n]$. Similarly, Teager-Kaiser energy operators [24]

$$\mathrm{TKE}_r\,(s[n]) = \sum_{n=0}^{N-r-1} s[n]^2 - s[n+r] \cdot s[n-r] \tag{4}$$

where $r \geq 0$ is the order of the TKE operator. Having defined equations for signal energy, we still need to obtain the variable describing noise that is present in the signal. Note that what we refer to as noise is not unwanted interference from an external source, but the irregular movement resulting from muscle contractions and irregularities caused by cognitive impairment as a result of PD. Therefore, the noise variance estimation method of [25] is used to estimate noise variance $\mathcal{N}(s[n])$. Then, signal-to-noise ratios are defined as

$$\mathrm{SNR}_{\mathrm{CE}} = \frac{\mathrm{CE}\,(s[n])\,/N}{\mathcal{N}\,(s[n])} \tag{5}$$

and

$$\mathrm{SNR}_{\mathrm{TKE_r}} = \frac{\mathrm{TKE_r}\,(\mathrm{s[n]})\,/\mathrm{N}}{\mathcal{N}\,(\mathrm{s[n]})} \tag{6}$$

In the equations above, $s[n]$ is the signal under evaluation, i.e. x or y coordinates of handwriting.

C. Intrinsic Features

1) Empirical Mode Decomposition: Empirical mode decomposition (EMD) is an intuitive data-dependent decomposition of a time series that allows decomposition of nonlinear and nonstationary data into intrinsic mode functions (IMF). EMD is conducted by the iterative extraction based on the local representation of the signal as the sum of a local oscillating component and a local trend. The first IMF, representing the oscillations of the entire signal, is extracted in the first iteration. The difference between the original signal and the IMF time series is the residual. The same procedure is then applied on the residual to extract the second IMF and so on. The entire procedure can be summarized as follows [26].

1) For a given signal $s[n]$, identify all the local extrema (both minima and maxima).
2) Construct upper envelope $s_{\text{upper}}[n]$ by using a cubic spline to connect all the local maxima. In the same way, the lower envelope $s_{\text{lower}}[n]$ is obtained by connecting all the local minima.
3) Compute the mean of the upper and lower envelope as

$$m[n] = \frac{s_{\text{upper}}[n] - s_{\text{lower}}[n]}{2} \tag{7}$$

and extract $m[n]$ from $s[n]$ to obtain

$$s[n] \leftarrow s[n] - m[n]. \tag{8}$$

4) Check if updated $s[n]$ conforms to the properties of IMF. Otherwise repeat all the above steps.
5) Subtract the IMF$[n]$ from $s[n]$ to get the residue $r_1[n]$ and then iterate on the residue.

The whole process is stopped if $\sum_n((\text{IMF}_j[n] - \text{IMF}_{j-1}[n])/(\text{IMF}_{j-1}[n]))^2$ is less than 0.2 or 0.3.

Given a time series $s[n]$, combining all the IMFs gives the original signal

$$s[n] = \sum_{j=1}^{N} \text{IMF}_j[n] + r_N[n] \tag{9}$$

where N is the total number of IMFs obtained from the time series and $r_N[n]$ is the residual. The number of IMFs depends on the nature and length of the signal. Each IMF satisfies two basic assumptions [26]: (1) The number of maxima, which are strictly positive, and the number of minima, which are strictly negative, for each IMF, are either equal or differ by no more than one. (2) The mean value of the envelope, as defined by the maxima and the minima, for each IMF, is zero.

2) Extracting Intrinsic Features: From several algorithms for EMD computation, we used the freely available algorithm [27]. A number of discriminative temporal and spectral features were extracted from IMFs obtained from both normal and PD-affected handwriting signals.

Intrinsic entropies $IH_S(X)$ and $IH_{R,r}(X)$ were obtained by computing the entropies defined in Section III-B of the first and second IMF. The motivation for using only the first two IMFs is, that in practice, the first few IMFs contain only time-varying high spectral components representing the noise [9], [26], [27]. Since for PD patients the noise represents underlying irregularities, it is anticipated that the first pair of IMFs carries significant discriminative information with regard to handwriting of control and PD subjects.

It is possible to compute intrinsic energies ICE and ITKE$_r$ similarly to intrinsic entropies by applying (5) and (6) to the first and the second IMF.

Based on the assumption that higher order IMFs contain a main trend, or useful signal, different SNR measures are introduced as

$$\text{SNR}_\Gamma = \frac{\sum_{k=3}^{k=N_{\text{imf}}} \Gamma\left(\text{IMF}_k[n]\right)}{\sum_{k=1}^{k=2} \Gamma\left(\text{IMF}_k[n]\right)}. \tag{10}$$

Here, Γ is a symbolic representation of the intrinsic entropies and energies described above and N_{imf} is the number of intrinsic mode functions that were extracted.

D. Feature Selection

In order to reduce the dimensionality of input data and remove the nonrelevant features, the first stage was a statistical analysis of the data using the Mann-Whitney U-test. The Mann-Whitney U test is nonparametric statistical test used to assess whether two independent groups are significantly different from each other. The features that passed the Mann-Whitney U-test with a significance level less than 0.05 level were kept. After this preprocessing stage, 203 features were selected as the candidate subset for further processing and classification.

Feature selection algorithms aim to choose a small subset of features that ideally are necessary and sufficient to describe the target concept. From many feature selection algorithms, we decided to use the Relief algorithm [28], which has been shown to achieve promising results in problems similar to ours [11].

Relief is a feature-weighting algorithm that relies entirely on statistical analysis and employs only a few heuristics. It selects most of the relevant features even though only a small number of them are necessary for prediction. In most cases it does not help with redundant features. Since we want all the relevant features to be included for prediction, even at the cost of higher dimensionality, Relief was a promising candidate.

E. Support Vector Machines

The effectiveness of the selected subset of features in classifying PD and non-PD subjects was evaluated using nonlinear SVM. The underlying idea of SVM classifiers is to calculate a maximal margin hyperplane separating two classes of the data. To learn nonlinearly separable functions, the data are implicitly mapped via nonlinear mapping $\varphi(x)$ to a higher dimensional space by means of a kernel function, where a separating hyperplane is found. The equation of the hyperplane separating two differential classes is given by the relation

$$y(x) = W^T\varphi(X) = \sum_{j=1}^{K} \omega_j\varphi_j(x) + \omega_0 = 0 \tag{11}$$

where $W = [\omega_0, \omega_1, \ldots, \omega_K]$ is the weight vector of the network. New samples are classified according to the side of the hyperplane they belong to. We used the Radial Basis Function (RBF) kernel [29]. The RBF kernel is defined as

$$K(x, x_i) = e^{\frac{-\|x-x_i\|^2}{2\gamma^2}} \tag{12}$$

TABLE II
FOURTEEN FEATURES WITH LARGEST RELEVANCE TO CLASS LABEL SORTED ACCORDING TO SPEARMAN'S CORRELATION COEFFICIENT. ADDITIONALLY, MI AND SVM PREDICTION FOR PARTICULAR FEATURES ARE DISPLAYED

Feature	$\|\rho\|$	MI	SVM [%]
$H^x_{R,2}$, std, task 2	0.455	4.94	68.9
$H^x_{R,2}$, percentile 99th - percentile 1st, task 2	0.448	5.02	69.0
SNR^y_{ICE}, scalar, task 7	0.445	3.55	**77.2**
$H^x_{R,3}$, percentile 1st, task 2	0.435	5.54	**73.2**
$H^x_{R,3}$, std, task 2	0.423	5.26	70.4
SNR^x, std, task 7	0.421	4.99	**76.2**
$H^y_{R,3}$, std, task 2	0.421	5.26	69.2
SNR^x_{ICE}, scalar, task 7	0.417	4.68	72.1
SNR^x_{ICE} (dB), scalar, task 7	0.417	5.79	72.1
$H^y_{R,3}$, percentile 99th - percentile 1st, task 2	0.411	5.51	62.6
stroke length, percentile 99th - percentile 1st, task 2	0.411	4.91	67.2
stroke length, std, task 2	0.407	4.82	66.2
$H^x_{R,2}$, percentile 1st, task 2	0.407	5.27	71.6
H^y_S,std, task 2	0.405	4.13	71.4

where γ controls the width of RBF function.

The parameters kernel gamma γ and penalty parameter C were optimized using a grid search of possible values. Specifically, we searched over the grid (C, γ) defined by the product of the sets $C = [2^{-7}, 2^{-6}, \ldots, 2^6, 2^7]$, $\gamma = [2^{-7}, 2^{-6}, \ldots, 2^6, 2^7]$.

IV. PRELIMINARY STATISTICAL ANALYSIS

To obtain some preliminary insight into the statistical properties of handwriting features, we computed Spearman's correlation coefficients ρ and mutual information (MI) between feature vectors and associated responses. MI is a measure of the amount of information shared by two random variables X and Y. It is defined as

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \cdot \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) \tag{13}$$

where x and y are possible variable values with a joint probability distribution function $p(x,y)$ and marginal distribution functions $p(x)$ and $p(y)$, respectively [30]. We computed MI by evaluating the marginal entropies $H(X)$, $H(Y)$ and joint entropy $H(X,Y)$ as $I(X;Y) = H(X) + H(Y) - H(X,Y)$, where entropies are defined as in (1).

Table II summarizes 14 handwriting measures with the largest relevance to response sorted according to an absolute correlation coefficient. All of the correlations are statistically significant ($p < 0.005$). Most of the listed features are newly proposed features, providing some initial confidence in the relevance of the selected features.

Additionally, every feature was used separately as an input to the SVM classifier to evaluate its classification accuracy in separating PD and HC. The resulting individual classification accuracies of the features listed in Table II are listed in last column of the table. Note that these classification accuracies represent only the discriminative power of a single feature to separate PD from healthy subjects; any possible combination of features or causal relationships among features are not taken into account. The features that are among the ten features with the highest classification accuracy are marked in bold font. As can be seen from Table II, the highest classification accuracy of a single feature is over 76%, indicating that the classification task has a good chance of success.

The probability density functions of the 12 most highly ranked features from Table II are shown in Fig. 2. The vertical axes are the probability densities of the normalized measures estimated using kernel density estimation with Gaussian kernels. The curves of the major part of the handwriting measures for PD show a clear difference from the probability densities of healthy subjects.

V. EXPERIMENTAL RESULTS

The classification test performance was determined by the computation of accuracy, sensitivity and specificity. The accuracy (P_{acc}), sensitivity (P_{sen}) and specificity (P_{spe}) are defined as

$$P_{\text{acc}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \cdot 100\% \tag{14}$$

$$P_{\text{spe}} = \frac{\text{TN}}{\text{TN} + \text{FP}} \cdot 100\% \tag{15}$$

$$P_{\text{sen}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \cdot 100\% \tag{16}$$

where true positive (TP) and false positive (FP) represent the number of correctly classified PD subjects and the number of

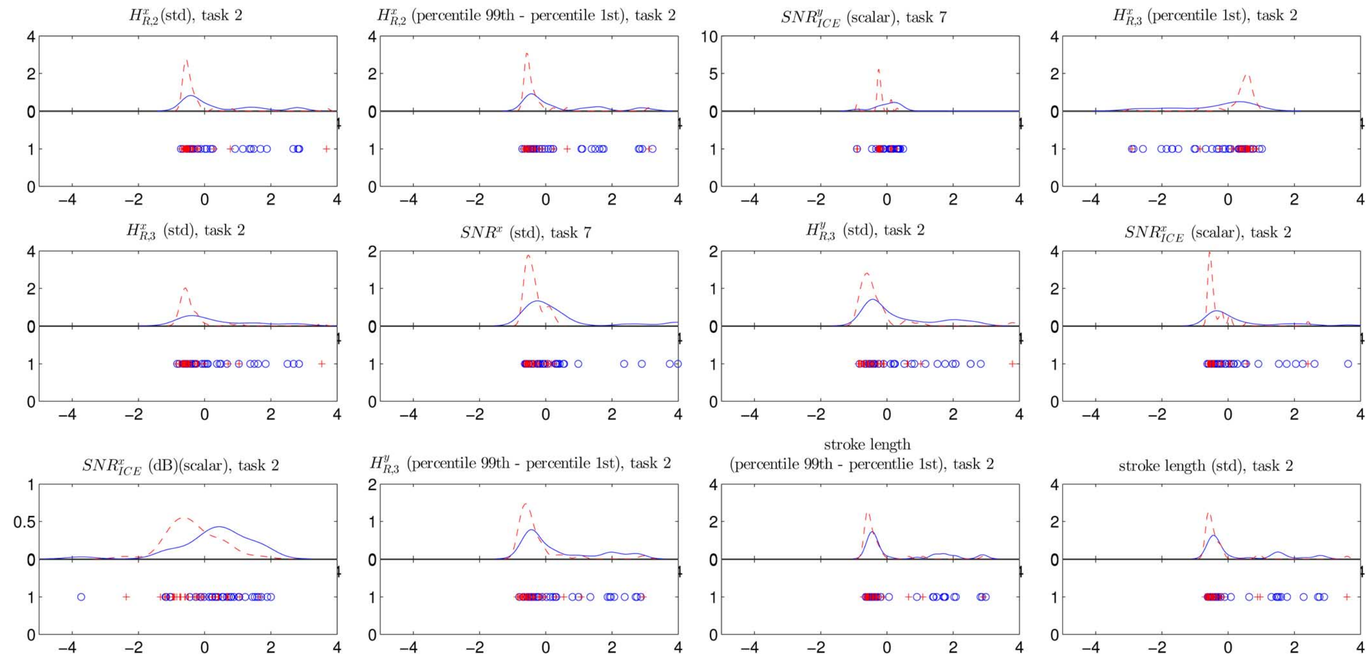


Fig. 2. Probability density functions for selected handwriting parameters. Solid blue lines are for PD patients, dashed red lines for HC subjects. Distribution of feature values is shown in bottom half of each figure.

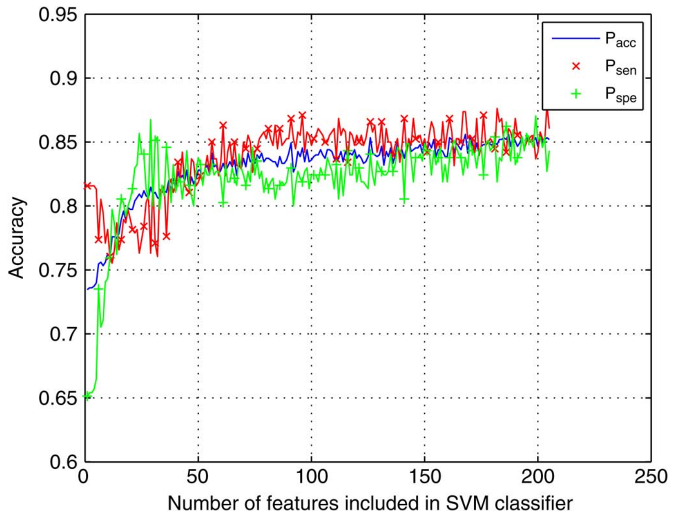


Fig. 3. Classification accuracy, sensitivity and specificity based on the features selected by Relief method. SVM with nonlinear kernel was used with 10-fold cross validation.

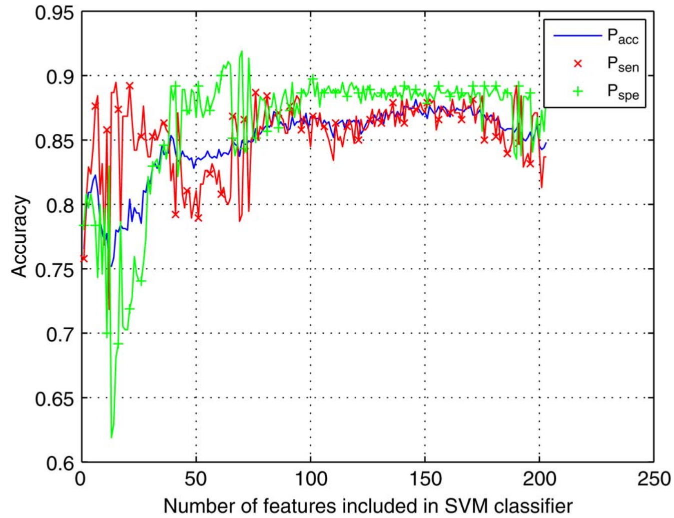


Fig. 4. Classification accuracy, sensitivity and specificity based on the features selected by SVM ranking method. SVM with nonlinear kernel was used with 10-fold cross validation.

actually healthy subjects diagnosed as PD, respectively. Similarly, true negative (TN) and false negative (FN) represent the total number of correctly classified healthy controls, and the PD patients incorrectly classified as healthy controls, respectively.

Classifier validation was conducted using stratified tenfold cross validation. The process was repeated ten times; in each repetition the original dataset was randomly permuted prior to splitting into training and testing subsets. Classification accuracy, sensitivity, and specificity over the ten repetitions were averaged. Training and testing features were normalized before classification on a per-feature basis to have zero mean and a standard deviation of one.

We computed classification accuracy, specificity, and sensitivity as the number of features is varied (Fig. 3). The features were selected by applying the Relief algorithm in each run. The highest accuracy $P_{\rm acc} = 85.6\%$ was achieved using 168 features. However, as can be seen from the figure, only slightly lower accuracy can be obtained when employing a considerably smaller number of features.

To provide another feature selection approach, we employed the strategy from Section IV. The features are ranked according to their classification accuracy, i.e., only n features that provided the highest individual classification accuracy are included in the SVM classifier. The highest classification accuracy using the SVM ranking approach was $P_{\rm acc} = 88.1\%$ for $N = 162$ features (see Fig. 4). As in the previous case, a smaller subset of features results in slightly worse classification accuracy.

VI. Discussion

Disturbances of motor control in PD, caused by the depletion of dopamine, affect the handwriting of PD patients. These disturbances involve the processing of motor planning, sequencing, movement initiation, and execution, and they result in hypometric movements, tremor and alterations in handwriting kinematics [13]. Modern technology makes it possible to capture handwriting so it can be incorporated into medical decision support systems. The acquisition of handwriting does not require any high quality controlled conditions and can be carried out at a clinic or at the patient's home. All of the data used in this study were collected in a clinic environment with a tablet connected to a notebook computer without any previous preparation in a room or special environment. The task performance is quite simple and natural and does not require any timing or exhaustive repetitions.

The proposed approach, based on handwriting, can be alternative or complementary to other approaches, such as currently popular speech assessment for PD classification [9], [10]. Recent studies show very encouraging results, providing as high as 97% PD classification accuracy. However, it should be noted that the first results in this area were significantly lower, identifying PD with around 90% accuracy [31]. The 90% classification accuracy presented by Little [31] is very similar to results presented in this study. We believe that further research will improve the performance of decision support systems based on handwriting. The application of this approach based on handwriting avoids the additional processing steps connected with speech processing, such as speech segmentation, noise removal, and requirements for a quality recording environment without external noises and disturbance. Other approaches, such as diagnosing PD from breath, do not achieve clinically relevant results and require dedicated sensors [32]. There is also huge potential in neuroimaging techniques, but these require expensive equipment and cannot be used for monitoring at a patient's home [33].

One of the main findings is that the features proposed in this study are relevant to PD diagnosis. They represent the majority of features that were ranked as the most correlated with class label. Additionally, when compared to our previous research [34], the inclusion of new features substantially improved classification accuracy. Since all of the PD subjects were on medication to correct symptoms of PD such as visible tremor, we had to focus on features that capture subtle signs of tremor. The tremor signals with irregularities are more similar to random signals. We used entropy to evaluate signals, where higher entropy means more irregular movement. We similarly computed the noise variance of a signal and obtained SNR indicating the amount of noise present in a signal. We assume that these features are related to the tremor; however, handwriting is a complex coordinated activity, where other factors can influence motor movement. In order to avoid or at least minimize other sources that could affect handwriting, only subjects without any history or presence of psychiatric symptoms or any disease affecting the central nervous system (other than PD in the PD cohort) and without impairment of the right upper extremity were included in this study. The fact that the proposed methodology is able to detect alterations in the handwriting of PD subjects on medication indicates the high sensitivity of the approach and its potential to discover even subtle changes in early stage PD. We did not find any significant correlations $(p < 0.005)$ between handwriting features and daily LED of medication.

The features with the highest relevance come mainly from two tasks: the second and the seventh ones. The seventh task (*Tramvaj dnes už nepojede.*) is specific in the sense that it cannot be written as one long stroke; it is also the longest task. It is well known that PD patient handwriting depends on a visual closed loop, whereas normal handwriting is automated and the movements are so fast that the normal feedback loop of visual perception and muscle control is disabled, resulting in an open loop configuration [17]. We assume that sequenced writing requires a higher programming load than more simple tasks. This probably poses more challenging conditions for PD patients and amplifies the influence of the disease on handwriting. The seventh task is probably the most representative task in terms of the observed clinical symptoms of PD. The second task (*le*) is similar to the tasks previously used in other studies to analyze handwriting. It usually consists of loops like *llll* and was proven to reflect handwriting alteration due to PD [13], [35]. When we compare the second task with two similar tasks, i.e., the first one and the third one, handwriting features obtained from the second task appear to be more contributing to overall classification performance. Even though these tasks are similar, there are significant differences that have to be considered. The second task contains letters of different sizes that may pose more demanding requirements for cognition. Comparing the second task to the other tasks, there is another important difference. The second task is a pseudoword not existing in Czech language, whereas tasks from 3 to 6 include words used in everyday language. Therefore, the performance of the second task is not automatized. The lateral premotor cortex is additionally activated when attention must be paid during a motor task. This notion may partially explain the fact that the second and the seventh tasks are more relevant to the class labels than the other tasks. This should be taken into consideration when preparing handwriting experiments, and besides conventional tasks such as ellipses or *llll* writing, one should include also more difficult tasks.

This study shows that handwriting is a promising biomarker that might be used as an early marker of PD. Our analysis showed that there is a link between the proposed features such as entropies and signal-to-noise ratios and motor symptoms of PD. Further studies in drug-naive, newly diagnosed cases of PD are warranted. Future studies performed in PD and atypical Parkinson syndromes will show whether handwriting could also serve as a possible biomarker for differential diagnosis of parkinsonism and help to explain how nonmotor features such as cognitive impairment relate to specific handwriting alterations. From an implementation point of view, there is no small subset of features for binary classification tasks as in the case of speech processing for PD diagnosis [11]. However, the results presented in Figs. 3 and 4 illustrate that most of the prediction performance comes from the group of 40–50 features. Adding more features increases prediction accuracy only by few perceptual points. This smaller group of features can be more conveniently implemented and used for monitoring. Especially entropy and signal-to-noise ratio features appear to sufficiently reflect alterations of PD handwriting. Finally, future studies could incorporate the prospective acquisition of samples from PD subjects to monitor the progression of handwriting impairment with the disease. Intraindividual variability (test-re-test variability) should also be evaluated.

VII. Conclusion

We proposed an assessment of the practical clinical value of kinematic and novel handwriting features for identifying subjects with PD by using simple, easy-to-perform handwriting tasks. Newly introduced handwriting measures show significant relevance to the PD-score, a binary score indicating whether a sample belongs to a person with PD. The accuracy using our method is over 88% with very similar values for specificity and sensitivity. We emphasize that all PD subjects were examined in their ON motor state while on their regular dopaminergic treatment. This confirms that handwriting is a medically relevant biomarker for classifying of Parkinson's disease. We believe our approach can be complementary to numerous speech or movement based discriminant analyses of PD patients and deserves further attention. Future studies should further explore the associations between cognitive/motor aspects of PD and handwriting parameters during sentence handwriting, both in the ON and OFF medication conditions, in order to shed further light on the sensitivity, specificity, and underlying mechanisms.

References

[1] S. von Campenhausen, B. Bornschein, R. Wick, K. Btzel, C. Sampaio, W. Poewe, W. Oertel, U. Siebert, K. Berger, and R. Dodel, "Prevalence and incidence of Parkinson's disease in Europe," *Eur. Neuropsychopharm.*, vol. 15, no. 4, pp. 473–490, 2005.

[2] L. M. de Lau and M. M. Breteler, "Epidemiology of Parkinson's disease," *Lancet Neurol.*, vol. 5, no. 6, pp. 525–535, 2006.

[3] A. J. Hughes, S. E. Daniel, Y. Ben-Shlomo, and A. J. Lees, "The accuracy of diagnosis of Parkinsonian syndromes in a specialist movement disorder service," *Brain*, vol. 125, no. 4, pp. 861–870, 2002.

[4] B. Mariani, M. Jiménez, F. Vingerhoets, and K. Aminian, "On-shoe wearable sensors for gait and turning assessment of patients with Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 155–158, Jan. 2013.

[5] L. Palmerini, S. Mellone, G. Avanzolini, F. Valzania, and L. Chiari, "Quantification of motor impairment in Parkinson's disease using an instrumented timed up and go test," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 4, pp. 664–673, Nov. 2013.

[6] A. Salarian, H. Russmann, C. Wider, P. Burkhard, F. J. G. Vingerhoets, and K. Aminian, "Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, pp. 313–322, Feb. 2007.

[7] A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of patients with Parkinson's disease," *Behav. Neurol.*, vol. 11, no. 3, pp. 131–137, 1998.

[8] K. M. Rosen, R. D. Kent, A. L. Delaney, and J. R. Duffy, "Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers," *J. Speech Lang. Hearing Res.*, vol. 49, no. 2, pp. 395–411, 2006.

[9] A. Tsanas, M. Little, P. McSharry, and L. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *J. R. Soc. Interface*, no. 8, pp. 842–855, 2011.

[10] B. Sakar, M. Isenkul, C. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE Biomed. Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013.

[11] A. Tsanas, M. Little, P. McSharry, J. Spielman, and L. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, May 2012.

[12] E. Nackaerts, G. Vervoort, E. Heremans, B. C. Smits-Engelsman, S. P. Swinnen, and A. Nieuwboer, "Relearning of writing skills in Parkinson's disease: A literature review on influential factors and optimal strategies," *Neurosci. Biobehav. Rev.*, vol. 37, no. 3, pp. 349–357, 2013.

[13] O. Tucha, L. Mecklinger, J. Thome, A. Reiter, G. L. Alders, H. Sartor, M. Naumann, and K. W. Lange, "Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease," *J. Neural Transm.*, vol. 113, no. 5, pp. 609–623, 2006.

[14] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "A new modality for quantitative evaluation of Parkinson's disease: In-air movement," in *Proc. 13th IEEE Int. Conf. BioInformatics and BioEngineering (BIBE)*, Chania, Greece, Nov. 2013.

[15] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, and I. Schlesinger, "Handwriting as an objective tool for Parkinson's disease diagnosis," *J. Neurol.*, pp. 1–5, 2013.

[16] H.-L. Teulings, J. L. Contreras-Vidal, G. E. Stelmach, and C. H. Adler, "Parkinsonism reduces coordination of fingers, wrist, arm in fine motor control," *Exp. Neurol.*, vol. 146, no. 1, pp. 159–170, 1997.

[17] K. W. Lange, L. Mecklinger, S. Walitza, G. Becker, M. Gerlach, M. Naumann, and O. Tucha, "Brain dopamine and kinematics of graphomotor functions," *Hum. Mov. Sci.*, vol. 25, no. 4–5, pp. 492–509, 2006.

[18] K. Stanley, J. Hagenah, N. Bruggemann, K. Reetz, L. Severt, C. Klein, Q. Yu, C. Derby, S. Pullman, and R. Sanders-Pullman, "Digitized spiral analysis is a promising early motor marker for Parkinson disease," *Parkinsonism Relat. Disord.*, vol. 16, no. 3, pp. 233–234, 2010.

[19] R. J. Martis, U. R. Acharya, J. H. Tan, A. Petznick, R. Yanti, C. K. Chua, E. Y. K. Ng, and L. Tong, "Application of empirical mode decomposition (EMD) for automated detection of epilepsy using EEG signals," *Int. J. Neural Syst.*, vol. 22, no. 06, 2012.

[20] C. L. Tomlinson, R. Stowe, S. Patel, C. Rick, R. Gray, and C. E. Clarke, "Systematic review of levodopa dose equivalency reporting in Parkinson's disease," *Mov. Disord.*, vol. 25, no. 15, pp. 2649–2653, 2010.

[21] S. Fahn and R. L. Elston, "Unified Parkinson's disease rating scale," in *Recent Developments in Parkinson's Disease*, C. D. Marsden, M. Goldstein, and C. D. Calne, Eds. New York, NY, USA: Macmillan, 1987, vol. 2, pp. 153–163.

[22] S. Folfstein, M. F. Folfstein, and P. McHugh, "Mini-mental state: A practical method for grading the cognitive state of patients for clinician," *J. Psychiatr. Res.*, vol. 12, pp. 189–198, 1975.

[23] C. Bidet-Ildei, P. Pollak, S. Kandel, V. Fraix, and J.-P. Orliaguet, "Handwriting in patients with Parkinson disease: Effect of l-dopa and stimulation of the sub-thalamic nucleus on motor anticipation," *Hum. Mov. Sci.*, vol. 30, no. 4, pp. 783–791, 2011.

[24] J. Kaiser, "On a simple algorithm to calculate the energy of a signal," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing . ICASSP-90*, 1990, pp. 381–384.

[25] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Comput. Stat. Data Anal.*, vol. 54, no. 4, pp. 1167–1178, 2010.

[26] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. A*, vol. 454, no. 1971, pp. 903–995, 1998.

[27] Y. Kopsinis and S. McLaughlin, "Investigation and performance enhancement of the empirical mode decomposition method based on a heuristic search optimization approach," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 1–13, Jan. 2008.

[28] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proc. Ninth Int. Workshop Machine Learning*, 1992, ser. ML92, pp. 249–256, San Francisco, CA, USA: Morgan Kaufmann.

[29] V. Vapnik, *Statistical Learning Theory*, 1st ed. London, U.K.: Wiley, 1998.

[30] T. M. Cover, *The Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[31] M. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.

[32] U. Tisch, I. Schlesinger, R. Ionescu, M. Nassar, N. Axelrod, D. Robertman, Y. Tessler, F. Azar, A. Marmur, J. Aharon-Peretz, and H. Haick, "Detection of Alzheimers and Parkinson's disease from exhaled breath using nanomaterial-based sensors," *Nanomed.*, vol. 8, no. 1, pp. 43–56, 2013.

[33] Y. Wu, W. Le, and J. Jankovic, "Preclinical biomarkers of Parkinson disease," *Arch. Neurol.*, vol. 68, no. 1, pp. 22–30, 2011.

[34] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Prediction potential of different handwriting tasks for diagnosis of Parkinsons," in *Proc. IEEE Int. Conf. e-Health and Bioengineering (EHB 2013)*, Iasi, Romania, Nov. 2013.

[35] A. W. A. Van Gemmert, C. H. Adler, and G. E. Stelmach, "Parkinsons disease patients undershoot target size in handwriting and similar tasks," *J. Neurology, Neurosurgery Psychiatry*, vol. 74, no. 11, pp. 1502–1508, 2003.

Peter Drotár was born in 1984 in Košice, Slovakia. He received the M.Sc. degree in electronics and telecommunications from Technical University of Košice, in 2007, and the Ph.D. degree from the same university in 2010.

From 2010 to 2012, he worked as a Scientist for Honeywell Advanced Technology Europe. Currently, he is a Postdoctoral Researcher at the Brno University of Technology, Czech Republic. His research interests include handwriting and biomedical signal processing. He also deals with machine learning and its application in decision support systems for biomedicine.

Jiří Mekyska was born in Holesov, Czech Republic. He received the Ph.D. degree in telecommunications Brno University of Technology (BUT), in 2014.

Currently, he is a Researcher and member of Signal Processing Laboratory (SPLab) at the Department of Telecommunications at BUT. His research interests include the field of speech signal analysis and its application in biomedicine. He deals with the hand-written text analysis, thermal image processing and information theory as well.

Irena Rektorová graduated from the Medical Faculty, Charles University, Prague, Czech Republic, in 1994.

She is a Professor of Neurology at the Department of Neurology, School of Medicine, Masaryk University, Brno, Czech Republic. She is Head of the Centre for Cognitive Impairment and Dementia at the Department of Neurology, St. Anne's Teaching Hospital, and Head of the Applied Neuroscience Research Group, Central European Institute of Technology, Masaryk University, Brno. She is a board-certified neurologist and her main interests lie in movement disorders and Parkinson's disease in particular, cognitive impairment and dementia.

Dr. Rektorová is a Secretary of the International Association of Parkinsonism and Related Disorders.

Lucia Masarová was born in Bojnice, Slovakia. Currently she is a student of the Medical Faculty of Masaryk University, Brno, Czech Republic.

Since 2012, she has been a Researcher at Applied Neuroscience Research Group, Central European Institute of Technology, Masaryk University, Brno. She is interested in neurodegenerative diseases especially Parkinson's disease.

Zdeněk Smékal (M'98–SM'06) received a professor degree in 2000 in the field of theoretical electrical engineering.

He is a Full Professor at the Faculty of Electrical Engineering and Communications, Brno University of Technology (BUT), Brno, Czech Republic. He is an author of three monographs and more than 150 papers for international conferences and journals.

Dr. Smékal is a member of the Audio Engineering Society. He is also a member of scientific councils (FEEC in BUT Brno and FEE in CTU Prague).

Marcos Faundez-Zanuy (M'99) was born in Barcelona, Spain. He received the B.Sc. degree in telecommunication in 1993 and the Ph.D. degree in 1998, both from the Polytechnic University of Catalunya.

He is now a Full Professor at EUP Mataro since 2010 and heads the Signal Processing Group there. Since 2009 he has been the Dean of Escola Universitaria Politecnica de Mataro (Polytechnic University of Catalunya) and responsible for Tecnocampus Research. His research interests lie in the fields of biometrics and speech coding. He is author of more than 30 papers indexed in ISI Journal citation report, more than 100 conference papers, ten books, and responsible for ten national and European research projects.

A.26 Hodnocení písma u pacientu s Parkinsonovou nemocí

Hodnocení písma u pacientů s Parkinsonovou nemocí

Assessing Handwriting in Patients with Parkinson's Disease

Souhrn

Cíl: Cílem naší studie bylo kvantitativně vyhodnotit poruchy písma u pacientů s Parkinsonovou nemoci (PN) ve srovnání s věkově a pohlavím vázanými zdravými kontrolami (ZK) pomocí digitalizačního tabletu. *Soubor a metoda:* Prospektivně jsme zařadili 40 pacientů s PN (průměrný věk 68,6 ± 11,36 let, délka trvání nemoci 8,02 ± 4,79 let) a 40 věkem a pohlavím vázaných ZK (průměrný věk 62,55 ± 11,22 let). Všichni jedinci byli praváci bez přítomnosti deprese či demence. Každý subjekt podstoupil sedm cvičení pro vyšetření písma a kresbu Archimédovy spirály a elips s pomocí digitalizačního tabletu. Byly hodnoceny rychlostní parametry mikrografie a kresby při pohybu pera po tabletu i nad tabletem. Pro statistickou analýzu dat jsme použili Mann-Whitneyho U test a Spearmanovy korelace s korekcí na opakovaná měření (Benjamini-Hochbergova metoda). *Výsledky:* U PN ve srovnání se ZK jsme při psaní na tabletu zjistili statisticky významné snížení v parametrech: okamžitá rychlost, okamžité zrychlení, okamžitá změna zrychlení v čase. Změny se zvýrazňovaly s délkou psaného segmentu. Ještě významnější byly rozdíly mezi oběma skupinami při hodnocení pohybu pera nad tabletem, tj. před vlastním zahájením psaní, při přípravě na pohyb. Zaznamenali jsme pokles sledovaných hodnot až o 20 % ve srovnání se ZK. *Závěr:* U pacientů s PN jsme prokázali specifické změny nejen při vlastním psaní, ale i ve fázi přípravy na psaní, které lze kvantifikovat pomocí digitalizačního tabletu. Výsledky studie mohou mít přímý klinický dopad: umožní nám studovat mikrografii jako možný časný klinický marker rozvoje PN.

Abstract

Aim: The aim of this study was to assess micrographia in patients with Parkinson's disease (PD) as compared to healthy controls (HC) using a digitizing tablet. *Methods:* We included 40 PD (mean 68.6 ± 11.36 years, duration of illness 8.02 ± 4.79 years) and 40 age- and sex-matched HC (mean 62.55 ± 11. 22 years). All subjects were right-handed, without the presence of depression or dementia. Each subject underwent seven exercises for writing and drawing of Archimedes spiral and ellipses using a digitizing tablet. The speed parameters of micrographia and drawing during the movement of a pen in the air and on the tablet were evaluated. The Mann-Whitney U test, Spearman correlation and Benjamini-Hochbergs method were used for statistical data analysis. *Results:* A statistically significant reduction in parameters of velocity, acceleration, and jerk was found when comparing both groups during writing. Changes were more pronounced with increased length of the written segment. The differences between the two groups were more pronounced when the in-air movements were assessed, i.e. during movement preparation. The values decreased up to 20% compared to HC. *Conclusion:* PD-specific changes assessed with a digitizing tablet were demonstrated not only during writing but also during preparation for writing. The results of the study may have a direct clinical impact: further research into its use as a clinical marker of early PD is likely to follow.

Seznam použitých zkratek

CNS	centrální nervová soustava
COMT	catechol-o-metyltransferáza
FDR	False Discovery Rate
LED	L-dopa Equivalent daily Dose
LHK	levá horní končetina
M	muži
MKN	mezinárodní klasifikace nemocí
MMSE	Mini-Mental State Examination
NDAT	Neurological Disorder Analysis Tool
PHK	pravá horní končetina
PN	Parkinsonova nemoc
UPDRS V	Unified Parkinson's Disease Rating Scale V – Modified Hoehn and Yahr Staging
ZK	zdravá kontrola
Ž	ženy

Autoři deklarují, že v souvislosti s předmětem studie nemají žádné komerční zájmy.
The authors declare they have no potential conflicts of interest concerning drugs, products, or services used in the study.
Redakční rada potvrzuje, že rukopis práce splnil ICMJE kritéria pro publikace zasílané do biomedicínských časopisů.
The Editorial Board declares that the manuscript met the ICMJE "uniform requirements" for biomedical papers.

L. Masárová[1], P. Drotár[3], J. Mekyska[3], Z. Smékal[3], I. Rektorová[1,2]

[1] I. neurologická klinika LF MU a FN u sv. Anny v Brně
[2] Výzkumná skupina Aplikované neurovědy, CEITEC – Středoevropský technologický institut MU, Brno
[3] Ústav telekomunikací, FEKT, VUT v Brně

✉
prof. MUDr. Irena Rektorová, Ph.D.
I. neurologická klinika
LF MU a FN u sv. Anny v Brně
Pekařská 53
656 91 Brno
e-mail: irena.rektorova@fnusa.cz

Přijato k recenzi: 13. 1. 2014
Přijato do tisku: 17. 4. 2014

Klíčová slova
mikrografie – Parkinsonova nemoc – písmo

Key words
micrographia – Parkinson's disease – writing

Podpořeno grantem IGA MZČR NT/ 13499: Řeč, její poruchy a kognitivní funkce u Parkinsonovy nemoci. Peter Drotár je finančně podporován projektem CZ.1.07/2.3.00/30.0005 Vysokého učení technického v Brně.

Úvod

Ručně psaný projev je charakteristickým rysem každého z nás a vyvíjejí se v průběhu celého života. V časném dětském věku, když se začíná dítě učit první motorické vzorce, píše/kreslí velmi pomalu, přičemž se soustřeďuje na každý detail [1]. V období dospělosti jsou už jemné motorické pohyby považovány za automatické, tzv. open loop, u lidí ve vyšším věku opět přestávají být motorické pohyby automatické, jsou tzv. closed loop, a starší lidé se musejí při psaní více soustředit [2–4]. Ručně psaný projev se mění také v důsledku onemocnění postihujících motorický systém, jejichž projevy je možné detekovat např. moderními digitalizačními tablety. Jedná se především o neurologická onemocnění z okruhu tzv. extrapyramidových onemocnění mozku nebo dle anglické klasifikace tzv. movement disorders.

Parkinsonova nemoc (PN) je progresivní neurodegenerativní onemocnění mozku, u něhož je typicky přítomna degenerace dopaminergních buněk v pars compacta substantiae nigrae, ale s progresí onemocnění dochází také k dalším neurotransmiterovým deficitům. PN se projevuje zejména hybnými příznaky, ke kterým patří klidový tremor hlavně akrálních částí horních a dolních končetin, svalová rigidita (ztuhlost), bradykineze (zpomalenost a snížení amplitudy pohybů) a posturální instabilita s poruchou stoje a chůze. Bradykineze je hlavním projevem každého pacienta s PN. Hybné projevy mají asymetrickou distribuci, tj. jsou vyjádřeny více na pravostranných nebo levostranných končetinách, a tato asymetrie v tíži symptomatiky zůstává i během progrese onemocnění.

Častým motorickým symptomem je také tzv. mikrografie – specifická porucha písma. Mikrografie je klinický příznak spojený hlavně s PN [5], ale může se vyskytnout např. i u dalších neurodegenerativních onemocnění mozku s parkinsonizmem označovaných jako onemocnění z okruhu „Parkinson plus", u Huntingtonovy choroby [6]. Mikrografie je definována jako zhoršení jemné motoriky projevující se zejména progresivní redukcí amplitudy (výšky) písma [7–9] a snížením tempa psaní. Mohou se objevit zubaté kontury a neobvyklé fluktuace v rychlosti psaní [10]. Písmo může být u části pacientů rušeno též třesem, dyskinezemi, únavou a mohou se objevovat symptomy typické pro poruchy chůze u PN – motorické zárazy (freezing), porucha iniciace pohybu (start hesitation) a redukce automatických pohybů [11–14]. Snižuje se schopnost naučit se nové automatické pohyby [15], což se může projevit třeba při kresbě elipsy. Mikrografie je způsobena zejména bradykinezí a svalovou rigiditou.

Předchozí studie zaměřené na studium písma u pacientů s PN hodnotily zejména následující úkoly: psaní „llllllll" a „lilili" [16], kreslení koncentrických kruhů [2]. Kromě digitalizačního tabletu byl ručně psaný projev analyzován také elektronickým perem, přičemž u nich byly nahrávány různé obrazce, smyčky, slova a celá věta [17]. Rovněž byla testována mikrografie s využitím standardního diametrického kuličkového pera na linkovaný papír při psaní písmen s ohledem na subjektivní vnímání mikrografie pacienty [5].

Cílem naší studie bylo zkoumat deficit grafomotorických poruch v oblasti formální složky psaného projevu bez ohledu na zkoumání poruch obsahových a také zjistit, které ze sledovaných paraklinických parametrů souvisejí s bradykinezí a v kterých konkrétních úlohách nejlépe odliší psaní/kresbu pacientů s PN na dopaminergní terapii od písma/kresby věkově a pohlavím vázaných ZK. Na rozdíl od předchozích studií jsme vyšetřili dostatečně velký počet subjektů (40 pacientů s PN, 40 ZK), použili jsme komplexní vyšetřovací protokol sestávající ze sedmi úloh hodnotících písmo a kresbu spirály a navíc jsme se zaměřili i na doposud nestudované parametry, které přinášejí informace o pohybu pera v momentě, kdy se nedotýká plochy tabletu, tedy když pacient nepíše/nekreslí (tzv. in-air movement), ale připravuje se na pohyb. Dále jsme studovali vztah mezi tíží mikrografie a dávkou dopaminergní medikace a délkou PN. Do studie byli zařazeni pouze praváci. Vedlejším cílem naší studie bylo testování hypotézy, zda pacienti s PN s počátkem nemoci vpravo budou mít výraznější příznaky ve srovnání s pacienty s počátkem PN vlevo.

Tab. 1. Charakteristika ZK a pacientů s PN.

	PN	ZK
počet	40	40
pohlaví	20 Ž, 20 M	20 Ž, 20 M
průměrný věk	68,6 ± 11,36 let	62,6 ± 11,2 let
dominantní končetina	pravá	
stav hybnosti	ON	
LED	1 373,4 ± 714 mg	
UPDRS V	2,25 ± 0,816 bodů	
délka PN	8,02 ± 4,79 let	
první příznak	LHK – 17 PHK – 20 symetricky – 3	

Soubor a metoda

Charakteristika ZK a pacientů s PN je uvedena v tab. 1. Prospektivní studie se zúčastnilo 80 subjektů, z toho 40 s diagnózou PN (20 mužů, 20 žen, průměrný věk 68,6 ± 11,36 let, délka trvání nemoci 8,02 ± 4,79 let) dle klinických diagnostických kritérií [18] a 40 věkem a pohlavím vázaných ZK (20 mužů, 20 žen, průměrný věk 62,55 ± 11,22 let). Mateřským jazykem všech byla čeština, ukončili alespoň základní vzdělání, všichni byli praváci bez postižení pravé horní končetiny a bez přítomnosti deprese či demence (dle MKN-10 kritérií, MMSE – Mini-Mental State Examination > 25 bodů).

Pacienti byli sledováni v Centru pro abnormální pohyby a parkinsonizmus při I. neurologické klinice LF MU ve FN u sv. Anny. Všichni byli během vyšetření na dopaminergní medikaci a v dobrém hybném stavu (on stavu) bez dyskinezí a bez třesu, cca 1–2 hod po podání dopaminergní medikace. Vybírali jsme pacienty bez tremor-dominantní formy onemocnění. Na základě lékařské dokumentace a pohovoru s pacientem jsme získali údaje o prvním příznaku PN a délce onemocnění. Tíže onemocnění jsme hodnotili pomocí standardizované škály UPDRS V (Unified Parkinson's Disease Rating Scale V – Modified Hoehn and Yahr Staging) [19] a každému pacientovi jsme vypočítali celkovou denní dávku dopaminergní medikace přepočtenou na ekvivalenty levodopy (LED, L-dopa Equivalent daily Dose) [20]. ZK sestávaly z hospitalizovaných pacientů I. neurologické kliniky FN u sv. Anny v Brně, jednalo se zejména o pacienty s vertebrogenním onemocněním. Žádný pacient ve skupině ZK neměl (ani anamnesticky) onemocnění CNS.

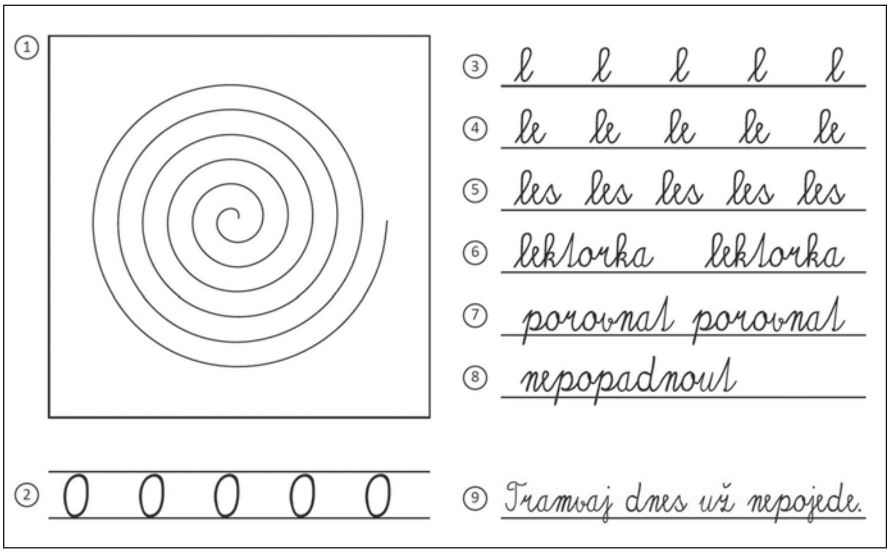
Popis metody

Všichni pacienti a kontroly byli nejprve vyšetřeni neurologem, obeznámeni s průběhem vyšetření ručně psaného projevu. Před vyšetřením podepsali informovaný souhlas, který byl schválen v rámci studie „Řeč, její poruchy a kognitivní funkce u pacientů s Parkinsonovou nemocí" etickou komisí FN u sv. Anny v Brně. Vyšetření probíhalo bez předchozího nácviku.

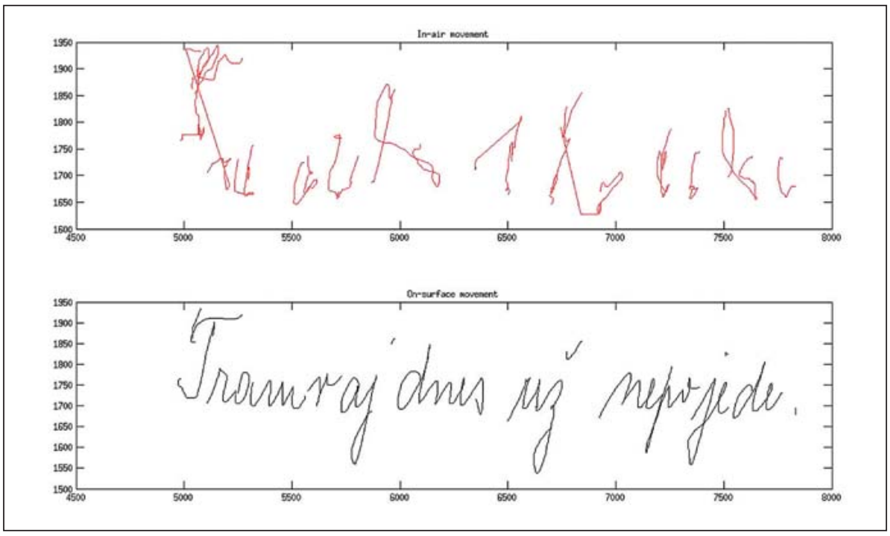
Každý subjekt podstoupil komplexní protokol obsahující devět cvičení (obr. 1), který jsme sestavili na základě dostupné literatury a jenž zahrnoval:

a) Archimédovu spirálu [21–25] – sledujeme tremor, dynamiku pohybu v horizontálním a vertikálním směru,
b) elipsu [2,26] – opět sledujeme hlavně dynamiku pohybu a schopnost udržet pohyb po delší čas (vícenásobné opisování spirál),
c) cvičení 3–5 [12,16,17,27–31] – kombinace slovních spojení s písmenem „l" – změny v rychlosti a akceleraci,
d) cvičení 6–8: slova – jedná se o přechod od jednoduchého po složité slovo – sledujeme schopnost psaní jedním tahem, tedy počet přerušení,
e) cvičení 9: [4,17,32] věta – sledujeme pohyb ruky nad tabletem, akceleraci, počet přerušení, dobu psaní.

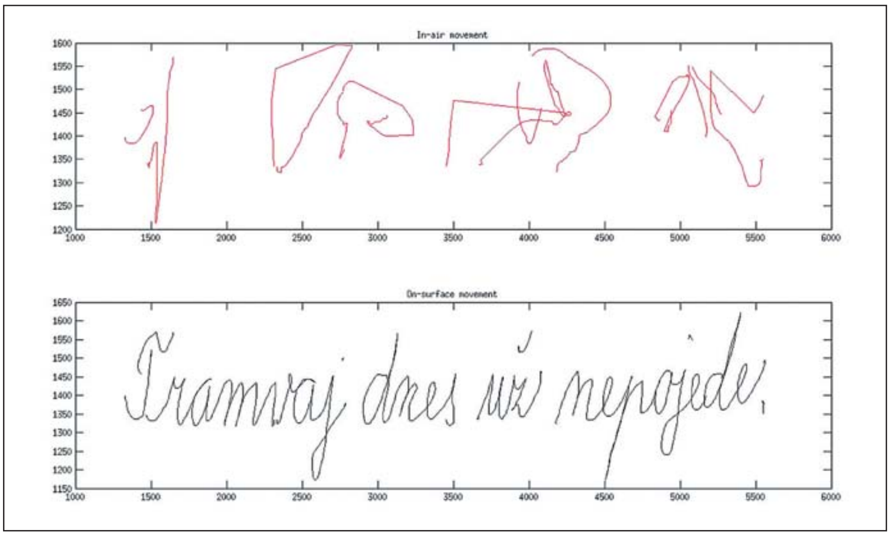
Samotný záznam písma probíhal pomocí digitalizačního tabletu Wacom, Intuos 4M – medium, sériové číslo: 2DBH016667. Výhodou tohoto typu tabletu je, že pacient k psaní využívá inkoustové pero, kterým píše na papír ve zcela přirozených podmínkách, a zároveň má



Obr. 1. Ukázka protokolu.



Obr. 2a) Ukázka pohybu nad tabletem u pacienta s PN.



Obr. 2b) Ukázka pohybu nad tabletem u ZK.

Tab. 2a) Opisování písmene „l“ (ZK vs PN).

Parametr	PN medián [× 10e-3]	ZK medián [× 10e-3]	Relativní rozdíl PN < ZK (v %)	p hodnota
průměrný počet změn v rychlosti	2,4	2,7	7,8	0,05
průměrný počet změn ve zrychlení	2,4	2,7	8,4	0,045
průměrná hodnota okamžité rychlosti	490	640	5,3	0,025
průměrné zrychlení	65,0	90,0	5,4	0,03
průměrný jerk (změna zrychlení v čase)	8,9	12,0	5,5	0,02

Tab. 2b) Opisování slabiky „les“ (ZK vs PN).

Parametr	PN medián [× 10e-3]	ZK medián [× 10e-3]	Relativní rozdíl PN < ZK (v %)	p hodnota
průměrná hodnota okamžité rychlosti	506	600	0,02	8,4
průměrné zrychlení	69,5	82,5	0,01	8,6
průměrný jerk	9,4	11,1	0,009	8,8
průměrná horizontální rychlost	107,3	125,3	0,05	5,0
průměrné horizontální zrychlení	14,4	16,8	0,04	4,8
průměrný horizontální jerk	1,9	2,3	0,04	5,0
průměrná vertikální rychlost	6,0	10,2	0,02	29,4
průměrné vertikální zrychlení	0,8	1,4	0,02	29,1
průměrný vertikální jerk	0,11	0,18	0,03	30,2

okamžitou vizuální odezvu. Jedná se o digitalizační tablet, jenž snímá s frekvencí 200 vzorků za sekundu horizontální a vertikální pozici hrotu pera, tlak na hrot pera, azimut a sklon pera. Snímaly se též informace o pohybu pera (horizontální a vertikální pozice hrotu pera), když se nedotýkal tabletu (in-air movement), přičemž subjekty nebyly poučeny o tom, že sledujeme i tyto pohyby. Ilustrace zachycených trajektorií je uvedena na obr. 2a (PN), obr. 2b (ZK). Soubor ze získané databáze se zpracovával nástrojem NDAT (Neurological Disorder Analysis Tool) na spolupracujícím pracovišti – Ústavu telekomunikací FEKT, Vysokého učení technického v Brně. Zaměřili jsme se na parametry související s bradykinezí a fluktuacemi tempa pohybů – tj. rychlostí a jejími derivacemi (změnou rychlosti v čase, tj. zrychlením a změnou zrychlení v čase, tj. tzv. jerk).

Statistické zpracování dat

Vzhledem k nenormálnímu rozložení dat jsme použili Mann-Whitneyho U test pro porovnání skupin a Spearmanovy korelace pro hodnocení asociací s vybranými demografickými a klinickými daty, viz též cíle studie. Data byla hodnocena na hladině významnosti p < 0,05 a byla provedena korekce na opakovaná měření (False Discovery Rate, FDR – Benjamini-Hochbergova metoda).

Výsledky

Všichni zařazení účastníci dokončili studii. Pacienti s PN byli ve fázi středně pokročilé nemoci s oboustranným hybným postižením bez výraznější poruchy rovnováhy (délka onemocnění 8,0 ± 4,8 let, UPDRS V: 2,2 ± 0,8 bodů). Žádný z pacientů neměl tremor-dominantní formu onemocnění, u všech převažovala bradykineze s rigiditou a poruchou chůze. Všichni byli na stabilní dopaminergní medikaci levodopou, někteří v kombinaci s inhibitory COMT (catechol-o-metyltransferázy) a/nebo agonisty dopaminových receptorů (přepočítaná denní LED = 1 373,4 ± 714 mg). Dvacet pacientů mělo počátek onemocnění vpravo a 17 vlevo, tři pacienti referovali oboustranný počátek nemoci. Pacienti s počátkem vpravo se signifikantně nelišili v demografických parametrech od skupiny s levostranným počátkem PN. Nezjistili jsme žádné signifikantní rozdíly mezi PN a ZK ve sledovaných parametrech při kresbě Archimédovy spirály (cvičení 1). Naopak obě skupiny se signifikantně lišily ve všech dalších cvičeních – pacienti s PN měli ve srovnání se ZK nižší průměrné hodnoty ve všech sledovaných parametrech (pro opisování písmene, slabiky a slova, viz tab. 2a–c). Uvedený relativní rozdíl mezi veličinami je definován následovně: $((H_{PN} – H_{ZK})/H_{ZK}) \times$ $\times 100$ %, kde H_{PN}, je hodnota veličiny pro PN skupinu a H_{ZK} je hodnota veličiny pro skupinu ZK. Při mnohonásobném obtahování elipsy jsme zjistili nejvýznamnější změny v průměrném počtu změn v rychlosti (relativní rozdíl 13,5 %) a v průměrném počtu změn ve zrychlení (relativní rozdíl 13,3 %) (tab. 3). Pro hodnocení pohybu pera nad tabletem jsme použili cvičení 9. PN ve srovnání se ZK měli signifikantně nižší všechny hodnoty vztahující se k okamžité rychlosti a jejím derivacím, a to zejména v horizontálním směru (relativní rozdíl PN < ZK až o 19,5 %) (tab. 4).

Pacienti s počátkem onemocnění vpravo měli ve srovnání s PN s počátkem vlevo signifikantně nižší hodnoty okamžité rychlosti, zrychlení a počtu jejich změn v čase (až o 17,4 %, p = 0,006) (tab. 5). U pacientů s PN jsme nezjistili signifikantní korelace mezi jednotlivými hodnocenými parametry mikrografie a délkou trvání PN a/nebo denní LED.

Diskuze

Mikrografie patří mezi velmi časné symptomy PN a významně ovlivňuje kvalitu života pacientů. Patofyziologický mechanizmus rozvoje tohoto příznaku není plně objasněn, na rozdíl od běžných hybných symptomů PN vyjádřených na končetinách nereaguje mikrografie na dopaminergní léčbu, nebo jen v omezené míře [11,12]. Tomu odpovídají i výsledky naší studie – nenašli jsme signifikantní korelace mezi tíží hodnocených příznaků mikrografie a celkovou denní dávkou dopaminergní medikace. Tíže mikrografie nekorelovala ani s délkou trvání nemoci, tedy hodnocení námi sledovaných parametrů písma s pomocí tabletu může sloužit spíš pro časnou diagnostiku a diferenciální diagnostiku PN nežli pro hodnocení progrese onemocnění v čase. K ověření této hypotézy bude zapotřebí prospektivní longitudinální studie u osob v riziku PN ještě před rozvojem charakteristických hybných symptomů. V souladu s naší hypotézou jsme zjistili, že pacienti s počátkem onemocnění vpravo, kteří měli i v průběhu klinického vyšetření výraznější rigiditu a bradykinezi na pravé horní končetině, měli výraznější příznaky mikrografie. Tento faktor tedy může ovlivnit výsledky budoucích studií a bude nutné brát jej v úvahu při hodnocení dat. Nejzajímavější výsledky jsme zaznamenali při porovnávání psaní u obou skupin. Výsledky ve všech cvičeních až na obkreslování Archimédovy spirály přinesly signifikantní rozdíly mezi oběma skupinami ve všech námi hodnocených parametrech mikrografie.

Tab. 2c) Opisování slova „lektorka" (ZK vs PN).

Parametr	PN medián [× 10e-3]	ZK medián [× 10e-3]	Relativní rozdíl PN < ZK (v %)	p hodnota
průměrná hodnota okamžité rychlosti	57,4	70,2	0,005	12,9
průměrné zrychlení	78,3	96,2	0,009	12,9
průměrný jerk	16,0	13,0	0,014	12,9
průměrná horizontální rychlost	122,6	156,2	0,03	8,8
průměrné horizontální zrychlení	16,5	21,0	0,03	8,8
průměrný horizontální jerk	2,2	2,8	0,02	8,9
průměrná vertikální rychlost	6,3	10,6	0,05	38,9
průměrné vertikální zrychlení	0,8	1,4	0,045	40,2
průměrný vertikální jerk	0,9	0,1	0,041	41,4

Tab. 3. Vícenásobné obtahování elips (ZK vs PN).

Parametr	PN medián [× 10e-3]	ZK medián [× 10e-3]	Relativní rozdíl PN < ZK (v %)	p hodnota
průměrný počet změn v rychlosti	13,500	1 500	13,5	0,01
průměrný počet změn ve zrychlení	13,500	1 500	13,3	0,02
průměrná hodnota okamžité rychlosti	563	653	6,8	0,03
průměrné zrychlení	75,9	88,4	6,8	0,04
průměrný jerk	10,2	11,9	6,8	0,04

Kresba Archimédovy spirály je velmi senzitivní pro diagnostiku a kvantifikaci zejména posturálního třesu u pacientů s esenciálním třesem [21]. U našich pacientů s PN převažovala rigidita a bradykineze nad tremorem. Pacienti byli v době vyšetření na dopaminergní terapii, tj. v dobrém hybném stavu a s minimálním tremorem. Pravděpodobně proto jsme nenašli žádné statisticky významné rozdíly při porovnání pacientů s PN a ZK při kresbě spirály.

V testu vícenásobného obtahovaní elips pacienti s PN vykazovali snížený počet změn v rychlosti a zrychlení oproti ZK, tj. jejich pohyby byly monotónnější, rigidnější, bez fyziologicky přítomného „švihu" během tahu perem. Tyto změny jsou pravděpodobně způsobeny jak rigiditou, tak i poruchou cirkumdukce v zápěstí a souhry mezi pohybem kloubů prstů a zápěstí [33]. Dále jsme pozorovali, že čím delší je psaný segment, tím výraznější jsou příznaky bradykineze během psaní; tedy projevuje se progresivní unavitelnost u těchto automatizovaných pohybů. Tento výsledek může souviset s vlastní bradykinezí [34]. Jako doposud první jsme zjistili, že signifikantní změny mezi PN a ZK jsou nejen při vlastním psaní, ale i při hodnocení pohybu pera nad tabletem. Námi hodnocenými parametry jsme u PN prokázali změny v dynamice pohybu pera ve vzduchu – snížení ve všech sledovaných parametrech. Tyto změny mohou souviset ovšem nejen s vlastní bradykinezí a rigiditou, ale také např. s plánováním pohybu, tj. kognitivním výkonem a zejména výkonem v exekutivních funkcích, poruchou iniciace nebo se zárazy pohybu. Doufáme, že další plánovaná detailní analýza našich výsledků nám umožní odpovědět na některé z těchto otázek.

Silnou stránkou této pilotní studie je fakt, že jsme studovali ve srovnání s lite-

Tab. 4. Opisování věty „Tramvaj dnes už nepojede." (ZK vs PN).

Parametr	PN medián [× 10e-3]	ZK medián [× 10e-3]	p hodnota	Relativní rozdíl PN < ZK (v %)
průměrné zrychlení	83,8	104,3	0,046	19,56
průměrná hodnota okamžité rychlosti	621,0	767,4	0,042	19,58
průměrný jerk	113,0	141,0	0,038	19,54

Tab. 5. Opisování slabiky „les" (porovnání pacientů s PN s pravostranným/ /levostranným počátkem onemocnění).

Parametr	Medián PHK [× 10e-3]	Medián LHK [× 10e-3]	p hodnota	Relativní rozdíl PHK > LHK (v %)
průměrný počet změn ve zrychlení	2,8	2,3	0,006	17,44
rychlost psaní jednotlivých úseků	678,8	439,9	0,018	11,53

raturou poměrně velké soubory pacientů s PN a ZK, které byly perfektně srovnatelné po stránce věku a pohlaví. Jedná se o pilotní výsledky, plánujeme analyzovat další parametry související s mikrografií u PN. Zcela nová je možnost kvantifikace pohybu pera nad tabletem, která nám může v budoucnu pomoci lépe objasnit některé motorické i kognitivní aspekty mikrografie u PN.

Závěr

Výsledky naší studie prokázaly, že u pacientů s PN ve srovnání se ZK existují specifické rychlostní změny nejen při vlastním psaní, ale i ve fázi přípravy na psaní, které lze kvantifikovat pomocí digitalizačního tabletu. Dalším plánovaným krokem je vyhodnotit senzitivitu a specificitu těchto zjištěných změn. Výsledky studie mohou mít přímý klinický dopad: umožní nám studovat mikrografii jakožto možný časný klinický biomarker rozvoje PN a mohou nám napomoci i v rámci zpřesnění časné diferenciální diagnostiky tohoto onemocnění. Kvantifikace mikrografie u PN je prvním krokem ke studiu patofyziologických mechanizmů tohoto axiálního symptomu a předpokladem pro hodnocení efektu jak terapie farmakologické, tak i chirurgické.

Literatura

1. Pangelinan, MM, Kagerer FA, Momen B, Hatfield BD, Clark JE. Electrocortical dynamics reflect age-related differences in movement kinematics among children and adults. Cereb Cortex 2011; 21(4): 737–747. doi:10.1093/cercor/bhq162.

2. Eichhorn TE, Gasser T, Mai N, Marquardt C, Arnold G, Schwarz J et al. Computational analysis of open loop handwriting movements in Parkinson's disease: a rapid method to detect dopamimetic effects. Mov Disord 1996; 11(3): 289–297.

3. Mergl R, Tigges P, Schröter A, Möller HJ, Hegerl U. Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives. J Neurosci Methods 1999; 90(2): 157–169.

4. Siebner HR, Ceballos-Baumann A, Standhardt H, Auer C, Conrad B, Alesch F. Changes in handwriting resulting from bilateral high-frequency stimulation of the subthalamic nucleus in Parkinson's disease. Mov Disord 1999; 14: 964–971.

5. Wagle Shukla A, Ounpraseuth S, Okun MS, Gray V, Schwankhaus J, Metzer WS. Micrographia and related deficits in Parkinson's disease: a cross-sectional study. BMJ Open 2012; 2(3): e000628. doi: 10.1136/bmjopen-2011-000628.

6. Iwasaki Y, Ikeda K, Shindoh T, Suga Y, Ishikawa I, Hara M et al. Micrographia in Huntington's disease. J Neurol Sci 1999; 162(1): 106–107.

7. Nakamura M, Hamamoto M, Uchida S, Nagayama H, Amemiya S, Okubo S et al. A case of micrographia after subcortical infarction: possible involvement of frontal lobe function. Eur J Neurol 2003; 10(5): 593–596.

8. Ho SL. Micrographia in Chinese characters. J Neurol Neurosurg Psychiatry 1995; 59(5): 471.

9. Yoshida T, Yamadori A, Mori E. A case of micrographia with the right hand due to left putaminal infarction. Rinsho Shinkeigaku 1989; 29(9): 1149–1151.

10. Gangadhar G, Joseph D, Chakravarthy VS. Understanding Parkinsonian handwriting through a computational model of basal ganglia. Neural Comput 2008; 20(10): 2491–2525. doi: 10.1162/neco.2008.03-07-498.

11. Rektorová I, Rektor I. Parkinsonský syndrom. In: Rektor I, Rektorová I et al (eds). Centrální poruchy hybnosti v praxi. Praha: Triton 2003: 21–70.

12. Tucha O, Mecklinger L, Thome J, Reiter A, Alders GL, Sartor H et al. Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease. J Neural Transm 2006; 113(5): 609–623. doi: 10.1007/s00702-005-0346-9.

13. Hoshiyama M, Kaneoke Y, Koike Y, Takahashi A, Watanabe S. Hypokinesia of associated movement in Parkinson's disease: a symptom in early stages of the disease. J Neurol 1994; 241(9): 517–521.

14. Kawakami O, Hoshiyama M, Hirayama M, Koike Y, Sobue G. Impaired associated movement in Parkinson's disease. Nihon Rinsho 1997; 55(1): 179–184.

15. Marsden CD. Which motor disorder in Parkinson's disease indicates the true motor function of the basal ganglia? Ciba Found Symp 1984; 107: 225–241.

16. Van Gemmert AW, Teulings HL, Stelmach GE. Parkinsonian patients reduce their stroke size with increased processing demands. Brain Cogn 2001; 47(3): 504–512. doi: 10.1006/brcg.2001.1328.

17. Unlu A, Brause R, Krakow K. Handwriting Analysis for Diagnosis and Prognosis of Parkinson's Disease. In: Proceedings of the 7th international conference on Biological and Medical Data Analysis. Berlin: Springer-Verlag 2006: 441–450.

18. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinicopathological study of 100 cases. J Neurol Neurosurg Psychiatry 1992; 55(3): 181–184.

19. Fahn S, Elston RL, members of the UPDRS Development Committee. Unified Parkinson's disease rating scale. In: Fahn S, Marsden CD, Goldstein M, Calne DB (eds). Recent developments in Parkinson's disease. Vol 2. New York: Macmillan 1987: 153–163.

20. Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. Mov Disord 2010; 25(15): 2649–2653. doi: 10.1002/mds.23429.

21. Louis ED, Yu Q, Floyd AG, Moskowitz C, Pullman SL. Axis is a feature of handwritten spirals in essential tremor. Mov Disord 2006; 21(8): 1294–1295. doi: 10.1002/mds.20915.

22. Pullman SL. Spiral analysis: a new technique for measuring tremor with a digitizing tablet. Mov Disord 1998; 13 (Suppl 3): 85–89.

23. Saunders-Pullman R, Derby C, Stanley K, Floyd A, Bressman S, Lipton RB et al. Validity of spiral analysis in early Parkinson's disease. Mov Disord 2008; 23(4): 531–537. doi:10.1002/mds.21874.

24. Wang H, Yu Q, Kurtis MM, Floyd AG, Smith WA, Pullman SL. Spiral analysis-improved clinical utility with center detection. J Neurosci Methods 2008; 171(2): 264–270. doi: 10.1016/j.jneumeth.2008.03.009.

25. Westin J, Ghiamati S, Memedi M, Nyholm D, Johansson A, Dougherty M et al. A new computer method for assessing drawing impairment in Parkinson's disease. J Neurosci Methods 2010; 190(1): 143–148. doi: 10.1016/j.jneumeth.2010.04.027.

26. Lepelley MC, Thullier F, Bolmont B, Lestienne FG. Age-related differences in sensorimotor representation of space in drawing by hand. Clin Neurophysiol 2010; 121(11): 1890–1897. doi: 10.1016/j.clinph.2010.04.024.

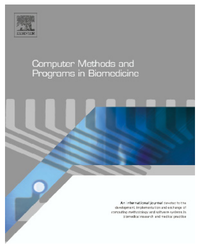
27. Bidet-Ildei C, Pollak P, Kandel S, Fraix V, Orliaguet JP. Handwriting in patients with Parkinson disease: effect of L-dopa and stimulation of the sub-thalamic

nucleus on motor anticipation. Hum Mov Sci 2011; 30(4): 783–791. doi: 10.1016/j.humov.2010.08.008.
28. Gallucci RM, Phillips JG, Bradshaw JL, Vaddadi KS, Pantelis C. Kinematic analysis of handwriting movements in schizophrenic patients. Biol Psychiatry 1997; 41(7): 830–833. doi: 10.1016/s0006-3223(96)00544-6.
29. Van Gemmert AW, Teulings HL, Contreras-Vidal JL, Stelmach GE. Parkinson's disease and the control of size and speed in handwriting. Neuropsychologia 1999; 37(6): 685–694.
30. Teulings HL, Contreras-Vidal JL, Stelmach GE, Adler CH. Parkinsonism reduces coordination of fingers, wrist and arm in fine motor control. Exp Neurol 1997; 146(1): 159–170. doi: 10.1006/exnr.1997.6507.
31. Teulings HL, Contreras-Vidal JL, Stelmach GE, Adler CH. Adaptation of handwriting size under distorted visual feedback in patients with Parkinson's disease and elderly and young controls. J Neurol Neurosurg Psychiatry 2002; 72(3): 315–324.
32. Ondo WG, Satija P. Withdrawal of visual feedback improves micrographia in Parkinson's disease. Mov Disord 2007; 22(14): 2130–2131. doi:10.1002/mds.21733.
33. Dounskaia N, Van Gemmert AW, Leis BC, Stelmach GE. Biased wrist and finger coordination in Parkinsonian patients during performance of graphical tasks. Neuropsychologia 2009; 47(12): 2504–2514. doi: 10.1016/j.neuropsychologia.2009.04.020.
34. Ling H, Massey LA, Lees AJ, Brown P, Day BL. Hypokinesia without decrement distinguishes progressive supranuclear palsy from Parkinson's disease. Brain 2012; 135(4): 1141–1153. doi: 10.1093/brain/aws038.

A.27 Analysis of in-air movement in handwriting: A novel marker for Parkinsons disease

COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 117 (2014) 405–411

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease

Peter Drotár [a], Jiří Mekyska [a], Irena Rektorová [b,*], Lucia Masarová [b], Zdenek Smékal [a], Marcos Faundez-Zanuy [c]

[a] Brno University of Technology, Technicka 12, Brno, Czech Republic
[b] First Department of Neurology, Masaryk University and St. Anne's Hospital, Pekarska 664, 656 91 Brno, Czech Republic
[c] Tecnocampus, Av. Ernest Lluch, 32, 08302 Mataro, Barcelona, Spain

ARTICLE INFO

Article history:
Received 17 March 2014
Received in revised form
11 August 2014
Accepted 22 August 2014

Keywords:
Handwriting
Disease classification
Parkinson's disease
Micrographia
In-air movement
Decision support systems

ABSTRACT

Background and objective: Parkinson's disease (PD) is the second most common neurodegenerative disease affecting significant portion of elderly population. One of the most frequent hallmarks and usually also the first manifestation of PD is deterioration of handwriting characterized by micrographia and changes in kinematics of handwriting. There is no objective quantitative method of clinical diagnosis of PD. It is thought that PD can only be definitively diagnosed at postmortem, which further highlights the complexities of diagnosis.
Methods: We exploit the fact that movement during handwriting of a text consists not only from the on-surface movements of the hand, but also from the in-air trajectories performed when the hand moves in the air from one stroke to the next. We used a digitizing tablet to assess both in-air and on-surface kinematic variables during handwriting of a sentence in 37 PD patients on medication and 38 age- and gender-matched healthy controls.
Results: By applying feature selection algorithms and support vector machine learning methods to separate PD patients from healthy controls, we demonstrated that assessing the in-air/on-surface hand movements led to accurate classifications in 84% and 78% of subjects, respectively. Combining both modalities improved the accuracy by another 1% over the evaluation of in-air features alone and provided medically relevant diagnosis with 85.61% prediction accuracy.
Conclusions: Assessment of in-air movements during handwriting has a major impact on disease classification accuracy. This study confirms that handwriting can be used as a marker for PD and can be with advance used in decision support systems for differential diagnosis of PD.

1. Introduction

Handwriting is a highly skilled and complex coordinated motor activity. Writing a sentence requires the dynamic interplay of the lower arm, wrist, and finger muscles. The accurate sequencing and online scaling of automated movements and the programming of subsequent strokes are also involved [1]. It has been well documented that handwriting

* *Corresponding author*. Tel.: +420 543182639.
E-mail address: irena.rektorova@fnusa.cz (I. Rektorová).

http://dx.doi.org/10.1016/j.cmpb.2014.08.007

is affected in Parkinson's disease (PD) and micrographia occurs in about 63% of PD patients as an early motor feature [2]. It is characterized by decreased letter size and by changes in kinematic aspects of movements [3,4]. Kinematic variables are sensitive measures for alterations of handwriting movements even with patients treated with dopaminergic medication [3,5].

Besides the PD, the alterations of the handwriting are connected with other diseases. Several authors investigated the temporal, kinematic, and dynamic aspects of handwriting movements to better characterize the handwriting difficulties of children with dysgraphia [6,7] or hyperactivity disorder [8]. Some aspects of the handwriting are also indicators for diagnosis of the Alzheimer disease – the most common neurodegenerative disease [9].

It has rarely been taken into account that hand movement during handwriting consists of two components: an on-surface component, comprising movements executed while exerting pressure on the writing surface, and an in-air component, comprising movements performed without touching the writing surface. The in-air movement has been mostly used for biometric applications [10], but some pilot data suggest that it could have meaningful applications for medical diagnostic purposes as well [11].

There were several attempts to design decision support systems for differential diagnosis of PD in recent years. These usually include speech assessment [12–14], gait monitoring [15,16,17] or tremor assessment [18]. Handwriting and especially in-air movement has not been explored so deeply even if there is proven relationship between symptoms of PD and handwriting. When compared to handwriting, both speech assessment and gait monitoring are more demanding in the terms of technical equipment and signal processing. Speech assessment requires high quality recording conditions without background noise and usually some further post-processing of recorded speech is necessary. This can include human operated speech segmentation that makes whole process much more tedious. Gait monitoring or tremor assessment techniques require specialized equipment such as accelerometers and gyroscopes. On the other hand, diagnosis of PD through the handwriting can be easily administrated at clinic or even patient's home. Handwriting acquisition is quite simple and natural, and does not require any timing or exhaustive repetitions.

Previous research has shown that there are some statistically significant differences between kinematics of PD patients and healthy controls. However extend to which any set of features could be useful in discriminating PD from HC was not given. The contribution of this work is twofold. First, we show that in-air movement has a significant role in diagnosis of PD providing together with on-surface movement clinically relevant classification accuracy. In addition, we proposed a classification model that can be used for automated differential diagnosis of PD. The achieved results indicate that in-air and on-surface trajectories can be used in decision support systems and assist in diagnosis of PD.

2. Materials and methods

2.1. Patients and data acquisition

Altogether, 37 PD patients (19 men/18 women; mean age 69.3 ± 10.9 years; mean disease duration 8.37 ± 4.8 years; UPDRS V score [19] 2.27 ± 0.84; daily levodopa equivalent dose 1373.4 ± 714 mg [20] and 38 age- and gender-matched HC (20 men/18 women; mean age 62.4 ± 11.3 years) were enrolled at the First Department of Neurology, St. Anne's University Hospital in Brno, Czech Republic. (UPDRS V provides information on disease stage based on motor clinical symptoms and their body distribution. UPDRS score 2.27 means that patients suffered from bilateral parkinsonism with mild postural instability). All subjects were right-handed, completed at least 10 years of education, and reported Czech as their first language. None of the subjects had a history or presence of any psychiatric symptoms or any disease affecting the central nervous system (other than PD in the PD cohort). The subjects were non-demented based on the clinician's judgment, caregiver's interview, and the MMSE [21] score (>27 points). PD patients were examined only in their ON-state while on dopaminergic medication, i.e. 1–2 h after taking their regular dose of dopaminergic medication. All patients were taking L-dopa dopamine agonist or COMT (catechol-o-methyl transferase) inhibitor. At the time of the study, their symptoms were successfully managed and they had no analgesic treatment. Age-matched healthy controls were examined and treated in St. Anne's University Hospital for cervical and/or back pain syndrome and had no speech problems and handwriting problems. All subjects signed an informed consent form that was approved by the ethics committee of St. Anne's Hospital in Brno.

Each subject wrote a Czech sentence: Tramvaj dnes už nepojede (the tram won't go today). Sentence was written into template form so there was upper and lower boundary that limited handwriting height. Subjects were instructed to write within the limits, but they did not need exactly match row height. For more information about full template see [22].

Writing of sentence allows to acquire in-air movement not only as interruptions during the writing on particular words, but also between words when subject is proceeding from one word to another. An ink writing pen was held in a normal fashion and subjects were asked to write a sentence at a self-determined comfortable size and speed. Patients had visual feedback of their on-surface writing only. All signals were acquired using Intuos 4M digitizing tablet (Wacom); in the terms of x-coordinate, $x(t)$; y-coordinate, $y(t)$; time stamp t and button status, $b(t)$. Button status is a binary variable, being 0 for pen-up state (in-air movement) and 1 for pen-down state (on-surface movement). The example signals are depicted in Fig. 1. Fig. 1(a) illustrates the handwriting sample of HC and Fig. 1(b) sample of PD patient, respectively.

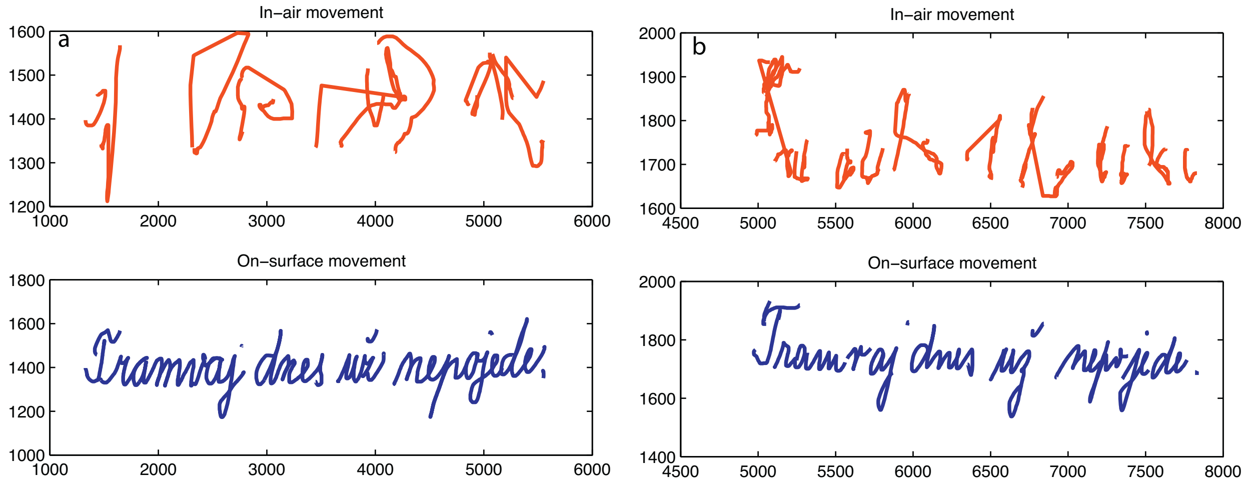


Fig. 1 – An example of the on-surface (blue solid line) and in-air (red dotted line) movement during writing of a sentence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2. *Handwriting features*

The *x* and *y* coordinates are segmented into on-surface and in-air strokes and analyzed in terms of handwriting measures. The feature calculation stage involves the extraction of kinematic features such as stroke speed, writing speed, velocity, acceleration, jerk (changes of acceleration with time), number of changes in velocity (NCV), number of changes in acceleration (NCA) and relative NCV/NCA. Relative NCV/NCA means that NCV/NCA was normalized by writing duration. Regarding the temporal features, we analyzed time spent in-air, i.e. in-air duration, on-surface duration and in-air to on-surface ratio. Complete feature description is provided in Table 1.

Finally, to obtain complete statistical representation of available features, 30 statistical functionals of the vector features were computed. These include means (arithmetic mean, geometric mean, trimmed means (5, 10, 20, 30, 40, 50)), percentiles (quartiles (25/lower, 75/upper), percentiles (1, 5, 10, 20, 30, 90, 95, 99)), moments (moments (3rd, 4th, 5th, 6th), kurtosis) and other (range, median, mode, standard deviation, outlier robust range (percentile 99th – percentile 1st)) statistical functionals.

2.3. *Preliminary statistical analysis*

To obtain some preliminary insight into the statistical properties of handwriting features, we followed the approach of Tsanas et al. [23] and computed Pearson correlation coefficients and mutual information between feature vectors and associated diagnosis (HC vs. PD). The Pearson correlation expresses measure of linear dependence between features vectors and associated response [24]. Mutual information (MI) is a measure of the amount of the information shared by two random variables. It is defined as:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \cdot \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right) \qquad (1)$$

where *x* and *y* are possible variable values with a joint probability distribution function $p(x, y)$ and marginal distribution functions $p(x)$ and $p(y)$, respectively [25]. We computed MI by evaluating the marginal entropies $H(X)$, $H(Y)$ and joint entropy

Table 1 – Description of handwriting features.

Feature	Description
Stroke speed	Trajectory during stroke divided by stroke duration
Speed	Trajectory during handwriting divided by handwriting duration
Velocity	Rate at which the position of a pen changes with time
Acceleration	Rate at which the velocity of a pen changes with time
Jerk	Rate at which the acceleration of a pen changes with time
Horizontal velocity/acceleration/jerk	Velocity/acceleration/jerk in horizontal direction
Vertical velocity/acceleration/jerk	Velocity/acceleration/jerk in vertical direction
Number of changes in velocity direction (NCV)	The mean number of local extrema of velocity ([3])
Number of changes in acceleration direction (NCA)	The mean number of local extrema of acceleration ([3])
Relative NCV	NCV relative to writing duration
Relative NCA	NCA relative to writing duration
In-air time	Time spent in-air during writing
On-surface time	Time spent on-surface during writing
Normalised in-air time	Time spent in-air during writing normalised by whole writing duration
Normalised on-surface time	Time spent on-surface during writing Normalised by whole writing duration
In-air/on-surface ration	Ratio of time spent in-air/on-surface

Table 2 – Description of calculated features.

Feature	Mutual information	Correlation coefficient
Stroke speed (on surface, standard dev.)	6.09	−0.388
Velocity (in air, standard dev.)	5.94	−0.387
Vert. jerk (in air, min.)	5.70	0.383
Acceleration (in air, standard dev.)	5.92	−0.380
Horz. jerk (in air, range)	5.72	−0.379
Jerk (in air, standard dev.)	5.96	−0.389
Horz. acceleration (in air, range)	5.81	−0.375
Horz. velocity (in air, range)	5.87	−0.371
Horz. velocity (on surface, quantile 75%)	4.46	−0.370
Vert. acceleration (in air, min.)	5.74	−0.369

$H(X, Y)$ as $I(X; Y) = H(X) + H(Y) - H(X, Y)$.The entropy is defined as

$$H_S(X) = -\sum_{x \in X} p(x)\log_2 p(x), \tag{2}$$

where $p(x)$ is probability density function computed using kernel density estimation with a Gaussian kernel.

Table 2 summarizes 10 handwriting features most strongly correlated with the target classification variable. All correlations are statistically significant ($p < 0.05$).

2.4. *Classification algorithm*

As an preprocessing step the data was analyzed using the Mann–Whitney U test for between-group (PD vs. HC) comparisons. The level of significance was set to $p < 0.05$. Features that did not pass Mann–Whitney test were discarded and were not used in further processing.

2.4.1. *Feature selection*

Our goal was to determine discriminative potential of handwriting and build predictive model only with relevant features. The most straightforward approach would be to try all possible feature combinations (brute force approach) and keep only those that contribute to correct prediction. However, this approach is computationally intractable, and requires huge amount of computational resources. An alternative is to use sequential forward feature selection (SFFS) that enables significant reduction of the computational complexity compared to that of a brute force search but still select relevant features. Since SFFS for large number of features is still computationally demanding, the minimum-redundancy-maximum-relevance (mRMR) [26] feature selection method was applied to reduce dimensionality to 50 features.

The first stage of mRMR, the maximum relevance method, selects the best individual features correlated to target classification variable [26]. Features selected according to the maximum relevance method could have a large redundancy. In order to remove redundancy among features, the minimum redundancy condition is introduced. Therefore, mRMR selects features that are mutually different from each other while still having a high correlation to yield well performing feature subset. Number of 50 features at the output of mRMR was decided as a trade-off between computational complexity and desire to include all relevant features in classification model.

2.4.2. *Support vector machines*

In order to develop a functional relationship to map handwriting measures to subject classification (PD vs. HC), we employed supervised machine learning algorithm support vector machines (SVM) [27,28] with nonlinear radial basis function (RBF) kernel.

The SVM minimizes the classification error and maximizes the margin by determining a separating hyperplane to identify different classes of data. For two-class support vector machine, we consider the following decision function [29]:

$$f(x) = \mathrm{sign}[w^T g(x) + b] \tag{3}$$

where w is the d-dimensional weight vector and b is a bias. To obtain w and b the following optimization problem with linear equality constraints is solved:

$$\text{minimize} \quad J(w, b, \xi_i) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{i=1}^{N}\xi_i^2 \tag{4}$$

$$\text{s.t.} \quad y_i[w^T g(x_i) + b] = 1 - \xi_i, i = 1, 2, \ldots, N. \tag{5}$$

In this minimization problem, N is the number of samples in the training data set, y_i is the target value of the training data set, γ is the regularization hyperparameter and ξ_i the slack variable.

After solving Lagrangian

$$L(w, b, \alpha_i, \xi_i) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{i=1}^{N}\xi_i^2 - \sum_{i=1}^{N}\alpha_i\{y_i[w^T g(x_x) + b] + \xi_i - 1\} \tag{6}$$

discriminant function of linear separating hyperplane is derived as

$$f(x) = \mathrm{sign}\left[\sum_{i=1}^{N}\alpha_i y_i K(x, x_i) + b\right] \tag{7}$$

where $\alpha_i \in R$ is Lagrangian multiplier and $K(x, x_i)$ is a kernel function [30]. We used radial basis kernel function, defined as $K(x, x_i) = \exp(-||x_i - x||^2/\sigma^2)$. The kernel parameter σ is referred to as the kernel width.

In general, SVM requires the specification of several internal parameters, and SVMs are known to be sensitive to the values of these parameters [28]. The performance of SVM with RBF kernel depends on three parameters: kernel width

(σ), penalty parameter (γ) and convergence epsilon (ϵ). These parameters were optimized using a grid search of possible values. Specifically, we searched over the grid (γ, σ, ϵ) defined by the product of the sets $\gamma = [10^{-5}, 10^{-4}, \ldots, 10^{3}, 10^{4}]$, $\sigma = [10^{-5}, 10^{-4}, \ldots, 10^{2}, 10^{3}]$ and $\epsilon = [10^{-5}, 10^{-4}, \ldots, 10^{2}, 10^{3}]$.

2.4.3. *Classifier validation*

Classifier validation was conducted using a leave-one-out approach. That is, we left out the sample of one individual as if it were an unseen individual; the remaining samples were used as a training dataset.

The whole process of feature selection and classification was repeated a total of 50 times, where in each repetition the original dataset was randomly permuted prior to splitting into the training and the testing subsets. The number of features at the SFFS output cannot be set to concrete number since algorithm itself evaluates the number of the most predictive features. In our experience fewer than ten features were usually selected during 50 repeated realizations. Classification accuracies over 50 repetitions were averaged. The standard deviation of averaged results is quite small indicating that feature selection (SFFS, mRMR) and classification are quite robust to initial conditions. The classification performance of the predictive model was evaluated for three different scenarios: using only features based on the in-air movement, using only features extracted from the on-surface movement, and using a combination of both groups of features. By combination, we mean that both feature groups were merged prior to the feature selection.

3. Results

On average, the ratio of time spent in-air to time spent on-surface is 0.77 for HC and 0.75 for PD subjects. Of the ten features that most strongly correlated with the diagnosis, nine were in-air movement-related features. This result provided an initial confirmation of our hypothesis that the in-air features contain information relevant for discriminating PD from HC. The features with the largest relevance to the diagnosis, sorted according to an absolute correlation coefficient are: range of in-air jerk ($R = -0.428$), range of in-air acceleration ($R = -0.424$), and minima of in-air acceleration ($R = 0.4148$), see the Table 2. The decreased range/standard deviation of relevant kinematic features in PD as compared to HC reflect the monotonous motor performance in PD caused by bradykinesia (increased slowness of movements) and rigidity rigidity (increased muscle tone).

To discriminate between PD patients and HC controls we used SVM, that showed good prediction performance in problems similar to ours [12] and in general perform well in various biomedical applications [31,32]. Classification employing features based on the in-air movements revealed an accuracy of $84.43 \pm 2.88\%$. Classification accuracy was $78.16 \pm 1.96\%$ when only features based on the on-surface movement were employed. The utilization of features from both modalities (in-air movement and on-surface movement) led to further improvement in classification accuracy: $85.61 \pm 1.72\%$. The results of classification accuracy together with sensitivity and specificity are provided in Table 3.

Table 3 – Classification accuracy, specificity and sensitivity of PD diagnosis using in-air and on-surface movement.

	Accuracy	Sensitivity	Specificity
In-air	84.43	87.47	82.89
On-surface	78.16	78.23	78.05
In-air + on-surface	85.61	85.95	85.26

4. Discussion

The writing of a sentence consists of different strokes elicited at a fine-tuned speed and acceleration that requires a high degree of simultaneous processing and may therefore have a higher programming load than a sequence of identical stroke [4]. This becomes important in the evaluation of handwriting in PD, since both motor program sequencing and concurrent processing have been shown to be disturbed in PD [4,5,33].

By applying mRMR, SFFS procedure, and SVM learning methods to separate PD from HC, we demonstrated for the first time that the assessment of in-air hand movements during sentence handwriting has a higher impact than the pure evaluation of on-surface movements, leading to classification accuracies of 84% and 78%, respectively. Interestingly, combining both in-air and on-surface kinematic features for identifying PD patients on dopaminergic medication improved classification accuracy by only 1% over the pure evaluation of in-air movements.

The binary SVM classifier was applied to segment data from healthy controls and patients with PD. We chose SVM since it allows to capture complex multivariate relationships in the data, it has good generalization properties and can deal with feature vectors of high dimensionality. In fact, SVM classifier was successfully applied to the individual classification of a variety diseases and medical conditions [12,34,35]. Because of the nonlinear propagation of the features we selected RBF kernel. Even thought we tuned SVM classifier to obtain highest classification accuracy, we believe that there may be still some space for improvement. Currently, there are plenty of new classifiers or improvements of existing classifiers that can further enhance classification performance. However, this is beyond the scope of this paper that focuses on demonstrating the idea of utilization of in-air movement for diagnosis of PD.

Handwriting in PD is thought to be impaired mainly due to hypokinesia (decreased amplitude of movements) and bradykinesia [4,5]. The underlying pathophysiological mechanisms probably involve inefficiency of the basal ganglia-thalamocortical circuits and particularly disturbed activation of the supplementary motor area, which is thought to be involved in "open-loop" performances, in which a motor task is run off automatically [36]. It has been shown that PD patients are able to compensate if the task is modified to involve "closed loop" performance. This can be done either by providing visual cues or by otherwise drawing attention to the task. The lateral premotor cortex is additionally activated when attention must be paid during a motor task. This notion may at least partially explain the fact that the in-air kinematic features reflected the impaired open-loop performance in PD better than the on-surface writing to which patients devoted more attention under the visual guidance. As compared to on-surface movements, the in-air movements elicited during handwriting of a sentence may involve additional cognitive processes such as motor planning, programming of the alternating motor sequences, and movement initiation that may also have impacted on the kinematic features and our results.

The presented results show that in-air movement possess significant amount of information relevant to diagnosis of PD and as such can be incorporated in decision support systems that are the important part of the next generation health-care. The main advantage of the proposed approach is that acquisition of handwriting signals at clinic or at home is relatively simple and easily administered. In fact, all of the data used in this study were collected in a clinic environment with a tablet connected to a notebook computer without any previous preparation in a room or special environment. We did not use any custom made hardware, only commercially available tablet that makes the whole approach very feasible. Relatively simple management of the test makes it possible to use it at a patient's home, e.g. for disease monitoring, that is important advantage over other approaches.

Handwriting assessment for PD diagnosis can serve as a complementary method to diagnosis made be clinician or other decision support tools [12,37]. We believe that further investigation, utilization of new handwriting features and further tuning of machine learning techniques can improve prediction accuracy and make this approach even more useful and competitive. Here, we assume that mainly new handwriting features can provide more insight into effects of PD on handwriting and can be beneficial for diagnosis.

Conflict of interest

None declared.

Acknowledgments

This work was supported by the project COSTIC 1206, NT13499 (Speech, its impairment and cognitive performance in Parkinson's disease), GAP102/12/1104, project CEITEC, Central European Institute of Technology: (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund and by FEDER and Ministerio de Economa y Competitividad TEC2012-38630-C04-03. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation. Peter Drotár was supported by the project CZ.1.07/2.3.00/30.0005 of Brno University of Technology.

REFERENCES

[1] H.-L. Teulings, J.L. Contreras-Vidal, G.E. Stelmach, C.H. Adler, Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control, Exp. Neurol. 146 (1) (1997) 159–170.

[2] E. Nackaerts, G. Vervoort, E. Heremans, B.C. Smits-Engelsman, S.P. Swinnen, A. Nieuwboer, Relearning of writing skills in Parkinson's disease: a literature review on influential factors and optimal strategies, Neurosci. Biobehav. Rev. 37 (3) (2013) 349–357.

[3] T. Eichhorn, T. Gasser, N. Mai, C. Marquardt, G. Arnold, J. Schwarz, W. Oertel, Computational analysis of open loop handwriting movements in Parkinson's disease: a rapid method to detect dopamimetic effects, Mov. Disord. 11 (3) (1996) 289–297.

[4] M. Broderick, A. Van Gemmert, H. Shill, G. Stelmach, Hypometria and bradykinesia during drawing movements in individuals with Parkinson's disease, Exp. Brain Res. 197 (3) (2009) 223–233.

[5] O. Tucha, L. Mecklinger, J. Thome, A. Reiter, G.L. Alders, H. Sartor, M. Naumann, K.W. Lange, Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease, J. Neural Transm. 113 (5) (2006) 609–623, http://dx.doi.org/10.1007/s00702-005-0346-9.

[6] S. Rosenblum, A.Y. Dvorkin, P.L. Weiss, Automatic segmentation as a tool for examining the handwriting process of children with dysgraphic and proficient handwriting, Hum. Mov. Sci. 25 (2006) 608–621.

[7] J. Danna, V. Paz-Villagrán, J.-L. Velay, Signal-to-noise velocity peaks difference: a new method for evaluating the handwriting movement fluency in children with dysgraphia, Res. Dev. Disabil. 34 (12) (2013) 4375–4384, http://dx.doi.org/10.1016/j.ridd.2013.09.012.

[8] I.-H. Shen, T.-Y. Lee, C.-L. Chen, Handwriting performance and underlying factors in children with attention deficit hyperactivity disorder, Res. Dev. Disabil. 33 (4) (2012) 1301–1309, http://dx.doi.org/10.1016/j.ridd.2012.02.010.

[9] K.E. Forbes, M.F. Shanks, A. Venneri, The evolution of dysgraphia in Alzheimer's disease, Brain Res. Bull. 63 (1) (2004) 19–24, http://dx.doi.org/10.1016/j.brainresbull.2003.11.005.

[10] E. Sesa-Nogueras, M. Faundez-Zanuy, J. Mekyska, An information analysis of in-air and on-surface trajectories in online handwriting, Cogn. Comput. 4 (2) (2012) 195–205, http://dx.doi.org/10.1007/s12559-011-9119-y.

[11] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh, I. Schlesinger, Handwriting as an objective tool for Parkinson's disease diagnosis, J. Neurol. 260 (2013) 1–5.

[12] A. Tsanas, M. Little, P. McSharry, J. Spielman, L. Ramig, Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease, IEEE Trans. Biomed. Eng. 59 (5) (2012) 1264–1271, http://dx.doi.org/10.1109/TBME.2012.2183367.

[13] M. Hariharan, K. Polat, R. Sindhu, A new hybrid intelligent system for accurate detection of Parkinson's disease, Comput. Methods Prog. Biomed. 113 (3) (2014) 904–913, http://dx.doi.org/10.1016/j.cmpb.2014.01.004.

[14] J. Rusz, R. Cmejla, H. Ruicková, J. Klempír, V. Majerová, J. Picmausová, J. Roth, E. Ruicka, Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test, Mov. Disord. 26 (10) (2011) 1951–1952, http://dx.doi.org/10.1002/mds.23680.

[15] M. Bächlin, M. Plotnik, D. Roggen, I. Maidan, J. Hausdorff, N. Giladi, G. Troster, Wearable assistant for Parkinson's disease patients with the freezing of gait symptom, IEEE Trans. Inf. Technol. Biomed. 14 (2) (2010) 436–446, http://dx.doi.org/10.1109/TITB.2009.2036165.

[16] E.E. Tripoliti, A.T. Tzallas, M.G. Tsipouras, G. Rigas, P. Bougia, M. Leontiou, S. Konitsiotis, M. Chondrogiorgi, S. Tsouli, D.I. Fotiadis, Automatic detection of freezing of gait events in patients with Parkinson's disease, Comput. Methods Prog. Biomed. 110 (1) (2013) 12–26, http://dx.doi.org/10.1016/j.cmpb.2012.10.016.

[17] E. Jovanov, E. Wang, L. Verhagen, M. Fredrickson, R. Fratangelo, deFOG: a real time system for detection and unfreezing of gait of Parkinson's patients, in: Engineering in

Medicine and Biology Society, 2009 (EMBC 2009). Annual International Conference of the IEEE, 2009, pp. 5151–5154, http://dx.doi.org/10.1109/IEMBS.2009.5334257.

[18] O. Geman, H. Costin, Automatic assessing of tremor severity using nonlinear dynamics, artificial neural networks and neuro-fuzzy classifier, Adv. Electr. Comput. Eng. 14 (1) (2014) 133–138.

[19] S. Fahn, R.L. Elton, Members of the UPDRS Development Committee, Unified Parkinson's disease rating scale, in: S. Fahn, C.D. Marsden, D.B. Calne, A. Lieberman (Eds.), Recent developments in Parkinson's disease., Macmillan Health Care Information, Florham Park, NJ, 1987, pp. 153–163.

[20] C.L. Tomlinson, R. Stowe, S. Patel, C. Rick, R. Gray, C.E. Clarke, Systematic review of levodopa dose equivalency reporting in Parkinson's disease, Mov. Disord. 25 (15) (2010) 2649–2653, http://dx.doi.org/10.1002/mds.23429.

[21] S. Folfstein, M.F. Folfstein, P. McHugh, Mini-mental state: a practical method for grading the cognitive state of patients for clinician, J. Psychiatr. Res. 12 (1975) 189–198.

[22] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, M. Faundez-Zanuy, Prediction potential of different handwriting tasks for diagnosis of Parkinson's, in: IEEE International Conference on e-Health and Bioengineering (EHB 2013), Iasi, Romania, 2013.

[23] A. Tsanas, M. Little, P. McSharry, L. Ramig, Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity, J. R. Soc. Interface (8) (2011) 842–855.

[24] D. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, 2nd ed., CRC Press, Boca Raton, FL, 2000.

[25] T.M. Cover, The Elements of Information Theory, Wiley, Hoboken, New Jersey, 1991.

[26] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1239.

[27] V. Vapnik, Statistical Learning Theory, 1st ed., John Willey & Sons, London, 1998.

[28] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer, New York, 2009.

[29] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300.

[30] Z. Yin, J. Zhang, Operator functional state classification using least-square support vector machine based recursive feature elimination technique, Comput. Methods Prog. Biomed. 113 (1) (2014) 101–115.

[31] M. Pirooznia, J. Yang, M. Yang, Y. Deng, A comparative study of different machine learning methods on microarray gene expression data, BMC Genomics 9 (Suppl. 1) (2008) 1–13, http://dx.doi.org/10.1186/1471-2164-9-S1-S13.

[32] T. Sun, J. Wang, X. Li, P. Lv, F. Liu, Y. Luo, Q. Gao, H. Zhu, X. Guo, Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set, Comput. Methods Prog. Biomed. 111 (2) (2013) 519–524, http://dx.doi.org/10.1016/j.cmpb.2013.04.016.

[33] P. Weiss, G. Stelmach, H. Hefter, Programming of a movement sequence in Parkinson's disease, Brain 120 (1997) 91–102.

[34] C. Cao, R. Tutwiler, S. Slobounov, Automatic classification of athletes with residual functional deficits following concussion by means of EEG signal using support vector machine, IEEE Trans. Neural Syst. Rehabil. Eng. 16 (4) (2008) 327–335.

[35] F.J. Badesa, R. Morales, N. Garcia-Aracil, J. Sabater, A. Casals, L. Zollo, Auto-adaptive robot-aided therapy using machine learning techniques, Comput. Methods Prog. Biomed. 116 (2) (2014) 123–130.

[36] R. Oliveira, J. Gurd, P. Nixon, J. Marshall, R. Passingham, Micrographia in Parkinson's disease: the effect of providing external cues, J. Neurol. Neurosurg. Psychiatry 63 (1997) 429–433.

[37] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M. Gilardi, A. Quattrone, Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and progressive supranuclear palsy, J. Neurosci. Methods 222 (2014) 230–237.

A.28 A Preliminary Study on Aging Examining Online Handwriting

A Preliminary Study on Aging Examining Online Handwriting

Marcos Faundez-Zanuy*, Enric Sesa-Nogueras*, Josep Roure-Alcobé*, Anna Esposito†,
Jiri Mekyska‡ and Karmele López-de-Ipiña§
*Tecnocampus
Avda. Ernest Lluch 32
08302 Matar (Barcelona), Spain
Email: {faundez, sesa, roure}@tecnocampus.cat
†Seconda Università di Napoli and IIASS, Italy
Email: iiass.annaesp@tin.it
‡Brno University of Technology, Brno, Czech Republic
Email: j.mekyska@phd.feec.vutbr.cz
§Basque Country University, Donostia, Spain
Email: karmele.ipina@ehu.es

***Abstract*—In order to develop infocommunications devices so that the capabilities of the human brain may interact with the capabilities of any artificially cognitive system a deeper knowledge of aging is necessary. Especially if society does not want to exclude elder people and wants to develop automatic systems able to help and improve the quality of life of this group of population (healthy individuals as well as those with cognitive decline or other pathologies).**

This paper tries to establish the variations in handwriting tasks with the goal to obtain a better knowledge about aging. We present the correlation results between several parameters extracted from online handwriting and the age of the writers. It is based on BIOSECURID database, which consists of 400 people that provided several biometric traits, including online handwriting. The main idea is to identify those parameters that are more stable and those more age dependent. One challenging topic for disease diagnose is the differentiation between healthy and pathological aging. For this purpose, it is necessary to be aware of handwriting parameters that are, in general, not affected by aging and those who experiment changes (increase or decrease their values) because of it. This paper contributes to this research line analyzing a selected set of online handwriting parameters provided by a healthy group of population aged from 18 – 70 years. Preliminary results show that these parameters are not affected by aging and therefore, changes in their values can only be attributed to motor or cognitive disorders.

***Keywords*—on-line handwriting; aging; BIOSECURID database**

I. Introduction

A challenging research topic is the differentiation between healthy and pathological individuals, considering that health is something that evolves with aging. While most interesting research is based on longitudinal studies, this is time consuming and difficult to implement, due to the impossibility to ensure the availability of participants during a long time period.

Aging can be defined as the accumulation of changes in people over time [1]. Aging is a multidimensional process of physical, psychological, and social change. Some dimensions of aging grow and expand over time, while others decline. Reaction time, for example, may slow with age, while knowledge of world events and wisdom may expand. Research shows that even late in life, potential exists for physical, mental, and social growth and development [2]. Aging is an important part of all human beings reflecting the biological changes that occur, but also reflecting cultural and societal conventions. Aging is among the largest known risk factors for most human diseases [3]. Roughly 100.000 people worldwide die each day of age-related causes [4].

While aging affects daily life activities, it also affects the interaction capabilities with other people as well as machines.

Cognitive infocommunications (CogInfoCom) [5]–[7] investigates the link between the research areas of infocommunications and the cognitive sciences, as well as the various engineering applications which have emerged as the synergic combination of these sciences.

The primary goal of CogInfoCom is to provide a systematic view of how cognitive processes can co-evolve with infocommunications devices so that the capabilities of the human brain may not only be extended through these devices, irrespective of geographical distance, but may also interact with the capabilities of any artificially cognitive system. This merging and extension of cognitive capabilities is targeted towards engineering applications in which artificial and/or natural cognitive systems are enabled to work together more effectively.

From this point of view some interesting possibilities appear, such as the improvement of those human beings experiencing some cognitive decline with the goal to provide a successful aging. This requires some measurement functions, which could be based on high level activities done by people (speech, handwriting, etc.).

The concept of successful aging can be traced back to the 1950s and was popularized in the 1980s. Successful ageing consists of three components [8]:

978-1-4799-7280-7/14/$31.00 ©2014 IEEE

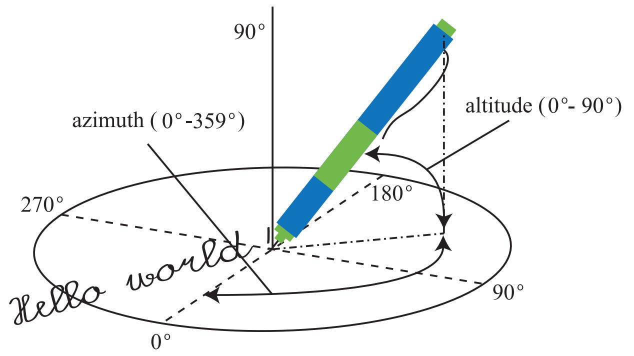


Fig. 1. Azimuth and inclination angles of the pen with respect to the plane of the graphic card

1) Low probability of disease or disability
2) High cognitive and physical function capacity
3) Active engagement with life.

In this paper we will use a simple approach, consisting of evaluating a selected set of online handwriting parameters extracted from the benchmarked handwriting database, BIOSECURID [9]. The database contains the handwriting of 400 subjects, which were required to perform specific handwriting tasks (among those their signature and copying a predefined paragraph). The handwritten was acquired by means of an Intuos Wacom 4 digitizing tablet plus an inkpen. From the users point of view, he is writing with a ordinary ball pen on an ordinary sheet of paper. In fact, the digitizing tablet is behind the sheet. The nice advantage of this online handwriting acquisition is the possibility to accurately measure handwriting timing, pressures and angles. In addition, the handwriting data are acquired in real time.

Other than the Intuos Wacom digitizing tablet, online handwriting data can be acquired through a stylus-operated PDAs. These devices can capture the following information:

1) Position in x-axis.
2) Position in y-axis.
3) Pressure applied by the pen.
4) Azimuth angle of the pen with respect to the tablet (see Fig. 1).
5) Altitude angle of the pen with respect to the tablet (see Fig. 1).

Using this set of dynamic data, further information can be inferred, such as handwriting acceleration, velocity, instantaneous trajectory angle, instantaneous displacement, tangential acceleration, curvature radius, centripetal acceleration, and more.

In order to establish a baseline for comparing healthy vs. pathological handwriting parameters (because of aging) it is necessary to have information on how they change along the time. To this aim, the authors collected a multimodal database of healthy people, which consists of handwriting data provided by university students, lecturers and administrative/support people. On the other hand, the acquisition of pathological samples requires:

- The participation of medical doctors able to label the samples.

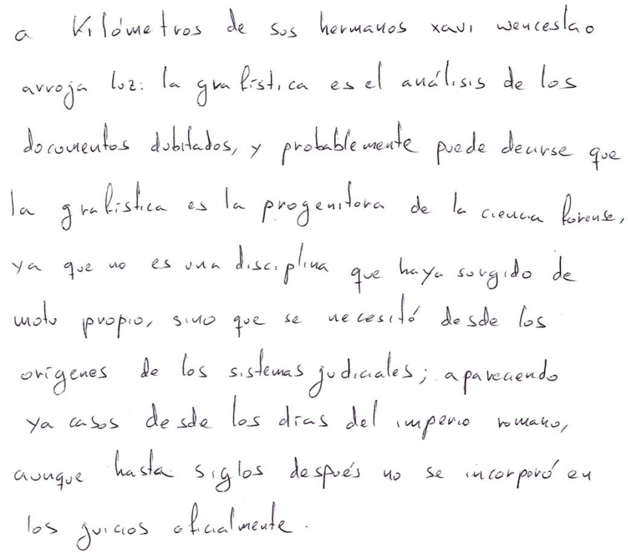


Fig. 2. Cursive handwriting task produced by a male subject

- The access to people affected by some pathology.

For most of the engineering teams it is hard to access this kind of samples, mostly because of ethical and privacy issues. For these reasons, to date, online pathological handwriting databases do not exist. Nevertheless, the differentiation between pathological and healthy samples requires a previous analysis of healthy population. This paper wants to contribute to this last issue.

II. Experimental Results

A. Biosecurid Database

In this paper we use a specific task of the BIOSECURID database [1], which consists of the copy of a predefined paragraph. Fig. 2 shows an example obtained from a masculine writer.

Using this task we have evaluated the parameters described in table I for the 400 users and for second acquisition session (the database consists of four different acquisition sessions suitable to analyze intra-user variations, which are not considered in this study). For sake of clarity, it must be said that the majority of the subjects involved in the data collection were university students with an average age of 20 years, as it can be seen from Fig. 3 that illustrate the subjects' age histogram. Nonetheless, the database also contains the handwriting of more aged subjects and their data were used for comparison.

B. Experimental Setup

A Pearson correlation between several features extracted from the online handwritten task and the age of the writer is applied, to check if some correlation exists.

In statistics, the Pearson product-moment correlation coefficient (sometimes referred to as the PPMCC or PCC or Pearson's r) is a measure of the linear correlation (dependence) between two variables X and Y, giving a value between $+1$ and -1 inclusive, where 1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation. It

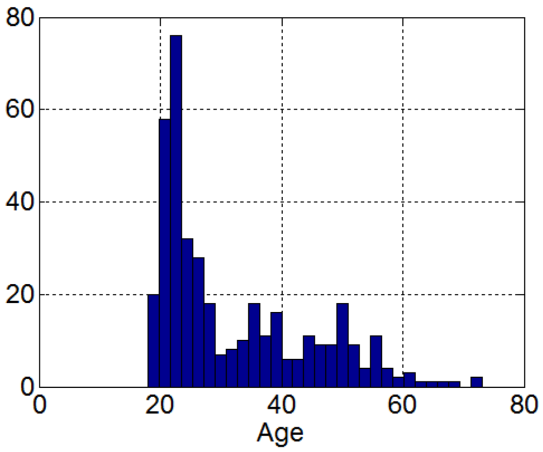


Fig. 3. Histogram of the participants' ages

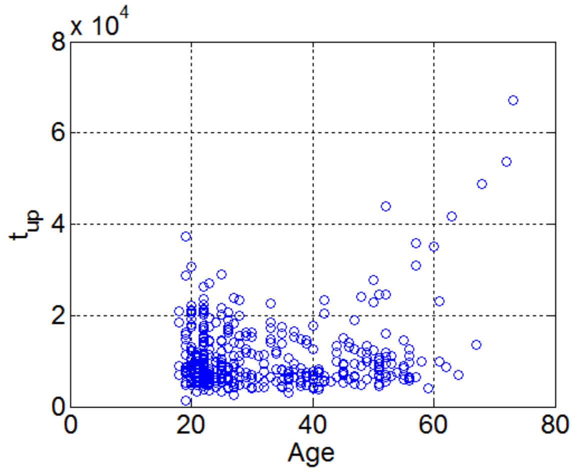


Fig. 4. Distribution of the pair of values (age, time up) for each writer

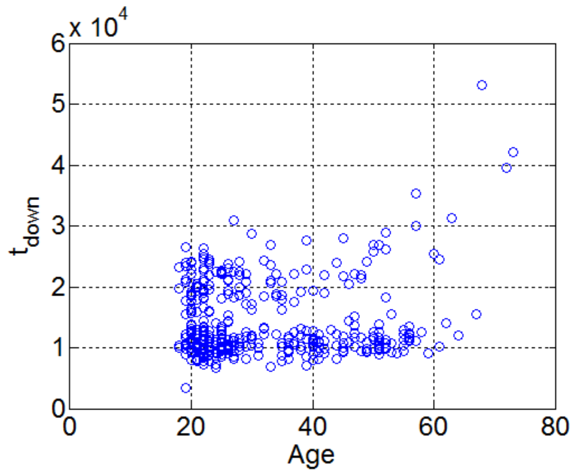


Fig. 5. Distribution of the pair (age, time down) for each writer

is widely used in sciences as a measure of the degree of linear dependence between two variables. In our case, the first variable is a measurement performed on the handwriting task (for instance time required on the paper surface to finish the task) and the second variable is the age of the writer. The interpretation of the correlation coefficient can be based on these ranges:

- High correlation: .5 to 1.0 or -0.5 to 1.0
- Medium correlation: .3 to .5 or -0.3 to .5
- Low correlation: .1 to .3 or -0.1 to -0.3

C. Experimental Results

Fig. 4 illustrates the distribution along the age of the handwriting parameter "time up", i.e. the time the pen was not on the sheet of paper (age, time up), for each writer. Fig. 5 represents the same for the "times down" parameter.

Table I shows the selected set of handwriting measured features, their Pearson correlation coefficinet and the p value, which is an index of the correlation significance. The selected handwriting features are: time up and down, pressure, speed of the trajectory, entropy, Zero crossing rate, number of strokes, normalized times, differential values obtained by the first (d) and second derivative (dd), pressure higher than a predefined threshold and Teager energy operators [10]. The *m* at the end of the parameter name stands for mean, *std* for standard deviation, and also reported are the median, mode, mean value (m) etc.

As it can be read from the "Rho" column reported in Table I, none of the selected handwriting parameters show a high correlation with the age, suggesting that age do not affect this daily functional activity. For those parameters where it would be possible to guess a weak correlation, the *p* value was too high to made it significant. A weak correlation with the age (near medium correlation) was found only for the "*nt_up*" feature (see Table I for the list of selected parameters), which has been defined by the authors as the normalized time up. It consists of the time up split by the number of strokes up in the air.

While our initial guess was that some of the selected features could have been affected by aging, the results of this research do not confirm it. However, these results have very good implications: once it has been established that the selected handwriting parameters are not (or weakly) affected by aging, they can be used as a baseline for discriminating between healthy/pathological subjects. Instead, if our initial guess was confirmed, the healthy/pathological classification of subjects through handwriting parameter would have be more challenging, suggesting that handwriting features were not useful for discriminating among healthy and pathological subjects.

Fig. 4 and 5 show some values in the region of the "more aged" area that seem to suggest an increase of time required to do the task, especially in the "*up in the air*" case. However, as it can be seen from the data reported in Table I, this is not confirmed by a strong Person's correlation coefficient. Clearly, these results must be confirmed with more data, in particular with more data from more aged subjects. Thus, we should consider these results preliminary.

To substantiate these results, one future work is to acquire a balanced database containing handwriting data from a large and balanced range of ages, and from individuals with similar education level, and maybe also from different culture.

TABLE I. Some extracted features, their Pearson's correlation and the p value.

features	Rho	p
t_upm	0.2	6.80E-05
t_downm	0.14	0.00448107
p_meanm	0.1	0.03966636
p_maxm	0.12	0.02060574
p_medianm	0.09	0.05866625
p_modem	0.08	0.12951908
p_stdm	0.06	0.22729248
speed_maxm	0.06	0.2198899
entropy_xm	-0.09	0.05843317
entropy_ym	0.12	0.01604943
entropy_pm	-0.1	0.05707732
ZCRm	-0.21	2.73E-05
NZCRm	-0.21	2.73E-05
strokes_dm	-0.21	2.73E-05
strokes_um	-0.21	2.73E-05
nt_up	0.29	4.50E-09
nt_down	0.22	5.86E-06
dp_meanm	0.04	0.404588
dp_maxm	0.19	0.00016188
ddp_maxm	0.18	0.00036086
entropy_dpm	-0.19	0.00016748
entropy_ddpm	-0.19	9.04E-05
entropy_accelerationm	-0.06	0.23452631
p100m	0.16	0.00157878
p200m	0.17	0.00069852
p300m	0.15	0.00188938
p400m	0.12	0.01359625
p500m	0.12	0.01319046
p600m	0.14	0.00625067
p700m	0.14	0.00471245
p800m	0.13	0.00774605
p900m	0.12	0.02098432
teagerxmax	0.04	0.42055342
teagerym	0.1	0.04343131
teagerymedian	-0.02	0.7450191
teagerymax	0.11	0.02860634
teagerpm	0.01	0.87595643
teagerpmedian	-0.01	0.90101579
teagerpmax	0.15	0.00216337

III. Conclusion

In this paper we have analyzed online handwriting features to search for age dependence. However, the Pearson correlation values for the parameters under examination revealed to be quite low, and not significant. While these results discourage the existence of a specific handwriting parameter indicative of age, we provided with these results a baseline for discriminating between healthy and pathological online handwriting parameters, considering that now we can hypothesize that handwriting is only affected by motor and cognitive disorders.

In addition, we should take into account that biological age does not need to be related to the "apparent age". Some people seem younger/older than they really are.

These findings also imply that the proposal of an automatic age estimator based on handwritten tasks is not a trivial problem, but it also implies that the alteration of some of these parameters could be related to health issues regardless of the age of the writer.

Future research works could be devoted to investigate the effects of age on other biometric traits also available in BIOSECURID, such as speech and face changes.

Acknowledgment

This work has been supported by FEDER and Ministerio de ciencia e Innovacin, TEC2012-38630-C04-03, COST IC 1206, GAP102/12/1104 and NT13499. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation.

References

[1] R. L. Bowen and C. S. Atwood, "Living and dying for sex," *Gerontology*, vol. 50, no. 5, pp. 265–290, 2004.

[2] D. E. Papalia, *Physical and Cognitive Development in Late Adulthood*. Mc-Graw Hill, 2009.

[3] A. Dillin, D. E. Gottschling, and T. Nystrom, "The good and the bad of being connected: the integrons of aging," *Current Opinion in Cell Biology*, vol. 26, pp. 107–112, 2014.

[4] A. D. N. J. de Grey, "Life span extension research and public debate: Societal considerations," *Studies in Ethics, Law, and Technology*, vol. 1, no. 1, 2007.

[5] P. Baranyi and A. Csapo, "Cognitive infocommunications: Coginfocom," in *Computational Intelligence and Informatics (CINTI), 2010 11th International Symposium on*, 2010, pp. 141–146.

[6] G. Sallai, "The cradle of cognitive infocommunications," *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 171–181, 2012.

[7] P. Baranyi and A. Csapo, "Definition and synergies of cognitive infocommunications," *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 67–83, 2012.

[8] J. W. Rowe and R. L. Kahn, "Successful aging," *The Gerontologist*, vol. 37, no. 4, pp. 433–440, 1997.

[9] J. Fierrez, J. Galbally, J. Ortega-Garcia, M. Freire, F. Alonso-Fernandez, D. Ramos, D. Toledano, J. Gonzalez-Rodriguez, J. Siguenza, J. Garrido-Salas, E. Anguiano, G. Gonzalez-de Rivera, R. Ribalda, M. Faundez-Zanuy, J. Ortega, V. Cardenoso-Payo, A. Viloria, C. Vivaracho, Q. Moro, J. Igarza, J. Sanchez, I. Hernaez, C. Orrite-Urunuela, F. Martinez-Contreras, and J. Gracia-Roche, "Biosecurid: a multimodal biometric database," *Pattern Analysis and Applications*, vol. 13, no. 2, pp. 235–246, 2010.

[10] J. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, vol. 1, 1990, pp. 381–384.

A.29 Biometric Applications Related to Human Beings: There Is Life beyond Security

Cogn Comput (2013) 5:136–151
DOI 10.1007/s12559-012-9169-9

Biometric Applications Related to Human Beings: There Is Life beyond Security

Marcos Faundez-Zanuy · Amir Hussain · Jiri Mekyska · Enric Sesa-Nogueras · Enric Monte-Moreno · Anna Esposito · Mohamed Chetouani · Josep Garre-Olmo · Andrew Abel · Zdenek Smekal · Karmele Lopez-de-Ipiña

Received: 1 December 2011 / Accepted: 10 July 2012 / Published online: 3 August 2012

Abstract The use of biometrics has been successfully applied to security applications for some time. However, the extension of other potential applications with the use of biometric information is a very recent development. This paper summarizes the field of biometrics and investigates the potential of utilizing biometrics beyond the presently limited field of security applications. There are some synergies that can be established within security-related applications. These can also be relevant in other fields such as health and ambient intelligence. This paper describes these synergies. Overall, this paper highlights some interesting and exciting research areas as well as possible synergies between different applications using biometric information.

Keywords Biometrics · Security · Healthcare · Ambient intelligence

Introduction

The term "biometrics" originates from the Greek words Bio (life) and metron (measure) and is defined as the science and technology of measuring and statistically analysing biological data. Although many people consider biometrics only relevant to security applications, in reality, the relevance of biometrics is very far reaching. This field has applications relevant to animals, plants and human beings. Some examples are:

M. Faundez-Zanuy (✉) · E. Sesa-Nogueras
Escola Universitària Politècnica de Mataró, Tecnocampus
Mataró Maresme, Avda. Ernest Lluch 32, 08302 Mataró, Spain
e-mail: faundez@eupmt.es

E. Sesa-Nogueras
e-mail: sesa@eupmt.es

A. Hussain · A. Abel
Department of Computing Science and Mathematics, University of Stirling, Stirling FK9 4LA, UK
e-mail: ahu@cs.stir.ac.uk

J. Mekyska · Z. Smekal
Department of Telecommunications, Brno University of Technology, Brno, Czech Republic
e-mail: j.mekyska@phd.feec.vutbr.cz

Z. Smekal
e-mail: smekal@feec.vutbr.cz

E. Monte-Moreno
TALP Research Center, UPC, Barcelona, Spain
e-mail: enric.monte@upc.edu

A. Esposito
IIASS (International Institute for Advanced Scientific Studies), Vietri sul Mare, Italy
e-mail: iiass.annaesp@tin.it

M. Chetouani
Pierre Marie Curie University, Paris, France
e-mail: Mohamed.chetouani@upmc.fr

J. Garre-Olmo
IAS (Sanitary Assistance Institute), Salt, Spain
e-mail: josep.garre@ias.scs.es

K. Lopez-de-Ipiña
EHU (Basque Country University), Donostia, Spain
e-mail: karmele.ipina@ehu.es

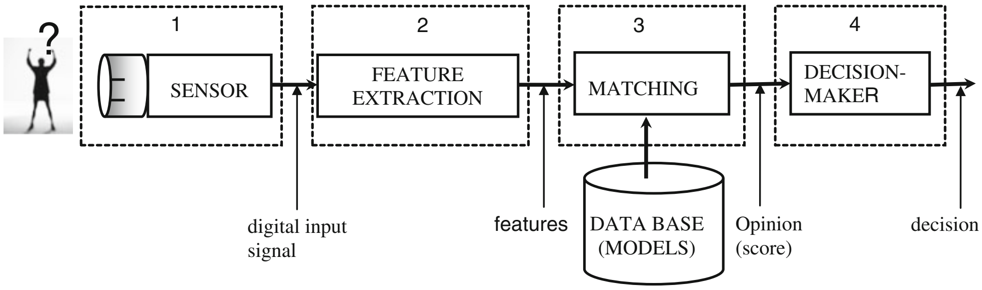


Fig. 1 Main blocks of a hypothetical biometric application system

- Statistical methods for the analysis of data from agricultural field experiments to compare the yields of different varieties of wheat.
- Analysis of data from human clinical trials evaluating the relative effectiveness of competing disease therapies.
- The analysis of biometric characteristics for animal/ human verification or identification.

The main components of a hypothetical biometric application system are shown in Fig. 1. The first block deals with the acquisition of input signals. Depending on the application and the kind of sensors, a variety of different signals may be obtained. Nowadays, most signals are acquired in a digital format or are converted to digital in order to make computerized analysis more feasible. While some signals can be acquired from both human beings and animals (such as iris and retinal analysis of the eye), others are specific to humans (such as speech, handwriting, etc.).

This paper is focused exclusively on applications that are relevant only to human beings. Therefore, we will limit discussion to only human-specific signals. The set of these signals can be split into two categories:

1) Behavioural biometrics: this category is based on the measurements and data derived from an action performed by a user and thus indirectly measures some characteristics of human body. Signature, gait, gesture and key stroking recognition belong to this category.
2) Morphological biometrics: this category is based on direct measurements of parts of the human body. Fingerprint, face, iris and hand-scanning recognition belong to this category.

However, this classification is quite artificial. For example, speech signals are dependent on behavioural traits such as semantics, diction, pronunciation, idiosyncrasy, etc. A speech signal might also be related to factors such as socio-economic status, education, place of birth, etc. Moreover, it is also dependent on individual speaker physiology, such as the shape of the vocal tract. On the other hand, physiological traits are also influenced by human behaviour, for example, the manner in which a user presents a finger and looks at a camera, etc.

Figure 2 summarizes possible biometric applications as well as the input signals that can be used for these applications. While a large set of signals can be utilized for biometric security applications, some offer much more potential in other fields, especially in the case of behavioural signals. For the remainder of this paper, we will concentrate exclusively on health and ambient intelligence applications.

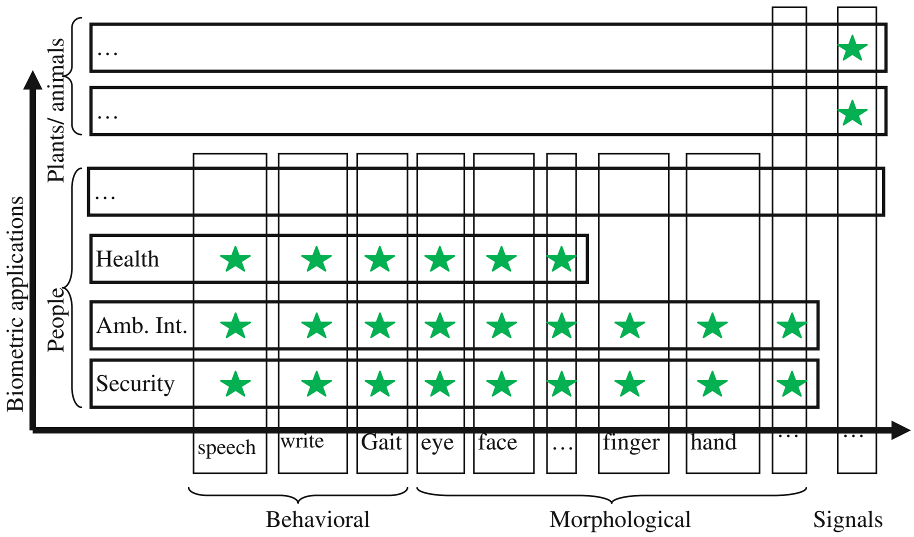


Fig. 2 Summary of main biometric applications and possible associated signals. Each star indicates the applicability of a given signal for a specific application

Health Applications

The skill level of humans is strongly related to their health state. An important example is the way our cognitive functions are related to the ageing process. Cognitive decline is a natural part of the ageing process. However, the extent of decline varies across subjects and across functions. For instance, handwriting and speech production is a fine motor control performed by our brain. When these signals are degraded, it is indicatory of health problems. Figure 3 shows the handwriting of one elder person as an example.

One important unsolved problem is how the dementia syndrome is associated with diseases such as Parkinson's and Alzheimer's, etc. In the case of Alzheimer's, it is estimated that the cost per year for a single patient is 35,000 USD in the USA. One in ten patients is below 60 years old. The incidence of Alzheimer's is doubled for every 5 years after 65, and beyond 85 years old, the incidence is between one-third and half of the amount of population. If a solution is not found, this problem will be unbearable for society. A relevant issue related to dementia is its diagnostic procedure. For example, Alzheimer's disease (AD) is the most common type of dementia and it has been pointed out that early detection and diagnosis may confer several benefits. However, intensive research efforts to develop a valid and reliable biomarker with enough accuracy to detect AD in the very mild stages or even in presymptomatic stages of the disease have not been conclusive. Nowadays, the diagnostic procedure includes the assessment of cognitive functions by using psychometric instruments such as general or specific tests that assess several cognitive functions. A typical test for AD is the clock drawing test (CDT) [84] that consists of drawing a circle and distributing the 12 h inside. An example of this is shown in Fig. 4. The top row shows the initial results produced by a person (baseline) on the left, and on the right, several samples of the same person after 6, 12 and 18 months of being damaged are also shown. This same test has also been used for detecting drug abuse, depression, etc. The bottom row of Fig. 4 shows a similar situation when copying two interlinking pentagons, which is one of the tasks of the mini-mental state examination (MMSE) [30]. The MMSE or Folstein test is a brief 30-point questionnaire test that is used to screen for cognitive impairment. It is also used to estimate the severity of cognitive impairment at a specific time and to follow the course of cognitive changes in an individual over time, thus making it an effective way to document an individual's response to treatment.

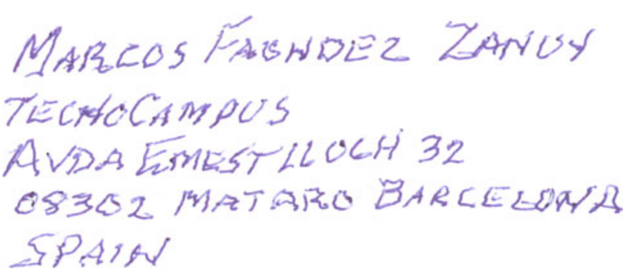


Fig. 3 Handwriting of an elder person

Research by Forbes et al. [31] showed the correlation between handwriting skill degradation and AD. Initially, it is possible to detect the disease using handwriting, especially in the case of cursive letters. Work by Neils-Strunjas et al. [60] established that some handwriting aspects are more open to vulnerabilities than others and thus can be good indicators for AD diagnosis.

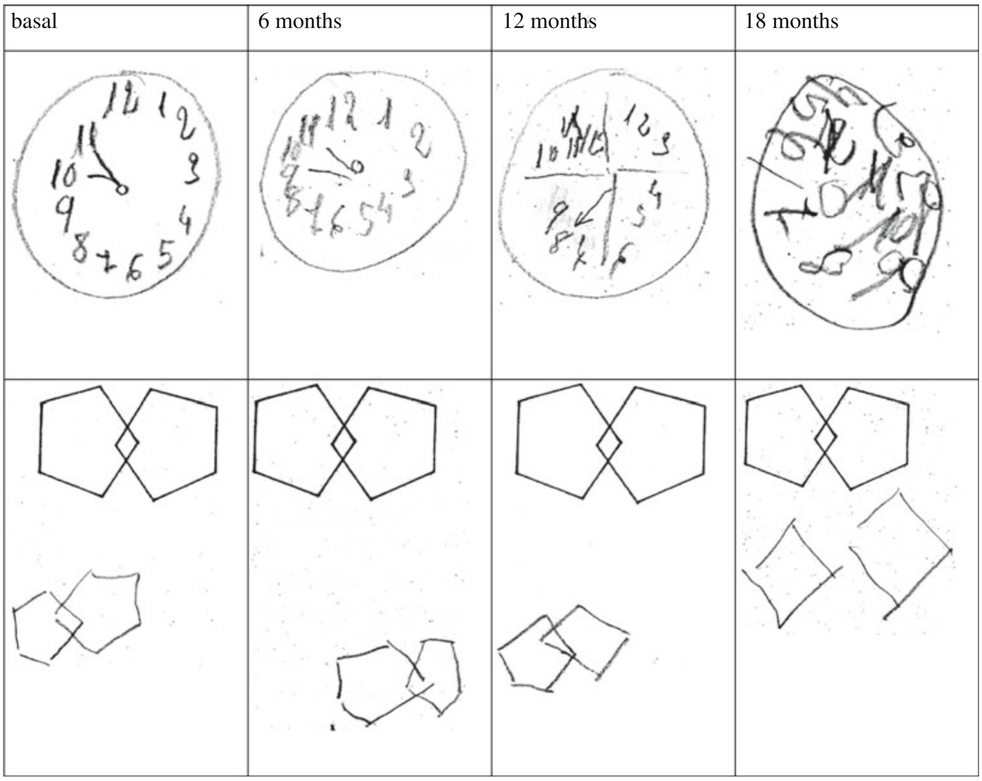
Handwriting tests are also very useful for determining the relevance of medication. For instance, Fig. 5 shows on the left the result of drawing an ellipsoid on a digitizing tablet. As can be seen, the *Y* plot, the velocity and acceleration of this coordinate are quite periodic for a healthy person (on the left). In the centre, we can see the results of a Parkinson disease (PD) patient and on the right a PD patient taking medication. It is evident that the medication permits the recovery to a large extent the skill of a healthy person. Obviously, this kind of analysis can be used for determining the dosage of drugs for a specific patient. This example has been extracted from [21]. Similar research line is exploited here [11].

There are similar experiences using the letter "ll" Tucha et al. [89, 90] and drawing an Archimedes spiral [75]. Werner et al. [94] showed the differences in handwriting between patients with mild AD and mild cognitive impairment. Ericsson et al. [23] evaluated the dictated handwriting and signature and observed that it remained unaltered longer than spontaneous writing. Heinik et al. [42] used the drawings for analysing depressive disorders in older people. Other interesting works using handwriting include:

a) Changes in handwriting due to Alcohol [27, 65]
b) Effects of caffeine on handwriting [90]
c) Effects of marijuana and alcohol [29]
d) Study of kids with perceptive/motor difficulties [48, 72]

Handwriting analysis using a digitizing tablet with an ink pen has an advantage over the classic method based on handwriting and posterior scanning, namely that the machine can acquire the information "in the air". That is, where there is no contact between pen and paper. Figure 6 shows the acquisition of the ten digits from 1 to 0 using an Intuos Wacom digitizing tablet (http://www.wacom.eu). The tablet acquired 100 samples per second including the spatial coordinates (x, y), the pressure, and a couple of

Fig. 4 Clock drawing test (*top*), pentagons of MMSE (*bottom*) for a person with AD, showing initial baseline on the *left*, and then from *left* to *right*, samples from the same person after 6, 12, and 18 months



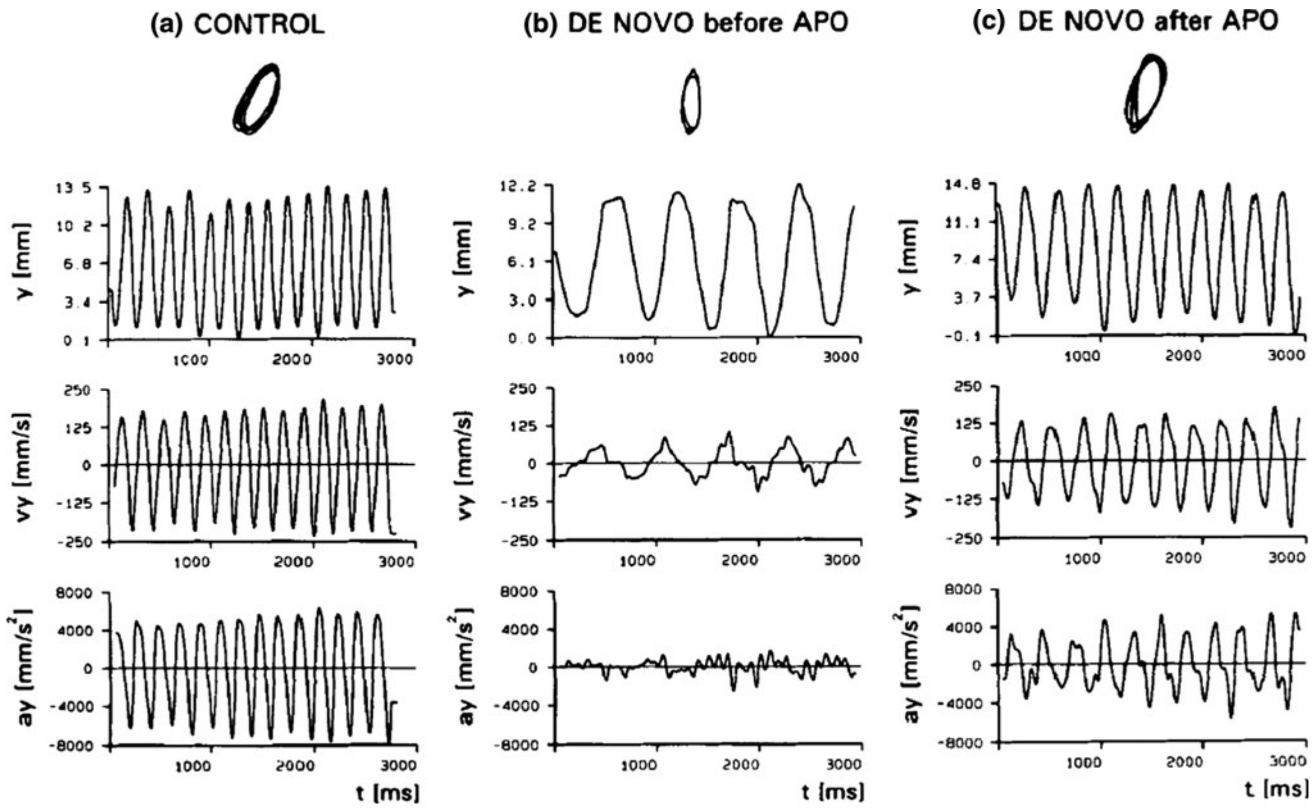
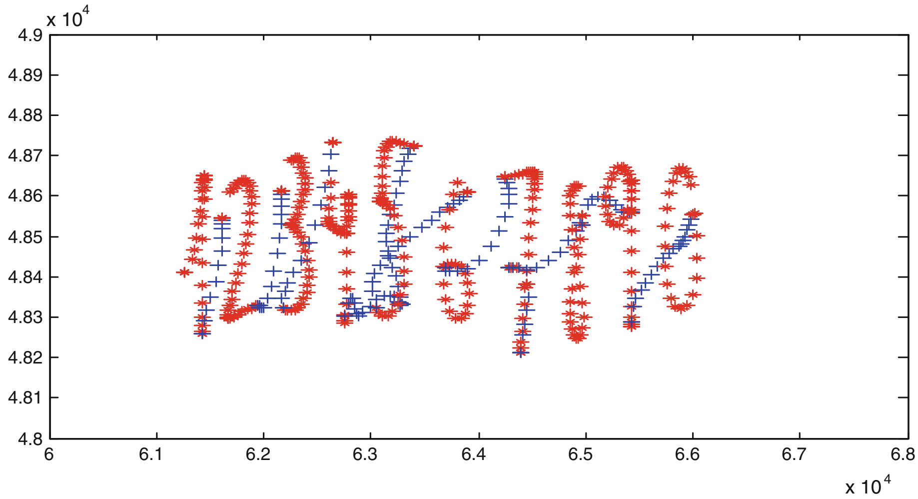


Fig. 5 Signals $y(t)$, $v_y(t)$, $a_y(t)$ (position, velocity and acceleration, respectively, of coordinate y) when drawing an ellipsoid by a healthy person (*left*), a PD patient (*centre*) and a PD patient taking apomorphine (APO)

Fig. 6 Example of handwritten numerical digits input onto a digitizing tablet. *Asterisks* (*) represent pen-down information and crosses (+) the pen-up



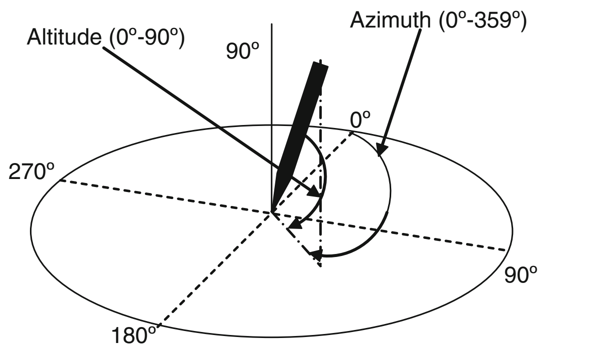
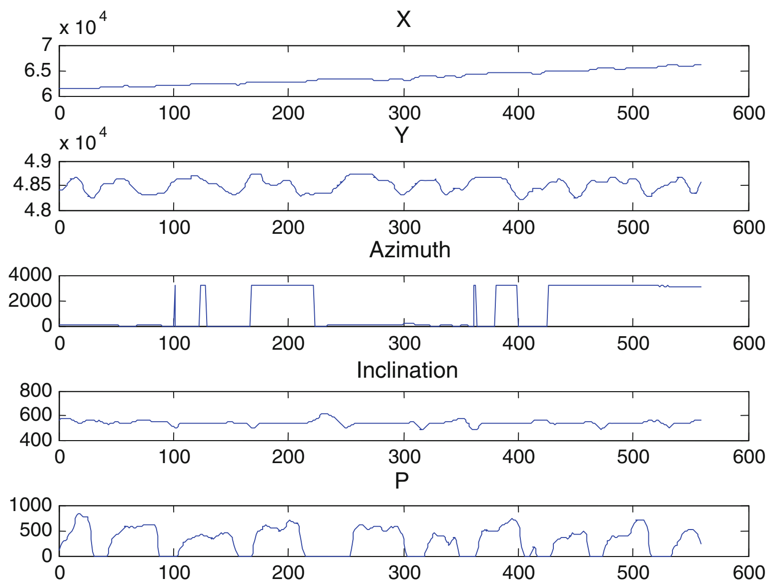


Fig. 7 Handwriting angle information acquired by the Intuos Wacom (*X*, *Y*, pressure, Azimuth, Altitude)

angles (see Figs. 7, 8). The pen-up information is represented in Fig. 6 using "+", while the pen-down is marked with "*". Our experiments on the biometric recognition of people reveal that these two kinds of information are complementary and in fact, contain a similar discriminative capability, even when using a database of 370 users [78, 79].

Speech signals represent another important possibility for health analysis. Hypokinetic dysarthria is a speech production alteration based on neurological problems [53]. There are multiple causes for this illness, such as brain paralysis, thrombosis, embolia, hemorrhagia, tumours and degenerative diseases (Alzheimer's, Parkinson's, Amyotrophic lateral sclerosis, etc.). Dysarthria affects speech quality (articulation, speech, intonation, speed, breath control, etc.) [57]. One possible analysis based on speech signals is of emotion analysis, because people affected by dementia display fewer emotions [80]. Moreau et al. [58] dealt with oral festination in PD. Festination is the tendency to speed up during repetitive movements. It appears

Fig. 8 Temporal evolution of the acquired parameters when drawing the numbers shown in Fig. 6



first with gait in order for sufferers to avoid falling down, and it subsequently appears in handwriting and speech. Ozsancak et al. [63] used speech signals to study PD. Ackermann et al. [5] analysed the trajectory of the lower lip when articulating speech signals, in order to study Parkinson's, Huntington's, cerebellum atrophia and pseudobulbar paralysis. Goberman and Coelho [36, 37], Nagulic et al. [59], Stewart et al. [83] used speech to evaluate the improvement of PD after treatment. [93] analysed the required time taken for sufferers to find the suitable word as well as time taken to articulate, and they found that AD specially affected the time taken to find the correct word, and to a lesser extent the articulation time. Rapcan et al. [69] used several measures (pitch, energy, etc.) for schizophrenia detection. They obtained promising results, which are especially interesting because there are no biological markers for this kind of disorder. Ferrand [28] used harmonic-to-noise ratio (HNR), jitter, fundamental frequency (F0), etc., and found that the most relevant of these parameters for studying the ageing process is the HNR.

Ringeval et al. [70] developed an automatic intonation recognition systems exploiting static (e.g. k-NN) and dynamic classifiers (e.g. HMMs) for the characterization of verbal productions of language-impaired children. The main results show that it is possible to characterize the prosodic abilities of those children and providing results in agreement with the clinical descriptions of the subjects' communication impairments.

Figure 9 shows a speech sentence pronounced by a healthy person and the same sentence pronounced by a PD affected person. It can be seen that the intonation is very flat for the PD sufferer, matching similar results as those reported in [41]. AD causes the changes in prosody [71]. The reason is based on the alteration of brain areas devoted to speech processing [44, 62]. In its initial stages, AD can be confused with multiple sclerosis, and speech analysis can differentiate between both [9]. Another classical

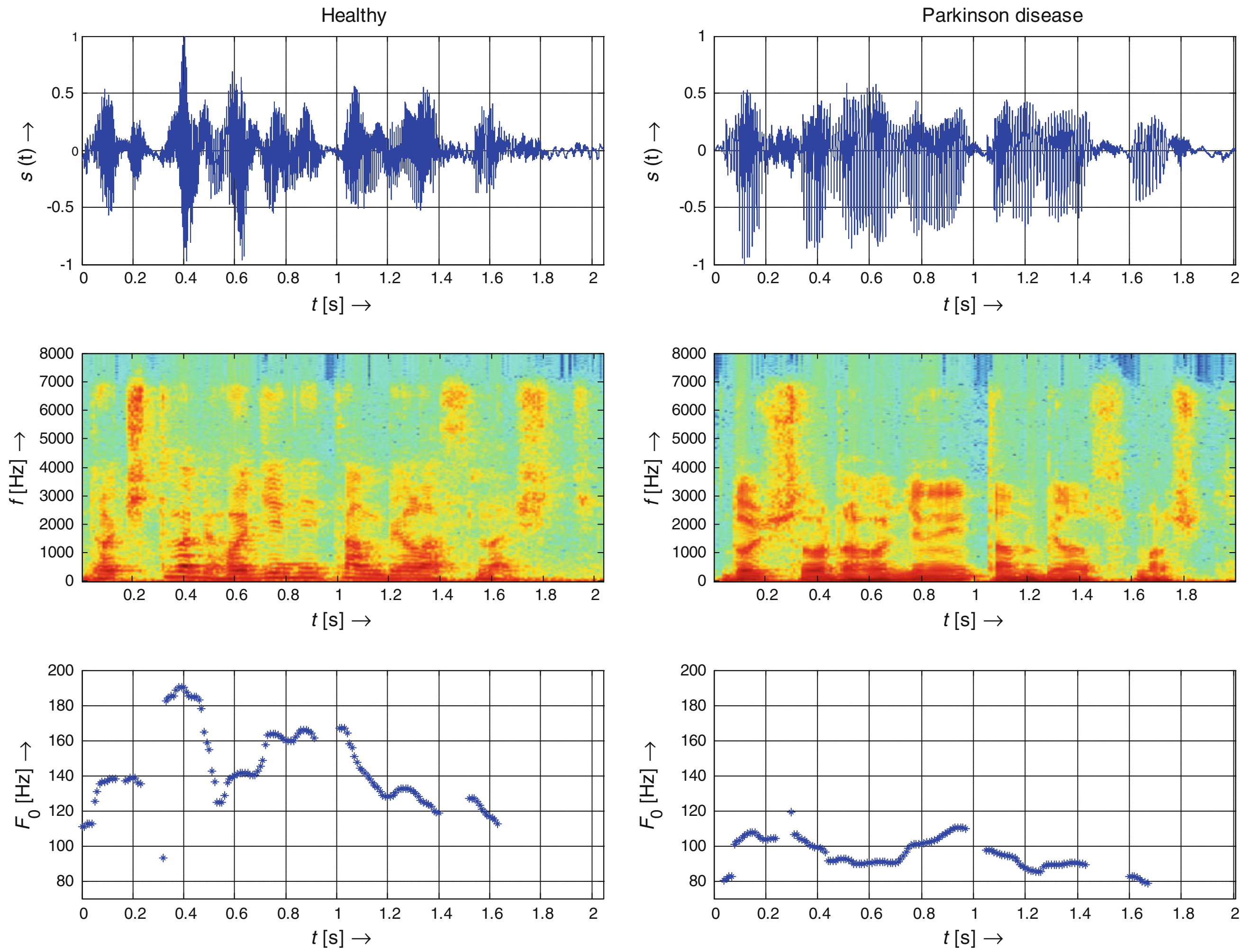


Fig. 9 Speech sentence uttered by a healthy person (*left*) and a PD person (*right*). Waveform (*top*), spectrogram (*middle*) and pitch frequency (*bottom*)

application of speech processing can be useful for dementia studies. For instance, AD patients exhibit a reduced vocabulary [19, 47]. Thus, speech recognition applications can be useful for evaluating the reduction of vocabulary in spontaneous speech.

Thus, speech signals offer significant potential for health analysis. Nevertheless, its acquisition can be more complicated than handwriting due to microphone position, recording level adjustment, etc. Some studies, such as [40], studied jointly handwriting and speech, although they were focused on lexical issues. Obviously, possibilities with other signals exist, such as the pupil reaction to light [32] and SPECT (single-photon emission computed tomography) [39], PET (positron emission tomography) [45], and MRI (magnetic resonance imaging) image analysis [88].

An interesting application of iris recognition might be to use this technology for characterizing the relationship of the change of pupil to the mood state of one person [33]. It is known that depressed patients manifest a shorter latency for constriction than control subjects, which is related to the fact that in depression, the activity of the neurotransmitters' decreases. Another application of biometric iris recognition technologies can be to predict the risk of age-related macular degeneration. Macular degeneration is one of the main causes of loss of sight in elderly people, and changes in iris colour are a sign of the risk of this illness [43]. Another extremely promising use of biometric technologies can be for a noninvasive estimate of cholesterol, through the changes in the iris of a patient [67].

Another potential use for biometric information is to develop the next generation of hearing aids. The previous audio-only developments in the field of speech enhancement (such as multi-microphone arrays and speech enhancement algorithms) have been developed academically and then been implemented into commercial hearing aids for the benefit of the hearing impaired community. In recent years, hardware has developed to an extent that very sophisticated multiple microphone hearing aids have been developed that exclusively exploit the audio modality. It is expected that in the future, conventional hearing aids will be transformed to make use of visual information with the aid of cameras for input in addition to conventional audio input, demonstrating that it is possible to combine audio and visual information to further improve the quality and intelligibility of speech in real-world noisy environments.

Speech is produced by vibration of the vocal cord and the configuration of the vocal tract that is composed of articulatory organs, and due to the visibility of some of these articulators such as tongue, teeth, and lips, there is an inherent relationship between the acoustic and visible properties of speech production. The speech perception connection between audio and visual aspects of communication has been established since pioneering works in 1954 [87] and subsequent developments such as the McGurk effect [56]. In addition, audiovisual speech correlation has been deeply investigated in the literature [7, 8, 74], including in the work by [4, 18], showing the connection between lip movement and acoustic speech and that this connection could be used for enhancing noisy speech.

Multimodal correlation is of interest because of the application of visual information to the speech enhancement domain. To the best knowledge of the authors, the first example of a functioning audiovisual speech filtering system was proposed in 2001 [35], and this was then followed by other related work [38, 81, 82]. The increased processing power of computers and the miniaturized and improved capability of relevant technical components such as video cameras and processors have made the concept of utilizing cameras for speech processing, possibly even as part of a hearing aid system, much more feasible. There are both strengths and weaknesses with the use of visual information for speech enhancement, but it has proved practical for further development. Following the pioneering work by Girin et al. [35], more recent work has focused on the use of visual information as part of a source separation-based system [81, 82]. In addition, [6] has made use of visual information as part of a Wiener filtering speech processing system. The use of visual biometric information, applied intelligently, has the potential to improve the quality of future hearing aid devices and aid the lives of those who suffer from hearing impairment.

Multimodal signal processing plays an important role in human communication analysis due to its integrative process. Indeed, correlations between speech and visual information (e.g. gestures, movements) make it possible to extract intra- and inter-coordination. In Delaherche and Chetouani [20], a general framework is proposed for the characterization of dyadic interactions for the automatic assessment of the interactional synchrony, which is considered as a measure of the quality of interaction.

Ambient Intelligence

In computing, ambient intelligence refers to electronic environments that are sensitive and responsive to the presence of people. According to [1, 96], it is characterized by systems and technologies that are:

- Embedded: many networked devices are integrated into the environment;
- Context aware: these devices can recognize you and your situational context;
- Personalized: they can be tailored to your needs;
- Adaptive: they can change in response to you;

- Anticipatory: they can anticipate your desires without conscious mediation.

While probably the highest level of intelligence that a machine can posses is the knowledge about the health condition of the human beings in front of machine, there is much other possible information that the machine can infer, such as

- Who is in front of the machine? (man/woman)
- How old are they? (child, elder, etc.)
- What is their emotional state? (angry/sad/happy, etc.)
- Who is speaking in a given room?

In a daily body-to-body interaction, emotional expressions play a vital role in creating social linkages, producing cultural exchanges, influencing relationships and communicating experiences. Emotional information is transmitted and perceived simultaneously through verbal (the semantic content of a message) and nonverbal (facial expressions, vocal expressions, gestures, paralinguistic information) communicative tools, and contacts and interactions are highly affected by the way this information is communicated/perceived by/from the addresser/addressee. Therefore, research devoted to the understanding of the relationship between verbal and nonverbal communication modes, and to investigate the perceptual and cognitive processes involved in the perception of emotional states, as well as the role played by communication impairments in their recognition, is particularly relevant in the field of human–human and human–computer Interaction both for building up and hardening human relationships and for developing friendly and emotionally coloured assistive technologies.

Emotions are considered as adaptive reactions to relevant changes in the environment, which are communicated through a nonverbal code from one organism to another [66]. This perspective is based on several assumptions, among which, the most important is that there exists a small set of universally shared discrete emotional categories from which other emotions can be derived [22, 46]. This small set of emotional categories includes happiness, anger, sadness and fear, which can be reliably associated with basic survival problems such as nurturing offspring, earning food, competing for resources, avoiding and/or facing dangers. In this context, basic emotions are brief, intense and adapted reactions to urgent and demanding survival issues. These reactions to goal-relevant changes in the environment require "readiness to act" and "prompting of plans" in order to appropriately handle (under conditions of limited time) the incoming event producing suitable mental states, physiological changes, feelings and expressions [34].

The categorization of emotions is, however, debated among researchers and different theories have been proposed for its conceptualization, among these dimensional models [73, 77]. Such models envisage a finite set of primary features (dimensions) in which emotions can be decomposed and suggest that different combinations of such features can arouse different affective states. Bringing the dimensional concept to an extreme, such theories suggest that, if the number of primary features extends along a continuum, it would be possible to generate an infinite number of affective states. This idea, even though intriguing, clashes with the principle of economy that seems to rule the dynamic of natural systems, since in this case, the evaluation of affective states may require an infinite computational time. Moreover, humans tend to categorize, since it allows for them to make associations, rapid recovery of information, and facilitates handling of unexpected events, and therefore, categories may be favoured in order to avoid excessive processing time. Furthermore, this discrete evolutionary perspective of basic emotions has been supported through several sources of evidence, such as the findings of (1) an emotion-specific autonomic nervous system's (ANS) activity[1] [50]; (2) distinct regions of the brain tuned to handle basic emotions [64]; (3) presence of basic emotional expressions in other mammalian species (as the attachment of infant mammals to their mothers) [61]; (4) universal exhibition of emotional expressions (such as smiling, amusement and irritability) by infants, adults, blind and sighted [61]; (5) universal accuracy in recognizing facial and vocal expressions of basic emotions by all human beings independently of race and culture [22, 46, 76].

Most of the relevant applications in information communication technologies exploit what are called the "expressions of emotions", that is, changes in expressions that allow interactants to perceive an emotional state during face-to-face interaction. In this sense, the perceptual appearance of emotional states is attributed to perceptual changes in the facial, vocal and gestural expressions [24–26].

In the field of human computer interface (HCI), the research objectives are to identify methods and procedures capable of automatically identifying human emotional states exploiting the multimodal nature of emotions. This requires the consideration of several key aspects, such as the development and the integration of algorithms and procedures for applications in communication, and for the recognition of emotional states, from gestures, speech, gaze and facial expressions, in anticipation of the implementation of intelligent avatars and interactive dialog systems

[1] It should be noticed that not all these findings proved to be strong enough, as, for example, [10, 13] disconfirmed the existence of an autonomic specificity and distinctive ANS's activity patterns for each basic emotion.

that could be exploited to improve the learning and understanding of emotional behaviour and facilitate the user's access to future communication services.

Emotional processes in disabilities and health disorders follow in some aspects the same paths exploited in typical normal conditions and are different in other aspects. Impairments and developmental disorders may change emotional expressions and needs with respect to normal emotional processes.

Emotional reactions may be different. Questions on how these differences are expressed, felt and relevant to social interaction are still open and can be considered to still be at a theoretical level. Comparing normative and disordered expressions of emotional states can be useful not only for implementing effective intelligent systems able to interact with disabled people, but also to improve the performance of these systems. To the best of our knowledge, very little research has been done up to now in this direction.

Given the complexity of the problem, there has been a branching of the engineering approach toward the improvement and the development of video–audio techniques, such as video and image processing, video and image recognition, synthesis and speech recognition, object and features extraction from audio and video, with the goal of developing new cutting edge methodologies for synthesizing, analysing and recognizing emotional states from faces, speech and/or body movements.

One example of an emerging ambient intelligence technology is the novel emotion and sentiment mining approach, termed sentic computing, that has been developed by Cambria and Hussain et al. [14], which aims to extract cognitive and affective information associated with natural language and, hence, better understand the current state of the user, including factors such as his/her emotional state, current needs and intent. Cambria et al. [14] also employed affective ontologies and common sense reasoning tools to analyse text not only at document, page or paragraph level, but also at sentence and clause level.

Sentic computing involves the use of AI and Semantic Web techniques, for knowledge representation and inference; mathematics, for carrying out tasks such as graph mining and multi-dimensionality reduction; linguistics, for discourse analysis and pragmatics; psychology, for cognitive and affective modelling; sociology, for understanding social network dynamics and social influence; and finally ethics, for understanding related issues about the nature of the mind and the creation of emotional machines.

In the field of health, in particular, sentic computing has been used for the development of patient-centred applications [15], which empower the real end-users of the health system by bridging the gap between unstructured and structured health-care data [16]. Sentic computing is also employed for the development of intelligent multimodal affective interfaces, in which many different technologies are concurrently applied and integrated, for example, a facial emotional classifier and a multimodal animation engine for managing virtual agents and 3D scenarios [17].

Different sensors usable in a home environment are nowadays available at reasonable prices. In the last few years, many efforts have been made to build different frameworks capable of integrating unstructured signals received from different sources. The main aim of such systems is to enhance everyday living (e.g. for home automation system) but also to allow people who require care to safely live in their home environment [55].

An interesting example of ambient intelligence has been given in Rantz et al. [68] where a number of sensors have been installed in an apartment within a retirement community. The sensor network (including bed, chair, stove temperature and motion sensors) passively collected data to detect the presence of the person in different rooms and to infer when the person is carrying out specific activities. Data from sensors are then aggregated for each patient and made available to clinicians and researchers; graphical representations of the activity level could help healthcare providers to detect any changes in activity patterns, after receiving automated alert from the system. It has been shown the potential of this kind of system for early detection of specific pathologies (e.g. for urinary treat infections).

There are a number of diverse applications for ambient intelligence-based technologies and systems that are expected to impact our daily life in the future. For example, consider the case of future intelligent transportation systems for tackling drink driving. One hypothetical example could be a scenario where a driver intends to drive his/her car after a night out drinking with friends, an embedded ambient intelligent system within his car will be able to automatically detect this situation and judge whether the level of alcohol is below the legal limit or not. If the system finds the driver might be illegally driving, it may send a signal to alert the driver and in an extra case it can stop the driver from starting his/her car. Another potential application of intelligent transport is where ambient intelligence can also help to alert the driver to be aware of speeding if the intelligent system can recognise the driving situation and detect the speed limit.

In adaptive cruise and steering control (ACC) of future cognitive or "smart" vehicles, ambient intelligence may also play a key role. Figure 10 illustrates such a typical system where an intelligent multivariable multi-controller approach is employed to realize speed tracking by using a longitudinal and lateral vehicle model and a switching strategy from one mode to another [2, 3]. One example is given in Fig. 11 showing how the intelligent system is able

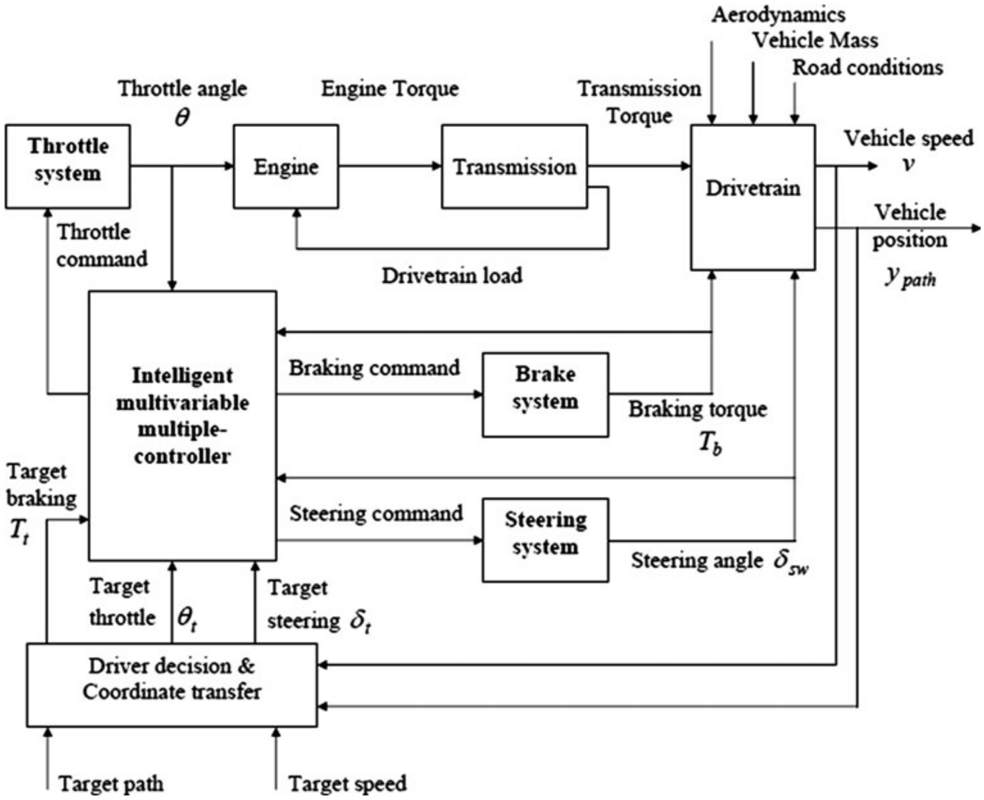


Fig. 10 Intelligent multivariable multi controller approach to adaptive cruise control

to track the target speed of vehicle by switching the controllers between two modes [2, 3].

Synergies and Interactions between Health and Security Applications

Ideally, a system with ambient intelligence should be able to detect the age of the persons in front of it, and their gender, health condition, emotional state, etc. This information should be inferred from the signals described in Section 2. Some security systems are also able to detect heart rates. If the heart rate is higher than a predetermined threshold, a silent alarm is activated because the system considers that the person is providing their biometric trait in a situation where they may be under duress.

One of the main concerns of biometrics applied to security is about privacy issues. Technological advances let to store, gather and compare a wide range of information on people. Using identifiers such as name, address, passport or social security number, institutions can search databases for individuals' information. This information can be related to salary, employment, sexual preferences, religion, consumption habits, medical history, etc. Though in most of the scenarios there should be no problem, there is a potential risk. Let us think, for instance, in sharing medical information. Obviously, in case of emergency, this sharing between hospitals would be beneficial. On the contrary, if this information is transferred to a personal insurance company or a prospective employer, the insurance or the job application can be denied. The situation is especially dramatic when biometric data collection is intended for security purposes but a third party tries to infer the health condition of the subject. For instance, in the case of retina and iris recognition, an expert can determine that a patient suffers from diabetes, arteriosclerosis, hypertension, etc.

For any biometric identifier, there is a portion of population for which it is possible to extract relevant information about their health, with similar implications to the ones described in previous paragraph, for example, speech disorders, hair or skin colour problems, etc. An important question is what exactly is disclosed when biometric scanning is used. In some cases, additional information not related to identification might be obtained. For instance, [95] presents a list of these cases that includes

- Some studies suggesting that fingerprints and finger images may disclose medical information about a person (chromosomal disorders such as Down syndrome, Turner syndrome and Klinefelter syndrome, and nonchromosomal disorders, such as chronic, intestinal pseudo-obstruction, leukaemia, breast cancer and Rubella syndrome).
- Several researchers reporting a link between fingerprints and homosexuality.

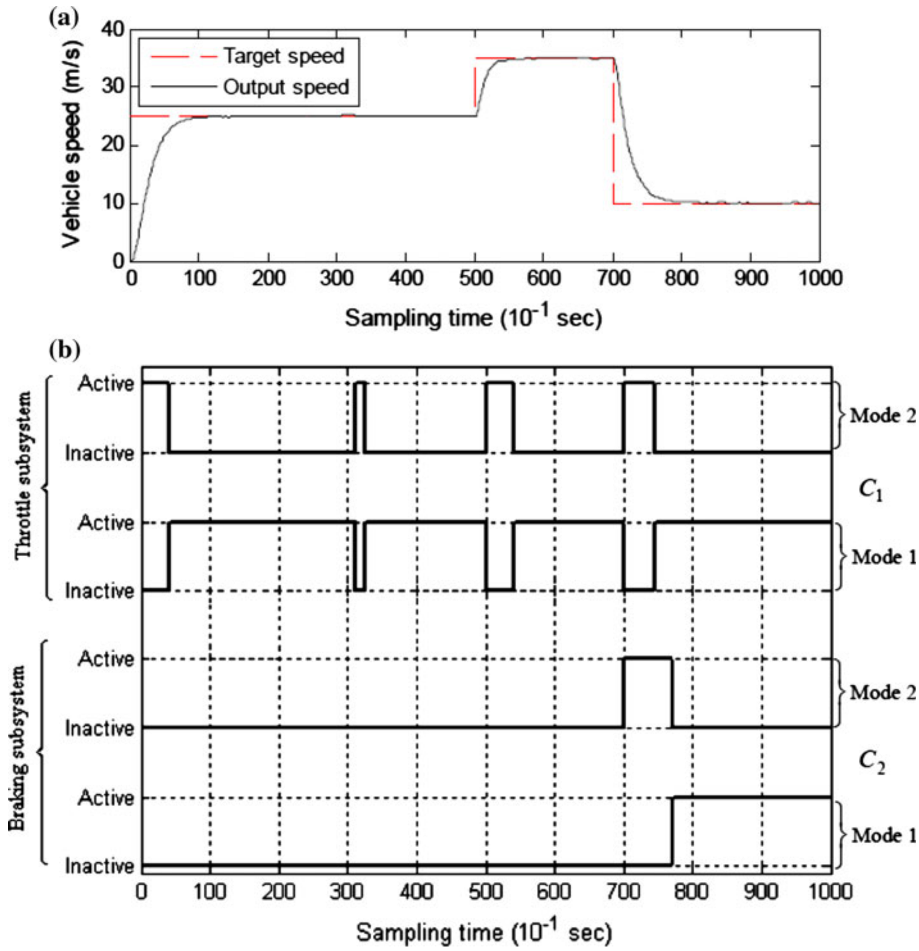


Fig. 11 Intelligent multiple controller in tracking target vehicle speed: (**a**) output speed trajectory, (**b**) multiple controller switching scheme among throttle and wheel brake subsystems

In (Maltoni et al. [54], p. 46), there is a set of references about statistical correlation between malformed fingers and certain genetic disorders.

While the relationship between genetic disorders and fingerprints may be possible, it is hard to believe that a fingerprint, which is fully formed at about 7 months of foetus development and does not change throughout the life of an individual (Maltoni et al. [54], p. 24), could be correlated with sexual preferences that can vary, or diseases that can appear and disappear during a lifetime.

Most biometric traits evolve through time. Feature extraction is a key point of classification, but nowadays there are no powerful studies about the evolution of different parameters: Which are more long lasting? Which phenomena affect these parameters? How can we use this information for a robust biometric security/health application? Is the person in front of the machine in good health condition and he/she can be responsible of his/her own acts? Fig. 12 shows a real case extracted from [92]. In this case, several women made an elder woman sign her name on blank sheets of paper (Fig. 13). Theoretically, it was to solve some issues related to medicines. When the elder person died, the other women took advantage of the signed sheets in order to write a rental agreement. The theoretical date of this agreement was 1985 (Fig. 12 on the bottom), but several documents signed in 1986 (Fig. 12 on the top) showed better control of calligraphic movements. In fact, the hesitantly written signature document signed in 1985 was closer in appearance to the blank sheets signed when the elder woman had dementia than to the 1986 document. Thus, it was demonstrated that in fact the rental document was not signed in 1985. It was signed later.

An interesting application of biometric system combined with ambient intelligence to health is the use of gait recognition for predicting falls of elderly people. From a healthcare perspective, different applications for exploiting ambient intelligence have been recently proposed. Among them, we can recall here an automated fall detection system [52] whose main aim is to promptly detect falls especially in older people to ensure a rapid medical intervention. In that study, falls are reported as the leading causes of accidental death in the US population over 65, with a large percentage of all people who died as a result of a fall being over 65. An inexpensive system based on Doppler radar sensors has been set up and a k-NN (nearest neighbour)-based classification system has been developed showing

Fig. 12 Documents signed in 1985 (hesitated) and 1986

excellent performance (with an AUC equal to 0.96) in detecting falls at home.

On a similar topic, [86] proposed a preliminary study on a depth camera device in home environments with a view of building a fall risk model. That study has shown how measurements of temporal and spatial gait parameters could be inexpensively and passively (i.e. without the active involvement of the person being observed) obtained by a depth camera (such the popular Microsoft Kinect) combined with a motion capture system for ground truth.

In biometric systems, the use of gait information has been used not only for recognizing the identity of people but also for indentifying gender [49]. The technology for gait recognition is based on deriving parameters from silhouettes, such as approximating ellipses, from which time-dependent features are extracted, which are fed into a classifier that gives as output the identity and/or the gender of the person in the image.

A straight forward extension of this idea might be to determine a specific gait sequence of a given person, and detect whether the gait process suffers changes. In the case of elderly people, it might be a good predictor of the probability of falling [91]. In this paper, the authors analysed the set of features that gave the best prediction of the risk of falling, which from a set of 7 gait markers, the feature that explained most of the variance was a slower gait speed. The technology for detecting gait anomalies does not need to be based on expensive video signal processing, but can be based on simple accelerometers [51], which can be implemented in a wrist wearable device or even on a mobile telephone. The same technology can be used for training and correcting the gait of elderly people and, therefore, diminishing the fall risk [85].

The use of the current technology on gait recognition not only can identify the danger of fall, but also can be used to train elderly people to reduce the risk of falling. As a matter of fact, the combination of biometric and ambient intelligence technologies may allow to improve the quality of living and autonomy of elderly or handicapped people, which improves the independence and auto-sufficiency and at the same time might lower the cost of attending a growing fraction of the population that needs specific care, but not on a 24 h basis.

A last straightforward question is about the physical "apparent" age and the real age. For instance, [12] reveal a loss of writing speed in later life, particularly in individuals suffering from senile psychoses. The differences in writing speed between senile subjects and "normal elderly" ones were less than the differences between normal elderly and young subjects. They also provide a plot that relates age with writing speed. Thus, theoretically, an apparent age estimation is possible looking at the writing speed, and some categorization of people could be possible: those with health condition below the average of those born the same year and those in better condition than the average. This classification could probably be considered very sensible and private data.

Conclusions

In this paper, we have discussed several applications of biometrics related to human beings beyond security applications. Mainly, we have investigated the possibilities in health and ambient intelligence, as well as the relationship between these applications. The most important issue is that the same signals used for security applications can be used for detecting diseases such as dementia, drug abuse, diabetes, arteriosclerosis, hypertension, genetic disorders, etc. This is a double-edged sword because biometric data can be used to assist in obtaining accurate and fast health diagnoses, but this information can also be illegally inferred without the consent of the user. Another synergy worth considering is when health issues are important for identity verification, that is, when the health

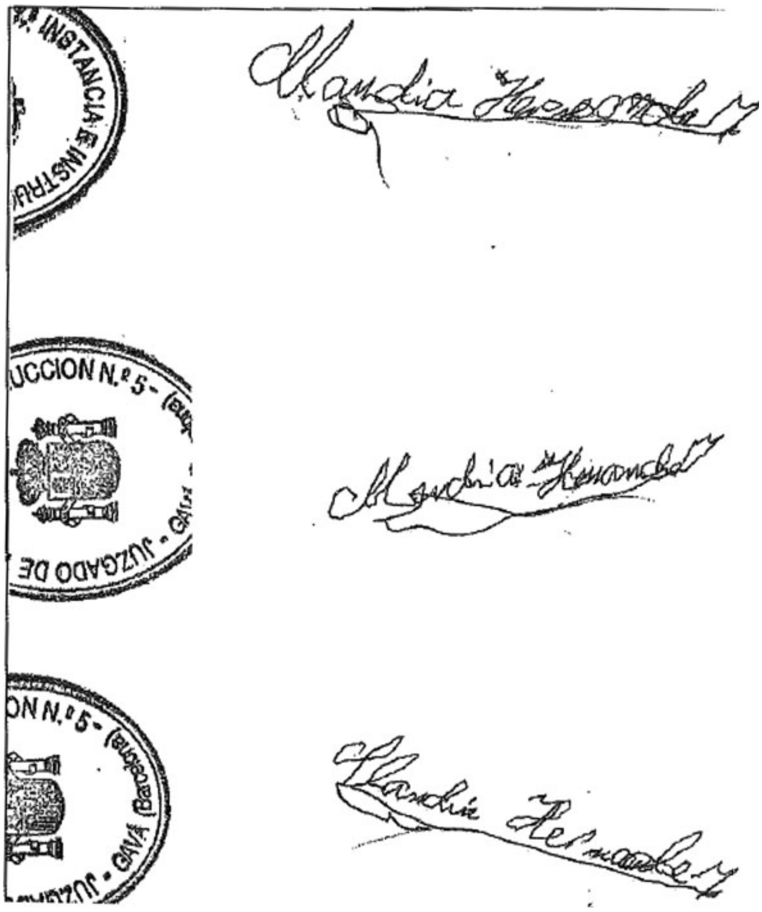


Fig. 13 Some signatures on blank sheets, when the elder woman was suffering dementia

state can change the validity of the authentication. These are the cases where the user has provided his biometric data, such as the signature, under pressure or affected by dementia.

Thus, this is a hot research topic that must be addressed, probably by signal processing teams cooperating with medical doctors and working on both biometric research fields: health and security.

Acknowledgments This work was supported by FEDER and MEC, TEC2009-14123-C04-04. SIX (CZ.1.05/2.1.00/03.0072), CZ.1.07/2.3.00/20.0094, VG20102014033, GACR 102/12/1104 and KONTAKT ME10123. We also thank Francesc Viñals and Mari Luz Puente for providing the examples in Figs. 12 and 13.

References

1. Aarts E, Harwig R, Schuurmans M. Ambient intelligence. In: Denning PJ, editor. The Invisible future: the seamless integration of technology into everyday life. New York: McGraw-Hill; 2001. p. 235–50.
2. Abdullah R. Intelligent methods for complex systems control engineering. PhD thesis with Dr Amir Hussain. UK: The University of Stirling; 2007.
3. Abdullah R, Hussain A, Warwick K, Zayed A. Autonomous intelligent cruise control using a novel multiple-controller framework incorporating fuzzy-logic-based switching and tuning. Neurocomputing. 2008;71:2727–41.
4. Abel A, Hussain A, Nguyen QD, Ringeval F, Chetouani M, Milgram M. Maximising audiovisual correlation with automatic lip tracking and vowel based segmentation. In: Biometric ID, editor. Management and multimodal communication. Berlin: Springer; 2009. p. 65–72.
5. Ackermann H, Hertich I, Daum I, Scharf G, Spieker S. Kinematic analysis of articularoty movements in central motor disorders. Mov Disord. 1997;12(6):1019–27.
6. Almajai I, Milner B. Effective visually-derived Wiener filtering for audio-visual speech processing. In: Proceedings of the interspeech. Brighton, UK; 2009.
7. Almajai I, Milner B. Maximising audio-visual speech correlation. In: Proceedings of the AVSP. 2007.
8. Barker JP, Berthommier F. Evidence of correlation between acoustic and visual features of speech. In: Proceedings of the ICPhS '99. 1999; p. 199–202.
9. Beatty WW, Orbelo DM, Sorocco KH, Ross ED. Comprehension of affective prosody in multiple sclerosis. Multiple Scler J. 2003;9(2):148–53.
10. Bermond B, Nieuwenhuyse B, Fasotti L, Schuerman J. Spinal cord lesions, peripheral feedback, and intensities of emotional feelings. Cognit Emot. 1991;5:201–20.
11. Bidet-Ildei C, Pollak P, Kandel S, Fraix V, Orliaguet J-P. Handwriting in patients with parkinson disease: effect of l-dopa and stimulation of the sub-thalamic nucleus on motor anticipation. Hum Mov Sci. 2011;30(4):783–91.
12. Birren JE, Botwinick J. The relation of writing speed to age and to the senile psychoses. J Consult Psychol. 1951;15(3):243–9.

13. Caccioppo JT, Klein DJ, Bernston GC, Hatfield E. The psychophysiology of emotion. In: Lewis JM, Haviland-Jones M, editors. Handbook of emotion. New York: Guilford Press; 1993. p. 119–42.
14. Cambria E, Hussain A, Havasi C, Eckl C. Sentic computing: exploitation of common sense for the development of emotion-sensitive systems. Lect Notes Comput Sci. 2010;5967:148–56.
15. Cambria E, Hussain A, Durrani T, Havasi C, Eckl C, Munro J. Sentic computing for patient centered applications. In: 10th IEEE international conference on signal processing (ICSP), 2010. p. 1279–82.
16. Cambria E, Hussain A, Eckl C. Bridging the gap between structured and unstructured health-care data through semantics and sentics. In: Proceedings of the ACM WebSci'11; 2011. p. 1–4.
17. Cambria E, Hupont I, Hussain A, Cerezo E, Baldassarri S. Sentic avatar: multimodal affective conversational agent with common sense. LNCS, vol. 6456. Berlin, Heidelberg: Springer; 2011. p. 81–95.
18. Cifani S, Abel A, Hussain A, Squartini S, Piazza F. An investigation into audiovisual speech correlation in reverberant noisy environments. Lect Notes Comput Sci. 2009;5641:331–43.
19. Chan D, Fox NC, Scahill RI, Crum WR, Whitwell JL, Leschziner G, Rossor AM, Stevens JM, Ciplolotti L, Rossor MN. Patterns on temporal lobe atrophy in semantic dementia and Alzheimer's disease. Ann Neurol. 2001;49:433–42.
20. Delaherche E, Chetouani M. Multimodal coordination: exploring relevant features and measures. In: SSPW '10 Proceedings of the 2nd international workshop on Social signal processing; 2010. p. 47–52.
21. Eichhorn TE, Gasser T, Mai N, Marquardt C, Arnold G, Schwarzy J, Oertel WH. Computational analysis of open loop handwriting movements in Parkinson's disease: a rapid method to detect dopamimetic effects. Mov Disord. 1996;11(3):289–97.
22. Ekman P. An argument for basic emotions. Cognit Emot. 1992;6:169–200.
23. Ericsson K, Forssell LG, Holmén K, Viitanen M, Winblad B. Copying and handwriting ability in the screening of cognitive dysfunction in old age. Arch Gerontol Geriatr. 1996;22:103–21.
24. Esposito A. The amount of information on emotional states conveyed by the verbal and nonverbal channels: some perceptual data. In: Stilianou Y, et al., editors. Progress in nonlinear speech processing. Lecture notes in computer science, vol. 4392. Berlin: Springer; 2007. p. 45–264.
25. Esposito A. Affect in multimodal information. In: Tao J, Tan T, editors. Affective information processing. Heidelberg: Springer; 2008. p. 211–34.
26. Esposito A. The perceptual and cognitive role of visual and auditory channels in conveying emotional information. Cogn Comput J. 2009;1(2):268–78.
27. Faruk A, Turan N. Handwritten changes under the effect of alcohol. Forensic Sci Int. 2003;132(3):201–10.
28. Ferrand C. Harmonics-to-noise ratio: an indexing of vocal aging. J Voice. 2002;16(4):480–7.
29. Foley RG, Miller L. The effects of marijuana and alcohol usage on handwriting. Forensic Sci Int. 1979;14(3):159–64.
30. Folstein MF, Folstein SE, McHugh PR. Mini mental state: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.
31. Forbes KE, Shanks MF, Venneri A. The evolution of dysgraphia in Alzheimer's disease. Brain Res Bull. 2004;63:19–24.
32. Fotiou DF, Stergiou V, Tsiptsios D, Lithari C, Nakou M, Karlovasitou A. Cholinergic deficiency in Alzheimer's and Parkinson's disease: evaluation with pupillometry. Int J Psychophysiol. 2009;73(2):143–9.
33. Fountoulakis K, Fotiou F, Iacovides A, Tsiptsios J, Goulas A, Tsolaki M, Ierodiakonou C. Changes in pupil reaction to light in melancholic patients. Int J Psychophysiol. 1999;31(2):121–8. ISSN 0167-8760.
34. Frijda NH. Moods, emotion episodes, and emotions. In: Lewis JM, Haviland-Jones M, editors. Handbook of emotion. New York: Guilford Press; 1993. p. 381–402.
35. Girin L, Schwartz JL, Feng G. Audio-visual enhancement of speech in noise. J Acoust Soc Am. 2001;109(6):3007–20.
36. Gobermana AM, Coelho C. Acoustic analysis of Parkinsonian speech I: speech characteristics and L-Dopa therapy. NeuroRehabilitation. 2002;17:237–46.
37. Gobermana AM, Coelho C. Acoustic analysis of Parkinsonian speech II: L-Dopa related fluctuations and methodological issues. NeuroRehabilitation. 2002;17:247–54.
38. Goecke R, Potamianos G, Neti C. Noisy audio feature enhancement using audio-visual speech data. In: Acoustics, speech, and signal processing, Proceedings (ICASSP'02), vol. 2. IEEE International Conference; 2002. p. 2025–2028.
39. Gorriz JM, Segovia F, Ramirez J, Lassl A, Salas-Gonzalez D. GMM based SPECT image classification for the diagnosis of Alzheimer's disease. Appl Soft Comput. 2011;11:2313–25.
40. Groves-Wright K, Neils-Strunjas J, Burnett R, O'Neill MJ. A comparison of verbal and written language in Alzheimer's disease. J Commun Disord. 2004;37(2):109–30.
41. Gustaw K, Gonet W. Speech disorders in multiple system atrophy of parkinson type. Clin Res. 2008;1(2):185–8.
42. Heinik J, Werner P, Dekel T, Gurevitz I, Rosenblum S. Computerized kinematic analysis of the clock drawing task in elderly people with mild major depressive disorder: an exploratory study. Int Psychogeriatr. 2010;22(3):479–88.
43. Holz FG, Piguet B, Minassian DC, Bird AC, Weale RA. Decreasing stromal iris pigmentation as a risk factor for age-related macular degeneration. Am J Ophthalmol. 1994;117(1): 19–23.
44. Iliadou V, Kaprinis S. Clinical psychoacoustics in Alzheimer's disease central auditory processing disorders and speech deterioration. Ann Gen Hospital Psychiatr. 2003;2:12.
45. Illán IA, Górriz JM, Ramírez J, Salas-Gonzalez D, López MM, Segovia F, Chaves R, Gómez-Rio M, Puntonet CG, the Alzheimer's Disease Neuroimaging Initiative. 18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis. Inf Sci. 2011;181(4):903–16.
46. Izard CE. Innate and universal facial expressions: evidence from developmental and cross-cultural research. Psycholog Bull. 1994;115:288–99.
47. Kempler D, Curtiss S. Catherine jackson "synthactic preservation in Alzheimer's disease". J speech Hearing Res. 1987;30:343–50.
48. Kushki A, Chau T, Anagnostou E. Handwriting difficulties in children with autism spectrum disorders: a scoping review. J Autism Dev Disord. 2011;41(12):1706–16.
49. Lee L, Grimson WEL. Gait analysis for recognition and classification. Automatic face and gesture recognition, 2002. In: Proceedings of the fifth IEEE international conference; 2002. p. 148–55.
50. Levenson RW. Human emotion: a functional view. In: Ekman PP, Davidson RJ, editors. The nature of emotion: fundamental questions. New York: Oxford University Press; 1994. p. 123–6.
51. Liu R, Zhou J, Liu M, Hou X. A wearable acceleration sensor system for gait recognition. In: 2nd IEEE Conference on industrial electronics and applications, ICIEA 2007; 2007. p. 2654–59.
52. Liu L, Popescu M, Rantz M, Skubic M, Cuddihy P, Yardibi T. Automatic fall detection based on Doppler radar motion signature. In: 5th International conference on pervasive computing technologies for healthcare; 2011. p. 222–5.

53. Llau Arcusa MJ, Gonzalez Alvarez J. Medida de la inteligibilidad en el habla disaártrica. Rev Logop Foniatr Audiol. 2004;24: 33–43.
54. Maltoni D, Maio D, Jain AK, Prabhakar S. Handbook of fingerprint recognition. 1st ed. New York: Springer; 2003.
55. Maternaghan C, Turner KJ. A component framework for telecare and home automation. In: CCNC'10 Proceedings of the 7th IEEE conference on consumer communications and networking conference; 2009. p. 886–870.
56. McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976;264:746–8.
57. Mekyska J, Smekal Z, Kostalova M, Mrackova M, Skutilova S, Rektorova I. Motor aspects of speech imparment in Parkinson's disease and their assessment. Cesk Slov Neurol N. 2011;74: 662–8.
58. Moreau C, Ozsancak C, Blatt J-L, Derambure P, Destee A, Defebvre L. Oral festination in parkinson's disease: biomechanical analysis and correlation with festination and freezing of gait. Mov Disord. 2007;22(10):1503–6.
59. Nagulic M, Davidovic J, Nagulic I. Parkinsonian voice acoustic analysis in real-time after stereotactic thalamotomy. Stereotact Funct Neurosurg. 2005;83(2–3):115–21.
60. Neils-Strunjas J, Groves-Wright K, Mashima P, Harnish S. Dyspgraphia in Alzheimer's disease: a review for clinical and research purposes. J speech Lang Hearing Res. 2006;49(6): 1313–30.
61. Oatley K, Jenkins JM. Understanding emotions. 2nd ed. Oxford: Blackwell; 2006.
62. Ohn TG, Braak H. Auditory brainstem nuclei in Alzheimer's disease. Neurosci Lett. 1989;2:60–3.
63. Ozsancak C, Auzou P, Jan M, Defebvre L, Derambure P, Destee A. The place of perceptual analysis of dysarthria in the differential diagnosis of corticobasal degeneration and Parkinson's disease. J Neurol. 2006;253(1):92–7.
64. Panksepp J. Emotions as natural kinds within the mammalian brain. In: Lewis JM, Haviland-Jones M, editors. Handbook of emotions. 2nd ed. New York: Guilford Press; 2000. p. 137–56.
65. Phillips JG, Ogeil RP, Muller F. Alcohol consumption and handwriting: a kinematic analysis. Hum Mov Sci. 2009;28: 619–32.
66. Plutchik R. Emotions as adaptive reactions: implications for therapy. Psychoanal Rev LIII. 1966;2:105–10.
67. Ramlee RA, Ranjit S. Using iris recognition algorithm, detecting cholesterol presence. In: International conference on information management and engineering; 2009. p. 714–7.
68. Rantz MJ, Skubic M, Koopman RJ, Phillips L, Alexander GL, Miller SJ, Guevara RD. Using sensor networks to detect urinary tract infections in older adults. In: Proceedings of the IEEE 13th international conference on e-health networking, applications and services. 2011.
69. Rapcan V, D'Arcy S, Yeap S, Afzal N, Thakore J, Reilly RB. Acoustic and temporal analysis of speech: a potential biomarker for schizophrenia. Med Eng Phys. 2010;32(9):1074–9.
70. Ringeval F, Demouy J, Szaszak G, Chetouan M, Robel L, Xavier J, Cohen D, Plaza M. Automatic intonation recognition for the prosodic assessment of language impaired children. IEEE Trans Audio Speech Lang Process. 2011;19(5):1328–42.
71. Roberts VJ, Ingram SM, Lamar M. Prosody impairment and associated affective and behavioral disturbances in Alzheimer's disease. Neurology. 1996;47:1482–8.
72. Rosenblum S, Parush S, Weiss PL. The in air phenomenon: temporal and spatial correlates of the handwriting process. Percept Mot Skills. 2003;96(3):933–54.
73. Russell JA. A circumplex model of affect. J Pers Soc Psychol. 1980;39:1161–78.
74. Sargin ME, Yemez Y, Erzin E, Tekalp AM. Audiovisual synchronization and fusion using canonical correlation analysis". IEEE Trans Multimed. 2007;9(7):1396–403.
75. Saunder-Pullman R, Derbym C, Stanley K, Floyd A, Bressman S, Lipton RB, Deligtisch A, Severt L, Qiping Yu, Kurtis M, Pullman SL. Validity of spiral analysis in early Parkinson's disease. Mov Disord. 2008;23(4):531–7.
76. Scherer KR, Banse R, Wallbott HG. Emotion inferences from vocal expression correlate across languages and cultures. J Cross Cult Psychol. 2001;32:76–92.
77. Schlosberg H. Three dimensions of emotion. Psychol Rev. 1953;61(2):81–8.
78. Sesa E, Faundez-Zanuy M. Biometric recognition using online uppercase handwritten text. Pattern Recognit. 2012;45(1): 128–44.
79. Sesa E, Faundez-Zanuy M, Mekyska J. An information analysis of in-air and on-surface trajectories in online handwriting. Cognit Comput. 2012;4:195–205.
80. Skodda S, Schlegel U. Speech rate and rhythm in parkinson's disease. Mov Disord. 2008;23(7):985–92.
81. Sodoyer D, Schwartz JL, Girin L, Klinkisch J, Jutten C. Separation of audio-visual speech sources: a new approach exploiting the audio-visual coherence of speech stimuli. EURASIP J Appl Signal Process. 2002;11(1):1165–73.
82. Sodoyer D, Girin L, Jutten C, Schwartz JL. Developing an audio-visual speech source separation algorithm. Speech Commun. 2004;44(1–4):113–25.
83. Stewart C, Winfield L, Junt A, Bressman SB, Fahn S, Blitzer A, Brin MF. Speech dysfunction in early Parkinson's disease. Mov Disord. 1995;10(5):562–5.
84. Sunderland T, Hill JL, Mellow AM, et al. Clock drawing in Alzheimer's disease: a novel measure of dementia severity. J Am Geriatr Soc. 1989;37:725–9.
85. Trombetti A, Hars M, Herrmann FR, Kressig RW, Ferrari S, Rizzoli R. Effect of music-based multitask training on gait, balance, and fall risk in elderly people: a randomized controlled trial. Arch Intern Med. 2011;171(6):525–33.
86. Stone EE, Skubic M. Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In: Proceedings of the 4th International conference on pervasive computing technologies for healthcare. Dublin; 2011.
87. Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. J Acoust Soc Am. 1954;26(2):212–5.
88. Tripolitia EE, Fotiadisb DI, Argyropoulou M. A supervised method to assist the diagnosis and monitor progression of Alzheimer's disease using data from an fMRI experiment. Artif Intell Med. 2011;53(1):35–45.
89. Tucha O, Mecklinger L, Thome J, Reiter A, Alders GL, Sartor H, Naumann M, Lange KW. Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease. J Neural Transm. 2006;113:609–23.
90. Tucha O, Mecklinger L, Walitza S, Lange KW. The effect of caffeine on handwriting movements in skilled writers. Hum Mov Sci. 2006;25(4–5):523–35.
91. Verghese J, Holtzer R, Lipton RB, Wang C. Quantitative gait markers and incident fall risk in older adults. J Gerontol Ser A Biol Sci Med Sci. 2009;64A(8):896–901.
92. Viñals Carrera F, Puente Balsells ML. "Grafología criminal", capítulo 3, alteraciones neurológicas y biológicas. Editorial Herder, 2009.
93. Warkentin S, Erikson C, Janciauskiene S. rCBF pathology in Alzheimer's disease is associated with slow processing speed. Neuropsychologia. 2008;46(5):1193–200.
94. Werner Perla, Rosenblum Sara, Bar-On Gady, Heinik J, Korczyn A. Handwriting process variables discriminating mild Alzheimer's

disease and mild cognitive impairment". J Gerontol. 2006;61B(4): 228–36.

95. Woodward J. Biometrics: identifying law and policy concerns. In: Jain AK, Bolle RM, Pankanti S, editors. Biometrics: personal identification in networked society. New York: Springer; 2005. p. 385–405.

96. Zelkha E, Epstein B, Birrell S, Dodsworth C. From devices to "ambient intelligence". Digital Living Room Conference, June 1998. http://www.epstein.org/brian/ambient_intelligence/DLR%20Final%20Internal.ppt.

A.30 Analysis of neurological disorders based on digital processing of speech and handwritten text

Analysis of Neurological Disorders Based on Digital Processing of Speech and Handwritten Text

Zdenek Smekal[1,2], Jiri Mekyska[1,2], Irena Rektorova[3,4], Marcos Faundez-Zanuy[5]
[1] Department of Telecommunications, Faculty of Electrical Engineering and Communication,
Brno University of Technology, Brno, Czech Republic
smekal@feec.vutbr.cz, j.mekyska@phd.feec.vutbr.cz
[2] SIX Research Centre, Faculty of Electrical Engineering and Communication,
Brno University of Technology, Brno, Czech Republic
[3] First Department of Neurology, Faculty of Medicine, St. Anne's University Hospital, Brno, Czech Republic
irena.rektorova@fnusa.cz
[4] Applied Neurosciences Research Group, Central European Institute of Technology, CEITEC, MU, Brno, Czech Republic
[5] EUP Mataró, Tecnocampus, Mataro, Barcelona, Spain
faundez@eupmt.es

Abstract— **The paper deals with the methods of non-invasive analysis of neurological disorders, focusing on speech signal processing and processing of handwritten text. The paper describes the whole procedure of the automated analysis of the disorder while the greatest attention is paid to a parameterization. In the case of speech signal analysis, the state-of-the-art features evaluating a presence of hoarseness, breathiness and hypernasality are mentioned. Nonlinear dynamic parameters and parameters derived from the empirical mode decomposition (EMD) are compared. Based on the tests, from the point of description of a noise component of signal, the best results were obtained using the approximation entropy, the largest Lyapunov exponent and parameters based on Teager-Kaiser energy operator, which is calculated from the first intrinsic mode function (IMF). In the case of handwritten text analysis, the most used exercises describing a tremor and movement dynamics are mentioned. The new approaches of hand movement analysis at a time when the pen tip does not touch the paper have been also proposed. Finally the paper discusses different applications of speech signal and handwriting text parameterization.**

I. INTRODUCTION

With development of information technologies, intensity of research of digital processing of biomedical signals has also increased. Thanks to sophisticated techniques and decreasing cost of hardware, it is possible to give physicians better support in their work and provide them important information, which help them to decide on next steps of treatment, or medical intervention, respectively.

Due to the changing way of life, a greater proportion of sedentary behaviour, stress, unhealthy diet, etc. different types of neurological disorders such as Parkinson's and Alzheimer's diseases, schizophrenia, dementia, dystonia, etc. become more and more in human populations. The analysis of speech utterances and handwritten text are frequently used by physicians as non-invasive methods that help them to diagnose different stages of neurological disorders and help them recognize the success or failure of the treatment. In particular, the analysis of speech and handwritten text is often used for diagnosis and analysis of Parkinson's disease. Human speech is produced by coordinated movement of speech organs using air flow generated from lungs. Handwritten text is actually formed by coordinated movement of fingers, wrist and elbow. Based on the analysis of these signals, we can find out different information about people related to their health, emotions, mental state, etc.

Parkinson's disease can manifest in speech dysarthria of patients [1], which affects the area of phonation, articulation, prosody and fluency. The voice tremor [2], hypophonia (reduced voice intensity [3]), dysphonia (degradation of speech quality [4]), hypernasality (increased nasality [5]), dysprosody [6], hesitation (unintentional introduction of pauses [7], palilalia (rapid repetition of words or syllables [7]), bradyphemia (sudden deceleration [8]) or tachyphemia (sudden acceleration in speech [8]) may be observed in patients.

In the case of handwritten text, slower movements (bradykinesia [9]), variations in speed of writing [10], decrease of the height of letters (micrographia [11]) or uncontrollable movements (dyskinesia [12]) can be seen. Sometimes muscles of the patient may be very stiff (rigidity [12]), or motionless (akinesia [9]), which results in the patient not able to write.

All above mentioned manifestations of the neurological diseases are able to describe and quantify by using carefully chosen 1D and 2D signal analysis methods. Based on the processing of these signals, the disease can be diagnosed, monitored its progression and watched an impact on individual body parts. Moreover, thanks to the parameterization, new medicines can be developed and the impact of medication on patients can be monitored. For these reasons, the methods are also used in the pharmaceutical industry. Analysis of speech and handwritten text can also be combined with other investigative methods such as functional magnetic resonance imaging (fMRI), ultrasound investigations, etc.

The aim of this paper is to describe modern approaches to non-invasive analysis of neurological disorders with emphasis on Parkinson's disease using speech signal and handwritten text

978-1-4673-6143-9/13/$31.00 ©2013 IEEE

processing. The paper is divided as follows: a general procedure for the automated analysis of neurological disorders is given in Sec. II, a description of the parameter extraction techniques from speech signals and handwritten text follows in Sec. III, Sec. IV lists the different use of non-invasive analysis of neurological disorders and the conclusion is in Sec. V.

II. Automated Analysis of Neurological Disorders

The analyzed signal of speech or handwritten text is usually acquired in a hospital or speech therapist's office. To record speech, high-quality condenser microphones are used and to register handwritten text a digital tablets are applied. Beside a current position, some tablets are also able to record pressure, azimuth and altitude. Depending on the specific disorder, it is possible to use various utterances and handwriting exercises. In the analysis of speech, the exercises of prolonged vocals, reading poems, specially adapted sentences, reading excerpts from books or exercises difficult for articulation are applied. For the analysis of handwritten text, the patients can draw ellipses, Archimedean spirals, connect points on the matrix, write selected words or phrases or draw different pictures (pointer clocks, houses, etc.). The recorded data are labelled (borders between utterances are set, written words or phrases to smaller parts are segmented, etc.) and from categorized data different databases are prepared. After the modification the data are processed automatically. Speech and handwritten text signals are further processed (re-sampling, filtering, etc.) and various parameters (features) are calculated. Based on chosen application the parameters are selected. The feature space is usually very large and it is necessary to select optimal parameters before further processing. For this reason, the mRMR (minimum Redundancy Maximum Relevance) method [13], the SFFS (Sequential Floating Forward Selection) method, the LASSO (Least Absolute Shrinkage and Selection Operator) method [14], the LLBFS (Local Learning-Based Feature Selection) method [15] or the RELIEF (Feature Weight based Algorithm Inspired by Instance-based Learning) method [16], etc. may be used. To develop a system that would identify one of the disorders, a classifier may be trained (e.g. the Gaussian Mixed Models-GMM, Support Vector Machine-SVM or Random Forests-RF). To select significant features, by which it would be possible to monitor a progression of disease or an influence of medication parametric and non-parametric statistical methods such as Student's t-test, Mann-Whitney U test, Wilcoxon paired test, etc. can be used.

III. Feature Extraction

The parameters that are extracted from speech record or handwritten text can be divided into basic, high-level and global features. The basic features are calculated directly from signal, and are expressed by scalars, vectors or are written into matrices (pitch frequency F_0, formant locations, speech intensity, etc.). If several features of one type are calculated from the same signal (pitch frequency F_0 is calculated for each speech segment with length of 20 ms), then it is suitable to select one parameter (representative) from this set to train the classifier or make statistical analysis in an easier way. From this set of features, we can calculate, for example, mean value, standard deviation, range, moments, percentiles, entropy, regression coefficients, etc. These parameters are called high-level features. Sometimes it is advantageous to calculate the global parameters, which combine individual features of different types (e.g. the first two formants calculated from different vowels). The calculation of these parameters is more difficult to automate, but the result provides a comprehensive and robust description of the disorder (e.g. possibility to describe the movement of tongue).

Based on an optimal extraction of parameters, the disorder can be appropriately quantified. It is usually the most important part of the analysis process and the improper selection of parameters can greatly distort the final result. Thanks to the optimal extraction of parameters, a physician can better quantify and objectively diagnose the disease and decide more precisely further treatment.

A. Speech Signal Analysis

In terms of analysis of patient speech records, the features which describe prosody, vocal cords function and function of speech organs (e.g. tip of the tongue) are primarily extracted. Information about vocal tract shape is not so important in this case. Disorders that are manifested in speech signal may include: Parkinson's disease, Alzheimer's disease, dementia, and schizophrenia. Speech features can be divided in terms of what they describe:

- Features describing phonation - pitch frequency F_0, jitter, glottal coefficients, etc.
- Features describing intensity of speech - shimmer, modulation energy, spectral flow, etc.
- Features describing speech fluency – index of rhythmicity, articulation rate, proportion of speech pauses, etc.
- Features describing tongue movement - location of formant frequencies and their bandwidths.
- Features describing speech quality - glottal signal to noise ratio, normalized noise energy, etc.

Segmental features such as mel frequency cepstral coefficients (MFCC) or perceptual linear prediction coefficients (PLP) are also used [17]. However, it is not clear which disorder they indicate. But, for example, regarding the MFCC coefficients Tsanas et al. [17] believe that they can indirectly describe motion of articulatory organs.

In this paper, our attention will be paid to features that are not well known in the area of speech analysis, but which can positively contribute to correct diagnosis. They are: non-linear dynamic features and parameters based on the empirical mode decomposition (EMD).

1) Non-linear Dynamic Features

Sometimes the patients with Parkinson's disease produce breathy voice, dysphonia and hypernasality so great that any features in voiced speech segments (e.g. pitch period) cannot be found and the speech signal seems to be completely chaotic [18, 19]. Speech is in this case so degraded by noise that no conventional linear methods give useful results. The features such as F_0, jitter, shimmer, etc. or formant characteristics in speech analysis completely fail. For this reason, chaos theory is increasingly applied. This theory helps determine how much speech is chaotic (stochastic), and thus it is possible to better assess speech properties.

One of the most useful feature is the correlation dimension (CD) [20]. This parameter describes geometry of the attractor in the phase space. The more the attractor is complicated, the more independent variables are necessary to describe it, and value of this attribute increases.

The largest Lyapunov exponent (LLE) is an often used feature that describes stability of a dynamic system as well as its sensitivity to initial conditions [19]. The more the system is stable (speech quality increases), the lower the exponent is.

To quantify complexity of the speech signal, the approximate entropy (AE) can also be used to describe a level of randomness [21]. Similar information can be extracted from the signal using the fractal dimension (FD) [22] or the Ziv-Lempel complexity (ZLC) [23]. In the case of the ZCL, it is a measure of regularity level in time series.

Other parameters are non-linear dynamic features based on the detrended fluctuation analysis (DFA) [24], features based on normalised recurrence time probability density entropy H_{norm} [18] and Hurst's exponent (HE) [25], which describes the relationship between the future and previous speech samples.

For the comparison of the nonlinear dynamic features, a simple test in which noise with the normal probability distribution is added to part of vowel [a] is performed. Such noise may represent stochastic part of speech that corresponds to hoarseness, breathiness or hypernasality. Power spectral density of noise was gradually increased, while the percentage of difference between new and original features` values was calculated. The results can be seen in Figure 1.

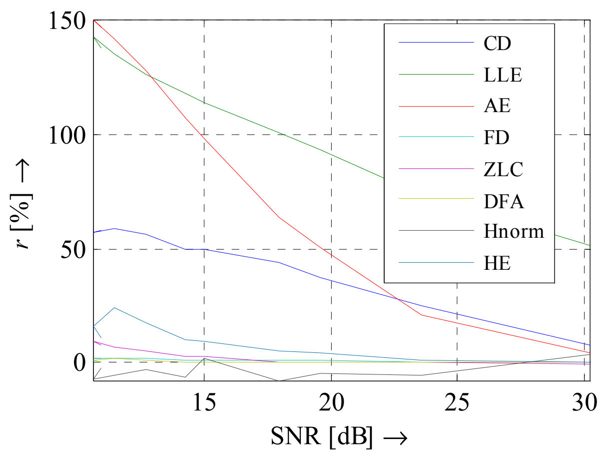


Figure 1: The comparison of sensitivity of the nonlinear dynamic features to additive noise with the normal probability distribution. Variable r expresses the difference between values when comparing to the situation when no noise has been added to the original signal. SNR expresses the signal-to-noise ratio.

The graph shows that the highest sensitivity to additive noise has the approximate entropy (AE) which increases exponentially with an increasing level of noise. In addition to the AE, the LLE and CD features make big differences, but their dependence is linear. The remaining features have no big changes.

2) Features Based on the Empirical Mode Decomposition

Recently, in speech processing new methods based on the EMD are used. Using the EMD it is possible to decompose the arbitrary nonlinear and time-varying signal into countable and usually a small number of the intrinsic mode functions (IMF). These functions are modulated in amplitude and frequency and their sum gives the original signal.

Tsanas et al. [17] proposed several speech parameters based on the IMF. Their idea is to represent the noise component of speech signal with the first few IMFs and stationary speech signal with the remaining IMFs. They designed a new method for SNR (signal-to-signal ratio) and NSR calculations (noise-to-signal ratio):

$$\mathrm{SNR} = \frac{\sum_{i=4}^{K} \mu_i}{\sum_{i=1}^{3} \mu_i}, \tag{1}$$

$$\mathrm{NSR} = \frac{\sum_{i=1}^{2} \tilde{\mu}_i}{\sum_{i=3}^{K} \tilde{\mu}_i}, \tag{2}$$

where μ_i is mean value of sequence, calculated from the original i-th IMF and K is total number of the IMF. $\tilde{\mu}_i$ is logarithmic mean value, which is calculated as mean value but for logarithm of the i-th IMF. In the end, it has to be defined which attributes are used for calculation of the original IMF. It can be the squared energy operator (SEO), Teager-Kaiser energy operator (TKEO), the entropy (H), or the zero crossing rate (ZCR).

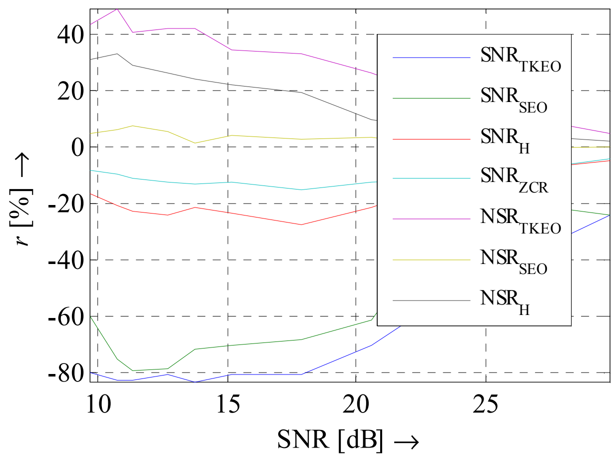


Figure 2: The comparison of the parameters' sensitivity derived from the IMF to additive noise with normal probability distribution. Variable r expresses the difference between values when comparing to the situation when no noise has been added to the original signal. SNR expresses the signal-to-noise ratio and NSR is noise-to-signal ratio.

To compare sensitivity of these features to additive noise, another example is prepared. The results of comparison can be shown in Figure 2. From the graph it can be seen that most of the features in dependence on noise level decrease (SNR) or increase (NSR) linearly. The best result provides SNR based on the squared energy and Teager-Kaiser energy operators. These features show the largest deviations from the original value.

Using Figs. 1 and 2, discriminatory efficiency between the nonlinear dynamical features and the EMD parameters can be compared. If we picked up top 3 candidates for noise part description, it would be the approximate entropy (AE), largest Lyapunov exponent (LLE) and SNR based on Teager-Kaiser energy operator (TKEO).

B. *Analysis of Handwritten Text*

In the case of analysis of handwritten text the feature selection is highly dependent on specific written exercises. In this paper, for example, four most commonly used exercises for analysis of Parkinson's disease are chosen.

1) *Archimedean Spiral*

Archimedean spiral is a curve defined by equation $\theta = k + ar^n$, where θ [rad] is an angle between a point on the spiral and a point on the origin of the spiral and r is distance from the spiral point to the origin point. Variables a, k, and n are real numbers. The parameters of this exercise are: number of changes in velocity (NCV) [26], number of changes in acceleration (NCA) [11], radial oscillation speed (ROS) [27] or parameters evaluating smoothness and density of spiral coils [28]. Archimedean spiral is also a good test to describe level of present tremor. Figure 3 shows the spiral in Cartesian and polar coordinates for healthy people and patients afflicted by Parkinson's disease. As it is illustrated by the figure, due to transformation to the polar coordinates, the 2D spiral curve can be expressed by the 1D curve. Using this spiral, healthy and ill people can be easily distinguished [29, 30]. It is also possible to calculate features based on the signal-to-noise ratio, or the chaos theory.

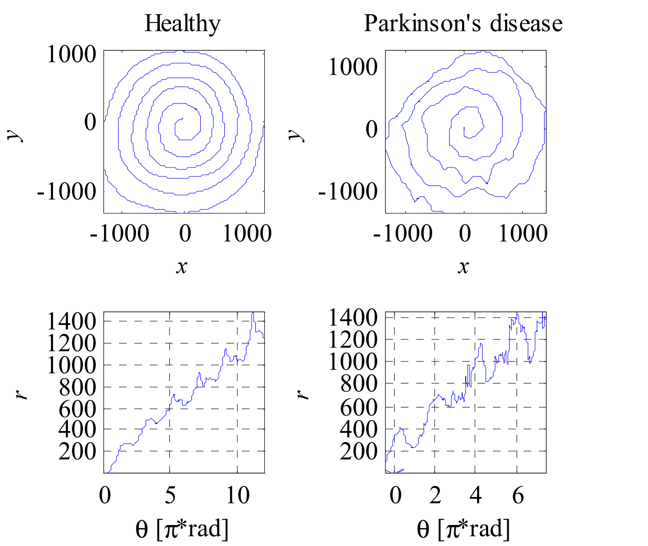


Figure 3: The comparison of Archimedean spirals, which are drawn by healthy person and patient afflicted by Parkinson's disease. In the upper part of the picture spiral in Cartesian coordinates can be seen, and on the bottom the same spiral is expressed in polar coordinates, where r denotes the distance from spiral points to centre of the spiral and θ is angle of straight line passing through this point and the origin point.

2) *Ellipses*

The ellipses are by patients quickly drawn over themselves, as it can be seen in Figure 5. When monitoring the vertical or horizontal movement, then dynamics of the movements can be well described. Using these curves the parameters as velocity (VE), acceleration (AC) or jerk (third derivative of length by time) can be calculated [31]. Other ways of ellipse analysis are discussed in Chap. C.

3) *Connected Letters "l" Written in Italics*

Using this exercise, the writing dynamics can be again seen clearly. The following parameters are usually extracted: stroke size (SS), stroke length (SL), stroke duration (SD), number of local extremes in velocity profile, the velocity (VE), acceleration (AC), and jerk (JE).

4) *Syllabes, Words, Sentences*

The speed of the writing and the presence of micrography can be better monitored on longer part of written text. Using these exercises we extract parameters such as: movement time (MT), writing pressure (WP), velocity (VE), acceleration (AC), jerk (JE), width and height of letters, tendency of decreasing letters over time, etc.

There are tablets which can record movements of the pen in time, when the tip does not touch the paper surface. Thanks to this information it is possible to see how patient's hand behaves at a time when the patient does not write. Figure 4 shows an example of Czech sentence "Tram will no longer go" written by a healthy person and a patient afflicted by Parkinson's disease. Blue line expresses movement of hand when the pen was touching the paper. Red line expresses movement of the pen when the tip does not touch the paper. If we focus only on pen movement on paper (blue line), we would not recognize so big differences in hand tremor between two tested participants. However, information about hand movement in air (red line) clearly indicates that one of participants has strong tremor. Using further analysis, frequency of this tremor, amplitude range, etc. can be estimated.

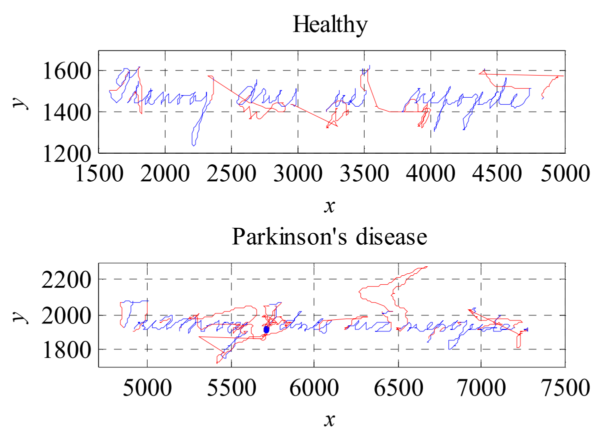


Figure 4: The comparison of Czech sentence "Tram will no longer go" written by a healthy person and a patient affected by Parkinson's disease. Blue line expresses movement of hand when the pen tip was touching the paper. Red line expresses movement of the pen when the tip doesn't touch the paper.

Authors in various publications often use lowpass filtering before processing to smooth the written text [32, 33]. However, such approach unnecessarily removes high-frequency components which are related to hand tremor. The high-frequency components should not be removed, but used for further analysis. It would be also helpful to combine parameters calculated from pressure the pen tip, azimuth and inclination.

C. Comparison of Speech and Handwritten Text Parameters

It is clear from previous chapter, that speech processing is very similar to processing of signals obtained from handwritten text. Fig. 5 demonstrates how the ellipse as a picture is transformed to 1D periodical signal (function of the vertical movement in time). 1D periodical signal obtained from 2D ellipse can be analysed similarly as voiced speech segments. Generally speaking, in analysis of handwritten text the features as jitter, shimmer, short-term energy, modulating energy, harmonic signal to noise ratio, the zero crossing rate, etc. can be applied. Moreover, as in an area of speech processing, even here it is possible to test nonlinear dynamic parameters and parameters based on the EMD analysis. This approach in analysis of handwritten text has not been used and it seems to bring good results.

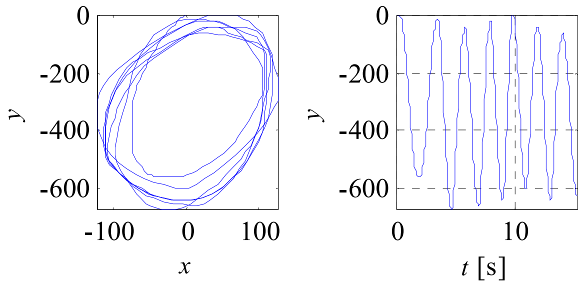


Figure 5: Monitoring the vertical movement in time when drawing the overlapping ellipses.

IV. Applications of Non-invasive Neurological Disorders Analysis

As it has been already mentioned, the most important part of the analysis of neurological disorders is quantification, i.e. feature extraction. When all patients' dysfunctions are appropriately described, it is possible to proceed to the next steps of data processing. Due to the objective description of the disorder, non-invasive analysis can be used as follows:

- Diagnosis/identification of neurological disorder - although it is possible to train a classifier which can perform the diagnosis without the presence of the doctors, this approach is not useful in practice. There is still a risk that the classifier determines inaccurately or bad diagnosis, which could have a very negative impact on the patient's health. The aim of the quantification is not to replace the doctors' work, but to make their work easier.
- To monitor disease progression - using some of the features, it is possible to follow development of the disorder over time,
- Identification of the first stages of the disorders,
- Development of new medication,
- Preparation of an individual plan of medication,
- Analysis of the effect of different stimulations in neurological disorders (e.g. the effect of repetitive transcranial magnetic stimulation rTMS on speech of patients afflicted by Parkinson's disease [34]),
- Monitoring the impact of various devices on the patient's condition (e.g. Duodopa or deep brain stimulation (DBS)).

V. Conclusion

The paper summarizes the modern methods of digital signal processing applied in the field of non-invasive analysis of neurological disorders. Emphasis is given on methods of feature extraction, which is the most important part of the whole procedure of analysis. Thanks to parameterization the disease can be adequately quantified, which can help doctors diagnose it more objectively, faster and more accurately. In increasing number of patients suffering from Alzheimer's or Parkinson's disease, an automated analysis has also positive economic impact (e.g. due to a possibility to diagnose more accurate and on time).

The paper addresses non-invasive methods of speech and handwritten text analyses. The advantage of handwritten text processing against speech processing is small influence of environment to recorded signal (low influence of noise and no influence of acoustic room properties). The main disadvantage of handwritten text processing is that some patients are not able to write anything but can speak. For this reason, it is advantageous to fuse information extracted from both signals.

Furthermore, the paper describes new methods of feature extraction which are not so much known. It is the non-linear dynamic parameters and parameters based on the empirical mode decomposition. Above introduced tests show that the attributes like the approximate entropy, the largest Lyapunov exponent or SNR derived from the squared energy operator of the IMF may well identify the hoarseness in speech, or tremor in writing.

Scientific research in this area aims to invent a system that could diagnose various neurological disorders at an early stage or to estimate their progression. Changes in speech in case of Parkinson's disease occur due to hypokinetic dysarthria in 60-90% of patients [35]. This means that error of disease identification by speech processing may be 10-40%. Moreover, each patient afflicted by Parkinson's disease may have different speech dysfunction. Therefore, first it is important to determine all patient's dysfunction (both speech and handwriting text), and then on the basis of the analysis results, the doctors could perform a reliable diagnosis. Therefore, it seems to be preferable to use artificial intelligence to recognize patient's dysfunction than to automatically determine diagnosis of the disease. Based on the knowledge of dysfunction, the doctors can diagnose the disease itself.

Acknowledgment

This work was supported by project NT13499 (Speech, its impairment and cognitive performance in Parkinson's disease) and project "CEITEC, Central European Institute of Technology": (CZ.1.05/1.1.00/02.0068) from the European Regional Development

Fund. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation.

References

[1] J. Mekyska, Z. Smekal, M. Kostalova, M. Mrackova, S. Skutilova, and I. Rektorova, "Motor aspects of speech imparment in Parkinson‘s disease and their assessment," Cesk Slov Neurol N, vol. 74, no. 6, pp. 662–668, 2011.

[2] A. M. Goberman, "Correlation between Acoustic Speech Characteristics and Non-speech Motor Performance in Parkinson Disease," Med Sci Monit, vol. 11, no. 3, pp. CR109–116, 2005.

[3] K. K. Baker, L. O. Ramig, E. S. Luschei, and M. E. Smith, "Thyroarytenoid muscle activity associated with hypophonia in Parkinson disease and aging," Neurology, vol. 51, no. 6, pp. 1592–1598, 1998.

[4] N. Roy, S. L. Nissen, C. Dromey, and S. Sapir, "Articulatory changes in muscle tension dysphonia: evidence of vowel space expansion following manual circumlaryngeal therapy," Journal of Communication Disorders, vol. 42, no. 2, pp. 124–135, 2009.

[5] K. A. Spencer and M. A. Rogers, "Speech motor programming in hypokinetic and ataxic dysarthria," Brain Lang, vol. 94, no. 3, pp. 347–366, 2005.

[6] S. Skodda, H. Rinsche, and U. Schlegel, "Progression of dysprosody in Parkinson's disease over time - a longitudinal study," Movement Disorders, vol. 24, p. 716–722, 2009.

[7] R. Moretti, P. Torre, R. M. Antonello, L. Capus, M. Gioulis, S. Zambito Marsala, G. Cazzato, and A. Bava, "Speech initiation hesitation following subthalamic nucleus stimulation in a patient with Parkinson's disease," Eur Neurol, vol. 49, no. 4, pp. 251–253, 2003.

[8] M. Gentil, P. Pollak, and J. Perret, "Parkinsonian dysarthria," Rev Neurol, vol. 151, no. 2, pp. 105–112, 1995.

[9] A. Unlu, R. Brause, and K. Krakow, "Handwriting analysis for diagnosis and prognosis of Parkinson's disease," in Proceedings of the 7th international conference on Biological and Medical Data Analysis, ser. ISBMDA'06. Springer-Verlag, 2006, pp. 441–450.

[10] G. Gangadhar, D. Joseph, A. Srinivasan, D. Subramanian, R. Shivakeshavan, N. Shobana, and V. Chakravarthy, "A computational model of Parkinsonian handwriting that highlights the role of the indirect pathway in the basal ganglia," Human Movement Science, vol. 28, no. 5, pp. 602–618, 2009.

[11] O. Tucha, L. Mecklinger, J. Thome, A. Reiter, G. L. Alders, H. Sartor, M. Naumann, and K. W. Lange, "Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease," Journal of Neural Transmission, vol. 113, pp. 609–623, 2006.

[12] J. Westin, S. Ghiamati, M. Memedi, D. Nyholm, A. Johansson, M. Dougherty, and T. Groth, "A new computer method for assessing drawing impairment in Parkinson's disease," Journal of Neuroscience Methods, vol. 190, no. 1, pp. 143–148, 2010.

[13] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, no. 8, pp. 1226–1238, 2005.

[14] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 73, no. 3, pp. 273–282, 2011.

[15] Y. Sun, S. Todorovic, and S. Goodison, "Local-learning-based feature selection for highdimensional data analysis," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 9, pp. 1610–1626, 2010.

[16] K. Kira and L. A. Rendell, "The feature selection problem: traditional methods and a new algorithm," in Proceedings of the tenth national conference on Artificial intelligence, 1992, pp. 129–134.

[17] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," Journal of The Royal Society Interface, vol. 8, no. 59, pp. 842–855, 2011.

[18] M. A. Little, P. E. Mcsharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," BioMedical Engineering OnLine, vol. 6, p. 23, 2007.

[19] G. Vaziri, F. Almasganj, and R. Behroozmand, "Pathological assessment of patients speech signals using nonlinear dynamical analysis," Computers in Biology and Medicine, vol. 40, no. 1, pp. 54–63, 2010.

[20] J. Shao, J. K. Maccallum, Y. Zhang, A. Sprecher, and J. J. Jiang, "Acoustic analysis of the tremulous voice: assessing the utility of the correlation dimension and perturbation parameters," J Commun Disord, vol. 43, pp. 35–44, 2010.

[21] M. Johnson, M. Straume, and M. Lampl, "The use of regularity as estimated by approximate entropy to distinguish saltatory growth," Ann Hum Biol., vol. 28, pp. 491–504, 2001.

[22] R. Esteller, G. Vachtsevanos, J. Echauz, T. Henry, P. Pennell, C. Epstein, R. Bakay, C. Bowen, and B. Litt, "Fractal dimension characterizes seizure onset in epileptic patients," in Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on, vol. 4, mar 1999, pp. 2343–2346.

[23] M. Borowska, E. Oczeretko, A. Mazurek, A. Kitlas, and P. Kuć, "Application of the Lempel-Ziv complexity measure to the analysis of biosignals and medical images," Sleep Rochester, vol. 50, pp. 31–30, 2005.

[24] K. Hu, P. C. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley, "Effect of trends on detrended fluctuation analysis," Physical Review E - Statistical, Nonlinear and Soft Matter Physics, vol. 64, p. 20, 2001.

[25] X. Navarro, A. Beuchee, F. Poree, and G. Carrault, "Performance analysis of Hurst's exponent estimators in higly immature breathing patterns of preterm infants," 2011, pp. 701–704.

[26] T. E. Eichhorn, T. Gasser, N. Mai, C. Marquardt, G. Arnold, J. Schwarz, and W. H. Oertel, "Computational analysis of open loop handwriting movements in Parkinson's disease: a rapid method to detect dopamimetic effects," Movement Disorders, vol. 11, no. 3, pp. 289–297, 1996.

[27] H.Wang, Q. Yu, M. M. Kurtis, A. G. Floyd, W. A. Smith, and S. L. Pullman, "Spiral analysis – improved clinical utility with center detection," Journal of Neuroscience Methods, vol. 171, no. 2, pp. 264–270, 2008.

[28] R. Saunders-Pullman, C. Derby, K. Stanley, A. Floyd, S. Bressman, R. B. Lipton, A. Deligtisch, L. Severt, Q. Yu, M. Kurtis, and et al., "Validity of spiral analysis in early Parkinson's disease," Movement disorders official journal of the Movement Disorder Society, vol. 23, no. 4, pp. 531–537, 2008.

[29] M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure-Alcobe, J. Garre-Olmo, J. Mekyska, K. Lopez de Ipina, and A. Esposito, "A preliminary study of online drawings and dementia diagnose," in Neural Nets and Surroundings, ser. Smart Innovation, Systems and Technologies, B. Apolloni, S. Bassis, A. Esposito, and F. C. Morabito, Eds. Springer Berlin Heidelberg, 2013, vol. 19, pp. 367–374.

[30] M. Faundez-Zanuy, A. Hussain, J. Mekyska, E. Sesa-Nogueras, E. Monte-Moreno, A. Esposito, M. Chetouani, J. Garre-Olmo, A. Abel, Z. Smekal, and K. L. de Ipina, "Biometric applications related to human beings: there is life beyond security," Cognitive Computation, 2012, (in press).

[31] C. Bidet-Ildei, P. Pollak, S. Kandel, V. Fraix, and J.-P. Orliaguet, "Handwriting in patients with Parkinson disease: effect of l-dopa and stimulation of the sub-thalamic nucleus on motor anticipation," Human Movement Science, vol. 30, no. 4, pp. 783–791, 2011.

[32] A. W. V. Gemmert, H.-L. Teulings, and G. E. Stelmach, "Parkinsonian patients reduce their stroke size with increased processing demands," Brain and Cognition, vol. 47, no. 3, pp. 504–512, 2001.

[33] J. H. Yan, S. Rountree, P. Massman, R. S. Doody, and H. Li, "Alzheimer's disease and mild cognitive impairment deteriorate fine movement control," Journal of Psychiatric Research, vol. 42, no. 14, pp. 1203–1212, 2008.

[34] I. Eliasova, J. Mekyska, M. Kostalova, R. Marecek, Z. Smekal, and I. Rektorova, "Acoustic evaluation of short-term effects of repetitive transcranial magnetic stimulation on motor aspects of speech in Parkinson's disease," Journal of Neural Transmission, pp. 1–9, 2012. [Online]. Available: http://dx.doi.org/10.1007/s00702-012-0953-1

[35] L. O. Ramig, C. Fox, and S. Sapir, "Speech treatment for Parkinson's disease," Expert Rev Neurother, vol. 8, no. 2, pp. 297–309, 2008.

A.31 A New Modality for Quantitative Evaluation of Parkinson's Disease: In-Air Movement

A New Modality for Quantitative Evaluation of Parkinson's Disease: In-Air Movement

Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal and Marcos Faundez-Zanuy

***Abstract*— Parkinsons disease (PD) is neurodegenerative disorder with very high prevalence rate occurring mainly among elderly. One of the most typical symptoms of PD is deterioration of handwriting that is usually the first manifestation of Parkinsons disease. In this study, a new modality - in-air trajectory during handwriting - is proposed to efficiently diagnose PD. Experimental results showed that analysis of in-air trajectories is capable of assessing subtle motor abnormalities that are connected with PD. Moreover, conjunction of in-air trajectories with conventional on-surface handwriting allows us to build predictive model with PD classification accuracy over 80%. In total, we compute over 600 handwriting features. Then, we select smaller subset of these features using two feature selection algorithms: Mann-Whitney U-test filter and relief algorithm, and map these feature subsets to binary classification response using support vector machines.**

I. INTRODUCTION

Parkinson's disease (PD) is progressive neurodegenerative disorder characterized by tremor, riginity, bradykinesia and loss of postural reflexes. PD usually affects people with the average age of 60, although 5% to 10% of patients may develop symptoms even before age 40 [1]. The particular causes of PD are not known, but there is ongoing research evaluating genetics, ageing and toxins. From the pathological point of view there is no objective quantitative method for clinical diagnosis. It is thought that PD can only be definitively diagnosed at postmortem that further highlights the complexities of diagnosis. Therefore there is intensive effort to develop expert systems and decision support systems for the assessment and diagnosis of PD.

Previous research has shown that one of the frequent syndromes of PD is significant vocal impairment such as dysphonia (impairment in the vocal production of normal sounds) and dysarthia (problems with normal articulation) [2],[3],[4]. These findings grasped attention of the speech processing community and motivated further research on link between PD and impaired speech. Several new and traditional voice measures has been proposed to discriminate healthy people from people with PD [5]. Recent studies for detection of PD with machine learning tools using acoustic measurement of voice impairment achieved different levels of PD prediction accuracy [6], [7]; where the latest reported results showed as high as 98% overall classification accuracy [8].

Not only speech, but also handwriting is affected by the PD [9],[10],[11],[12],[13]. Parkinson's Disease patients tend to move more slowly than healthy subjects and reduce movement amplitude when they are required to make movement with upper extremities. Slowness of movement and reductions in movement amplitude in clinical observations of PD patients are called bradykinesia and hypometria, respectively. Several studies have documented that handwriting provide numerous features that display statistically significant differences between healthy subjects and subjects with PD [11]. Statistical significance only is not sufficient, as this does not provide a complete picture of the extent to which any one measurement or set of measurements is useful in predicting and diagnosis of PD. Therefore we propose classification model for diagnosis of PD and test it on relatively large dataset consisting of 75 individuals. In addition, minimal subset of the most predictive features is selected.

The fact, that has been rarely taken into account is, that hand movement during handwriting a text consist of two components: an on-surface component, comprising the movements executed while exerting pressure on the writing surface, and an in-air component, comprising the movements performed without touching the writing surface. The amount of information is similar in both types of trajectories and, even if they share some information, in-air and on-surface trajectories appear to be notably non-redundant [14]. In-air movement has been so far used only for biometric application, but here we show that it has meaningful application also for medical analysis.

The rest of the paper is organized as follows. In Section 2., the database of handwriting samples is introduced and described, followed by initial feature analysis. Application of feature selection and machine learning methods to problem of PD classification is described in Section 3. Finally, conclusions are drawn in the last section.

II. DATA AND METHODS

A. *Parkinson's Dataset*

37 Parkinsonian patients (19 men/18 women) and 38 (20 men/18 women) age matched healthy controls took part in this study. Dominant hand of all participants was the right hand. Parkinsonian patients completed the session in the ON state (under medication by L-DOPA). Mean and standard deviation of age, Unified Parkinsons Disease Rating Scale-Part V., score and disease duration are summarized in Table I.

P. Drotár, J. Mekyska and Z. Smékal are with Department of Telecommunications, Brno University of Technology, Technická 12, 61200 Brno, Czech Republic `drotarp@feec.vutbr.cz`

I. Rektorová and L. Masarová are with the First Department of Neurology, Faculty of Medicine, St. Ann's University Hospital , Pekarska 664, 66591 Brno, Czech Republic and Central European Institute of Technology, Masaryk University, Brno, Czech Republic

Marcos Faundez-Zanuy is with Escola Universitaria Politecnica de Mataro, Tecnocampus, Avda. Ernest Llunch 32, 08302, Mataro, Spain

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

TABLE I

PARKINSON'S HANDWRITING DATASET CHARACTERISTICS

	Age		UPDRS (part V)		Years since diag.	
	mean	*std*	*mean*	*std*	*mean*	*std*
PD	69.3	10.9	2.27	0.84	8.37	4.8
H	62.4	11.3	-	-	-	-

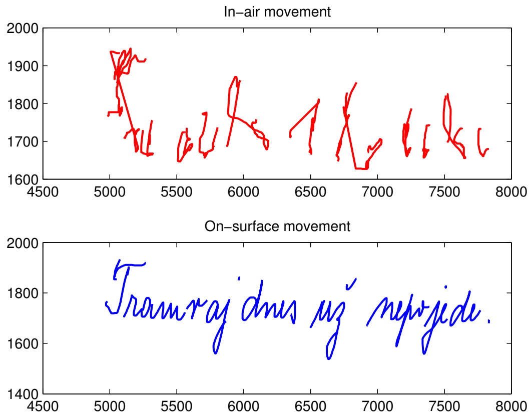


Fig. 1. Handwriting sample of PD patient.

Each subject was asked to write sentence in Czech language (native language of participants) "Tramvaj dnes u nepojede" (The tram won't go today). Handwritten signals were acquired using digitizing tablet Intuos 4M (Wacom technology) in the x-y plane, and in the pressure axis. An inked writing pen was held in a normal fashion without constraints to allow for full visual feedback during writing. As was already mentioned, signals were acquired not only during movements executed while exerting pressure on the writing surface, but also during movement performed without touching the writing surface. Fig. 1 and Fig. 2 show example of on-surface and in-air trajectories taken from executions of the sentence performed by PD patient and healthy control, respectively.

B. Measured feature sets

The recordings starts when the pen touched the surface of digitizer and finishes when task is completed. Digitazing tablet captures following dynamic features (time-sequences): *x*-coordinate, *x(t)*; *y*-coordinate,*y(t)*; time stamp, *s(t)* and button status, *b(t)*. Button status is binary variable being 0 for pen-up(in-air movement) and 1 for pen-down(on-surface movement), this means that tablet captures pen movement while on surface, but also in close proximity of surface - in-air. The x and y components are segmented into on-surface and in-air strokes and analyzed in terms of handwriting measures. The feature calculation stage involves the application of the traditional and nonstandard measurement methods to all handwriting signals. Each method produce either a single value or vector of numbers for each of 75 signals. List of computed features is provided in Tab. II, where single value features are denoted as s and vector features are denoted as v. Additionally 30 statistical functionals of the vector features were computed. These include minima, maxima, range, outlier robust range(percentile *99th* - percentile *1st*), geometrical mean, median, mode, mean, standard deviation, statistical moments (3, 4, 5, 6), trimmed means (5, 10, 20, 30, 40, 50), percentiles(1, 5, 10, 20, 30, 90, 95, 99), quartiles(25/lower, 75/upper), kurtosis.

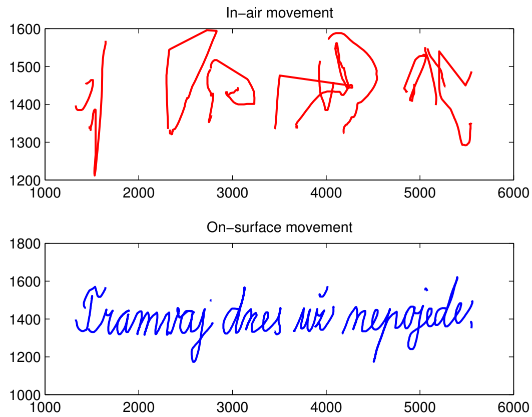


Fig. 2. Handwriting sample of healthy control.

C. Feature analysis

Previous processing stages produce together more than six hundred features for in-air and on-surface movement. In order to obtain some preliminary insight into statistical properties of handwriting features we computed Pearson correlation coefficients and mutual information between feature vectors and associated response. Pearson correlation express measure of linear dependence between features vectors and associated response. Mutual information is a measure of the amount of information shared by two random variable X and Y. It is defined as:

$$I(X;Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\cdot\log_2\left(\frac{p(x,y)}{p(x)p(y)}\right) \quad (1)$$

where x and y are possible variable values with a joint probability distribution function $p(x,y)$ and marginal distribution functions $p(x)$ and $p(y)$, respectively [16].

Table III sumarizes ten handwriting measures with largest relevance to response sorted according absolute correlation coefficient. All correlations are statistically significant ($p < 0.05$). Eight of ten features are in-air movement related features, that give us some initial confirmation of our hypothesis that in-air features contain information relevant for predicting PD. The Mann-Whitney test indicated significant differences ($p < 0.05$) between control group and PD group for all features listed in table.

III. CLASSIFICATION RESULTS

A. Selection of candidate feature set for classification

After removing all features that did not pass the Mann-Whitney U test for significant differences there are still 262

TABLE II

PLEASE WRITE YOUR TABLE CAPTION HERE

Feature	(s)/(v)	Description
stroke speed	v	trajectory during stroke divided by stroke duration
speed	s	trajectory during handwriting divided by handwriting duration
velocity	v	rate at which the position of a pen changes with time
acceleration	v	rate at which the velocity of a pen changes with time
jerk	v	rate at which the acceleration of a pen changes with time
horizontal velocity/acceleration/jerk	v	velocity/acceleration/jerk in horizontal direction
vertical velocity/acceleration/jerk	v	velocity/acceleration/jerk in vertical direction
number of changes in velocity direction (NCV)	s	the mean number of local extrema of velocity [15]
number of changes in acceleration direction (NCA)	s	the mean number of local extrema of acceleration [15]
relative NCV	s	NCV relative to writing duration
relative NCA	s	NCA relative to writing duration
in-air time	s	time spent in-air during writing
on-surface time	s	time spent on-surface during writing
normalised in-air time	s	time spent in-air during writing normalised by whole writing duration
normalised on-surface time	s	time spent on-surface during writing normalised by whole writing duration
in-air/on-surface ration	s	ratio of time spent in-air/on-surface

TABLE III

DESCRIPTION OF CALCULATED FEATURES

Feature	Mutual Information	Correlation Coefficient
stroke speed (on surface, standard dev.)	6.09	-0.388
velocity (in air, standard dev.)	5.94	-0.387
vert. jerk (in air, min.)	5.7	0.383
acceleration (in air, standard dev.)	5.92	-0.38
horz. jerk (in air, range)	5.72	-0.379
jerk (in air, standard dev.)	5.96	-0.389
horz. acceleration (in air, range)	5.81	-0.375
horz. velocity (in air, range)	5.87	-0.371
horz. velocity (on surface, quantile 75%)	4.46	-0.37
vert. acceleration (in air, min.)	5.74	-0.369

candidate features left. Even if many classification algorithms are fairly robust to the inclusion of potentially irrelevant features, their performance in speed (due to high dimensionality) and predictive accuracy (due to irrelevant information) may be severely degraded. Feature selection algorithms aim to choose a small subset of features that ideally is necessary and sufficient to describe target concept. From many feature selection algorithms we decided to use Relief algorithm [17], that has been shown to achieve promising results in problems similar to ours [8]. Relief is feature weighting algorithm that relies entirely on statistical analysis and employs only few heuristics. It selects most of the relevant features even though only a small number of them is necessary for prediction. In most cases it does not help with redundant features. Since we want all relevant features to be included for prediction even at the cost of higher dimensionality Relief appears to be promising candidate.

B. Support Vector Machines

The underlying idea of SVM classifiers is to calculate a maximal margin hyperplane separating two classes of the data. To learn non-linearly separable functions, the data are implicitly mapped to a higher dimensional space by means of a kernel function, where a separating hyperplane is found. New samples are classified according to the side of the hyperplane they belong to. We used RapidMiner Java implementation of the mySVM with radial kernel. The parameters kernel gamma γ, penalty parameter C and convergence epsilon ϵ were optimized using grid search of possible values. Specifically, we searched over the grid (C, γ, ϵ) defined by the product of the sets $C = [10^{-5}, 10^{-4}, \ldots, 10^{3}, 10^{4}]$, $\gamma = [10^{-5}, 10^{-4}, \ldots, , 10^{2}, 10^{3}]$ and $\epsilon = [10^{-5}, 10^{-4}, \ldots, 10^{2}, 10^{3}]$. Classifier validation was conducted using a leave-one-out approach. That is, we left out the sample of one individual to be used for validation as if it is an unseen individual. The process was repeated a total of 50 times, where in each repetition the original dataset was randomly permuted prior to splitting into training and testing subsets. Training and testing features were normalized to have zero mean and a standard deviation of one on a per-feature basis before classification.

C. Numerical Results

Classification performance for different number of features was computed for three different scenarios: using only features based on in-air movement; using only features extracted from on-surface movement and using fusion of both groups of features. By fusion we mean that both feature groups were merged prior to feature selection. Fig.3 shows prediction accuracy of PD using SVM classifier for increasing number of features. Features were selected by application of Relief algorithm. Classification features based on in-air movement provide classification accuracy similar or higher then accuracy of features based on on-surface movement.

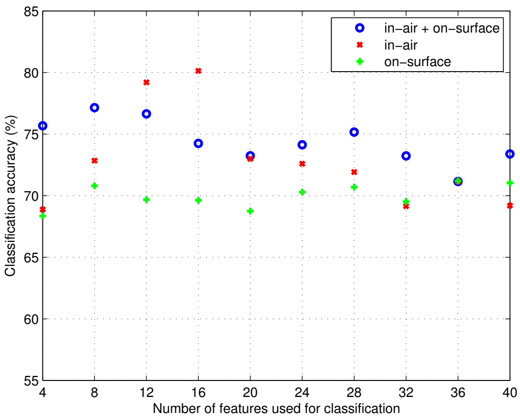


Fig. 3. Classification accuracy of SVM for different modalities.

This confirms our initial hypothesis that in-air movement holds significant information with regards to diagnosis of PD. The highest classification accuracy, 80.09%, was achieved for 16 features selected from in-air. Merging of both modalities brings in most of the cases improvement in classification accuracy indicating amount of non-redundant information in in-air and on-surface movement. As can be seen from Fig. 3 increasing number of features is not always beneficial.

IV. Conclusion

It was shown that proposed scheme can be used for diagnosis of PD with classification accuracy over 80%. Besides conventional on-surface handwriting also in-air trajectories during writing were utilized for PD prediction task. Results indicate that novel in-air features outperform conventional on-surface features in separating healthy controls from subjects with PD. Conjunction of both modalities to built predictive model can be used for quantitative recording for the treating doctor in order to detect and predict long term changes in the individual disease history. Beside the PD classification and disease tracking the handwriting analysis can be also used during an evaluation of modern non-invasive treatment methods such as high-frequency repetitive transcranial magnetic stimulation (rTMS), see e.g. [16]. In our future work, we will analyse new features that can more efficiently capture tremor, micrographia and other medically relevant information. We believe that merging handwriting features with e.g. voice features can further improve diagnosis, evaluation and tracking of PD.

Acknowledgment

This work was supported by the project NT13499 (Speech, its impairment and cognitive performance in Parkinsons disease), CZ.1.07/2.3.00/20.0094, project CEITEC, Central European Institute of Technology: (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund and by FEDER and Ministerio de Economa y Competitividad TEC2012-38630-C04-03. The described research was performed in laboratories supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation. Peter Drotár was supported by the project CZ.1.07/2.3.00/30.0005 of Brno University of Technology.

References

[1] L. M. de Lau and M. M. Breteler, "Epidemiology of Parkinson's disease," *The Lancet Neurology*, vol. 5, no. 6, pp. 525 – 535, 2006.

[2] A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of patients with Parkinson's disease," *Behav. Neurol.*, vol. 11, no. 3, pp. 131 – 137, 1998.

[3] K. M. Rosen, R. D. Kent, A. L. Delaney, and J. R. Duffy, "Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers," *J Speech Lang Hear Res*, vol. 49, no. 2, pp. 395–411, 2006.

[4] J. Mekyska, Z. Smekal, M. Kostalova, M. Mrackova, S. Skutilova, and I. Rektorova, "Motor aspect of speech impairment in parkinsons disease and their assesment," *Czech and Slovak Neurology and Neurosurgery*, vol. 74, no. 6, pp. 662–668, 2011.

[5] M. Little, P. McSharry, E. Hunter, J. Spielman, and L. Ramig, "Suitability of dysphonia measurements for telemonitoring of parkinson's disease," *Biomedical Engineering, IEEE Transactions on*, vol. 56, no. 4, pp. 1015–1022, 2009.

[6] P.-F. Guo, P. Bhattacharya, and N. Kharma, "Advances in detecting parkinsons disease," in *Medical Biometrics*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, vol. 6165, pp. 306–314.

[7] C. Sakar and O. Kursun, "Telediagnosis of parkinsons disease using measurements of dysphonia," *Journal of Medical Systems*, vol. 34, no. 4, pp. 591–599, 2010.

[8] A. Tsanas, M. Little, P. McSharry, J. Spielman, and L. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 5, pp. 1264–1271, 2012.

[9] C. Marsden, "Slowness of movement in parkinson's disease," *Movement Disorders*, vol. 4, no. 1, pp. 26–37, 1989.

[10] R. Saunders-Pullman, C. Derby, K. Stanley, A. Floyd, S. Bressman, R. B. Lipton, A. Deligtisch, L. Severt, Q. Yu, M. Kurtis, and S. L. Pullman, "Validity of spiral analysis in early parkinsons disease," *Movement Disorders*, vol. 23, no. 4, pp. 531–537, 2008.

[11] M. Broderick, A. Van Gemmert, H. Shill, and G. Stelmach, "Hypometria and bradykinesia during drawing movements in individuals with parkinsons disease," *Experimental Brain Research*, vol. 197, no. 3, pp. 223–233, 2009.

[12] M. Wang, B. Wang, J. Zou, L. Chen, F. Shima, and M. Nakamura, "A new quantitative evaluation method of parkinson's disease based on free spiral drawing," in *Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on*, vol. 2, 2010, pp. 694–698.

[13] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Prediction potential of different handwriting tasks for diagnosis of Parkinsons disease," in *IEEE International Conference on e-Health and Bioengineering (EHB 2013)*, Iasi, Romania, November 2013.

[14] E. Sesa-Nogueras, M. Faundez-Zanuy, and J. Mekyska, "An information analysis of in-air and on-surface trajectories in online handwriting," *Cognitive Computation*, vol. 4, no. 2, pp. 195–205, 2012.

[15] T. Eichhorn, T. Gasser, N. Mai, C. Marquardt, G. Arnold, J. Schwarz, and W. Oertel, "Computational analysis of open loop handwriting movements in parkinsons disease: a rapid method to detect dopamimetic effects," *Movement Disorders*, vol. 11, no. 3, pp. 289–297, may 1996.

[16] T. M. Cover, *The elements of Information Theory*. Wiley, 1991.

[17] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proceedings of the ninth international workshop on Machine learning*, ser. ML92. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1992, pp. 249–256.

A.32 A Preliminary Study of Online Drawings and Dementia Diagnose

A Preliminary Study of Online Drawings and Dementia Diagnose

Marcos Faundez-Zanuy[1], Enric Sesa-Nogueras[1],
Josep Roure-Alcobe[1], Josep Garre-Olmo[2], Jiri Mekyska[3],
Karmele Lopez-de-Ipiña[4], and Anna Esposito[5]

[1] Escola Universitària Politècnica de Mataró, Tecnocampus
Barcelona, Spain
{faundez,sesa,roure}@tecnocampus.com
[2] Institut d'assistència sanitària, Salt
Girona, Spain
josep.garre@ias.scs.es
[3] SPLab, Department of Telecommunications, FEEC,
Brno University of Technology Brno, Czech Republic
j.mekyska@phd.feec.vutbr.cz
[4] EHU (Basque Country University)
Donostia, Spain
karmele.ipina@ehu.es
[5] IIASS (International institute for advanced scientific studies)
Vietri sul Mare, Italy
iiass.annaesp@tin.it

Abstract. In this paper we present preliminary results about on-line drawings acquired by means of a digitizing tablet, and performed by control population (left and right hand) as well as pathological subjects using their dominant hand. Experimental results reveal a clear difference between both groups, specially on the on-air movements. Although the acquired samples are not enough to extract significant conclusions we think that this preliminary results encourage the experimentation in this research line. Thus, the main purpose of this paper is to attract the attention of the scientific community.

Keywords: On-line handwriting, Drawings, Dementia.s

1 Introduction

Information and communication technologies are converging to health applications and a great improvement of health diagnose and recover will be possible if the signal processing community collaborates with medical doctors. The authors strongly believe that for the biometric security community this would be quite straightforward, and this paper points out in this direction.

Biometrics has been successfully applied to security applications for some time. However, the extension of other potential applications with the use of

B. Apolloni et al. (Eds.): *Neural Nets and Surroundings*, SIST 19, pp. 367–374.
DOI: 10.1007/978-3-642-35467-0_36

biometric information is a very recent development. This paper summarizes the field of biometrics and investigates the potential of utilizing biometrics beyond the presently limited field of security applications.

The term "biometric" originates from the Greek words Bio (life) and metron (measure), and is defined as the science and technology of measuring and statistically analyzing biological data. Although many people consider biometrics only relevant to security applications, in reality, the relevance of biometrics is much more far reaching. This field has applications relevant to animals, plants and human beings. Some examples are:

- Statistical methods for the analysis of data from agricultural field experiments to compare the yields of different varieties of wheat.
- The analysis of data from human clinical trials evaluating the relative effectiveness of competing disease therapies.
- The analysis of biometric characteristics for animal/human verification or identification.

While some signals can be acquired from both human beings and animals (such as images of iris and retina), others are specific to humans (such as speech, handwriting, etc.).

This paper is focused exclusively on applications which are relevant only to human beings, and more precisely on on-line handwritten drawings. Therefore, we will limit discussion to only human specific signals. The set of these signals can be split into two categories:

1. Behavioral biometrics: this category is based on the measurements and data derived from an action performed by a user, and thus indirectly measures some characteristics of the human body. Signature, gait, gesture and key stroking recognition belong to this category.
2. Physiological biometrics: this category is based on direct measurements of parts of the human body. Fingerprint, face, iris and hand-scanning recognition belong to this category.

The skill level of humans is strongly related to their health state. An important example is the way our cognitive functions are related to the aging process. Cognitive decline is a natural part of the aging process. However, the extent of decline varies across subjects and across functions. For instance, handwriting and speech production is a fine motor control performed by our brain. When these signals are degraded, it is indicatory of health problems.

2 On-Line Handwriting

In the past, the analysis of handwriting had to be performed in an offline manner. Only the writing itself (strokes on a paper) were available for analysis. Nowadays, modern capturing devices, such as digitizing tablets and pens (with or without ink) can gather data without losing its temporal dimension. When spatiotemporal information is available, its analysis is referred as online. Modern digitizing tablets not only gather the x-y coordinates that describe the movement of the writing device as it changes its position, but it can also collect other data, mainly the pressure exerted by the writing device on the writing surface and also the azimuth, the angle of the pen in the horizontal plane, and the altitude, the angle of the pen with respect the vertical axis (Fig. 1).

A very interesting aspect of the modern online analysis of handwriting is that it can take into account information gathered when the writing device was not exerting pressure on the writing surface. Thus, the movements performed by the hand while writing a text can be split into two classes:

a) *On-surface trajectories* (pen-downs), corresponding to the movements executed while the writing device is touching the writing surface. Each of these trajectories produces a visible stroke.
b) *In-air trajectories* (pen-ups), corresponding to the movements performed by the hand while transitioning from one stroke to the next. During these movements the writing device exerts no pressure on the surface.

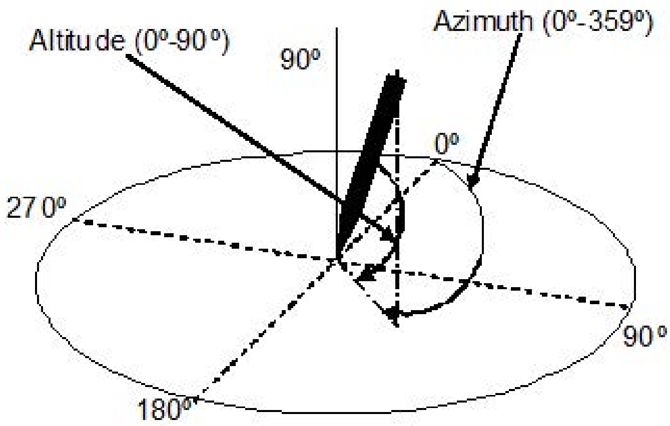


Fig. 1. Azimuth and altitude angles of the pen with respect to the plane of the writing surface

Fig. 2 shows the acquisition of the ten digits from 1 to 0 using an Intuos Wacom digitizing tablet. The tablet acquired 100 samples per second including the spatial coordinates (x, y), the pressure, and a couple of angles (see Fig. 1). The pen-up information is represented in Fig 1 using "+" while the pen- down is marked with "*". Fig. 3 shows the temporal evolution of the signals acquired while handwriting the digits in Fig. 2.

Our experiments on the biometric recognition of people reveal that these two kinds of information are complementary [8] and in fact, contain a similar discriminative capability, even when using a database of 370 users [7].

3 On-Line Drawings Applied to Health Analysis

In the medical field, the study of handwriting has proven to be an aid to diagnose and track some diseases of the nervous system. For instance, handwriting skill

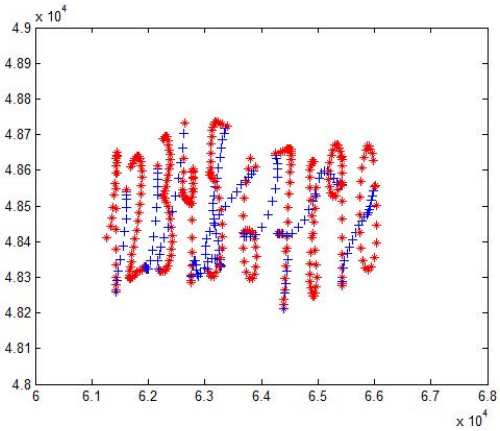


Fig. 2. Example of handwritten numerical digits input onto a digitizing tablet. Asterisks (*) represent pen-down information and cross (+) the pen-up.

degradation and Alzheimer's disease (AD) appear to be significantly correlated [3] and some handwriting aspects can be good indicators for its diagnosis [4] or help differentiate between mild Alzheimer's disease and mild cognitive impairment [10]. Also, the analysis of handwriting has proven useful to assess the effects of substances such alcohol [1] [5], marijuana [2] or caffeine [9]. Aided by modern acquisition devices, the field of psychology has also benefitted from the analysis of handwriting. For instance in [6], Rosenblum et al. link the proficiency of the writers to the length of the in-air trajectories of their handwritings.

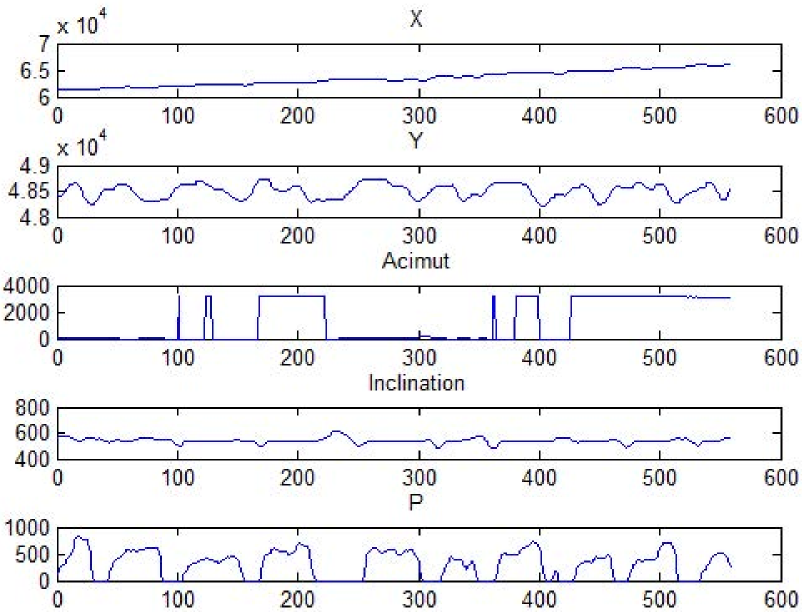


Fig. 3. Temporal evolution of the acquired parameters when drawing the numbers shown in Fig. 1

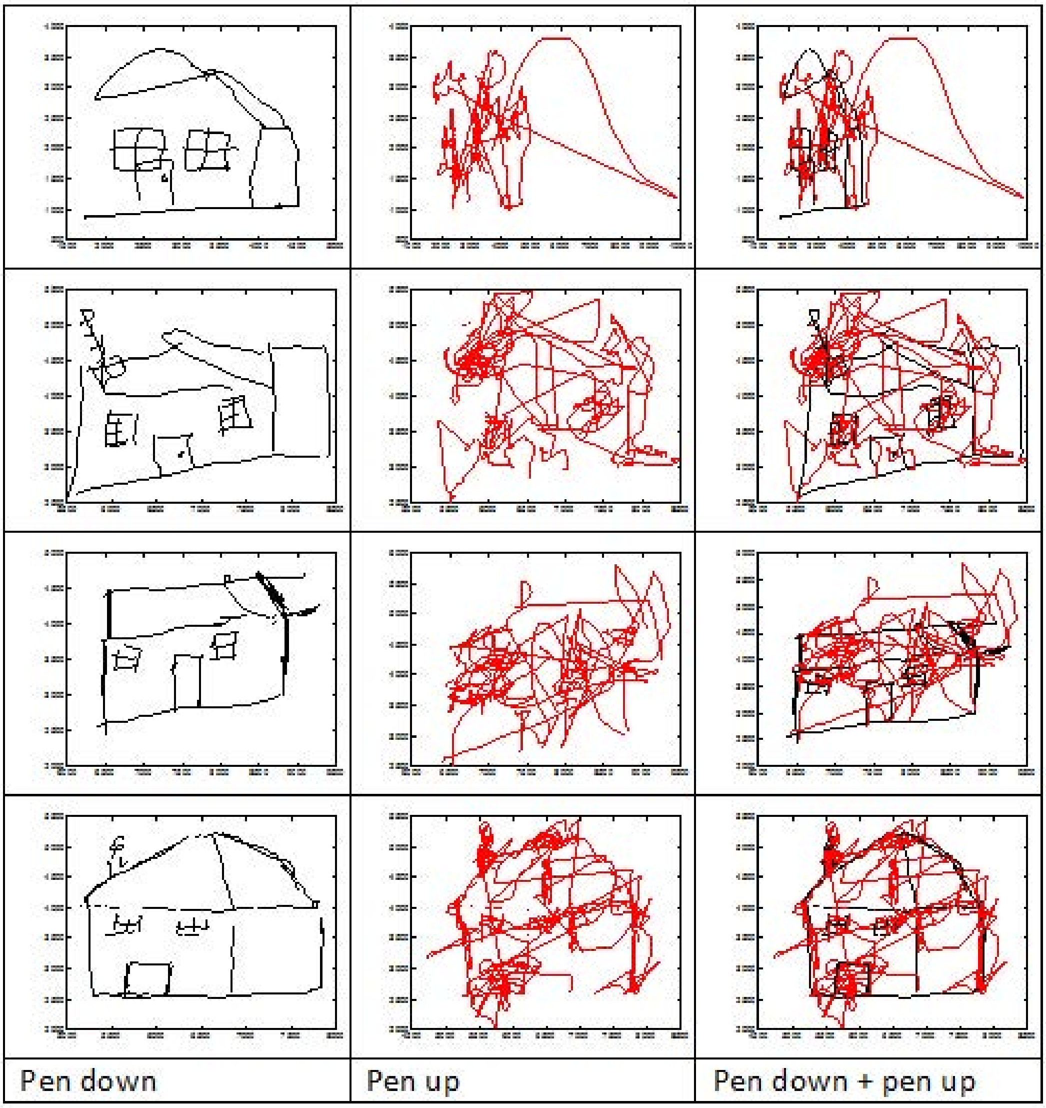


Fig. 4. House drawing performed by four individuals with Alzheimer's disease (one per row). Each column corresponds to pen-down, pen-up and both simultaneously.

Table 1. Statistical analysis/descriptives from the drawings shown in Fig. 4, 5 and 6

	Control		Pathological
Measurement	Dominant hand	Non-dominant hand	Dominant hand
Time in-air	8334	10927	61008
Time on-surface	9680	22177	31521
Total time	18014	33104	92259

In the Fig. 4 we present one complex drawing with three dimensions performed by individuals with AD of different clinical severity. The visual inspection of the pen down image suggest a progressive degree of impairment, where drawing becomes more disorganized and the three dimensions effect is only achieved in the mild case. The visual information provided by the pen up drawing between AD

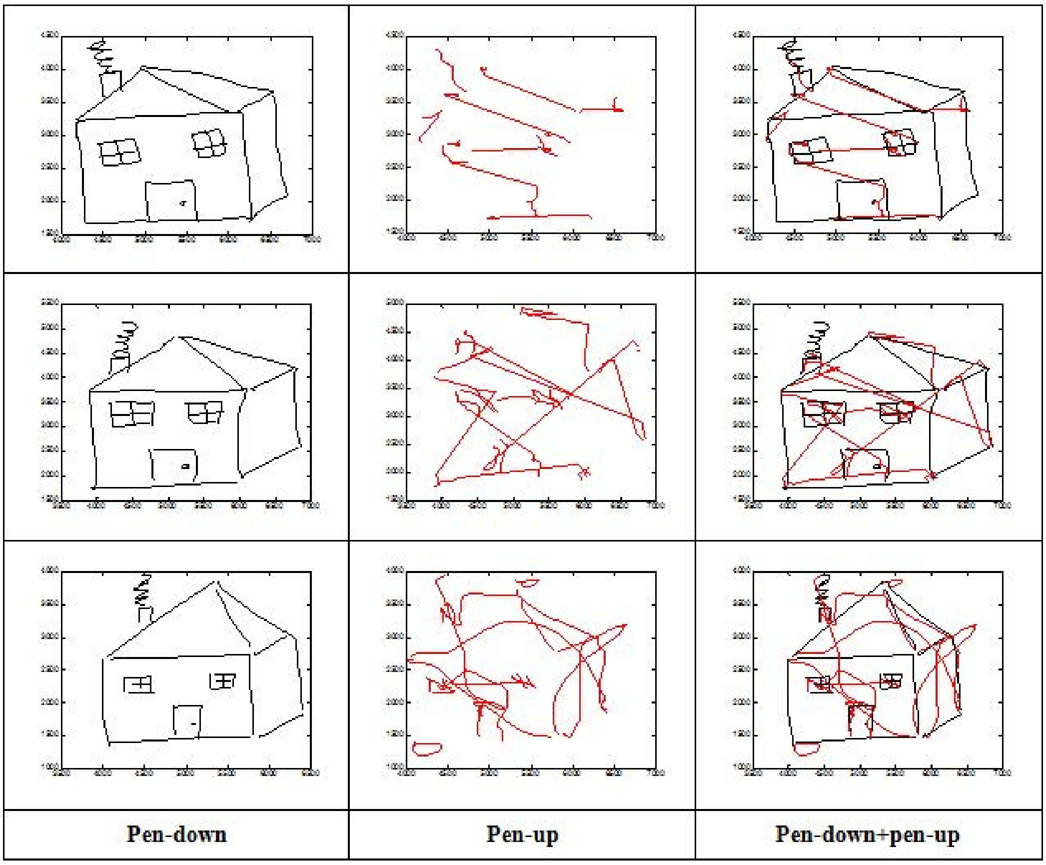


Fig. 5. House drawing performed by three control people (one per row). Each column corresponds to pen-down, pen-up and both simultaneously, performed with the dominant hand.

individuals also indicates a progressive impairment and disorganization when the individuals try to plan the drawing. It is also important to note that the comparison of the pen-up drawing between the mild case of AD and the control (Fig. 5 and 6) also shows important differences. Besides the increased time on air, there is an increased number of hand movements before decide to put the pen in the surface to drawn. We consider that these graphomotor measures applied to the analysis of drawing and writing functions may be a useful alternative to study the precise nature and progression of the drawing and writing disorders associated with several neurodegenerative diseases. Table 1 summarizes some experimental measures of the drawings shown in Fig. 4, 5 and 6.

Looking at the experimental results of Table 1 it is evident the higher time for in-air movements for the AD group, which are around 7 times longer. On the contrary, the time on surface is just around 3 times longer. Thus, there are more differences between control and AD groups when looking at in-air movements.

When comparing the non-dominant movements performed by the control group we obtained a 5.6 ratio and 1.5. Again, the in-air times are significantly higher for the AD group than the control group.

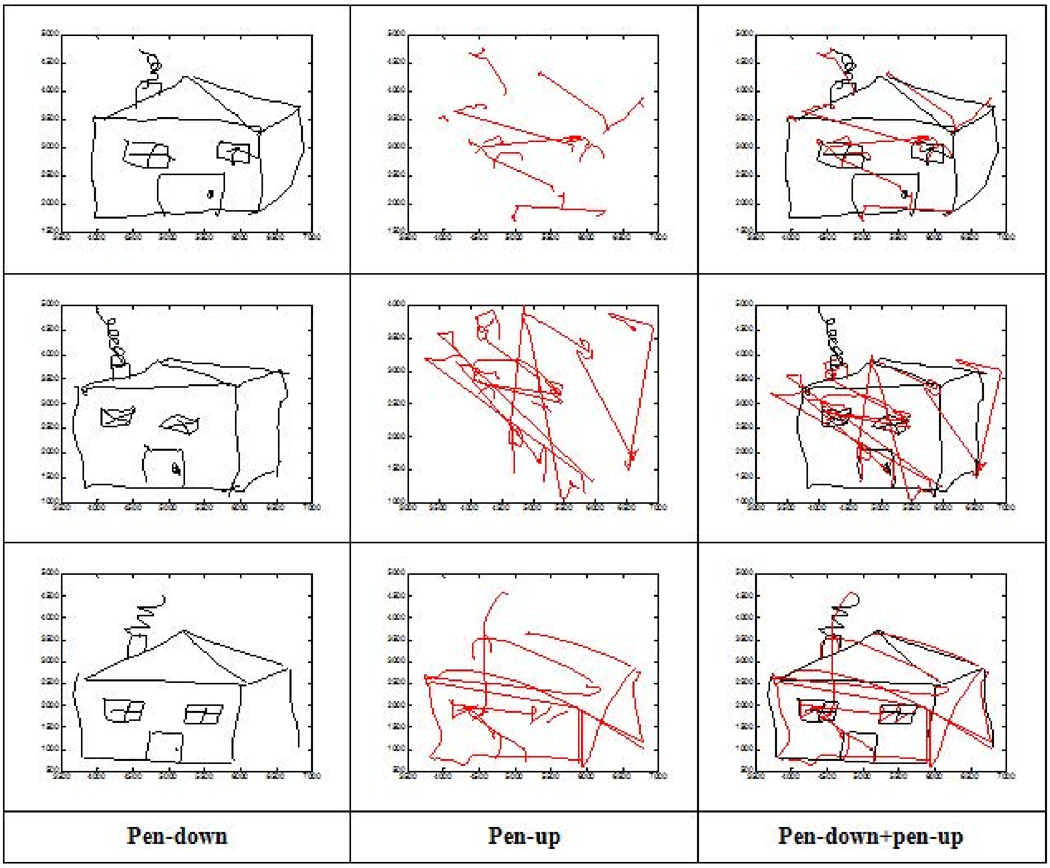


Fig. 6. House drawing performed by three control people (one per row). Each column corresponds to pen-down, pen-up and both simultaneously, performed with the non-dominant hand.

4 Conclusions

Although some pathological drawings may look “normal” according to pen-down information, the pen-up information looks quite entangled and should permit easier diagnose. This observation points out the convenience of online handwriting analysis, which can outperform the classic offline mode, mainly due to the larger amount of available information.

The differences between control and pathological group do not seem to be related to some physical problem, because the control group, even when using the non-dominant hand performs less entangled pen-up movements.

Future work will include a more exhaustive experimental section, with a larger database.

Acknowledgments. This research has been supported by Saiotek, FEDER and MEC, TEC2009-14123-C04-04, KONTAKT-ME 10123, VG20102014033, CZ.1.-07/2.3.00/20.0094, GACR 102/12/1104 and CZ.1.05/2.1.00/03.0072.

References

1. Asicioglu, F., Turan, N.: Handwriting Changes Under the Effect of Alcohol. Forensic Science International 132(3), 201–210 (2003)
2. Foley, R.G., Miller, A.L.: The effects of marijuana and alcohol usage on handwriting. Forensic Science International 14, 159–164 (1979)
3. Forbes, K.E., Shanks, M.F., Venneri, A.: The evolution of dysgraphia in Alzheimers disease. Brain Research Bulletin 63(1), 19–24 (2004)
4. Neils-Strunjas, J., Groves-Wright, K., Mashima, P., Harnish, S.: Dysgraphia in Alzheimer's Disease: A Review for Clinical and Research Purposes. J. Speech Lang. Hear. Res. 49(6), 1313–1330 (2006)
5. Phillips, J.G., Ogeil, R.P., Mller, F.: Alcohol Consumption and Handwriting: A Kinematic Analysis. Human Movement Science 28(5), 619–632 (2009)
6. Rosenblum, S., Parush, S., Weiss, P.: The In Air Phenomenon: Temporal and Spatial Correlates of the Handwriting Process. Percept. Mot. Skills 96, 933–954 (2003)
7. Sesa-Nogueras, E., Faundez-Zanuy, M.: Biometric Recognition Using Online Uppercase Handwritten Text. Pattern Recogn. 45(1), 128–144 (2012)
8. Sesa-Nogueras, E., Faundez-Zanuy, M., Mekyska, J.: An Information Analysis of In-Air and On-Surface Trajectories in Online Handwriting. Cognitive Computation, 1–11 (2012)
9. Tucha, O., Walitza, S., Mecklinger, L., Stasik, D., Sontag, T.A., Lange, K.W.: The Effect of Caffeine on Handwriting Movements in Skilled Writers. Human Movement Science 25(4-5), 523–535 (2006)
10. Werner, P., Rosenblum, S., Bar-On, G., Heinik, J., Korczyn, A.: Handwriting Process Variables Discriminating Mild Alzheimer's Disease and Mild Cognitive Impairment. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 61(4), P228–P236 (2006)

A.33 Prediction potential of different handwriting tasks for diagnosis of Parkinson's

The 4th IEEE International Conference on E-Health and Bioengineering - EHB 2013
Grigore T. Popa University of Medicine and Pharmacy, Iaşi, Romania, November 21-23, 2013

Prediction Potential of Different Handwriting Tasks for Diagnosis of Parkinson's

Peter Drotár, Jiří Mekyska, Zdeněk Smékal
Dept. of Telecommunications
Brno University of Technology
Brno, Czech Republic

Irena Rektorová, Lucia Masarová
First Department of Neurology
Faculty of Medicine, St. Ann's University Hospital
Brno, Czech Republic

Marcos Faundez-Zanuy
Escola Universitaria Politecnica de Mataro
Tecnocampus
Mataro, Spain

***Abstract*—One of the most frequent clinical hallmarks of Parkinson's disease (PD) is micrographia. Micrographia in PD is characterized by the decreased letter size and by changes in the kinematic aspects including increased movement time, decreased velocities and accelerations, and increased number of changes in velocity and acceleration. Based on the literature survey we proposed template to acquire handwriting during different tasks. In addition to well established tasks for PD diagnosis such as Archimedean spiral, we designed new tasks to acquire all aspects of micrographia. The database consists of eight different handwriting samples from seventy-five subjects. The presented results shows almost 80% overall classification accuracy**

Keywords*—handwriting, Parkinson's disease, decision support systems, micrographia, SVM.

INTRODUCTION

Parkinson's disease (PD) is a complex neurodegenerative disease affecting large portion of population worldwide [1]. The PD influences a part of the brain known as the substantia nigra, which controls movement in the body. Unlike healthy people, patients with Parkinsons disease exhibit disruption in the execution of the practiced skills such as handwriting [2], [3], [4], [5] and speech [6], [7]. They have severe difficulties in coordinating components of the motor sequence movement. They tend to perform sequential movements in a more segmented fashion. When the handwriting or another motor task have to be produced continuously they are slower as if they are executed separately. Hesitations and pauses are often observed between the components of the sequence [8]

Several handwriting tasks were proposed to analyse handwriting of PD patients and to obtain insight into motor disruption aspects of PD. Spiral drawing has been used for the assessment of the impact of therapy on motor performance in various movement disorders including PD [3], [9], [10]. Words containing one or multiple repetitions of cursive letter l are also frequently used to evaluate handwriting samples [8], [11].

Based on literature survey, we have developed the complex template to acquire handwriting samples from PD subjects. In addition to already mentioned tasks, our template contain also new handwriting tasks: simple words and one sentence. These words and sentence were selected because of easy syntax- every word can be written with one stroke- i.e. continuous contact between pen and the surface can be preserved during writing. The orthography is also rather simple to minimize cognitive effort during the writing task.

Several studies have documented that handwriting provide numerous features that display statistically significant differences between healthy subjects and subjects with PD [11], [4], [12], [13]. Statistical significance only is not sufficient, as this does not provide a complete picture of the extent to which any one measurement or set of measurements is useful in predicting and diagnosis of PD. Therefore we propose classification model for diagnosis of PD and test it on relatively large dataset consisting of 75 individuals. Additionally, our aim is not only to propose classification model for diagnosis of PD, but also to compare handwriting tasks that are frequently used in literature to analyse handwriting of PD subjects.

The rest of the paper is organized as follows. In Section II, the database of handwriting samples is introduced and described, followed by methods in III section. PD classification and obtained results are given in Section IV. Finally, conclusions are drawn in the last section.

PARKINSON'S DISEASE DETECTION HANDWRITING DATASET

The Parkinsons handwriting dataset consist of multiple handwriting samples from 37 parkinsonian patients (19 men/18 women) and 38 gender and age matched controls (20 men/18 women). All PD participants were recruited from the Movement Disorders Center at the First Department of Neurology, Masaryk University and St. Annes Hospital in Brno, Czech Republic. Mean age was 69.3 ± 10.9 for parkinsonian patient and 62.4 ± 11.3 for control subjects, respectively. All subjects used their dominant right hand. For parkinsonian patients the mean value of Unified Parkinson's Disease Rating Scale-Part V. was 2.27 ± 0.84 and all patients completed the tasks under medication L-DOPA.

978-1-4799-2373-1/13/$31.00 ©2013 IEEE

TABLE I.

OVERVIEW OF HANDWRITING FEATURES

Feature	(s)/(v)	Description
stroke speed	v	trajectory during stroke divided by stroke duration
speed	s	trajectory during handwriting divided by handwriting duration
velocity	v	rate at which the position of a pen changes with time
acceleration	v	rate at which the velocity of a pen changes with time
jerk	v	rate at which the acceleration of a pen changes with time
horizontal velocity/acceleration/jerk	v	velocity/acceleration/jerk in horizontal direction
vertical velocity/acceleration/jerk	v	velocity/acceleration/jerk in vertical direction
number of changes in velocity direction (NCV)	s	the mean number of local extrema of velocity [14]
number of changes in acceleration direction (NCA)	s	the mean number of local extrema of acceleration [14]
relative NCV	s	NCV relative to writing duration
relative NCA	s	NCA relative to writing duration
writing duration/length	s	time duration (in seconds)/ length (in points) of writing
stroke height/width	v	width and height of stroke

Each subject was asked to complete handwriting task according to the prepared template. The filled task sheet is depicted in Fig.1. Template consist of eight different tasks. First task, Archimedean spiral, is established task used for analysis of PD or essential tremor [10], [9]. In the tasks 2-4 participants wrote cursive letters or bi/tri-grams of letters.

The tasks of similar type(letter l - or its repetitions) are also commonly used for handwriting analysis [8]. Next three tasks are words, that can be written as one long stroke i.e. during writing of these words is writing device in continuous contact with surface. Words are written in Czech language (native language of participants) with following translation to English: lektorka - lector(female), porovnat - to compare, nepopadnout - do not catch. Finally, the last task is longer sentence, that allows to capture also effect of fatigue during writing (Tramvaj dnes už nepojede - The tram won't go today).

Handwritten signals were acquired using digitizing tablet Intuos 4M (Wacom technology) in the x-y plane, and in the pressure axis. An inked writing pen was held in a normal fashion without constraints to allow for full visual feedback during writing. Digitized signals were acquired during movements executed while exerting pressure on the writing surface. The recordings start when the pen touched the surface of digitizer and finishes when task is completed. Digitizing tablet captures following dynamic features (time-sequences): x-coordinate, x(t); y-coordinate, y(t); time stamp, s(t) and button status, b(t). Button status is binary variable being 0 for pen-up state (in-air movement) and 1 for pen-down state (on-surface movement).

I. METHODS

Our goal is not only to propose classification model for diagnosis of PD, but also to compare handwriting tasks that are frequently used to analyse handwriting of PD subjects. Therefore, we have analysed every task individually to see predictive potential of particular task. Then, all task were merged to make use of task diversity and build predictive model for diagnosis of PD.

Feature Calculation

The same procedure is applied to all handwriting tasks. The x and y components are segmented into strokes and analyzed in terms of handwriting features. The feature calculation stage involves the application of the traditional and nonstandard measurement methods to all handwriting signals. Each method produces either a single value or vector for each of seventy-five signals. List of computed features is provided in Tab. I, where single valued features are denoted as s and vector features are denoted as v. Additionally, six "basic"

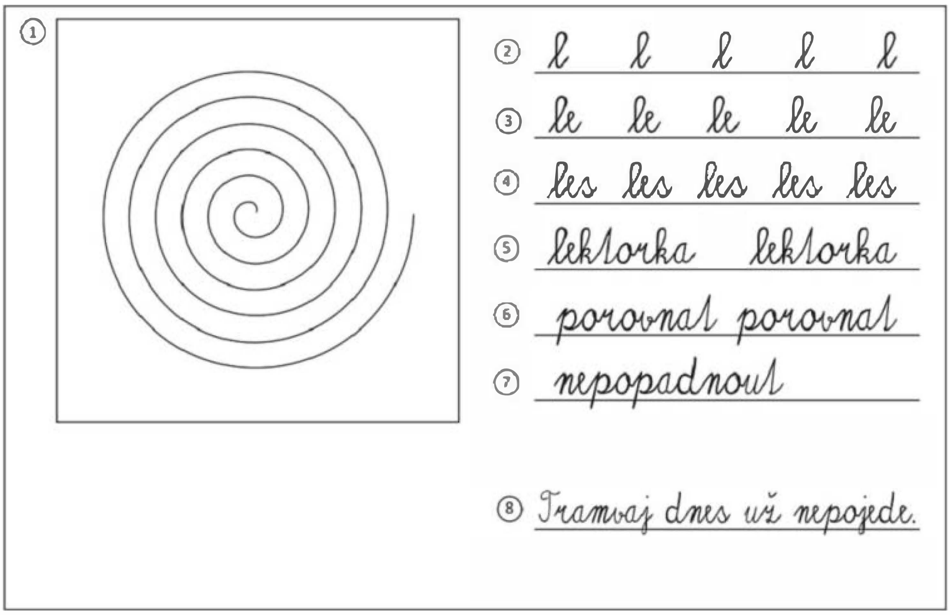


Figure 1. Illustration of filled template (not actual handwriting sample)

functionals (mean, median, standard deviation, 1st percentile, 99th percentile, 99th percentile 1st percentile) of the vector features were computed. Here, the 99th percentile - 1st percentile represents outlier robust range, and 1st and 99th percentiles are outlier robust minimum and maximum, respectively.

Support Vector Machines

The underlying idea of SVM classifiers is to calculate a maximal margin hyperplane separating two classes of the data. To learn non-linearly separable functions, the data are implicitly mapped to a higher dimensional space by means of a kernel function, where a separating hyperplane is found.

New samples are classified according to the side of the hyperplane they belong to. We used Radial Basis Function (RBF) kernel [15]. The RBF kernel is defined as

$$K(x, x_i) = e^{\frac{-\|x-x_i\|^2}{2\gamma^2}} \qquad (2)$$

where γ controls the width of RBF function.

Statistical Analysis

In order to obtain some preliminary insight into statistical properties of handwriting features we analysed relationship between feature vectors and associated response (i.e. label indicating medical diagnosis PD/H). Table II sumarizes six handwriting measures with largest relevance to response sorted according SVM classification accuracy using single feature. Table shows feature name, as defined in Table I, number of task, that was used to obtain feature and statistical functional computed from feature vector. Additionally, we computed Pearson correlation coefficients between feature vectors and associated response. Depicted features are strongly correlated to response (highest $\rho =$ 0.4, $p < 0.05$). SVM classification accuracy (70.6 % for feature stroke speed) is also indicating strong association between handwriting and subject diagnosis. These findings suggest that the extraction of handwriting features accentuate relationship between handwriting and PD.

Probability densities of the top four handwriting features from Table II are shown in Fig. 2. The vertical axes are the probability densities of the normalized measures, estimated using kernel density estimation with Gaussian kernels. Differences between probability densities of PD and healthy control are visible also in this case.

II. Results

The parameters kernel gamma and penalty parameter C were optimized using grid search of possible values. Specifically, we searched over the grid (C, γ) defined by the product of the sets $C = [2^{-6}, 2^{-5}, \cdots, 2^{7}, 2^{8}]$, $\gamma = [2^{-5}, 2^{-4}, \cdots, 2^{8}, 2^{9}]$. Classifier validation was conducted using stratified tenfold cross-validation. The process was repeated a total of ten times, where in each repetition the original dataset was randomly permuted prior to splitting into training and testing subsets. Classification accuracy, sensitivity and specificity over the ten repetitions were averaged. Training and testing features were normalized before classification on a per-feature basis to have zero mean and a standard deviation of one.

TABLE II.
SVM PREDICTION ACCURACY AND CORRELATION WITH PD/H DIAGNOSIS OF MOST REDICTOVE FEATURES

Feature	SVM prediction accuracy [%]	Correlation Coefficient
stroke speed (task 8, standard dev.)	70.6	-0.39
stroke width (task 3, percentile 1st)	68.3	-0.40
horz. velocity (task 8, percentile 99th)	68.2	0.33
stroke width (task 3, mean)	66.7	-0.24
stroke length (task 3, percentile 1st)	65.3	-0.24
stroke length (task 3, standard dev.)	64	0.32

The classification test performance was determined by the computation of accuracy, sensitivity and specificity. The accuracy (P_{acc}), sensitivity (P_{sen}) and specificity (P_{spe}) are defined as

$$P_{acc} = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100\% \qquad (2)$$

$$P_{spe} = \frac{TN}{TN + FP} \cdot 100\% \qquad (3)$$

$$P_{sen} = \frac{TP}{TP + FN} \cdot 100\% \qquad (4)$$

where TP (true positive) and FP (false positive) represents the number of correctly decided PD subject and number of subject diagnosed as PD, but being healthy. Similarly, TN (true negative) and FN (false negative) represent the total number of correctly decided healthy control, and PD patients evaluated as healthy control.

Mann-Whitney U test showed statistically significant differences between patients with PD and healthy controls in terms handwriting features for data obtained from tasks 2, 3, 5, 6, 7 and 8. For task 1 and 4, i.e. spiral drawing and repetitive writing of les world, were found no differences between those two groups.

The discrimination potential of every task was evaluated individually with results depicted in Table III. Additionally, we merged all tasks displaying statistical differences between PD and H (i.e. 2, 3, 5, 6, 7 and 8) to compute overall classification accuracy, sensitivity and specificity.

As an input to SVM classifier we employed only features that passed the Mann-Whitney U test for significant differences. In the case of 1st and 4th task, where no statistically significant features exist, we used all computed features.

The best classification performance of 79.4 % was reached in a combination of all tasks including all relevant handwriting exercises. The maximal classification accuracy using simple task was 78.7 % for 8th task. Comparing this number to SVM classification accuracy yielded by evaluation of other tasks, it is clearly visible, that the most of the SVM prediction power comes from features acquired during 8th task

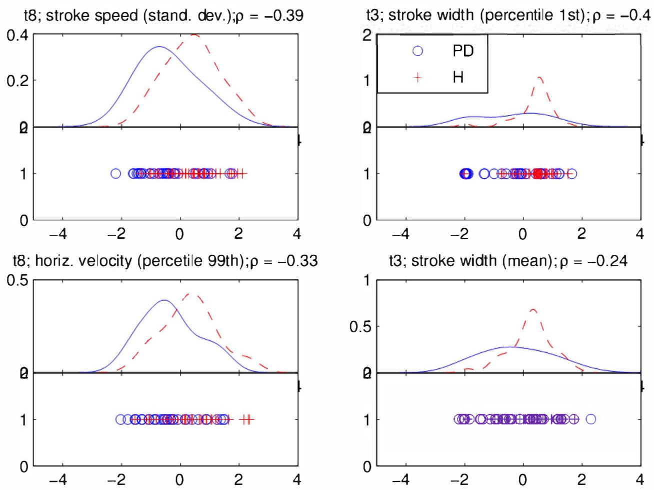


Figure 2 Distribution of feature values (bottom half of figures) and estimated probability density functions (top half of figures) of four features, for the top 4 features from Table II.

TABLE III.

CLASSIFICATION RESULTS USING SVM CLASSIFIER

Evaluated task	P_{acc} [%]	P_{spe} [%]	P_{sen}[%]
task 1	65.4	32.4	95.5
task 2	70.0	59.2	80.1
task 3	72.3	68.7	75.3
task 4	65.4	61.9	68.9
task 5	66.7	68.9	64.5
task 6	67.7	69.7	65.8
task 7	67.1	48.7	85
task 8	78.7	75.1	82.1
overall	79.4	78.9	80

III. CONCLUSION

We proposed methodology for building the predictive model of PD from kinematic handwriting features obtained from different conventional and novel handwriting tasks. The accuracy obtained using our method is 79.4 % with very similar values for specificity and sensitivity. Introduced results show that analysis of handwriting has promising potential for computer based decision support tools for the next-generation health-care. With the recent advent of technology and ubiquitous electronic devices, such a tablet or different touch-screen devices, acquisition and evaluation of handwritten signal is becoming really simple, once proper methodologies are established.

IV. ACKNOWLEDGEMENTS

This work was supported by the project NT13499 (Speech, its impairment and cognitive performance in Parkinsons disease), CZ.1.07/2.3.00/20.0094, project CEITEC, Central Euro-pean Institute of Technology: (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund and by FEDER and Ministerio de Economa y Competitividad TEC2012-38630-C04-03. The described research was performed in laborato-ries supported by the SIX project; the registration number CZ.1.05/2.1.00/03.0072, the operational program Research and Development for Innovation. Peter Drotár was supported by the project CZ.1.07/2.3.00/30.0005 of Brno University of Technology.

REFERENCES

[1] L. M. de Lau and M. M. Breteler, “Epidemiology of Parkinson’s disease,” The Lancet Neurology, vol. 5, no. 6, pp. 525 – 535, 2006.

[2] C. Marsden, “Slowness of movement in parkinson’s disease,” Movement Disorders, vol. 4, no. 1, pp. 26–37, 1989.

[3] R. Saunders-Pullman, C. Derby, K. Stanley, A. Floyd, S. Bressman, R. B. Lipton, A. Deligtisch, L. Severt, Q. Yu, M. Kurtis, and S. L. Pullman, “Validity of spiral analysis in early parkinsons disease,” Movement Disorders, vol. 23, no. 4, pp. 531–537, 2008..

[4] M. Broderick, A. Van Gemmert, H. Shill, and G. Stelmach, “Hypometria and bradykinesia during drawing movements in individuals with parkinsons disease,” Experimental Brain Research, vol. 197, no. 3, pp. 223–233, 2009 [5] G.B. Moody, *et. al.*, „Predicting the onset of paroxysmal atrial fibrillation: the Computers in Cardiology Challenge 2001”, *Computers in Cardiology*, **28**, pp. 113-116, 2001.

[5] O. Geman and H. Costin, “Parkinsons disease prediction based on multistate markov models,” Int. J. Computers. Commun. Control, vol. 8, no. 4, pp. 525–537, 2013.

[6] J. Mekyska, Z. Smekal, M. Kostalova, M. Mrackova, S. Skutilova, and I. Rektorova, “Motor aspect of speech impairment in parkinsons Disease and their assesment,” Czech and Slovak Neurology and Neurosurgery, vol. 74, no. 6, pp. 662–668, 2011.

[7] A. Tsanas, M. Little, P. McSharry, J. Spielman, and L. Ramig, “Novel speech signal processing algorithms for high-accuracy classification of Parkinson’s disease,” Biomedical Engineering, IEEE Transactions on, vol. 59, no. 5, pp. 1264–1271, 2012.

[8] C. Bidet-Ildei, P. Pollak, S. Kandel, V. Fraix, and J.-P. Orliaguet, “Handwriting in patients with parkinson disease: Effect of l-dopa and stimulation of the sub-thalamic nucleus on motor anticipation,” Human Movement Science, vol. 30, no. 4, pp. 783 – 791, 2011.

[9] P. H. Kraus and A. Hoffmann, “Spiralometry: Computerized assessment of tremor amplitude on the basis of spiral drawing,” Movement Disorders, vol. 25, no. 13, pp. 2164–2170, 2010. [Online]. Available: http://dx.doi.org/10.1002/mds.23193

[10] K. stanley, J. Hagenah, N. Bruggemann, K. Reetz, L. Severt, C. Klein, Q. Yu, C. Derby, S. Pullman, and R. Sanders-Pullman, “Digitized spiral analysis is a promising early motor marker for parkinson disease,” Parkinsonism and Related Disorders, vol. 16, no. 3, pp. 233–234, 2010.A.C.C. Yang, H.W. Yin, “Prediction of Paroxysmal Atrial Fibrillation by Footprint Analysis”, *Computers in Cardiology*, **28**, pp. 401-404, 2001.

[11] H.-L. Teulings, J. L. Contreras-Vidal, G. E. Stelmach, and C. H. Adler, “Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control,” Experimental Neurology, vol. 146, no. 1, pp. 159 – 170, 1997.

[12] P. Drotár, J. Mekyska, I. Rektorov á, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, “A new modality for quantitative evaluation of parkinsons disease:in-air movement,” in 13th IEEE International Conference on BioInformatics and BioEngineering (BIBE), Chania, Greece, November 2013

[13] H. Costin and C. Rotariu, “Tremor assessment by means of handscript analysis and fuzzy logic,” Biomedical Softcomputing & Human Sciences Journal, vol. 7, no. 1, pp. 39 – 45, 2001.

[14] T. Eichhorn, T. Gasser, N. Mai, C. Marquardt, G. Arnold, J. Schwarz, and W. Oertel, “Computational analysis of open loop handwriting movements in parkinsons disease: a rapid method to detect dopamimetic effects,” Movement Disorders, vol. 11, no. 3, pp. 289–297, may 1996.

[15] V. Vapnik, Statistical Learning Theory, 1st ed. London: John Willey & Sons, 1998.

A.34 An Information Analysis of In-Air and On-Surface Trajectories in Online Handwriting

Cogn Comput (2012) 4:195–205
DOI 10.1007/s12559-011-9119-y

An Information Analysis of In-Air and On-Surface Trajectories in Online Handwriting

Enric Sesa-Nogueras · Marcos Faundez-Zanuy · Jiří Mekyska

Received: 2 September 2011 / Accepted: 18 November 2011 / Published online: 6 December 2011

Abstract This paper is aimed at analysing, from an information theory perspective, the gestures produced by human beings when handwriting a text. Modern capturing devices allow the gathering of data not only from the on-surface movements of the hand, but also from the in-air trajectories performed when the hand moves in the air from one stroke to the next. Our past research with isolated uppercase words clearly suggests that both types of trajectories have a biometric potential to perform writer recognition and that they can be effectively combined to enhance the recognition accuracy. With samples from the BiosecurID database, we have analysed the entropy of each kind of trajectories, as well as the amount of information they share, and the difference between intra- and inter-writer measures of the mutual information. The results show that when pressure is not taken into account, the amount of information is similar in both types of trajectories. Furthermore, even if they share some information, in-air and on-surface trajectories appear to be notably non-redundant.

Keywords Handwriting · Biometrics · Information theory

E. Sesa-Nogueras (✉) · M. Faundez-Zanuy
EUP Mataró, Tecnocampus Av. Ernest Lluch 32, 08302 Mataró, Spain
e-mail: sesa@eupmt.cat

M. Faundez-Zanuy
e-mail: faundez@eupmt.es

J. Mekyska
Faculty of Electrical Engineering and Communication, Department of Telecommunications, Brno University of Technology, Brno, Czech Republic
e-mail: xmekys01@stud.feec.vutbr.cz

Introduction

The hand movements carried out when handwriting a text present two components: an on-surface component, comprising the movements executed while exerting pressure on the writing surface, and an in-air component, comprising the movements performed without touching the writing surface. The on-surface movements have been extensively studied in different fields of research, whereas the in-air trajectories have seldom been taken into account. Do in-air movements contain useful information? Is this information redundant with respect to information in the on-surface trajectories? This paper aims at answering these questions with an approach based on information theory.

From a general perspective, the complex movements performed by the hand while handwriting a text can be regarded as *gestures*. These gestures have been learnt at school and thus are mainly conventional. Different types of factors influence the production of these gestures: the muscular movements involved in handwriting are controlled by the central nervous system and are partly outside the conscious control of the writer. The long- and short-term conditions within the central nervous system have an effect on the handwriting. Biomechanical factors such as the structure and size of the hand, the arm and the shoulder or the state of the muscles (stiffness, elasticity) also influence the produced handwriting. Finally, the cultural background determines the kind of characters written (Asian, Western, Arabic…). The resulting graphemes on a paper allow certain measures of these gestures that have proven useful in different areas. In the forensic field, questioned document examination (QDE) aims at answering questions about documents in dispute in a court of law [1]. These questions are mainly related to authorship. In the medical field, the study of handwriting has proven to be an aid to diagnose and track some diseases of the nervous system. For instance, handwriting skill degradation and Alzheimer's disease appear to be significantly correlated [2] and some handwriting aspects can be good indicators for its diagnosis [3] or help differentiate between mild Alzheimer's disease and mild cognitive impairment [4]. Also, the analysis of handwriting has proven useful to assess the effects of substances such alcohol [5, 6], marijuana [7] or caffeine [8]. Aided by modern acquisition devices, the field of psychology has also benefitted from the analysis of handwriting. For instance in [9], Rosenblum et al. link the proficiency of the writers to the length of the in-air trajectories of their handwritings. In a more controversial field, graphology aims at drawing conclusions about different psychological traits of the writer based, on traits of their handwriting [10, 11].

In recent years, one of the fields that has most benefited from handwriting analysis is that of biometric security and, more precisely, biometric recognition [12]. As handwriting analysis is based on measurements and data derived from an action performed by the writer, handwriting-based biometric recognition is a type of *behavioural biometrics*. Signature-based recognition is the most well-know approach to biometric writer recognition because signature, having a long history as a method to prove one's identity (legal documents, bank transactions), has been the method of choice for an important number of handwriting-based recognition schemes. The interested reader can find extensive surveys of the research results in this specific field in [13–16].

In the past, the analysis of handwriting had to be performed in an *offline* manner. Only the writing itself (strokes on a paper) was available for the analysis. Nowadays, modern capturing devices, such as digitizing tablets and pens or online whiteboards, can gather data without losing its temporal dimension. When spatiotemporal information is available, its analysis is referred as *online*. A typical modern digitizing tablet (Fig. 1) not only gathers the *x*–*y*-coordinates that describe the movement of the writing device as it changes its position, but it can also collect other data, mainly the pressure exerted by the writing device on the writing surface and also the azimuth, the angle of the pen in the horizontal plane, and the altitude, the angle of the pen with respect the vertical axis (Fig. 2). From now on, *x*–*y*-coordinates, pressure, azimuth and altitude will be referred as *features of the handwriting*.

A very interesting aspect of the modern online analysis of handwriting is that it can take into account the information gathered when the writing device was not exerting pressure on the writing surface. Thus, the movements performed by the hand while writing a text can be split into two classes:

(a) *On-surface trajectories* (pen-downs), corresponding to the movements executed while the writing device is touching the writing surface. Each of these trajectories produces a visible stroke.

(b) *In-air trajectories* (pen-ups), corresponding to the movements performed by the hand while transitioning from one stroke to the next. During these movements, the writing device exerts no pressure on the surface.

Fig. 1 Intuous WACOM A5 digitizing tablet and pen

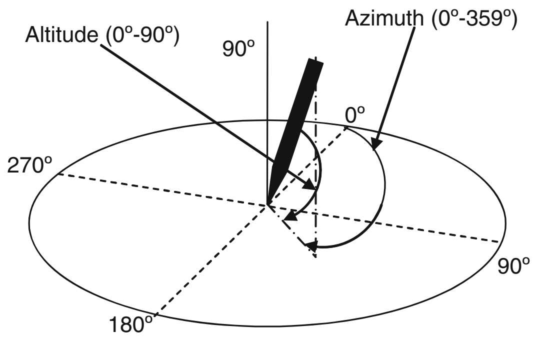


Fig. 2 Azimuth and altitude angles of the pen with respect to the plane of the writing surface

Figure 3 shows two examples of on-surface and in-air trajectories taken from two executions of the word BIODEGRADABLE performed by two different writers.

All offline databases (those acquired by a conventional scanner once the writing process has finished or by any other equivalent process) lack the information regarding in-air trajectories.

In the field of automated applications based on handwriting analysis, on-surface trajectories have received considerable attention mainly due to optical character recognition applications and signature-based writer recognition systems (involving both identification and verification). On the other hand, in-air trajectories have received almost no attention at all, even in online approaches where spatiotemporal information is available.

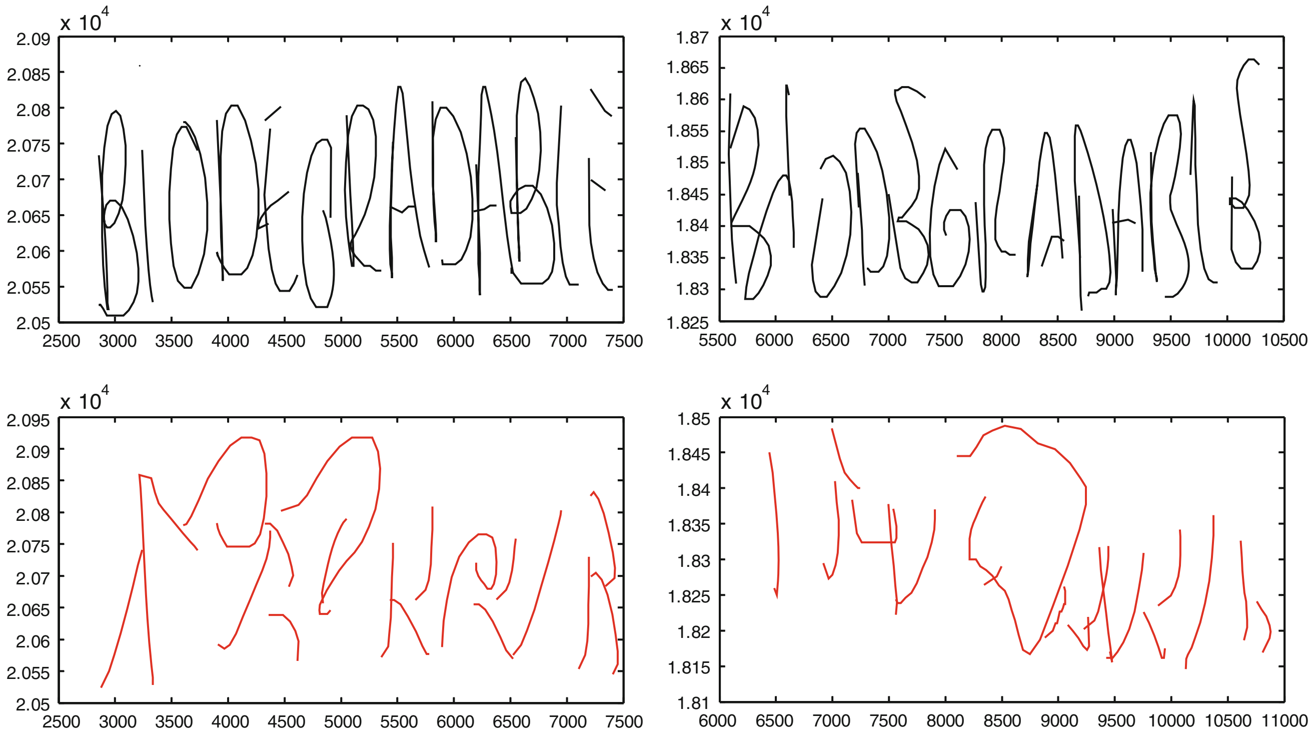


Fig. 3 On-surface (*top*) and in-air (*bottom*) trajectories from two executions of the word BIODEGRADABLE performed by two different writers

In the past, we have successfully used the information in the in-air trajectories to perform writer recognition. In [17], we presented a biometric writer recognition system based on isolated uppercase words. One of the particularities of this system is that it considers on-surface (pen-down strokes) and in-air (pen-up strokes) trajectories separately. Although the measures obtained from both types of trajectories are combined in a final step in order to obtain an improved word-level measure, the aforesaid separation allows comparing the *discriminative power* of each type of the movement. In the context of this paper, discriminative power refers to the ability to distinguish, in a set of writers, a particular writer from the rest, given a sample of their handwriting. Thus defined, discriminative power arises from *writer individuality*. Writer individuality refers to the hypothesis that each individual has consistent handwriting that is distinct from the handwriting of other individuals. In the past, several authors have scrutinized this hypothesis in the particular case of short sequences of text (isolated characters, isolated words and sentences comprising a small number of words) and reached the conclusion that handwriting is an individual trait, both in the offline case [18, 19] and in the online one [20, 21]. Results in [22] not only add further evidence to support the individuality of handwriting even when only isolated uppercase words are considered, but extend the reach of the hypothesis to the invisible part of the handwriting.

Table 1 contains the identification rates and verification error rates obtained with the writer recognition system presented in [17] when using a dataset from the BiosecurID database that comprises 16 different words from 100 different users (See "Database" for further details about the BiosecurID database). Identification rate (IDR) refers to the percentage of correctly identified users, whereas verification error rate (VER) refers to the average rate of false rejections (users deemed as impostors when they are not) and false acceptances (users deemed as authentic when they are not). Figure 4 graphically depicts the values obtained with each type of trajectory and with their combination.

Quite surprisingly, in-air trajectories show a discriminative power that often surpasses that of pen-down strokes produced in on-surface trajectories except for words 2, 9, 11, 12 and 13, and pen-up strokes produced during in-air trajectories perform better in verification (lower error) than pen-down strokes. When it comes to identification, pen-up strokes perform better in identification (higher identification rate) in words 4, 5, 7, 8, 9, 10 and 14.

Not only data gathered from in-air trajectories perform well in identification and verification, but when combined with on-surface data, the overall performance is enhanced. In all cases, the identification rate for the combination outperforms the best identification rate obtained with one type of strokes only. The same can be said of the

Table 1 Recognition (identification and verification) results obtained with the 16 words of the BiosecurID database

Word	Text	Length	In-air trajectories		On-surface trajectories		Combination	
			IDR (%)	VER (%)	IDR (%)	VER (%)	IDR (%)	VER (%)
W1	BIODEGRADABLE	12	86	**3.51**	88	4.17	99	2.81
W2	DELEZNABLE	10	74	7.95	81	4.40	93	3.28
W3	DESAPROVECHAMIENTO	18	92	**1.68**	93	3.82	98	1.64
W4	DESBRIZNAR	10	**78**	**5.43**	75	5.94	90	4.54
W5	DESLUMBRAMIENTO	15	**93**	**2.69**	85	5.19	97	2.99
W6	DESPEDAZAMIENTO	15	92	**4.01**	92	4.90	97	3.67
W7	DESPRENDER	10	**78**	**5.42**	74	8.30	87	5.20
W8	ENGUALDRAPAR	12	**79**	**5.88**	70	6.69	89	4.51
W9	EXPRESIVIDAD	12	**88**	6.03	86	4.65	94	2.91
W10	IMPENETRABLE	12	**81**	**4.38**	75	6.94	93	3.43
W11	INEXPUGNABLE	12	82	4.51	86	4.10	94	2.74
W12	INFATIGABLE	11	87	4.34	90	3.87	94	1.71
W13	INGOBERNABLE	12	80	5.03	87	4.44	95	3.27
W14	MANSEDUMBRE	11	**73**	**6.41**	71	7.18	85	4.67
W15	ZAFARRANCHO	11	73	**6.32**	79	6.48	86	5.05
W16	ZARRAPASTROSA	13	78	**5.77**	81	6.07	93	4.57

Bold value indicates in-air trajectories outperform on-surface trajectories

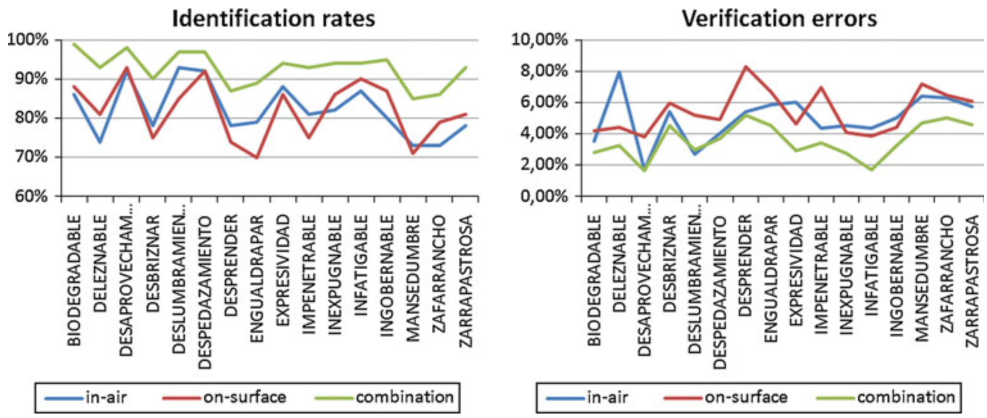


Fig. 4 Identification rates and verification errors for the 16 words in the BiosecurID database

verification error rates, except for word 5 where the figure obtained for the combination is slightly worse than that obtained from in-air data only.

The results summarized in Table 1 suggest that:

(a) Data from in-air trajectories may not contain significantly less information than data from the on-surface trajectories. Otherwise, they would not show an equal, sometimes better, performance in the recognition tasks.
(b) A significant amount of this information may be non-redundant, thus explaining why the combination outperforms both types of trajectories.

The goal of this paper is to study both kinds of trajectories from the perspective of the information theory [23]. Specifically, the following points will be analysed:

(a) The amount of information contained in each feature of each trajectory: x-coordinate, y-coordinate, azimuth and elevation in in-air trajectories and the same features, plus pressure, in on-surface trajectories.
(b) The redundancy between in-air and on-surface trajectories. This redundancy will be analysed in a per-feature basis.
(c) The inter-writer and intra-writer variability for each kind of trajectory and feature. This variability assesses the biometric potential of the information in the trajectories for the recognition task.

The approach taken is similar to that of [24] and [25]. In [24], the complementarities between thermal, visible and near infrared images used for face recognition were studied. In [25], the potential of each spectral band for biometric data fusion and cross-sensor operation (a model is trained in one spectral band while testing is performed with samples from a different band) was assessed.

Background on Information Theory

In order to facilitate the understanding of the following sections, this section provides a very brief introduction to information theory. The most relevant aspects, connected to the results shown in this paper, are highlighted. The interested reader can find in the literature much more in-depth treatments of the topic (e.g. [26, 27])

If X is a random variable with several possible values x and a marginal probability distribution function $p(x)$, the *entropy* of X, measured in bits, is defined as

$$H(X) = -\sum_{x \in X} p(x) \cdot \log_2(p(x))$$

$H(X)$ is a measure of the uncertainty associated with X. If X is a source of data or a message, then $H(X)$ measures the average information content in X. Other equivalent interpretations are also possible. For instance, $H(X)$ is the average number of bits (binary symbols) required to encode all the possible outcomes (values) of X.

For two random variables, X and Y, with possible values x and y, a joint probability distribution function $p(x,y)$ and marginal distribution functions $p(x)$ and $p(y)$, respectively, the following measures are often considered:

(a) *Conditional entropy*, often called equivocation in information theory, quantifies the remaining entropy (i.e. the uncertainty) of one of the variables when the value of the other one is know. It is defined as:

$$H(Y|X) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \cdot \log_2(p(y|x))$$

(b) *Joint entropy*, a measure of the amount of information in the joint system of X and Y. Its definition is:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \cdot \log_2(p(x, y))$$

(c) *Mutual information*, a measure of the amount of information shared by X and Y. It is defined as:

$$I(X; Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \cdot \log_2\left(\frac{p(x, y)}{p(x) \cdot p(y)}\right)$$

Intuitively, a low value for $I(X; Y)$ means that X and Y provide different, non-redundant, information. Notice that $I(X; Y) = 0$ if and only if X and Y are independent (the knowledge of one has no effect whatsoever on the knowledge of the other).

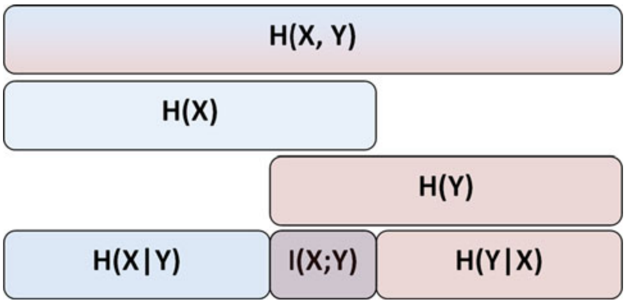


Fig. 5 Relations among the individual entropies ($H(X)$, $H(Y)$), the conditional entropies ($H(X|Y)$, $H(Y|X)$), the joint entropy ($H(X,Y)$) and the mutual information ($I(X;Y)$)

These three measures are tightly related to each other:

$$I(X; Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y)$$

Figure 5 graphically depicts the relations among conditional entropy, joint entropy and mutual information, as shown in [27]

In order to facilitate the comparison of amounts of mutual information obtained from different pairs of random variables, $I(X;Y)$ can be expressed relative to $H(X,Y)$:

$$I'(X; Y) = I(X; Y)/H(X, Y)$$

Thus, *relative mutual information* $I'(X; Y)$ is the proportion of the joint entropy that is shared by both random variables.

Experimental Measures and Results

Database

The results shown in this section were obtained using a subset of 100 writers from the BiosecurID database [28]. The BiosecurID database comprises 8 biometric traits, including handwritten text. Data were collected during 4 different sessions in a time span of 4 months. Regarding handwritten text, each writer was requested to write 16 different Spanish words in uppercase, each one in a single line, without corrections or crossing outs. These words are the same ones that are shown in Table 1. When designing the database, uppercase words were preferred to lowercase words because they pose a more challenging writer recognition problem (less differences among different writers) and because they tend to be the preferred writing method in online-gathering capable devices such as tablets and PDAs.

The acquisition was carried out with a WACOM INTOUS A4 USB pen tablet. The following dynamic data were captured at 100 samples per second: x-coordinate, y-coordinate, time stamp, button status, azimuth, altitude and pressure. Each execution of a word is given as seven dynamic features (time-sequences): $x(t)$, the x-coordinate; $y(t)$, the y-coordinate; $ts(t)$, a time stamp value; $bs(t)$, the button status value (0 for pen-up, 1 for pen-down); $az(t)$, the azimuth; $al(t)$, the altitude and $pr(t)$, the pressure. All features have the same length, varying from execution to execution. Thus, the execution of a word can be formally described as a matrix $[x(t), y(t), ts(t), bs(t), az(t), al(t), pr(t)]$ with $t \in [1, N]$ where N is the length (number of sampling units) of the execution. The in-air and on-surface trajectories of each execution can be straightforwardly separated thanks to the $bs(t)$ feature ($pr(t)$ could also have been used). The on-surface part of an execution is described by $[x(t), y(t), az(t), al(t), pr(t)]$,, while the in-air part is described by $[x(t), y(t), az(t), al(t)]$. Notice that in both cases, features $ts(t)$ and $bs(t)$ have been removed. These two features are not used in the recognition system [17] from which the results shown in Table 1 were obtained.

Entropy of Each Feature

The average entropy for each word, type of trajectory and feature has been computed. The results are shown in Table 2. Each figure was obtained averaging 400 executions of each word from 100 different writers (100 writers, 4 sessions per writer).

All entropies are expressed in *bits* (i.e. $\log_2$ is considered when computing $H(X)$). Figure 6 provides a summarized view of data in Table 2.

The following facts are worth noticing:

(a) If pressure is not taken into account, the global amount of information (considering all the other features) is almost the same in both types of trajectories. Figure 3 already suggested that for the x- and y-coordinates, the amount of information in the in-air case could have not been much lower than in the on-surface case because both cases do not appear to have very different degrees of *complexity*.

(b) Pressure and coordinates contain much more information than writing angles. For instance, in the in-air case, the entropy of the x-coordinate is about 7.5 bits. This represents a $2^{7.5} \approx 180$ different *states* for this feature. For the azimuth, the entropy is about 4.1 bits, representing $2^{4.1} \approx 17$ different *states*.

(c) X-coordinate and y-coordinate, especially the latter, tend to have higher entropies in on-surface trajectories

(d) On the other hand, both azimuth and altitude have higher entropy in in-air trajectories

(e) Variability (measured by the standard deviation of the entropy of the 400 executions considered) is low. This means that, in average, there are not great differences among users and sessions.

Table 2 Entropy in bits (average–avg- and standard deviation–std-) of each feature

Word	Pressure				*X*-coordinate				*Y*-coordinate				Azimuth				Altitude			
	In-air		On-surf.		In-air		On-surf.		In-air		On-surf.		In-air		On-surf.		In-air		On-surf.	
	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std
W1	n/a	n/a	7.7	0.2	7.6	0.4	7.7	0.3	7.1	0.4	7.2	0.3	4.1	0.5	4.0	0.5	3.1	0.6	2.6	0.5
W2	n/a	n/a	7.6	0.2	7.3	0.4	7.1	0.3	6.9	0.4	6.9	0.3	3.9	0.5	3.8	0.5	3.0	0.6	2.5	0.5
W3	n/a	n/a	7.9	0.2	8.0	0.4	7.9	0.3	7.4	0.3	7.4	0.3	4.3	0.5	4.1	0.5	3.1	0.5	2.6	0.5
W4	n/a	n/a	7.6	0.2	7.1	0.5	7.4	0.3	6.7	0.4	7.1	0.3	3.9	0.6	3.9	0.6	3.0	0.6	2.7	0.5
W5	n/a	n/a	7.8	0.2	7.6	0.5	7.8	0.3	7.1	0.4	7.4	0.3	4.1	0.5	4.0	0.6	3.1	0.6	2.7	0.5
W6	n/a	n/a	7.8	0.2	7.8	0.4	7.7	0.3	7.2	0.4	7.3	0.3	4.2	0.6	4.1	0.6	3.0	0.6	2.7	0.5
W7	n/a	n/a	7.6	0.2	7.1	0.5	7.3	0.3	6.7	0.5	7.0	0.3	3.9	0.5	3.9	0.6	2.8	0.6	2.6	0.5
W8	n/a	n/a	7.7	0.2	7.4	0.5	7.6	0.3	6.9	0.4	7.2	0.3	4.1	0.6	4.1	0.5	3.1	0.6	2.6	0.5
W9	n/a	n/a	7.6	0.2	7.4	0.4	7.3	0.3	6.9	0.4	7.0	0.3	4.1	0.5	4.1	0.5	3.0	0.5	2.7	0.5
W10	n/a	n/a	7.7	0.2	7.5	0.4	7.5	0.3	7.0	0.4	7.1	0.3	4.2	0.6	4.1	0.6	3.0	0.6	2.7	0.5
W11	n/a	n/a	7.6	0.2	7.5	0.5	7.4	0.3	7.0	0.4	7.0	0.3	4.2	0.5	4.1	0.5	3.1	0.6	2.7	0.6
W12	n/a	n/a	7.6	0.2	7.3	0.4	7.2	0.4	6.9	0.3	6.9	0.4	4.0	0.5	4.0	0.6	3.0	0.6	2.5	0.6
W13	n/a	n/a	7.6	0.2	7.3	0.5	7.6	0.3	6.9	0.4	7.2	0.3	4.2	0.5	4.1	0.5	3.1	0.6	2.6	0.5
W14	n/a	n/a	7.6	0.2	7.3	0.5	7.5	0.3	6.8	0.4	7.2	0.3	4.2	0.5	4.2	0.5	3.2	0.6	2.7	0.5
W15	n/a	n/a	7.6	0.2	7.3	0.5	7.5	0.4	6.8	0.6	7.1	0.5	4.2	0.5	4.1	0.6	3.2	0.6	2.6	0.5
W16	n/a	n/a	7.7	0.2	7.6	0.4	7.8	0.3	7.0	0.8	7.3	0.8	4.3	0.5	4.2	0.5	3.3	0.6	2.8	0.6

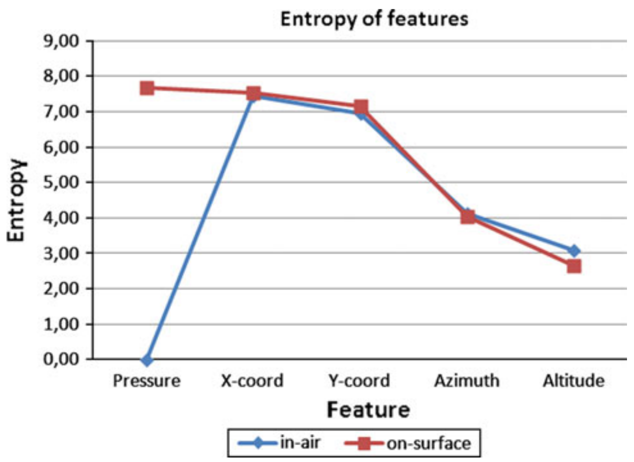


Fig. 6 Entropy of each feature. The values shown are the averages among the words of the entropies shown in Table 2

Redundancy Between In-Air and On-Surface Trajectories

The fact that in-air and on-surface trajectories are different does not imply that they convey entirely different information. In order to evaluate the degree of redundancy between them, we propose to use the mutual and the relative mutual information between pairs of measures of the same feature, taken from in-air and on-surface trajectories.

For a given word and feature, f, fa_u^e and fs_u^e, respectively, denote the in-air and on-surface values of that feature for the e-th execution of this word performed by writer u. $\text{Joint}^{\text{air-surface}}$ denotes the average joint entropy between pairs of measures (one in-air, one on-surface) of that feature taken from the same user and execution. Analogously, $\text{Mutual}^{\text{air-surface}}$ and $\text{RMutual}^{\text{air-surface}}$ denote the average mutual information and relative mutual information between pairs of measures of a feature, taken from the same user and execution:

$$\text{Joint}^{\text{air-surface}} = \operatorname*{avg}_{\forall e, \forall u}\left(H\left(fa_u^e, fs_u^e\right)\right)$$

$$\text{Mutual}^{\text{air-surface}} = \operatorname*{avg}_{\forall e, \forall u}\left(I\left(fa_u^e; fs_u^e\right)\right)$$

$$\text{RMutual}^{\text{air-surface}} = \operatorname*{avg}_{\forall e, \forall u}\left(I'\left(fa_u^e; fs_u^e\right)\right)$$

Table 3 contains the average values obtained for $\text{Joint}^{\text{air-surface}}$, $\text{Mutual}^{\text{air-surface}}$, and $\text{RMutual}^{\text{air-surface}}$. In the case of $\text{RMutual}^{\text{air-surface}}$, the standard deviation is also shown. All figures were obtained from the same dataset that in the previous section.

Figure 7 provides a summarized view of the proportion between joint entropy and mutual information.

The following facts are worth noticing:

(a) X-coordinate shows a redundancy of about 6.5 bits (with relative mutual information around 0.8 that is about 80%), while y-coordinate shows a redundancy of slightly less than 6 bits (70%). Although in both cases redundancy is high, there is still a significant

Table 3 Relations between in-air and on-surface trajectories measured by their joint, mutual and relative mutual information

Word	*X*-coordinate				*Y*-coordinate				Azimuth				Altitude			
	Joint	Mutual	RMutual		Joint	Mutual	RMutual		Joint	Mutual	RMutual		Joint	Mutual	RMutual	
	avg	avg	avg	std	avg	avg	avg	std	avg	avg	avg	std	avg	avg	avg	std
W1	8.4	6.8	0.81	0.03	8.4	5.8	0.70	0.04	6.5	1.6	0.25	0.1	5.2	0.5	0.10	0.1
W2	8.0	6.4	0.80	0.03	8.0	5.6	0.71	0.04	6.2	1.6	0.25	0.1	4.9	0.5	0.10	0.1
W3	8.7	7.1	0.82	0.02	8.7	6.0	0.68	0.04	6.8	1.6	0.23	0.1	5.2	0.5	0.09	0.1
W4	8.0	6.4	0.79	0.03	8.0	5.6	0.70	0.04	6.1	1.6	0.26	0.1	5.0	0.6	0.12	0.1
W5	8.5	6.8	0.80	0.03	8.5	5.9	0.69	0.04	6.5	1.7	0.26	0.1	5.1	0.6	0.11	0.1
W6	8.5	6.9	0.81	0.02	8.5	5.9	0.69	0.04	6.6	1.7	0.26	0.1	5.1	0.6	0.10	0.1
W7	8.0	6.4	0.79	0.04	8.0	5.6	0.70	0.04	6.1	1.7	0.28	0.1	4.9	0.6	0.12	0.1
W8	8.3	6.6	0.80	0.03	8.3	5.7	0.69	0.04	6.4	1.8	0.28	0.1	5.1	0.6	0.11	0.1
W9	8.1	6.5	0.80	0.03	8.1	5.8	0.71	0.04	6.3	1.9	0.29	0.1	5.0	0.6	0.12	0.1
W10	8.3	6.6	0.80	0.03	8.3	5.8	0.70	0.04	6.4	1.8	0.28	0.1	5.1	0.6	0.12	0.1
W11	8.2	6.6	0.80	0.03	8.2	5.7	0.70	0.04	6.4	1.9	0.29	0.1	5.1	0.6	0.12	0.1
W12	8.0	6.4	0.80	0.03	8.0	5.7	0.71	0.04	6.2	1.8	0.29	0.1	4.9	0.6	0.13	0.1
W13	8.3	6.6	0.80	0.03	8.3	5.7	0.69	0.04	6.4	1.9	0.30	0.1	5.1	0.7	0.13	0.1
W14	8.2	6.5	0.79	0.03	8.3	5.8	0.69	0.04	6.4	1.9	0.29	0.1	5.1	0.6	0.12	0.1
W15	8.2	6.5	0.79	0.03	8.2	5.7	0.69	0.05	6.4	1.9	0.29	0.1	5.1	0.6	0.12	0.1
W16	8.4	6.8	0.81	0.03	8.4	5.9	0.70	0.04	6.6	1.9	0.28	0.1	5.4	0.6	0.11	0.1

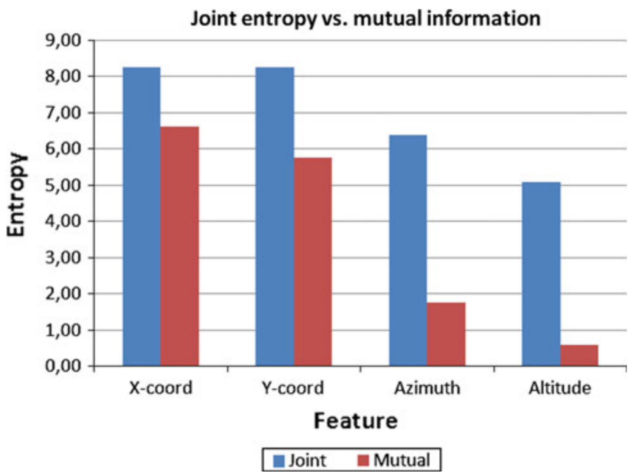


Fig. 7 Comparison of joint entropy and mutual information. The values shown are the averages among the words of the values shown in Table 3

amount of non-redundant information (20–30%). The reader should notice that entropy being a logarithmic measure, one bit of difference amounts to a multiplication factor of 2 in the number of states. Thus, a difference of 1.5 bits between the joint entropy and the mutual entropy (as in the *x*-coordinate) amounts to a multiplication factor of $2^{1.5} \approx 2.83$ in the number of states.

(b) On the other hand, azimuth and altitude, especially the latter, show a very low redundancy. In the case of azimuth, it is less than 2 bits (25–30%), and when it comes to altitude, it is less than 1 bit (about 10%)

(c) Variability, measured by the standard deviation, is low (azimuth and altitude) and very low (*x*-coordinate, *y*-coordinate). This means that, in average, all users/sessions have similar behaviour with respect to redundancy.

Inter-Writer and Intra-Writer Difference

From a biometric recognition point of view, the fitness of a feature to perform recognition does not only depend on the amount of information it contains but also on the difference between the intra-writer and the inter-writer case. Given a feature *f*, it is highly desirable that different measures of *f* taken from the same writer are more alike to each other than measures taken from different writers. From an information theory perspective, it would be desirable that the amount of mutual information was higher when considering the same writer (intra-writer) than when considering different writers (inter-writer). We propose to use the average difference between both cases as a mean to evaluate the potential usefulness of a given feature.

As in the previous section, for a given word and feature, f, fa_u^e and fs_u^e, respectively, denote the in-air and on-surface values of that feature for the *e*-th execution of this word performed by user *u*.

For a given word and feature, $\mathrm{Intra}_u^{\mathrm{air}}$ and $\mathrm{Intra}_u^{\mathrm{surface}}$ denote the average values for all measures of the relative mutual information between different executions of this word performed by writer *u*.

$$\mathrm{Intra}_u^{\mathrm{air}} = \operatorname*{avg}_{i \neq j}\left(I'\left(fa_u^i; fa_u^j\right)\right)$$

$$\mathrm{Intra}_u^{\mathrm{surface}} = \operatorname*{avg}_{i \neq j}\left(I'\left(fs_u^i; fs_u^j\right)\right)$$

Analogously, for a given word and feature, $\mathrm{Inter}_u^{\mathrm{air}}$ and $\mathrm{Inter}_u^{\mathrm{surface}}$ denote the average value of the relative mutual information between executions of this word performed by writer *u* and any other writer.

$$\mathrm{Inter}_u^{\mathrm{air}} = \operatorname*{avg}_{u \neq v}\left(I'\left(fa_u^*; fa_v^*\right)\right)$$

$$\mathrm{Inter}_u^{\mathrm{surface}} = \operatorname*{avg}_{u \neq v}\left(I'\left(fs_u^*; fs_v^*\right)\right)$$

where * means any execution.

Finally, $\mathrm{Diff}_u^{\mathrm{air}}$ and $\mathrm{Diff}_u^{\mathrm{surface}}$ denote the differences between the inter-writer and the intra-writer measures for the in-air and on-surface cases, respectively:

$$\mathrm{Diff}_u^{\mathrm{air}} = \mathrm{Intra}_u^{\mathrm{air}} - \mathrm{Inter}_u^{\mathrm{air}}$$

$$\mathrm{Diff}_u^{\mathrm{surface}} = \mathrm{Intra}_u^{\mathrm{surface}} - \mathrm{Inter}_u^{\mathrm{surface}}$$

Should $\mathrm{Diff}_u^{\mathrm{air}} > 0$ and $\mathrm{Diff}_u^{\mathrm{surface}} > 0$, this would mean that, in average and relative to their joint entropies, the executions from writer u share more information among them than they share with executions from other writers.

Table 4 shows, for each word and feature, the average values of $\mathrm{Diff}_u^{\mathrm{surface}}$ and their standard deviation. The reader will notice that the averages are all positive but quite close to zero. In order to determine whether these average values are significantly positive, they have been put to a Student's unilateral paired t-test with the following parameters: null hypothesis $H_0 : \operatorname*{avg}_{\forall u}\left(\mathrm{Diff}_u^{\mathrm{surface}}\right) = 0$; alternative hypothesis $H_1 : \operatorname*{avg}_{\forall u}\left(\mathrm{Diff}_u^{\mathrm{surface}}\right) > 0$; degrees of freedom: 99. For each feature, the third column (*p*-val) contains the *p*-value of the test. The *p*-value is the probability of obtaining an average value for $\mathrm{Diff}_u^{\mathrm{surface}}$ as extreme as the one that was actually obtained, assuming that the null hypothesis is true.

Table 5 shows the same results for the in-air trajectories. Figure 8 summarizes and compares the *p*-values obtained for both types of trajectories.

Notice that, for both types of trajectories, with a significance level of $\alpha = 0.01$, the null hypothesis would be

Table 4 Differences in relative mutual information between the inter-writer and intra-writer case for on-surface trajectories

Word	Pressure			*X*-coordinate			*Y*-coordinate			Azimuth			Altitude		
	avg	std	*p*-val	avg	std	*p*-val	avg	std	*p*-val	avg	std	*p*-val	avg	std	*p*-val
W1	0.003	0.02	6.E−02	0.018	0.02	1.E−13	0.017	0.02	7.E−12	0.019	0.05	2.E−04	0.017	0.04	6.E−06
W2	0.004	0.02	3.E−02	0.025	0.02	2.E−19	0.023	0.02	6.E−17	0.021	0.05	4.E−05	0.020	0.04	1.E−07
W3	0.003	0.02	9.E−02	0.017	0.02	1.E−15	0.016	0.02	3.E−12	0.019	0.05	1.E−04	0.017	0.03	2.E−06
W4	0.003	0.02	8.E−02	0.018	0.02	6.E−12	0.017	0.02	1.E−10	0.025	0.06	1.E−05	0.024	0.04	1.E−07
W5	0.003	0.02	6.E−02	0.018	0.02	5.E−14	0.015	0.02	5.E−09	0.022	0.06	9.E−05	0.021	0.04	5.E−06
W6	0.003	0.02	1.E−01	0.016	0.02	3.E−12	0.016	0.02	1.E−10	0.022	0.06	4.E−04	0.020	0.04	9.E−07
W7	0.004	0.02	2.E−02	0.018	0.02	8.E−13	0.018	0.02	1.E−11	0.026	0.07	5.E−05	0.024	0.04	1.E−08
W8	0.004	0.02	2.E−02	0.018	0.02	1.E−13	0.014	0.02	9.E−10	0.024	0.06	4.E−05	0.017	0.04	3.E−05
W9	0.005	0.02	1.E−02	0.020	0.02	8.E−18	0.019	0.03	6.E−12	0.022	0.05	5.E−05	0.025	0.05	1.E−06
W10	0.003	0.02	8.E−02	0.018	0.02	5.E−14	0.017	0.02	3.E−12	0.023	0.06	1.E−04	0.021	0.04	4.E−06
W11	0.003	0.02	6.E−02	0.021	0.02	5.E−16	0.018	0.02	2.E−11	0.021	0.05	7.E−05	0.021	0.05	1.E−05
W12	0.004	0.02	4.E−02	0.024	0.02	6.E−18	0.023	0.03	4.E−15	0.019	0.05	2.E−04	0.023	0.04	2.E−07
W13	0.003	0.02	7.E−02	0.019	0.02	1.E−11	0.017	0.02	2.E−10	0.024	0.06	3.E−05	0.020	0.05	8.E−06
W14	0.003	0.02	1.E−01	0.020	0.02	1.E−16	0.017	0.02	3.E−11	0.019	0.05	2.E−04	0.023	0.05	6.E−06
W15	0.003	0.02	6.E−02	0.019	0.02	6.E−16	0.019	0.02	2.E−13	0.022	0.06	4.E−04	0.013	0.04	1.E−03
W16	0.003	0.02	7.E−02	0.015	0.02	7.E−12	0.014	0.02	3.E−08	0.020	0.05	3.E−04	0.018	0.04	2.E−05

Table 5 Differences in relative mutual information between the inter-writer and intra-writer case for in-air trajectories

Word	*X*-coordinate			*Y*-coordinate			Azimuth			Altitude		
	avg	std	*p*-val	avg	std	*p*-val	avg	std	*p*-val	avg	std	*p*-val
W1	0.011	0.02	5.E−06	0.011	0.02	4.E−06	0.017	0.05	1.E−03	0.015	0.04	1.E−04
W2	0.010	0.03	2.E−04	0.010	0.02	6.E−05	0.015	0.05	2.E−03	0.014	0.04	5.E−04
W3	0.009	0.02	2.E−04	0.009	0.03	2.E−04	0.015	0.05	2.E−03	0.010	0.03	2.E−03
W4	0.014	0.03	4.E−06	0.012	0.03	1.E−04	0.022	0.05	2.E−05	0.018	0.04	4.E−05
W5	0.013	0.03	2.E−06	0.013	0.02	4.E−07	0.018	0.05	6.E−04	0.013	0.04	7.E−04
W6	0.011	0.03	3.E−05	0.010	0.02	5.E−05	0.017	0.06	3.E−03	0.013	0.04	5.E−04
W7	0.019	0.03	2.E−09	0.018	0.03	8.E−11	0.025	0.07	1.E−04	0.021	0.05	1.E−05
W8	0.011	0.03	2.E−04	0.011	0.03	6.E−05	0.021	0.06	2.E−04	0.014	0.04	9.E−05
W9	0.013	0.03	1.E−06	0.013	0.03	2.E−06	0.016	0.05	1.E−03	0.018	0.05	8.E−05
W10	0.011	0.02	3.E−06	0.011	0.02	2.E−06	0.018	0.06	2.E−03	0.017	0.04	7.E−05
W11	0.014	0.02	7.E−08	0.013	0.02	3.E−07	0.018	0.05	3.E−04	0.015	0.04	1.E−04
W12	0.012	0.03	2.E−05	0.013	0.02	3.E−07	0.019	0.06	8.E−04	0.019	0.04	3.E−06
W13	0.012	0.03	3.E−06	0.013	0.03	1.E−06	0.020	0.06	4.E−04	0.017	0.05	2.E−04
W14	0.014	0.03	4.E−06	0.014	0.03	5.E−07	0.018	0.05	6.E−04	0.017	0.04	1.E−04
W15	0.012	0.03	2.E−06	0.013	0.03	1.E−06	0.021	0.06	5.E−04	0.015	0.04	2.E−04
W16	0.010	0.02	1.E−04	0.010	0.03	1.E−04	0.016	0.06	6.E−03	0.010	0.03	2.E−03

rejected in all cases except for pressure. This means that from a purely statistical point of view, all features but pressure exhibit, in average, a significant difference between the intra-user and the inter-user case. When it comes to pressure, even if the average difference is positive for all words, the variability (standard deviation) is high enough to prevent a clear rejection of the null hypothesis (*p*-values range from 0.01 to 0.1)

Conclusions and Further Research

The experimental results presented in the previous section clearly support the claim that in-air trajectories contain as much information as on-surface trajectories. For four of the five features considered (all except pressure), the difference between the on-surface and the in-air case is lower than a single bit, usually some tenths of a bit (see Table 2). In

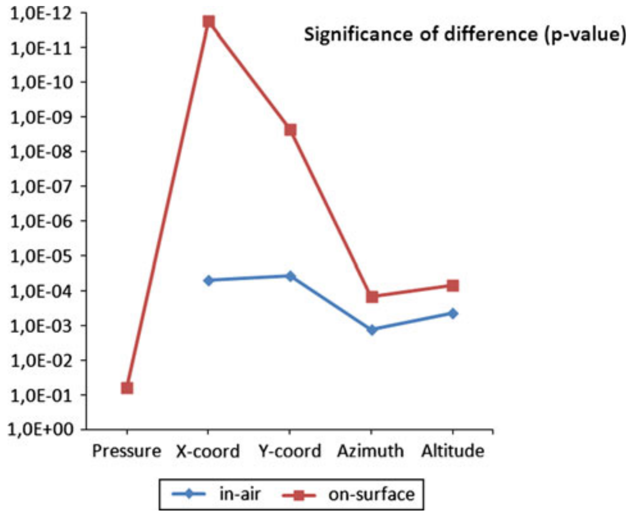


Fig. 8 *p*-values for the differences in relative mutual information between inter-writer and intra-writer measures (average values among the words in Tables 4 and 5)

fact, the only substantial difference between the two types of trajectories lies in the information provided by pressure, the fifth feature under consideration. When it comes to redundancy, the results show that although it is noticeable in the case of the *x*-coordinate and the *y*-coordinate, it is low and very low in the case of azimuth and altitude, respectively (see Table 3). From a global perspective, it cannot be said that in-air and on-surface trajectories are entirely non-redundant. Nevertheless, although a certain amount of redundancy is present, it is far from seeming to be enough to deem the in-air trajectories as superfluous. Entropy and redundancy, the latter measured by mutual information and relative mutual information, show, for all the analysed words and features, a low variability. This fact is important because it somehow implies that the obtained results are valid for a great majority of writers since they show a similar behaviour with respect to these measures.

When both aspects, amount of information in each type of trajectory and non-superfluousness of the in-air trajectories, are considered together, there seems to be no need to discard the information contained in in-air trajectories, as it is often done in handwriting-based biometric recognition systems. What is more, it may be advisable to gather and process this information on its own. Research results presented in [17, 22] had already given support to the notion that in-air trajectories were rich in information. Now, from the information theory perspective, we have further evidence to support this notion.

Regarding the biometric potential of both types of trajectories measured by the difference between the intra-user and the inter-user cases, the results are not conclusive. On the one hand, the differences are always positive and, except for pressure, statistically significant. On the other hand, these differences are very close to zero which may prevent their use as a score of the similitude between different executions of the same word. Nevertheless, this lack of conclusiveness does not imply that handwriting words do not possess a considerable biometric potential; it just means that the information theory, and more precisely the selected measures and the experiments performed, cannot prove the existence of this potential. Fortunately, past research in this field does show that words and short sequences of text can perform well in biometric recognition tasks.

The results of the information analysis of the writing trajectories also suggest some possibilities of further research. The fact that the amount of redundancy between in-air and on-surface trajectories is highly dependent on the feature under consideration could be exploited in the recognition field by focusing on the less redundant features in order to improve the recognition performance. Also, a future analysis of the redundancy between different features could help improve the selection of the most relevant ones.

The writing angles, azimuth and altitude, only available when online data are gathered, contain an amount of information that deserves not to be disregarded. Although their use in the intra-user vs. inter-user problem may be limited (their differences yield the lowest significances when compared to the other features, except pressure,), they may still prove useful in other classification problems such as gender classification.

Acknowledgments This work has been supported by FEDER and MEC, TEC2009-14123-C04-04. We also want to acknowledge the SIX (CZ.1.05/2.1.00/03.0072), CZ.1.07/2.3.00/20.0094, VG20102014033 and KONTAKT ME10123 projects for providing Jiri's support.

References

1. Huber RA, Headrick AM. Handwriting identification: facts and fundamentals. Boca Raton: CRC Press; 1999.
2. Forbes KE, Shanks MF, Venneri A. The evolution of dysgraphia in Alzheimer's disease. Brain Res Bull. 2004;63:19–24.
3. Neils-Strunjas J, Groves-Wright K, Mashima P, Harnish S. Dysgraphia in Alzheimer's disease: a review for clinical and research purposes. J Speech Lang Hear Res. 2006;49:1313–30.
4. Werner P, Rosenblum S, Bar-On G, Heinik J, Korczyn A. Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment. J Gerontol B Psychol Sci Soc Sci. 2006;61:P228–36.
5. Asıcıoglu F, Turan N. Handwriting changes under the effect of alcohol. Forensic Sci Int. 2003;132:201–10.
6. Phillips JG, Ogeil RP, Müller F. Alcohol consumption and handwriting: a kinematic analysis. Hum Mov Sci. 2009;28:619–32.
7. Foley RG, Lamar Miller A. The effects of marijuana and alcohol usage on handwriting. Forensic Sci Int. 1979;14:159–64.
8. Tucha O, Walitza S, Mecklinger L, Stasik D, Sontag T, Lange KW. The effect of caffeine on handwriting movements in skilled writers. Hum Mov Sci. 2006;25:523–35.

9. Rosenblum S, Parush S, Weiss P. The in air phenomenon: temporal and spatial correlates of the handwriting process. Percept Mot Skills. 2003;96:933.
10. Viñals Carrera F, Puente Balsells ML. Grafología criminal. Herder, Barcelona, 2009.
11. Puente Balsells ML, Viñals Carrera F. Grafología y ciencia: validación con ciento cincuenta tesis doctorales, 1a en lengua caellana ed., Uoc, Barcelona, 2010.
12. Vielhauer C. Biometric user authentication for it security from fundamentals to handwriting. Boston, MA: Springer Science + Business Media Inc; 2006.
13. Plamondon R, Lorette G. Automatic signature verification and writer identification: the state of the art. Pattern Recogn. 1989;22:107–31.
14. Leclerc F, Plamondon R. Automatic signature verification: the state of the art 1989–1993. Int J Pattern Recogn Artif Intell. 1994;8:643.
15. Plamondon R, Srihari S. On-line and off-line handwriting recognition: a comprehensive survey. IEEE Trans Pattern Anal Mach Intell. 2000;22:63–84.
16. Impedovo D, Pirlo G. Automatic signature verification: the state of the art. IEEE Trans Syst., Man and Cybernetics, Part C (Applications and Reviews). 2008;38:609–35.
17. Sesa-Nogueras E, Faundez-Zanuy M. Biometric recognition using online uppercase handwritten text. Pattern Recogn. 2012;45:128–44.
18. Srihari S, Sung-Hyuk C, Sangjik L, Establishing handwriting individuality using pattern recognition techniques. In: Proceedings of the sixth international conference on document analysis and recognition. 2001;1195–204.
19. Zhang B, Srihari S, Analysis of handwriting individuality using word features. In Proceedings of the seventh international conference on document analysis and recognition. 2003;1142–6.
20. Hook C, Kempf J, Scharfenberg G, A novel digitizing pen for the analysis of pen pressure and inclination in handwriting biometrics, biometric authentication workshop, Prague 2004, Lecture Notes in Computer Science. 2004;3087:283–94.
21. Chapran J. Biometric writer identification: feature analysis and classification. Int J Pattern Recogn Artif Intell. 2006;20:483–503.
22. Sesa-Nogueras E, Faundez-Zanuy M. Writer recognition by means of stroke categorization based on self-organizing maps. In: Apolloni B, Bassis S, Esposito A, Morabito CF, editors. Neural nets WIRN11. Volume 234 of Frontiers in Artificial Intelligence and Applications. IOS Press; 2011. ISBN:978-1-60750-971-4. http://www.iospress.nl/book/neural-nets-wirn11/.
23. Shannon CE. A mathematical theory of communication. New York: American Telephone and Telegraph Co.; 1948.
24. Espinosa-Duró V, Faundez-Zanuy M, Mekyska J. Beyond cognitive signals. Cogn Comput. 2011;3:374–81.
25. Espinosa-Duro V, Faundez-Zanuy M, Mekyska J, Monte-Moreno E. A criterion for analysis of different sensor combinations with an application to face biometrics. Cogn Comput. 2010;2:135–41.
26. Cover TM, Thomas JA. Elements of information theory. New York: Wiley; 1991.
27. MacKay DJC. Information theory, inference, and learning algorithms. Cambridge: Cambridge University Press; 2003.
28. Fierrez-Aguilar J, Galbally J, Ortega-Garcia J, Freire M, Alonso-Fernandez F, Ramos D, et al. BiosecurID: a multimodal biometric database. Pattern Anal Appl 2010;13.