

VĚDECKÉ SPISY VYSOKÉHO UČENÍ TECHNICKÉHO V BRNĚ

Edice Habilitační a inaugurační spisy, sv. 520

ISSN 1213-418X

Daniel Schwarz

**NEUROIMAGING DATA ANALYSIS
AND RECOGNITION
FOR BRAIN RESEARCH**

BRNO UNIVERSITY OF TECHNOLOGY
Faculty of Electrical Engineering and Communication

Ing. Daniel Schwarz, Ph.D.

**NEUROIMAGING DATA ANALYSIS
AND RECOGNITION FOR BRAIN RESEARCH**

ANALÝZA A ROZPOZNÁVÁNÍ NEUROZOBRAZOVACÍCH DAT
PRO VÝZKUM MOZKU

SHORT VERSION OF HABILITATION THESIS



BRNO 2015

KEYWORDS

biomedical engineering, magnetic resonance imaging, neuroimaging, computational neuroanatomy, machine learning, brain morphometry, schizophrenia, image registration, pattern recognition, classification, linear discriminant analysis, support vector machines, principal component analysis, wavelet transform

KLÍČOVÁ SLOVA

biomedicínské inženýrství, magnetická resonance, neurozobrazování, výpočetní neuroanatomie, strojové učení, morfometrie mozku, schizofrenie, registrace obrazů, rozpoznávání vzorů, klasifikace, lineární diskriminační analýza, algoritmus podpůrných vektorů, analýza hlavních komponent, vlnková transformace

Originál habilitační práce je dostupný na Vědeckém oddělení děkanátu FEKT VUT v Brně, Technická 10, Brno, 616 00.

The original version of the habilitation thesis is available at the Research department of FEEC Brno University of Technology, Technická 10, Brno, 616 00.

TABLE OF CONTENTS

CURRICULUM VITAE.....	4
1 INTRODUCTION.....	5
2 STATE OF THE ART	5
2.1 IMAGE REGISTRATION	5
2.2 WHOLE-BRAIN AUTOMATED MORPHOMETRIC METHODS	5
2.3 PATTERN RECOGNITION IN NEUROPSYCHIATRIC RESEARCH.....	7
2.4 CONCERNS, DOUBTS AND PROBLEMATIC AREAS	9
3 DATA, METHODS AND PROGRESS BEYOND THE STATE OF THE ART.....	10
3.1 SCHIZOPHRENIA PATIENTS AND HEALTHY CONTROLS.....	10
3.2 IMAGE PRE-PROCESSING AND SPATIAL NORMALIZATION	11
3.3 APPROACH I: COMBINING FEATURES.....	11
3.4 APPROACH II: WAVELET FEATURES	13
3.5 APPROACH III: COMBINING FEATURES AND CLASSIFIERS	16
4 RESULTS	21
5 DISCUSSION	25
5.1 CLASSIFICATION OF FIRST EPISODE SCHIZOPHRENIA.....	25
5.2 NEUROIMAGING & PEDAGOGY	30
6 CONCLUSIONS	31
BIBLIOGRAPHY	32
ABSTRACT.....	37

CURRICULUM VITAE

Name, degrees Schwarz, Daniel, Ing, Ph.D.
Date and place of birth 20th June 1977 in Karvina, Czech Republic
Department, faculty, university Institute of Biostatistics and Analyses,
Faculty of Medicine, Faculty of Science
Masaryk University
Current position Assistant professor



Education 2005: Ph.D. degree in Biomedical electronics and biocybernetics (Brno University of Technology)
2000: Ing. (eq. MSc.) degree in Electronics and communication technologies (Brno University of Technology)

Professional experience 2005-now: Masaryk University
2002-2004: SOVA NET, Ltd.

Teaching activities Linear and Adaptive Data Processing, Project in Mathematical Biology, Advanced methods in Biostatistics

Research activities Image processing and analysis, Computational neuroanatomy
SELECTED PROJECTS:

2012-2015: IGA No. 13359/4 - Advanced Methods for Recognition of MR brain images for Computer Aided Diagnosis of Neuropsychiatric Disorders; (principal investigator)

2009-2011: IGA No. 10347/3 - Modern pattern recognition methods for image data analysis in neuropsychiatric research; (principal investigator)

2009-2011: IGA No. 10347/3 - Prediction of the initial phases of the schizophrenia using brain morphology imaging methods (co-investigator)

2007-2008: GACR No. 102/07/P263 - Nonlinear multimodal registration for automatic morphometry of MRI brain images based on anatomically constrained spatial deformations; (principal investigator)

Academic activities Member of the committee for simulations at the Faculty of Medicine, Masaryk University (since 2014).

Member of the Czech Society for Clinical Neurophysiology ČLS JEP, HuBraM – section for functional mapping of brain (since 2013).

Editor-in-chief of the MEFANET Journal (since 2013)

Member of the editorial board at the Faculty of Medicine, Masaryk University (since 2007)

Member of the coordinating committee of the educational network of Czech and Slovak medical faculties – MEFANET (since 2007)

Chairman or co-chairman of annual MEFANET national and international conferences (since 2007)

1 INTRODUCTION

The last two decades have witnessed an explosive growth in the ability to “understand the human brain” – a key to progress in neuroscience, to promote and protect brain health, and to develop treatments for restoring, regenerating, and repairing diseased brain functions. Computational neuroanatomy is a growing field of powerful applications of imaging modalities and computational techniques in neuroscience. It promises an automated methodology to characterize neuroanatomical configuration of structural magnetic resonance imaging (MRI) brain scans. One of the crucial techniques in this methodology is image registration. Together with techniques adopted from inferential statistics and hypothesis testing, it allows to uncover brain regions with significant morphological differences between normal and clinical populations. Such techniques have been already used also in modern psychiatry research to seek for biomarkers and neurobiology of various mental diseases.

The real challenge for psychiatry would be, however, to move from group analysis between patients and healthy volunteers to computer-aided diagnostics on the level of an individual patient. Although pioneering works employing machine learning techniques have recently borne fruit in case of neurological diseases, this is extremely difficult in mental diseases.

2 STATE OF THE ART

This chapter summarizes research and developments in the neuroimaging field during the last two decades. It focuses on those image analysis methods, which are relevant to the original work presented in the chapters 3–5 of this thesis, i.e. “progress beyond”. The technique which underlies the research presented here is registration of images. The second part explains basics of automated brain morphometry methods and introduces voxel-based and deformation-based morphometry. Both these analytical methods contain a spatial normalization step – highly dependent on image registration. Besides surveys of their use in neuro-psychiatric research, quantitative results of the comparison of their performance, reported in (Schwarz and Kasperek, 2011), are shown here. The final parts are devoted to pattern recognition techniques for the analysis of neuroimaging data, which have recently been very helpful to gain novel biological insights about neuropsychiatric disorders. Some concerns, doubts and pitfalls which have appeared in the literature lately, are discussed at the end.

2.1 IMAGE REGISTRATION

Image registration is a process of estimating a spatial transformation which maps each point of an image onto its physically corresponding point of another image (Rohr, 2000). One of its main application fields is biomedical imaging, where the major challenges include finding correspondences between image data from different sensors and from image databases. The spectrum of geometric differences between images is very broad including nonlinear image distortions caused by different modalities, time-varying processes or anatomical variability among different subjects (Schwarz, 2005).

A universal method does not exist due to the diversity of registration tasks. Various approaches to the classification of image registration methods might be found in general surveys, such as (Maintz and Viergever, 1998; Rohr, 2000; Zitová and Flusser, 2003) or in the review of registration approaches applied to the field of computational neuroanatomy (Gholipour et al., 2007). More recent surveys are focused on more elaborate deformable image registration algorithms (Rueckert and Aljabar, 2010; Sotiras et al., 2012).

2.2 WHOLE-BRAIN AUTOMATED MORPHOMETRIC METHODS

Analysis of brain morphology using neuroimaging data is an important area of research in neuroscience. At first volumetric approaches based on manual delineation of regions of interest

(ROI) were used, later followed by several computational approaches. These were designed to overcome limitations of volumetry that is labor intensive, i.e. limits the number of subjects in a study, requires a prior anatomical hypothesis for region selection, is prone to errors that arise from subjectivity of boundaries detection, i.e. limits reliability and inter-center comparability of the results. The first implementations of computational neuroanatomic approaches are methods for voxel- and deformation-based morphometry (Ashburner et al., 1998; Ashburner and Friston, 2000).

Voxel-based morphometry (VBM) is based on the assumption that after the removal of general shape differences during image registration, local misregistrations remain, resulting in between-subject differences in local brain tissue content. Usually, the brain intensity image is segmented into different brain tissue compartments which are then analyzed separately. These local differences in tissue content are then explained by a disease effect. Besides tissue segmentation and spatial normalization, VBM algorithms usually contain also a step referred to as modulation, in which normalized tissue maps are scaled by the macroscopic deformations to preserve local volumes. The VBM approach has been validated several times – corresponding findings are obtained using both VBM and ROI-based volume calculations (Giuliani et al., 2005; Gong et al., 2005; Keller et al., 2002). However, the idea of VBM is also criticized for its proneness to errors and false positive results due to imprecise and possibly erroneous image registrations (Bookstein, 2001). The heterogeneity of the results coming from VBM analyses is still very high, as no gold-standard configuration of the parameters for all preprocessing steps inside the VBM pipeline, such as modulation or smoothing, exists. Experimental validation of the modulation step in VBM is provided in (Radua et al., 2014) – the effects of modulation on the efficacy to detect cortical thinning are assessed. Surprisingly, the modulation step in the VBM pipeline is shown to be associated to a decrease of the sensitivity to detect abnormalities.

The magnitude of voxel size changes during the registration process is encoded in the relevant deformations or displacement fields. Their analysis is the core principle of deformation-based morphometry (DBM). It is able to detect changes in brain shape and volume irrespective of the brain compartment in which they occur, in contrast to VBM. In general, DBM approaches differ in the registration method used, mainly in terms of the spatial deformation model. In the initial works (Ashburner et al., 1998; Ashburner and Friston, 2000), smooth parametric transforms with low-frequency sine basis functions are used. Therefore it is not possible to encode all anatomical variability, including subtle differences, into the spatial transforms (low-resolution DBM). A complex description of brain morphology has been possible since methods for high-resolution deformable registration were introduced (high-resolution DBM). These methods include spatial deformation models based on high-dimensional parametric transforms or models inspired by similarity to continuum mechanics. DBM approach is also compared to traditional ROI-based volume calculations and yields corresponding results (Gaser et al., 2012, 2001).

There are several ways of statistical analysis of deformations, among them a univariate analysis applied to Jacobian determinants, which represent the factors by which the deformation expands or shrinks volumes at the respective voxels. The analysis of Jacobian determinants allows for the detection of local volume changes in the brain. In short, DBM analyzes how much the volume of voxels changed during subject image registration to the template image, in contrast to VBM which focuses on the residual image variability after its transformation. The finer the image transformation, the higher resolution of the deformation field, the more anatomical information is encoded in the deformation field, and the smaller the residual differences in tissue content. The high-resolution DBM could, therefore, encode local anatomical changes; moreover, it focuses on changes in spatial arrangement of images, not on the residual misregistrations, and, therefore, high-resolution DBM could overcome VBM limitations.

The application of high-resolution DBM in (Schwarz et al., 2007) is developed with the deformable registration method based on multimodal point similarity measures and the spatial

deformation model allowing for large deformations while preserving the topology of the images. Indirect comparison of results obtained using VBM and the high-resolution DBM method shows that DBM is able to detect changes in first-episode schizophrenia (Schwarz et al., 2007) that are analogous to those detected with VBM in (Kasperek et al., 2007).

The utility of mass-univariate approaches is questioned in the literature – the issues of sensitivity or the ability to correctly characterize inherently multivariate brain morphology are raised in (Davatzikos, 2004; Friston and Ashburner, 2004) and it is proposed that multivariate techniques may provide more valid information about brain morphology.

2.2.1 Quantitative comparison of DBM and VBM

The aim of the simulation study (Schwarz and Kasperek, 2011) is direct comparison of high-resolution DBM with widely used VBM analysis. Two sets of spatial deformations are generated: (i) simulations of normal anatomical variability and (ii) simulations of local volume changes at particular stereotaxic coordinates. The nonlinear spatial transformations, which represent normal anatomical variability, are computed in the model by natural neighbor scattered data interpolation from random forces pointed in the volume delimited by a binary head mask. In addition, 20 images contain three volume expansions of different extent in three exactly defined locations, together with the simulated normal anatomical variability. The extent and shape of the volume expansions in each image are randomized to simulate the variability of volume changes in pathological processes. The other 30 images are generated with the use of deformations which contain only the simulated normal anatomical variability.

The simulation results show superior performance of DBM that is able to detect all simulated local tissue expansions with very high precision – with the smallest simulated volume expansion at the scale of 600 mm³. VBM is not able to detect any of the three expansions - it is able to uncover tissue density change in near vicinity of the largest expansion – at the scale of 4000 mm³. The poor performance of VBM, especially in the case of detection of subtle local changes, may be caused by the preprocessing steps: a substantial portion of variability is removed with nonlinear registration of the images to the template as well as with Gaussian smoothing of the binary tissue segments. In contrast, when using DBM, one tries to make all variability encoded in the deformation fields. Thus, no trade-off between removing variability with registration and detecting variability itself is necessary.

2.3 PATTERN RECOGNITION IN NEUROPSYCHIATRIC RESEARCH

Schizophrenia is a disabling psychiatric disorder, manifesting in a variety of symptoms ranging from misinterpretation of reality and delusions to disorganization of thinking and behavior. It is associated with progressive altered brain functions during the course of the illness (Van Haren et al., 2012). Findings in different areas of the brain are published in numerous reviews and meta-analyses (Ellison-Wright et al., 2008; Honea et al., 2005; Shenton et al., 2001; J. Sun et al., 2009). However, many of these findings are inconsistent or even contradictory, which could indicate the heterogeneity of this severe disorder (Nenadic et al., 2012).

Recent efforts in schizophrenia research lean toward searching for new image analysis approaches in order to apply brain images in computer-aided diagnostics, since early and accurate diagnosis can significantly improve patient recovery rates and their overall prognosis (Perkins et al., 2005). It is known that schizophrenia patients show significant group differences in brain morphology in comparison to healthy people, as it is demonstrated in chapter 2.2. At individual levels, however, brain-imaging measurements in schizophrenia show considerable overlap with the normal range (D. Sun et al., 2009). The diagnosis of schizophrenia based on individual brain-image data is, therefore, much more challenging than uncovering brain regions with morphological differences between patients and healthy people (Zarogianni et al., 2013).

In addition to the morphometric analyses described in chapter 2.2, MR image data has been recently also used for uncovering spatially complex patterns which distinguish patients suffering from neuropsychiatric disorders from healthy volunteers. Assuming that the classification algorithms were of high accuracy, MRI-based prediction of a neuropsychiatric disorder would be possible at individual level. Such classification accuracy has already achieved clinically meaningful values in Alzheimer's dementia – over 95% correctly classified subjects simply on the basis of features deduced from structural brain images (Thomaz et al., 2007b). Alzheimer's dementia is, however, accompanied with significant changes in the anatomy. In case of disorders with less prominent anatomical changes – such as schizophrenia – the classification accuracy is between 70 % and 90 % in most of the studies (Davatzikos et al., 2005; Kasperek et al., 2011; Kawasaki et al., 2007; Leonard et al., 1999; Nakamura et al., 2004). Of special interest are the cases of early manifestation of the disorder where timely detection of the subjects at risk of developing schizophrenia (Koutsouleris et al., 2009) or subjects with first-episode (Kasperek et al., 2011) represents an urgent clinical priority.

The methods of automated brain morphometry are often used for extracting interesting areas and features that are subsequently used for classification: MR image segmentation is involved (Fan et al., 2008, 2007) where the feature vectors are formed by tissue densities for GM, WM and CSF, and a nonlinear support vector machine (SVM) classifier is then constructed from the data of schizophrenia patients and healthy controls. The SVM algorithm, originally invented by (Vapnik, 1999), is the most common classifier used for recognition of schizophrenia patients based on their MRI data. The values of accuracy achieved by using this classifier vary between 66 % and 90 % in recent studies (Castellani et al., 2012; Fan et al., 2008, 2007; Ingalhalikar et al., 2010; Liu et al., 2011, 2004; Shen et al., 2010). It is worth to note that a substantially lower accuracy of only 70 % is reported on a significantly greater dataset (277 subjects) (Nieuwenhuis et al., 2012). SVM has been also used recently for classification of patients with first episode of schizophrenia (FES) – with accuracy reported from 54 % to 73 % (Mourao-Miranda et al., 2012; Zanetti et al., 2013).

An alternative approach to feature extraction is presented in (Pohl and Sabuncu, 2009). The authors derive the features from parameters of optimal affine transforms which mapped selected anatomical structures between the images and a digital brain atlas. They also use SVM for classification and report a leave-one-out cross-validation accuracy of up to 90 % but without the sensitivity and specificity values; moreover, this result is achieved on a very small number of training patterns. Similar problematic results are analyzed in the review (Demirci et al., 2008), which warns against common errors in reporting prediction accuracies for various numbers of training patterns in each class and the poorly chosen validation techniques. The next chapter 2.4 unveils also other problematic issues and pitfalls, which naturally occur in analyses performed on high-dimensional datasets with a limited number of subjects.

Other popular methods for brain-image classification of schizophrenia patients and healthy controls are based on linear discriminant analysis (LDA) with reported accuracies ranging from 70.7% to 82.9% in classification based on MRI data (Karageorgiou et al., 2011; Leonard et al., 1999; Nakamura et al., 2004; Ota et al., 2012; Takayanagi et al., 2010) and from 80% to 96% in classification based on diffusion tensor images (Ardekani et al., 2011; Caprihan et al., 2008). However, (Thomaz et al., 2007a) warn against using LDA in the classification of image data if there is a limited number of subjects for analysis, as the LDA results may be unstable. Instead, they propose a modification of LDA, called the maximum uncertainty linear discriminant analysis (MLDA), to overcome the mentioned problem. MLDA is, for example, used in the classification of MR images of preterm infants (Thomaz et al., 2007a), MRI data of patients with Alzheimer disease (Thomaz et al., 2007b) and in functional MR images of healthy controls that are classified into groups of younger or older subjects (Sato et al., 2008). In Chapter 3.5 and in (Janoušová et al., 2015), MLDA is further modified to improve the classification results and is further referred to as modified MLDA (mMLDA).

2.3.1 Ensemble strategies

Combining classifiers is rapidly growing and enjoying a lot of attention from pattern recognition and machine learning communities. By combining classifiers, they are aiming at a more accurate classification decision at the expense of increased complexity (Kuncheva, 2004).

There are three types of reasons why a classifier ensemble might be better than a single classifier. (i) Statistical: “averaging” outputs of classifiers in an ensemble will diminish the risk of picking an inadequate single classifier, although the ensemble might not be better than the single best classifier. (ii) Computational: assuming that the training process of each individual classifier may lead to different local extrema of the error function, some form of aggregating will produce a better approximation to the optimal classifier than any single one. (iii) Representational: restricting the space of possible classifiers to linear classifiers only, the optimal classifier for a nonlinear problem will not belong in this space. However, an ensemble of linear classifiers can approximate any decision boundary (Dietterich, 2000; Kuncheva, 2004).

2.4 CONCERNS, DOUBTS AND PROBLEMATIC AREAS

It has been claimed and demonstrated that many of the conclusions drawn from biomedical research are probably false (Ioannidis, 2005). The reasons for this include using flexible study designs and flexible statistical analyses and running small studies with low statistical power.

2.4.1 Low statistical power in neuroscience studies

The recent study (Button et al., 2013) illustrates that low statistical power is an endemic problem also in neuroscience and discuss the implications of this for interpreting the results of individual studies.

2.4.2 Generalizability of pattern recognition in neuroimaging

Machine learning or pattern recognition in neuroimaging provides results for individual subjects, rather than results related to group differences. This is a more complicated endeavor that must be approached more carefully and efficient methods should be developed to draw generalized and valid conclusions out of high dimensional data with a limited number of subjects, especially for heterogeneous disorders whose pathophysiology is unknown (Demirci et al., 2008). The difficulties, compared to conventional pattern recognition, come from at least four following sources. (i) The feature-to-instance ratio is extremely large, in the order of 5000:1, while in a typical pattern recognition problem it is expected to be much smaller than 1. Thus, algorithms for brain-image classification often cope with the problem known as the “curse of dimensionality” or the so-called “small sample size problem” - a well-known and described problem in the domain of computational neuroscience (Lemm et al., 2011). (ii) There is a spatial relationship between the features that needs to be taken into account. (iii) The signal-to-noise ratio is low. (iv) There is great redundancy in the feature set (Kuncheva et al., 2010).

In order to obtain an unbiased estimate of the generalization performance of a particular model, careful definition of the techniques used both in designing algorithms and reporting prediction or classification accuracies must come into place. Cross-validation (CV) represents a set of techniques estimating the out-of-sample error rates (i.e. generalizability) of a particular model. There is a multitude of CV schemes, where the process of splitting the sample in two (training subset and testing subset) is repeated several times using different partitions of the sample data. Subsequently, the resulting validation errors are averaged across the multiple rounds of CV. The miscellaneous CV schemes differ by the way they split up the sample data. The most widely used method is K -fold CV (Lemm et al., 2011). In order to achieve a good compromise between bias and variance the use of 10-fold or 5-fold CV are often recommended (James et al., 2013; Kohavi, 1995). A stratified K -fold CV is a variation of the k -fold method which uses stratified folds: each set contains the same percentage of samples of each target class as the complete dataset. A special

type of K -fold CV is the so-called leave-one-out cross-validation (LOO-CV) which uses all but a single data point of the original sample for training the model. The estimated model is then validated on the single observation left out. This procedure actually performs a complete 1-fold CV. It provides a nearly unbiased estimate of the generalization error (Lemm et al., 2011). CV should be properly applied in every stage of the analysis rather than only during the performance evaluation. In other words, the design of a classification study has to ensure that testing set cannot be seen at any stage of the training. This is, unfortunately, not always true in papers reporting classification accuracy in the field – often, the full set of data is used during the design of the model and not built from scratch for each different test group or fold.

2.4.3 Positive-outcome bias

An increase in the selective publication of some results against some others is worrying because it can lead to enhanced bias in meta-analysis and hence to a distorted picture of the evidence for or against a certain hypothesis (Pautasso, 2010). Concerns about the veracity of the published results may be supported by the scientometric study (Fanelli, 2012) which focuses on worsening of positive-outcome bias in the scientific literature in general. The scientific discipline relevant to this thesis – Neuroscience and Behaviour – shows a slight decline in positive results over the years.

3 DATA, METHODS AND PROGRESS BEYOND THE STATE OF THE ART

This section describes a dataset containing T1-weighted MR images and proposes three approaches to an analysis performed on it with the use of novel methods combining automated brain morphometry and pattern recognition. The presented methods and results summarize the original work which has been done in developing new analytical methods for neuroimaging data mainly in the field of schizophrenia research. They have been published or at least submitted recently – in 2014, reporting results obtained on the same dataset or its subset. The three approaches differ in feature extraction, feature selection and classification for recognition of patients with first-episode schizophrenia from healthy volunteers. The features in the first approach (Schwarz and Kasperek, 2014) are extracted from the results of two different automated whole-brain morphometric methods – VBM and DBM – in a way that keeps in balance the need for data dimensionality reduction and the richness of information provided on brain morphology. The second approach to the analysis (Dluhoš et al., 2014) uses only the VBM results and reduces the data dimensionality thanks to its sparse representation in the wavelet domain. The third approach (Janoušová et al., 2015) is based on combinations of various classifiers employing machine learning on features selected by PCA performed on VBM and DBM results as well as on T1-weighted intensities.

3.1 SCHIZOPHRENIA PATIENTS AND HEALTHY CONTROLS

Fifty-two patients (mean age 24, SD 5.1 years) admitted to the all-male unit of the Department of Psychiatry, University Hospital Brno, for first episode of schizophrenia were recruited. The first episode schizophrenia patients are further referred to as FES.

Fifty-two healthy subjects (matched for age – mean age 24, SD 3.7 years - gender and handedness) were recruited from the community, the local staff, and medical students. The healthy subjects are referred to as the group of normal controls (NC).

A subset of 39 patients and the same number of controls took part in the previous study (Kasperek et al., 2011). One of the three studies presented here (Janoušová et al., 2015) comprises 49 patients and 49 healthy controls recruited from the same clinical workplace. The numbers of subjects differ due to the prospectivity in the study design.

3.2 IMAGE PRE-PROCESSING AND SPATIAL NORMALIZATION

The dataset contained 104 T1-weighted images of the entire head obtained with the 1.5 T MR device (sagittal tomographic plane thickness was 1.17 mm, the in-plane resolution was 0.48 mm x 0.48 mm, 3-D field of view contained $160 \times 512 \times 512$ voxels). GM tissue segments were obtained from all images after correction for bias-field inhomogeneity, spatial normalization and segmentation (Ashburner and Friston, 2005) with the use of VBM8 toolbox implemented in SPM8 software package for Matlab. Spatial normalization steps involved affine registration to standard SPM T1 template, followed by fast diffeomorphic registration algorithm.

GM tissue segments were modulated with the determinant of Jacobian matrices of the deformations to account for registration related changes in local volumes. The modulated GM segment images were finally smoothed with 8 mm FWHM Gaussian kernel to enable intersubject comparisons and to render data distribution more normal.

Spatial normalization steps for DBM included the same affine registration algorithm as for VBM. After transforming all bias-corrected images into stereotaxic space, the original high-dimensional deformable registration technique (Schwarz et al., 2007) was used to compute vector displacement fields which maximized the normalized mutual information between the images and the high-resolution single-subject template retrieved from the database of International Consortium for Brain Mapping (ICBM). The registration algorithm involved calculation of local forces in each voxel and their regularization with the use of modified Rogelj's elastic-incremental spatial deformation model. The resulting 3-D displacement vector fields were converted into scalar fields by computing their respective determinants of Jacobian matrices at each voxel of the stereotaxic space. After logarithmic transformation, the resulting scalar values are positive for local volume expansions caused by the deformation, negative for local volume contractions caused by the deformation or zero for no effect of the deformation.

3.3 APPROACH I: COMBINING FEATURES

The proposed MRI-based procedure for recognizing schizophrenia patients and for predicting the outcome of the acute treatment is shown in Figure 3.1. The image pre-processing phase includes the VBM and DBM methods, each employing a different registration algorithm. The SVM and k -NN classifiers are trained and tested on the features derived from each morphometric method separately as well as on the features extracted from significant results of both morphometric methods. The same recognition procedure is then applied on the problem of distinguishing between schizophrenia patients who responded well to the instituted acute treatment, from the patients who had no or poor improvement of the symptoms.

The modulated and smoothed images from VBM reflect spatial distribution of GM density. Voxel-wise two-sample t -tests of GM density means in FES and NC groups with the subject's age as single covariate are used to compute the distribution of Student's t -statistics in the stereotaxic space. The map of t -statistic is thresholded at the significance level of $p < 0.001$ without correction for multiple comparisons, consequent to which voxel clusters less than 100 mm^3 are filtered out. The resulting binary mask is used to select the classification features. These features, extracted with the use of VBM, in fact represent statistically significant reduction in GM density which might be interpreted, for a brain region localized by a cluster of significant voxels, as loss of GM tissue volume in FES patients when compared to NC subjects.

Scalar fields from DBM represent volume reduction of the voxels in which the Jacobian determinant is less than one, and volume expansion of the voxels in which the Jacobian determinant is greater than one. The data is log-transformed and normally distributed reductions (negative values) and expansions (positive values) are obtained. Statistical analysis, as in the case of VBM, is used to detect significant differences between the FES and NC groups. The resulting

binary mask is used to select the second set of features for classification. These DBM features reflect significant local volume changes in FES patients when compared to NC subjects.

Two well-known methods, namely k -nearest neighbours (k -NN) and SVM with linear kernel are compared on the above described dataset. SVM classifier was chosen because of its good performance reported in MRI data classification, see Chapter 2.3. The k -NN algorithm was chosen as a very simple and one of the most popular method in machine learning, which may outperform SVM in selected tasks (Colas and Brazdil, 2006). The same three training sets of feature vectors are used in both classification algorithms: (i) GM densities obtained from VBM, (ii) local volume changes obtained from DBM and, (iii) union of the two feature sets. The quality of classification was determined in a LOO-CV procedure on 104 subjects from FES and NC groups. For each experiment 103 subjects are used for training and the remaining one for testing. Sensitivity, specificity, precision, overall accuracy and error are assessed based on classification of testing subjects.

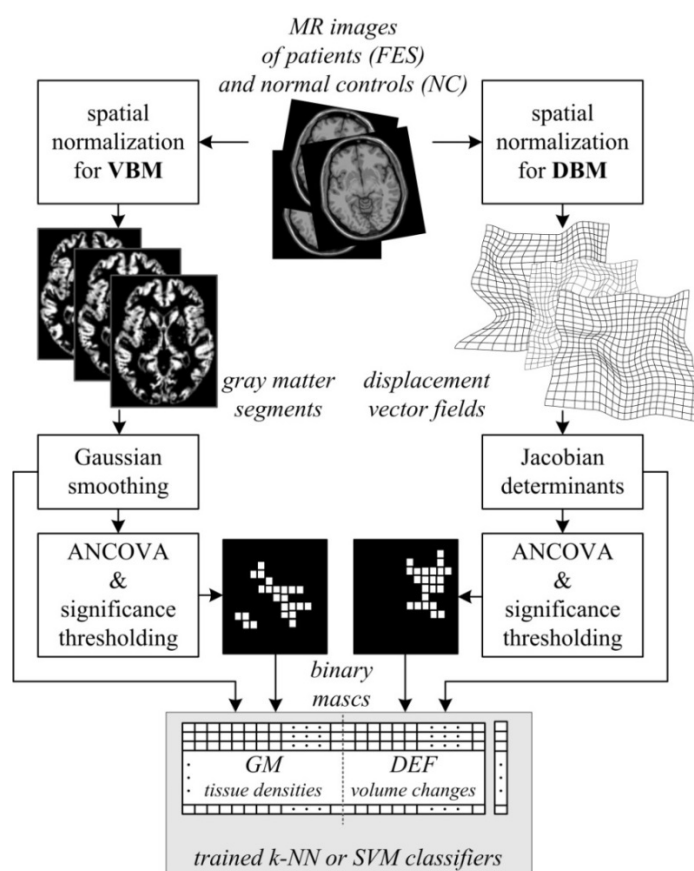


Figure 3.1: Training phase of the recognition algorithm. Feature vectors for classification of schizophrenia patients and normal controls are obtained by selecting the most statistically significant local volume changes and gray matter densities detected with VBM and DBM – both employing analysis of covariance (ANCOVA). The same features are further used for predicting reduction in schizophrenia symptoms. (Schwarz and Kasparek, 2014)

The same sets of features, but only for 52 FES patients, are further used for predicting the outcome of the acute treatment instituted for the first schizophrenia episode. PANSS is used for rating the symptoms of schizophrenia. A reduction in the initial PANSS total score is used here as cut-off to define two subgroups of FES patients. The group with higher percentage PANSS total score reduction contains FES patients who responded well to the treatment and the group with lower percentage PANSS score reduction contains FES patients with no or poor improvement. The

interval of all the PANSS score reduction values is sampled 50 times equidistantly and the true positive rate as well as the false positive rate are evaluated for each cut-off.

3.4 APPROACH II: WAVELET FEATURES

The proposed algorithm for recognizing schizophrenia patients from healthy subjects based on their structural MRI brain images consists of three main steps. Firstly, the images are transformed into a domain providing sparse representation. Secondly, the best discriminating features in the new domain are selected. And lastly, a supervised classifier is applied to the selected features. In this study, several variants for each step of the proposed classification scheme are implemented and then systematic experiments are performed, in order to find a setting showing the best classification results.

3.4.1 Sparsity and Wavelet transform

Generally, a signal is called sparse if most of its samples are equal to zero. Natural signals such as images are usually not sparse in the space domain. However, they can be often transformed into a suitable representation, in which they are sparse or at least weakly sparse in the sense that most of the coefficients in the new domain are almost zero (Starck et al., 2010). For natural images, one of the transforms producing such behavior is based on wavelets.

Wavelet transform decomposes a signal into a weighted sum of wavelets - functions of certain form (Misiti et al., 2007). This new representation captures not only the time course of the signal, but also its properties in the frequency domain. A small fraction of the representation coefficients with the highest magnitudes retains the major part of information contained in the signal. Moreover, it is usually the substantial part of the information, because noise tends to be contained mainly in the small coefficients (Misiti et al., 2007). For practical applications on discrete signals, the discrete wavelet transform (DWT) was developed, having originated in the Mallat's multiresolution decomposition scheme (Mallat, 1989). The signal is iteratively decomposed into detail and approximation coefficients by combination of two operations: (i) convolution with special finite response filters and (ii) subsampling. The approximation coefficients are then taken as input for a new level of decomposition. The output of this procedure is several sequences of coefficients describing details of the signal at different levels and one sequence of coefficients composing its rough approximation. DWT can be easily generalized into more dimensions. For 2-D images, one dimensional DWT is applied on the rows, columns and diagonals leading to three sets of detail coefficients for each level. Similarly, seven sets of detail coefficients are generated for 3-D images.

The number of wavelet coefficients approximately corresponds to the number of voxels in the transformed image, which is around 2 million in the GM density images computed from MRI data as described above. In order to reduce the noise contained in the data and to lower its dimensionality, the coefficients from all levels of DWT decomposition are sorted according to their maximum magnitude among all subjects and those below a certain threshold are removed. The optimal value of this threshold is one of the parameters which has to be determined experimentally, since it represents the trade-off between lower dimensionality and better noise reduction on the one hand and lesser losses of potentially useful information on the other hand. This operation leads to a reduction in the number of coefficients by the factor of 5-100, depending on the selected threshold. The remaining coefficients continue to the next steps as features describing the subjects.

Systematic optimization of the wavelet family and the level of decomposition used for DWT is not performed due to high computational demands of the experimental procedure. Based on the results of preliminary experiments, sym5 wavelet from the Symlet family and four levels of decomposition were chosen in the computations. This wavelet family has been shown to provide good results in natural image compression (Kumari and Vijay, 2012).

3.4.2 Supervised feature selection

After the feature extraction using DWT, a limited number of features with the best discriminative power are selected. Further reduction of the feature space dimensionality helps to match better the number of subjects in the dataset as well as to avoid the features carrying only low information about the differences between the studied groups. Several criteria for determining the discriminative power of individual features are examined, while testing the effect of the number of the best features selected for subsequent classification. The studied criteria taken from the literature are:

Fisher's discriminant ratio (Bishop, 2006):

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}, \quad (3.1)$$

Bhattacharyya distance (Kailath, 1967):

$$Bha = \frac{1}{4} \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} + \frac{1}{2} \ln \left(\frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1\sigma_2} \right), \quad (3.2)$$

and divergence (Wang et al., 2011):

$$Div = \frac{1}{2} \left(\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2} - 2 \right) + (\mu_1 - \mu_2)^2 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right). \quad (3.3)$$

For each feature, μ_1 and μ_2 represent the mean values of this feature in the first and second group and σ_1 and σ_2 represent the variances of the feature values in each group. Apart from the criteria $FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$, (3.1)- $Div = \frac{1}{2} \left(\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2} - 2 \right) + (\mu_1 - \mu_2)^2 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)$. (3.3), three others are proposed and tested. Two of them are modifications of FDR designed for better robustness in case of non-normally distributed data:

$$medFDR = \frac{(med_1 - med_2)^2}{\sigma_1^2 + \sigma_2^2}, \quad (3.4)$$

$$quantileFDR = \frac{(med_1 - med_2)^2}{(\sigma_1^*)^2 + (\sigma_2^*)^2}, \quad (3.5)$$

where med_1 and med_2 are medians and σ_1^* and σ_2^* are estimates of standard deviations by interquartile range $\sigma^* = Q_{84} - Q_{16}$. The last criterion designed is:

$$variances = \frac{\Sigma^2}{\sigma_1^2 + \sigma_2^2}, \quad (3.6)$$

where Σ^2 stands for the variance of the tested feature among all subjects. High values of variances are expected for the features, which show a high variance in the whole dataset and are homogenous inside the studied groups at the same time.

Features extracted and selected in the previous steps are used for training the SVM. Three implementations of the SVM classifier from the PRTools¹ toolbox for MATLAB are tested. They

¹ PRTools is a Matlab toolbox for pattern recognition: <http://prtools.org>.

differ in kernel functions (linear: SVC, NuSVC and radial basis functions: RBSVC are used²) and in the regularization method. The entire procedure of predicting a class for a new subject is illustrated in Figure 3.2 and works as follows: MRI image is preprocessed as described above, the computed images representing GM density is transformed into the wavelet domain. The coefficients with magnitude under a threshold, computed over the whole data set, are removed. Then the most discriminative coefficients are selected as features for the classification. The discrimination criteria computations and training of the classifier are performed only over the training subjects. The class for the unknown subject is finally predicted using its values of the selected features.

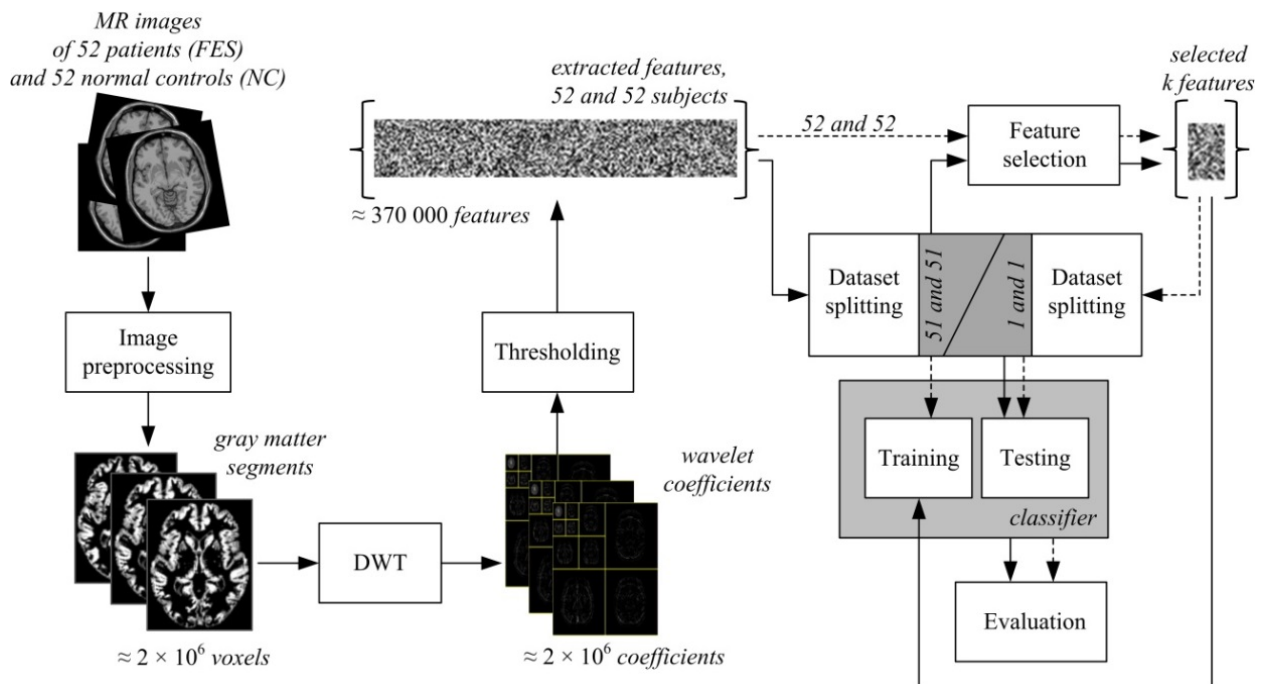


Figure 3.2: The scheme of the proposed recognition algorithm with correct (solid lines) and incorrect (dotted lines) CV. MRI images are pre-processed and the gray matter is segmented. The resulting images are transformed into DWT coefficients. Only the coefficients with magnitude greater than a chosen threshold are extracted as potential features. In case of correct CV, the dataset is divided into testing and training subsets and further steps are performed repeatedly with the subjects in the training subset only. A limited number of the most discriminant features is selected and used for training a classifier. The performance of the classification is tested on subjects in the testing subset. Incorrectness of the dotted variant lies in the reversed order of dataset splitting and feature selection, as the feature selection step relies on the information about subjects in the testing subset. (Dluhoš et al., 2014)

If the most discriminative features were computed on the whole data set, see the dotted path in the Figure 3.2, the results would be biased towards more correct classification. Separation into training and testing sets has to be done just prior to selecting the most discriminative features. This way of CV corresponds to a real application of the algorithm for prediction. The stratified 52-fold CV is used rather than the more frequent LOO-CV approach, in order to avoid possible bias caused by uneven proportions of subjects from different classes in the training and testing subsets

² SVC and Nu-SVC are mathematically equivalent, but vary in the implementation of SVM with the linear kernel (Nu-SVC uses a parameter to control the number of support vectors.). RBSVC is an implementation of SVM with RBF as the kernel.

(Kohavi, 1995). In one inner loop of the stratified 52-fold CV, the number of the most discriminative features k is determined, with the use of three different algorithms: (i) nested CV with a 3-, 17- and 51-folds; (ii) bootstrap with testing subset sizes of 2, 6, 10 and 34 subjects; (iii) voting of independent classifiers trained progressively on the first 1, 2, ..., k features, $k = 20, 100, 1000$.

3.5 APPROACH III: COMBINING FEATURES AND CLASSIFIERS

Three classifiers: mMLDA, the centroid method, and the average linkage method are used to solve a task of distinguishing FES patients from NC subjects based on their MRI brain data preprocessed in three different ways and reduced using PCA. Three types of features are used: (i) MR intensities in each voxel, (ii) GM densities obtained by VBM, and (iii) local volume changes obtained by DBM.

There are four hypotheses: (i) mMLDA yields an improved classification rate compared with MLDA; (ii) classifiers based on features extracted using automated and more computation-intensive image preprocessing of the MRI data perform better; (iii) keeping less than $N - 1$ eigenvectors during data reduction leads to better classification results pursuant to the removed noise; (iv) combining classifiers using the majority vote leads to higher classification performance than in the case of using single classifiers.

3.5.1 Intersubject PCA for reduction of feature sets

Removal of the background and non-brain voxels in the preprocessing phase reduces the number of features in each of the three feature sets from $7 \cdot 10^6$ million to $2 \cdot 10^6$. However, this number is still very high and leads to the problems mentioned in Chapter 2.4.2, related to curse of dimensionality. To avoid the small sample-size problem, PCA can be used for feature reduction. It is a multivariate data-reduction technique which is based on the assumption that there are correlations among the features in large datasets, rendering part of the data redundant (Wallisch, 2014). The goal of PCA is to decrease data dimensionality while keeping as much original data variability as possible via transformation of the original features to new uncorrelated features, called as the principal components. If the original variables are correlated, it is sufficient to choose only a few principal components to preserve most of the original variability. The PCA algorithm is based on calculation of eigenvalues and eigenvectors of a covariance matrix of features (Jolliffe, 2002). However, as the number of imaging features is equal to $2 \cdot 10^6$, the covariance matrix of features would be of size $2 \cdot 10^6 \times 2 \cdot 10^6$ which leads to enormous computer memory requirements. Thus, modification of PCA based on intersubject covariance matrix (Demirci et al., 2008; Thomaz et al., 2007a; Wang et al., 2006) is used for image data reduction here and is termed as intersubject PCA (isPCA) because eigenvectors of the covariance matrix of features are computed using transformation of the eigenvectors of the intersubject covariance matrix. Thus, calculation of the huge covariance matrix of features is avoided.

An input into the isPCA is the imaging data matrix \mathbf{X} of size of $N \times n$, where N is the number of subjects and n is the number of features (each of the three types of imaging features is processed separately). According to the linear algebra rules, nonzero eigenvalues of covariance matrix of features $\mathbf{X}^T \mathbf{X}$ and covariance matrix of subjects (intersubject covariance matrix) $\mathbf{X} \mathbf{X}^T$ are the same and eigenvectors corresponding to the higher dimensional covariance matrix can be derived from the eigenvectors of the smaller one by:

$$\mathbf{v}_j = \frac{\mathbf{X}^T \mathbf{w}_j}{\sqrt{q_j(N-1)}}, \quad (3.7)$$

where \mathbf{v}_j and \mathbf{w}_j are the j^{th} eigenvector of the covariance matrix of features and the covariance matrix of subjects, respectively, \mathbf{X}^T is the transposed imaging data matrix, and q_j is the j^{th}

eigenvalue of the covariance matrix of subjects. The proof of the transformation can be found in (Fukunaga, 1990) and (Demirci et al., 2008). The entire isPCA algorithm can be described in the following steps:

1. Calculate $N \times N$ covariance matrix of subjects \mathbf{C}_S of data matrix \mathbf{X} by $\mathbf{C}_S = (1/(N-1))(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T$, where $\bar{\mathbf{X}}$ is the matrix with all rows equal to the mean image $\bar{\mathbf{x}}$ which is defined by $\bar{\mathbf{x}} = (1/N)\sum_{i=1}^N \mathbf{x}_i$, where $\mathbf{x}_i, i = 1, \dots, N$ are rows of matrix \mathbf{X} .
2. Find eigenvalues q_j and normalized eigenvectors $\mathbf{w}_j, j = 1, \dots, N$, of \mathbf{C}_S .
3. Select all $m = N - 1$ eigenvectors that correspond to all nonzero eigenvalues.
4. Compute eigenvectors $\mathbf{v}_j, j = 1, \dots, m$, of the covariance matrix of features by
$$\mathbf{v}_j = \frac{\mathbf{X}^T \mathbf{w}_j}{\sqrt{q_j(N-1)}}, \quad (3.7).$$
5. Construct $n \times m$ projection matrix \mathbf{V}_{isPCA} with column-wise computed eigenvectors $\mathbf{v}_j, j = 1, \dots, m$.
6. Compute the reduced data matrix \mathbf{Y} with the size $n \times m$ by $\mathbf{Y} = (\mathbf{X} - \bar{\mathbf{X}}) \cdot \mathbf{V}_{isPCA}$.

The isPCA allows using all $N - 1$ eigenvectors with non-zero eigenvalues for data reduction, enabling the preservation of all dataset variance; thus, it allows one to maintain the whole signal important for classification. It is not clear, however, if keeping less than $N - 1$ principal components would not lead to better classification results pursuant to the removed noise. So, an experiment is performed, in which eigenvectors are removed systematically corresponding to the smallest eigenvalues from the matrix \mathbf{V}_{isPCA} as well as eigenvectors corresponding to the largest ones during data reduction.

To further examine the isPCA results, features whose loadings contribute most to the principal components can be extracted and visualized in the image space. The loadings can be easily calculated using correlation of the desired principal component (i.e. the desired column of the matrix \mathbf{Y}) with all original feature vectors (i.e. all columns of the matrix \mathbf{X}). For visualization, the loadings are thresholded at 30% of their maximum absolute value.

3.5.2 Classification

The first of the three classifier used in the recognition algorithm is modified MLDA (mMLDA). The original MLDA is based on the maximum entropy covariance selection method. With the use of MLDA, all features are reduced into one number, a discriminative score. An image of a new subject is then classified into one of the groups depending on whether the discriminative score falls above or below the boundary computed as an arithmetic mean of the mean discriminative scores of both groups of subjects. It is expected, however, that if the variability of the groups is incorporated while calculating the boundary, similarly to the authors of (Culhane et al., 2002), who used a mean weighted by standard deviations of discriminative scores of the groups as the threshold in microarray data analysis, the classification performance can be improved.

Description of mMLDA requires explanation of the MLDA algorithm firstly. All MLDA algorithm steps are fully described elsewhere (Fujita et al., 2008; Thomaz et al., 2007b) and can be briefly summarized as follows.

1. Let the within-class scatter matrix \mathbf{S}_w be defined as $\mathbf{S}_w = \sum_{i=1}^g \sum_{j=1}^{N_i} (\mathbf{y}_{i,j} - \bar{\mathbf{y}}_i)(\mathbf{y}_{i,j} - \bar{\mathbf{y}}_i)^T$ and the between-class scatter matrix \mathbf{S}_b be defined as $\mathbf{S}_b = \sum_{i=1}^g N_i (\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^T$, where g is the total number of groups (here $g = 2$), vector $\mathbf{y}_{i,j}$ is j^{th} subject from the group π_i described by reduced m -dimensional feature vector, N_i is the number of training subjects

from group π_i , vector \bar{y}_i is a sample mean of group π_i (i.e. mean subject from group π_i) and \bar{y} is the overall mean vector.

2. Find eigenvalues λ_j and eigenvectors $\Phi_j, j=1, \dots, m$, of matrix S_p , where $S_p = S_w / (N - g)$.
3. Calculate the average eigenvalue $\bar{\lambda}$ of matrix S_p by $\bar{\lambda} = \text{tr}(S_p) / m$, where $\text{tr}(S_p)$ is a trace of matrix S_p .
4. Construct a new matrix of eigenvalues based on the following largest dispersion criterion $\Lambda^* = \text{diag}[\max(\lambda_1, \bar{\lambda}), \dots, \max(\lambda_m, \bar{\lambda})]$.
5. Form the modified within-class scatter matrix S_w^* by $S_w^* = (\Phi \Lambda^* \Phi^T)(N - g)$, where Φ is a matrix with column-wise computed eigenvectors $\Phi_j, j = 1, \dots, m$.
6. Calculate the projection matrix V_{MLDA} with column-wise eigenvectors of matrix S , where $S = S_w^{*-1} S_b$; the projection matrix V_{MLDA} maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix (Fishers' criterion).
7. Multiply the reduced matrix Y by the projection matrix V_{MLDA} to compute matrix Z with the size $N \times (g - 1)$ containing discriminative scores by $Z = Y \cdot V_{MLDA}$.

As $g = 2$ groups are classified here, Z has the size $N \times 1$ and so every input vector of imaging features from one subject is now represented just by one discriminative score. Each of the two groups (FES and NC) can be now represented by a mean discriminative score of subjects from the corresponding group. In MLDA, the boundary between the two groups is computed using an arithmetic mean of the mean discriminative group scores. Here, however, the following formula for the weighted mean is used to calculate the boundary:

$$z_{mMLDA} = \frac{\frac{\bar{z}_1}{SD_1} + \frac{\bar{z}_2}{SD_2}}{\frac{1}{SD_1} + \frac{1}{SD_2}} = \frac{\bar{z}_1 SD_2 + \bar{z}_2 SD_1}{SD_1 + SD_2}, \quad (3.8)$$

where \bar{z}_1 is the mean discriminative score of FES patients, \bar{z}_2 is the mean discriminative score of NC subjects, and SD_1 and SD_2 are the standard deviations of discriminative scores of the group of

$$z_{mMLDA} = \frac{\frac{\bar{z}_1}{SD_1} + \frac{\bar{z}_2}{SD_2}}{\frac{1}{SD_1} + \frac{1}{SD_2}} = \frac{\bar{z}_1 SD_2 + \bar{z}_2 SD_1}{SD_1 + SD_2}$$

patients and controls, respectively. It is evident from

(3.8) that smaller standard deviation in a group indicates higher weight of the mean discriminative score of the group and thus the classification boundary gets shifted towards that group. Expect a lower variance in the NC group, one can anticipate that the classification boundary will be shifted closer to the group of controls and thus the sensitivity of mMLDA will be higher in comparison with MLDA. After the classification boundary estimation, a subject who is supposed to be classified based on the feature vector x is assigned to one of the groups depending on whether the discriminative score $z = (x - \bar{x}) \cdot V_{isPCA} \cdot V_{MLDA}$ falls above or below the classification boundary z_{mMLDA} .

Due to classification of the $g = 2$ groups, V_{MLDA} has also the size $N \times 1$. This projection vector can be mapped back to the image space by multiplication with the isPCA projection matrix V_{isPCA} . The resulting map then shows to what extent each voxel contributes to the discrimination of the

two groups of subjects. Such map with discriminative coefficients of the voxels can be thresholded to show just the most discriminative voxels.

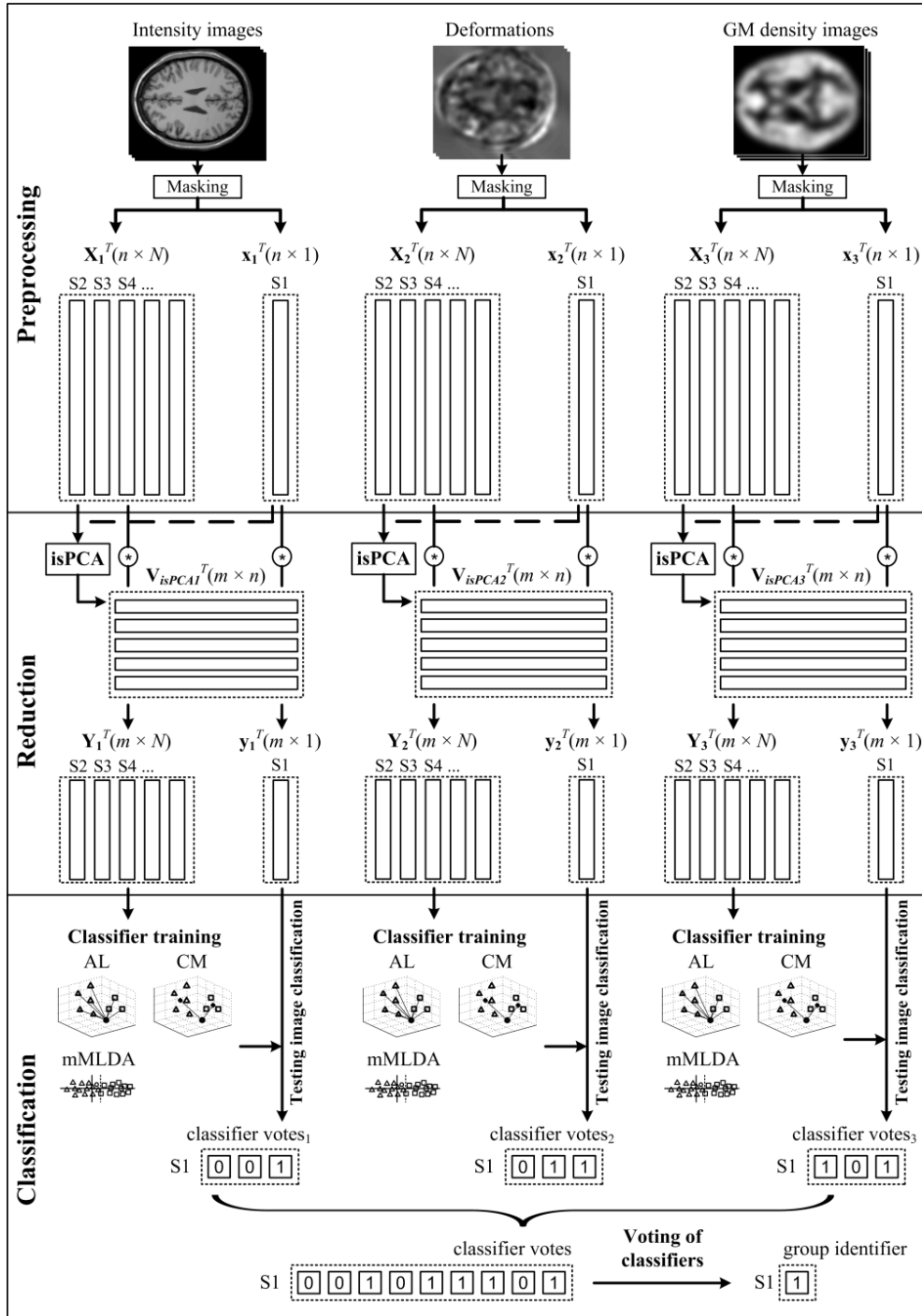


Figure 3.3: The scheme of the recognition algorithm. Three sets of features are obtained from the MR intensity images, the deformations and the GM density images, from which the background and non-brain areas are removed using a binary mask. The datasets of features are split into training and testing subsets, and stored as 1-D vectors in an imaging data matrix \mathbf{X} and a vector \mathbf{x} , respectively. The training matrix \mathbf{X} is used for calculation of a projection matrix \mathbf{V}_{isPCA} using *isPCA*, reduced and then serves for training *mMLDA*, *CM* and *AL* classifiers. The testing vector \mathbf{x} is reduced using the projection matrix \mathbf{V}_{isPCA} and classified into the FES group (ones) or NC group (zeros). The votes of all classifiers for all types of imaging features serve as input into the voting algorithm with the result of the majority group identifier for each subject. The whole

process is repeated N -times as each image is chosen for one test in LOO-CV. The scheme also shows an incorrect implementation of LOO-CV (the dashed line) – with all images involved in the computation of the projection matrix $\mathbf{V}_{\text{isPCA}}$ (Janoušová et al., 2015).

The centroid method (CM) is used as the second classifier. It is one of the clustering methods where subjects are classified into groups according to distances in the feature space. Here, the number of imaging features was reduced by isPCA before; so the feature space has m dimensions. Distances of a new subject from centroids of the groups are computed and the subject is then classified into the group represented by the nearest centroid (Legendre and Legendre, 1998).

The third classifier used here is the average linkage method (AL), which seeks for the shorter of the average distances of the new subject from all images of the FES group and NC group (Legendre and Legendre, 1998).

The three classification algorithms mMLDA, CM and AL are used for classifying the three types of features which are extracted from MRI data and then reduced using isPCA. The resulting nine classifiers are further combined to achieve more robust and possibly also more accurate classification performance. Combinations of odd numbers of classifiers are performed to avoid ties. The entire classification scheme which includes the LOO-CV technique to avoid biased results is depicted in Figure 3.3. Sequentially, each of the N subjects is selected as a testing subject and the sets of features of the remaining $N - 1$ training subjects are reduced using isPCA and then used for training the classifier. The features of the testing subject is reduced using isPCA eigenvectors calculated during reduction of the training feature sets and classified into the FES or NC class. In case of classifier combination, a testing subject is assigned to the group of FES or NC according to the majority vote of classifiers. Then, the resulting class is compared to the true classification label. Finally, the classification performances for all subjects are put together in order to create the overall classification performance measures, namely accuracy, sensitivity and specificity. Comparison of the classification results with the classification by chance is performed using one-tailed one sample binomial test.

4 RESULTS

The first approach to feature selection was based on standard statistical parametric maps resulting from mass-univariate statistical tests involved in the automated brain morphometry methods. VBM selected 14,700 significant GM densities and DBM selected local volume changes in 11,568 significant voxels. The united feature set was represented by the total number of 26,268 features. The description of each subject was reduced to 1.36 % of the data when compared to all 1,924,670 voxels covered by the brain tissues of one subject in the stereotaxic space.

Table 4.1: Comparison of selected classification results obtained with the use of gray matter density features and with features based on local volume changes. LOO-CV procedure was executed on 104 subjects from FES and NC classes (Schwarz and Kasperek, 2014).

Features and classifier		Accuracy [%]	Sensitivity [%]	Specificity [%]	Precision [%]	Error [%]
GM	SVM linear	65.4	65.4	65.4	65.4	34.6
	11-NN euclidean	74.0	76.9	71.2	72.7	26.0
	21-NN euclidean	77.9	84.6	71.2	74.6	22.1
DEF	SVM linear	76.0	73.1	78.8	77.6	24.0
	11-NN euclidean	82.7	78.8	86.5	85.4	17.3
	21-NN euclidean	84.6	76.9	92.3	90.9	15.4
GM \cup DEF	SVM linear	83.7	84.6	82.7	83.0	16.3
	11-NN euclidean	87.5	88.5	86.5	86.8	12.5
	21-NN euclidean	86.5	84.6	88.5	88.0	13.5

DEF = local volume changes caused by deformations (DBM results); GM = gray matter densities (VBM results); SVM = support vector machines; k -NN = k -nearest neighbour.

Complete NC-FES classification results including sensitivity, specificity, precision and error in selected classifiers only are shown in Table 4.1.

In the prediction experiment, the interval of the PANSS score reduction in 52 FES subjects acquired values of $TP_{change} \in < 0.990; +0.097 >$. The true positive rates and the false positive rates of the SVM and 11-NN classifiers were computed for all TP_{change} cut-offs and for all three cases – with the use of both sets of features individually as well as with their union – see Figure 4.1.

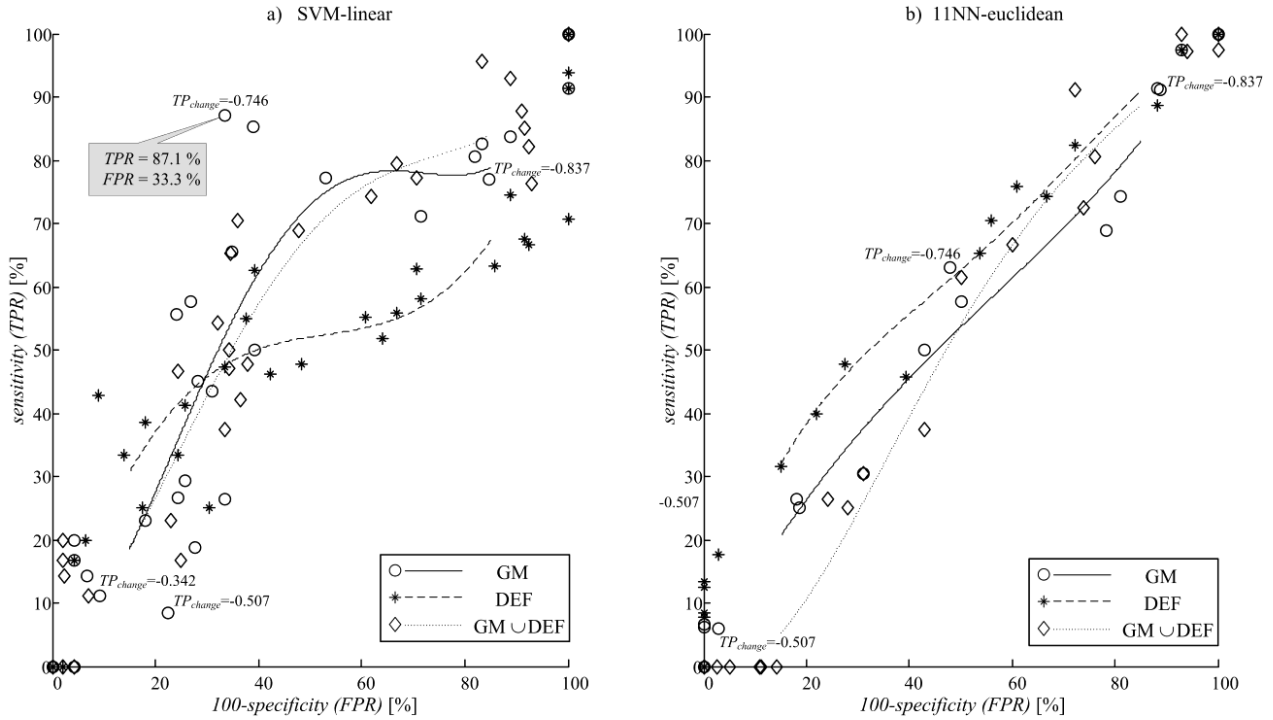


Figure 4.1: The false positive rates (FPR) and the true positive rates (TPR) of predictions with a) SVM classifier and b) k -NN classifier with 11 neighbours. The rates were obtained for three different sets of features and for various cut-offs in reduction in PANSS total score. The TP_{change} cut offs were used to divide 52 FES subjects into two groups. The fitted curves are shown only for better orientation in the x - y plots (Schwarz and Kasparek, 2014).

With the second approach to classification based on GM densities as features transformed to the wavelet domain, several series of experiments were performed, in order to find the best parameters for each step of the whole classification algorithm.

The parameters were: (i) usage of approximation coefficients {YES, NO}, (ii) the threshold for removing small magnitude coefficients {0, 0.001, 0.01, 0.1, 1}, (iii) the criterion for evaluating the discriminative power of the features, (iv) the algorithm for choosing the number of selected features {nested CV; bootstrap selection; voting of independent classifiers}, and (v) the SVM classifier implementation {SVC, NuSVC, RBSVC}. The quality of classification for each parameter setting was evaluated using stratified 52-fold CV on the whole dataset. All runs were repeated one hundred times to improve robustness of the estimates. As testing all 1,800 possible combinations of the setting parameters would not be computationally feasible, a default setting was chosen and the effects of the parameters were tested in one-at-a-time manner. The default setting was: using the approximation coefficients, removing coefficients with values below 0.01, FDR criterion, voting of classifiers with 1-1000 best features and the SVC implementation of the SVM classifier. The results of the experiments are summarized in Table 4.2.

Table 4.2: Effects of different settings on the accuracy, sensitivity and specificity of the classification algorithm based on supervised learning in the wavelet domain. Best values for each parameter are highlighted in bold. Each value was estimated by averaging the results of 100 independent runs of stratified 52-fold CV on the whole dataset. Both SVC and NuSVC implementations train a SVM classifier with linear kernel and differ only in the regularization method. The RBSVC implementation searches optimal kernel in the form of radial basis functions by nested CV (Dluhoš et al., 2014).

Parameter	Value	Sensitivity [%]	Specificity [%]	Accuracy [%]
approximation coefficients	YES	71.1	74.8	72.9
	NO	71.6	72.7	71.9
thresholding coefficients	0	64.2	67.8	66.0
	0.001	64.7	69.9	67.3
	0.01	71.1	74.8	72.9
	0.1	59.4	66.7	63.0
	1	59.5	64.9	62.2
discrimination criterion	FDR	71.1	74.8	72.9
	Bhattacharyya	71.3	74.8	73.1
	Divergence	65.2	72.1	68.7
	medFDR	57.8	58.3	58.1
	quantileFDR	53.8	58.2	56.0
	Variances	71.7	74.7	73.2
number of features	51-fold	56.2	52.5	54.3
	17-fold	53.0	54.1	53.6
	3-fold	53.0	56.9	55.0
	bootstrap-2	58.0	48.2	53.1
	bootstrap-6	54.7	52.1	53.4
	bootstrap-10	52.9	54.1	53.5
	bootstrap-34	51.5	56.4	53.8
	voting 1-20	57.5	66.3	61.9
	voting 1-100	60.2	64.3	62.3
	voting 1-1000	71.1	74.8	72.9
SVM classifier implementation	SVC	71.1	74.8	72.9
	NuSVC	62.9	74.0	68.4
	RBSVC	64.2	75.0	69.6

Confirmation of the four hypotheses related to the third approach to recognition of FES patients from NC subjects required performing four experiments. Firstly, the three types of imaging features reduced by isPCA were classified using mMLDA and the results were compared with the classification using MLDA. Table 4.3 shows that mMLDA enabled classification with slightly higher accuracy than MLDA. As expected, the sensitivity of mMLDA was higher than of MLDA as the classification boundary calculated as weighted mean was shifted closer to the group of controls in case of mMLDA, see Figure 4.2.

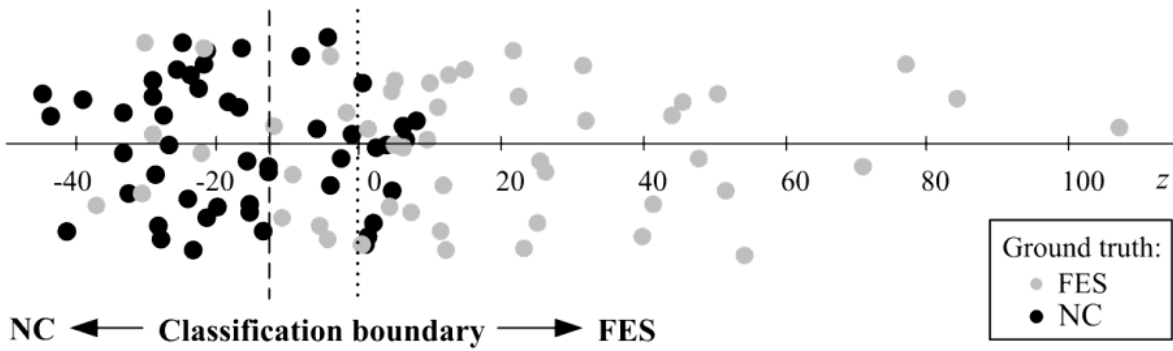


Figure 4.2: Visualization of MLDA and mMLDA used for recognition of FES patients from NC subjects. Discriminative scores of FES and NC are depicted by grey and black dots, respectively. The dotted and dashed lines represent the classification boundary of MLDA and mMLDA, respectively. The dots to the left of a boundary are classified as NC and the dots placed on the right side of a boundary are classified as FES (Janoušová et al., 2015).

The difference between sensitivity of mMLDA and MLDA was 14.3 % for local volume changes, 22.5 % for MR intensities and only 2.0 % for GM densities. Furthermore, Table 4.3 reveals that the accuracy of classification based on the MR intensities was much lower than in case of the features extracted by automated brain morphometry methods – VBM and DBM. Further, comparison of the performance of the classification based on all three types of imaging features – reduced using isPCA – and based on the three classification methods (mMLDA, CM and AL) was accomplished. The results are given in Table 4.4

Table 4.3: Comparison of classification efficiency of MLDA and mMLDA based on various types of features extracted from MR images of FES patients and NC subjects (Janoušová et al., 2015).

Features and classifier		Accuracy [%]	Sensitivity [%]	Specificity [%]
DEF	MLDA	75.5	71.4	79.6
	mMLDA	76.5	85.7	67.3
INT	MLDA	62.2	57.1	67.3
	mMLDA	65.3	79.6	51.0
GM	MLDA	77.6	79.6	75.5
	mMLDA	78.6	81.6	75.5

DEF = local volume changes caused by deformations (DBM results); GM = gray matter densities (VBM results); INT = MR intensities; MLDA = maximum uncertainty linear discriminant analysis; mMLDA = modified MLDA.

Regarding the types of imaging features, the best classification results were achieved in classification based on the local volume changes and the worst results when using the MR intensities. The comparison of the classification methods reveals that the AL method led to very low sensitivity of classification with all feature types and to the lowest accuracy in case of local volume changes and MR intensities and thus seemed least appropriate for the classification. The highest sensitivity (85.7 %) was obtained in classification with the local volume changes using mMLDA. The highest accuracy (80.6 %) was achieved in classification with the local volume changes using CM. As CM seems to have high overall performance, mMLDA high sensitivity and AL high specificity and it is intended to use as much of information captured in the imaging data as possible, a simple ensemble strategy was tried and assessed whether voting of the classifiers could further improve the classification performance.

Table 4.4: Classification performance with various classifiers and feature types (Janoušová et al., 2015).

Features and classifier		Accuracy [%]	Sensitivity [%]	Specificity [%]
DEF	mMLDA	76.5	85.7	67.3
	CM	80.6	71.4	89.8
	AL	67.3	34.7	100.0
INT	mMLDA	65.3	79.6	51.0
	CM	61.2	59.2	63.3
	AL	54.1	18.4	89.8
GM	mMLDA	78.6	81.6	75.5
	CM	64.3	63.3	65.3
	AL	69.4	49.0	89.8

DEF = local volume changes caused by deformations (DBM results); GM = gray matter densities (VBM results); INT = MR intensities; mMLDA = modified maximum uncertainty linear discriminant analysis; CM = centroid method; AL = average linkage.

All possible combinations of odd numbers of classifiers were performed and summarized. The best 20 combinations achieving highest classification performance are displayed in Table 4.5. Combining the classifiers improved classification results only very slightly compared to the results of single classifiers (Table 4.4). The highest accuracy (81.6 %) was achieved in classification with the use of combination of five classifiers (mMLDA, CM and AL with the local volume changes, mMLDA with the MR intensities, and mMLDA with the GM densities). This classification performance was statistically significantly higher than the classification by chance ($p < 0.001$, one sample binomial test, one-tailed).

5 DISCUSSION

The first part of this section discusses the results and findings which have been demonstrated in the Chapter 4 and points out the limitations of the study with three different algorithms for recognition of schizophrenia patients from healthy volunteers. The second part then describes the usability of this work and lessons learned for the pedagogical practice.

5.1 CLASSIFICATION OF FIRST EPISODE SCHIZOPHRENIA

Schizophrenia causes considerable adverse socioeconomic effects resulting from necessity to provide a long-term social help and from losing economic productivity in working-aged subjects. The possibility to uncover potentially high-risk patients around the time of the illness onset may enable searching for preventive and therapeutic strategies, which would be able to eliminate or minimize the adverse effects of the disease. Studying the first-episode populations in the schizophrenia research brings the advantage of controlling many potential confounding factors such as long term medication, possible progression of morphological changes or unfavorable course of the illness. Thus, the changes present at the beginning of the illness may reflect its primary pathology. On the other hand, the changes in the brain morphology are only subtle during the first episode, and hence often undistinguishable even by an experienced psychiatrist.

Usefulness of the whole-brain automated morphometry methods in research on the neurobiology of schizophrenia and other neuropsychiatric diseases has been demonstrated repeatedly by many authors – see the survey in Chapter 2.2. The goal of the methods and approaches presented here is not to uncover specific regions with significant differences in brain

morphology and neither was the goal to compare the findings made using VBM and DBM, as it was explored in (Schwarz and Kasperek, 2011).

Table 4.5: The top 20 combinations of 3 classifiers and 3 types of imaging features in classification of FES patients and NC subjects. The results are sorted in descending order of accuracy. The meaning of the symbols in the cells: ● the best classification result with higher accuracy than the best single classifier (CM-VOL); ■ the results with the same accuracy as the best individual classifier but having higher sensitivity; ◆ the results with the same or worse performance than the best individual classifier (Janoušová et al., 2015).

order	mMLDA DEF	CM DEF	AL DEF	mMLDA INT	CM INT	AL INT	mMLDA GM	CM GM	AL GM	Accuracy [%]	Sensitivity [%]	Specificity [%]
1	●	●	●	●			●			81.6	75.5	87.8
2	■	■					■			80.6	77.6	83.7
3	■	■						■		80.6	77.6	83.7
4	■	■		■			■		■	80.6	77.6	83.7
5	■	■							■	80.6	75.5	85.7
6	■		■				■			80.6	75.5	85.7
7	■	■	■				■		■	80.6	73.5	87.8
8	■	■	■	■					■	80.6	73.5	87.8
9		◆								80.6	71.4	89.8
10	◆	◆	◆							80.6	71.4	89.8
11		◆	◆				◆			80.6	69.4	91.8
12	◆	◆	◆			◆	◆			80.6	69.4	91.8
13	◆	◆		◆	◆		◆			79.6	81.6	77.6
14	◆		◆	◆			◆		◆	79.6	75.5	83.7
15	◆	◆		◆		◆	◆			79.6	75.5	83.7
16	◆	◆	◆	◆				◆		79.6	75.5	83.7
17	◆	◆	◆				◆	◆		79.6	73.5	85.7
18	◆	◆	◆	◆		◆	◆		◆	79.6	71.4	87.8
19		◆	◆	◆						79.6	67.3	91.8
20	◆	◆	◆			◆			◆	79.6	61.2	98.0

DEF = local volume changes caused by deformations (DBM results); GM = gray matter densities (VBM results); INT = MR intensities; mMLDA = modified maximum uncertainty linear discriminant analysis; CM = centroid method; AL = average linkage.

Only male patients were included to the study; this may decrease the power to generalize the results to all schizophrenia patients. On the other hand, the intention was to have a homogeneous data set as there are gender differences in structural brain changes in schizophrenia. Accordingly, analyses were performed for males and females subjects separately in several schizophrenia classification studies (Demirci et al., 2008; Fan et al., 2007; Nakamura et al., 2004; Takayanagi et al., 2010).

The advantage of all three presented approaches is that the recognition algorithms are based on imaging features extracted from the results of whole-brain morphometry methods. Many other previously published works used only selected parameters related to anatomical structures. Focusing on selected structures of the brain may increase the classification power in a small sample size setting, but tends to limit the clinical understanding of a complex behavior inherent in brain disorders.

The presented algorithms for recognition schizophrenia patients from normal controls are similarly to (Fan et al., 2008, 2007) based on a combination of automated whole-brain morphometric methods and pattern recognition algorithms. VBM for all three brain tissues was employed in (Fan et al., 2008, 2007) and local volume changes were then characterized by three variables reflecting density of the tissues. The method to extract local volume changes used here is based on Jacobian determinants computed directly from the spatial transformations resulting from the high-dimensional deformable registration. The analyzed parameter (change of local volume) has a clear biological meaning. On the other hand, in VBM the meaning of tissue density is much less evident.

5.1.1 Approach I: Combining features

Table 4.1 showed that recognition based on the united set of features selected from VBM and DBM provided better results – sensitivity 88.5 % and specificity 86.5% – than in the case of using only features selected with one morphometric method. Therefore, the more information on brain morphology was used for classification purposes, the better was the classifier performance.

One might assume that these well discriminating features for FES-NC recognition selected on the basis of group-level analyses are closely linked to the pathology of schizophrenia, and thus might well discriminate the groups of patients with different treatment outcomes. This assumption was confirmed only in the case of features selected by VBM, with which the linear-kernel SVM classifier predicted 74.6 % reduction in PANSS total score with sensitivity and specificity of 87.1 % and 66.7 % respectively. Given that such large reduction occurred in 21 out of 52 FES patients in the investigated group, the result might be considered as worthy of note; although clinical studies focused on evaluating the effect of treatment operate with PANSS total score reduction cut-off of 30 % and 50 % (Leucht et al., 2007). Surprisingly, the recognition procedure based only on the local volume changes, which seemed to discriminate FES and NC groups almost as well as the union of the two feature sets, gave the worst results in the prediction of treatment outcome.

The features were selected only based on statistical significance thresholding. This is a very naive approach, which might be, however, less prone to overfitting than selection of features with the highest discrimination power derived from correlation with the classification outcome as in (Fan et al., 2008, 2007).

The main limitation of the presented approach lies in the bias of the CV scheme – the model behind selection of the features (i.e. the t-test with one covariate) was not built from scratch for each different test subject or fold. Hence, the reported accuracy was overestimated – similarly to the case of (Fan et al., 2008, 2007) and other early attempts for pattern recognition in neuroimaging data. For a defense of the first naive approach, one might argue that the features were selected to be used in the subsequent prediction experiment – as it is usually done in validation schemes based on independent datasets.

The other two approaches presented below use also supervised learning to select the most discriminative features. Their designs, however, includes always a proper CV scheme which prevent overestimated accuracy.

5.1.2 Approach II: Wavelet features

The results in Table 4.2 show that almost all parameters in the algorithm based on GM features transformed to the wavelet domain had a significant impact on the quality of classification. In the first step, the best results were achieved with using only the detail coefficients with values above 0.01. Looser thresholds left too much noise in the data while the more stringent ones probably removed information necessary for discriminating between FES and NC subjects. In the second step, the criteria FDR, Bhattacharyya and variances performed similarly well and significantly better than the three remaining criteria. Quite surprisingly, voting methods were superior to the methods based on selection via nested validation cycle. The reason may be the high heterogeneity of schizophrenia manifestations in the images (Nenadic et al., 2012), in combination with the relatively small dataset. This might have caused for the optimal number of features selected on a smaller group in a nested cycle not to be generalized for all subjects in the testing set. Ensemble strategies, such as voting, might perform better in such situations, as was described in Chapter 2.3.1. This is probably also the reason for the rather inferior results of RBSVC implementation of the SVM classifier, because the radial basis kernel functions were optimized with 5-fold CV.

The overall best quality of classification was achieved with the default configuration of all parameters, except for the discriminating criterion for which the *Variances* showed the best results – accuracy 73.2 % (SD 2.1 %), sensitivity 71.7 % (SD 3.0 %) and specificity 74.7 % (SD 2.6 %).

The computations that led to the optimal setting parameters of the classification algorithm took several days due to thousands of repetitions of the whole classification procedure (100 repetitions □ 52 validation runs □ best feature selection □ classifier optimization). For this reason, it was not feasible to test the other two key parameters of DWT – the wavelet mother function and the level of decomposition. The fourth level of decomposition was preset based on preliminary results and the wavelet sym5 was selected based on the results in the study (Kumari and Vijay, 2012) as well as from the previous studies of the same authors, who reported a good performance of the wavelets from Symlet family in coding of natural images.

5.1.3 Approach III: Combining features and classifiers

The recognition algorithm, which combined three types of features, reduced using isPCA, and three classifiers (mMLDA, CM, AL) achieved the classification accuracy of 81.6 % (with 75.5 % sensitivity and 87.8 % specificity). This performance was slightly higher than in case of single classifiers and was significantly better than classification by chance.

MLDA was proposed in (Thomaz et al., 2007a) to improve classification performance on limited sample size problems. Here, the method was modified to further improve the accuracy. The modification in mMLDA consist of calculation the classification boundary as an average discriminative score weighted by standard deviations of the two groups as opposed to MLDA, where an unweighted average serves as the boundary. Moreover, mMLDA was combined here with two other classification algorithms to obtain even slightly better classification results. In the previous works related to Thomaz's MLDA, MR intensity features were used to distinguish between preterm infants and term controls (Thomaz et al., 2007a) and to classify patients with Alzheimer disease from healthy controls (Thomaz et al., 2007b). Functional MRI data were used in (Sato et al., 2008) for classification of young and old healthy subjects.

The classification performances of MLDA and mMLDA were compared and it was observed that mMLDA enabled classification with higher sensitivity than MLDA in case of local volume changes and MR intensities. However, accuracy was only very slightly and insignificantly higher in mMLDA compared to MLDA results. The higher sensitivity can be attributed to the classification boundary, computed as weighted average, getting shifted towards the control group due to larger variability (and thus larger standard deviation) of the discriminative scores of the patients. On the contrary, shifting the classification boundary causes worsening of the balance

between sensitivity and specificity, which might be unwanted in some experiments. In the identification of subjects with neuropsychiatric and neurodegenerative disorders, however, higher sensitivity than specificity is welcomed as the cost associated with further investigating those subjects who are subsequently found not to have the disease is outweighed by the benefit of identifying a considerably larger number of individuals in the earliest stages of the disease (O'Bryant et al., 2008).

In the second experiment with the number of eigenvectors used in isPCA for data reduction, very heterogeneous results were received. In mMLDA-based classification, reducing the number of eigenvectors did not improve the performance at all. In AL-based classification, considerably better results were obtained when local volume changes and MR intensities but not GM densities were reduced by the first few components. In CM-based classification, the results based on the lower number of eigenvectors were more or less comparable with the performance based on all components. This heterogeneity is caused by the fact that while decreasing the number of eigenvectors, not only the noise but also the signal important for classification gets removed. In all the experiments, the first eigenvector was able to account for most of the accuracy, even if it accounted only for a fraction of the original variance and missed some of the important discriminative areas in case of local volume changes or included non-informative areas in case of MR intensities and GM densities (as discussed in more detail below). Adding second and few other eigenvectors, however, did not improve the results significantly. Thus, to receive good classification results while reducing the data using isPCA, one of the two following options can be used. (i) Searching for the best number of principal components used for reduction for each type of features and each classification method. The drawback of such approach lies in considerably increased computational requirements. (ii) Selection of those components that best discriminate between the groups using two-sample t-test (Ardekani et al., 2011; Yoon et al., 2007). In such an approach, it is necessary to ensure that LOO-CV is performed correctly, i.e. the data from the test subject are not used in the component creation and selection. (iii) Using all eigenvectors to avoid losing any signal important for classification. In case of single classifiers, this approach can however decrease the robustness of the classification algorithm due to the present noise in the data. To increase the robustness, a combination of classifiers into an ensemble is recommended (Liu et al., 2012).

The third experiment concerning the recognition of FES patients from NC subjects using the three classifiers and the three types of imaging features revealed that the highest sensitivity was achieved in the classification using mMLDA, the second highest sensitivity using CM and the lowest using AL. In terms of accuracy, the best result was obtained in the CM-based classification. However, it cannot be claimed that the CM classifier would always enable classification with higher accuracy than mMLDA or AL, because the classification performance is also influenced by data reduction. If a different method for data reduction was used instead of isPCA, the classification results could have been slightly different. Unlike the CM classifier, the AL classifier did not allow achieving high classification results. The balance between sensitivity and specificity was poor and in favour of specificity, which may be due to higher variability in the FES group than in NC group. The performance of AL-based classification, however, could be improved by reducing the number of principal components. The comparison of the types of imaging features showed that MR intensities were less appropriate for the classification than local volume changes and GM densities. The classifiers based on automated and more computation-intensive image preprocessing steps perform better as they reflect the disease more clearly. It is, therefore, worth to extract grey matter from the intensity images as the grey matter has already been shown to be affected by schizophrenia (Honea et al., 2005; Shenton et al., 2001). Regarding the local volume changes caused by deformations, they show how the brain anatomy of a diagnosed subject differs from the normal template anatomy in terms of local volume expansions and contractions (Gaser et

al., 2001; Schwarz et al., 2007) and thus, they are also highly appropriate for distinguishing schizophrenia patients from healthy individuals.

When voting of the classifiers was performed, the classification performance was only slightly higher than the best single classifier. Anyway, the ensemble of classifiers is supposed to be more robust than single classifiers and therefore, preferred as was shown in Chapter 2.3.1. The best classification performance (accuracy of 81.6 %, sensitivity of 75.5 %, and specificity of 87.8 %) was achieved using the majority vote of five classifiers (mMLDA, CM and AL with local deformations, mMLDA with MR intensities and mMLDA with GM densities). This specific combination of voting classifiers was found among all possible odd numbers of classifiers without a priori assumption that the specific combination of variables would be the most informative. This classification performance was statistically significantly higher than the classification by chance and was superior to or comparable with the results obtained in other state-of-the-art studies concerning schizophrenia research, in which the accuracy was around 75% (range from 67% to 86%) (Castellani et al., 2012; Karageorgiou et al., 2011; Kawasaki et al., 2007; Leonard et al., 1999; Mourao-Miranda et al., 2012; Nakamura et al., 2004; Ota et al., 2012; D. Sun et al., 2009; Takayanagi et al., 2010). Nevertheless, the efficiency of the methods proposed in some of the mentioned studies could be biased due to incorporation of images of chronic schizophrenia patients for classification. Chronic schizophrenia might be, due to the combined effects of the pharmacotherapy and the long-term disease, accompanied with greater changes in the brain morphology than in the early stages of the disease. As an example, the accuracy was 91.8 % for female subjects and 90.8 % for male subjects while discriminating a mixed group of FES and ChS patients from NC subjects using the COMPARE algorithm (Fan et al., 2007), compared to the accuracy of 73.4 % when the same algorithm was used for classification of only FES and NC individuals (Zanetti et al., 2013).

5.2 NEUROIMAGING & PEDAGOGY

The research presented in this thesis has taken part also in the author's pedagogy-oriented academic activities. Several bachelor theses in the field have been completed and defended under author's supervision, e.g.: Statistical methods for segmentation in MRI brain images (Janoušová, 2008); Mathematical modelling of tissue deformations in MRI images (Kuhn, 2010); Metaheuristic optimization methods for magnetic resonance image registration (Dluhoš, 2011). Several master theses in the field have been completed and defended under author's supervision, e.g.: Modern methods for image data analysis in neuropsychiatric research (Janoušová, 2010); Visual stimuli recognition based on data from functional magnetic resonance imaging (Kuhn, 2012); Multiresolution feature selection for recognition in magnetic resonance brain images (Dluhoš, 2013).

Some of the theories presented here have been included in the compulsory subjects of the Computational Biology curriculum (Linear and Adaptive Data Processing and Analysis – Masaryk University) and Biomedical Engineering & Bioinformatics curriculum (Advanced Methods in Biostatistics – Brno University of Technology). Further, the topics of author's research were at focus of two summer schools organized for students of Computational Biology, Biomedical Engineering and Neuroscience: Analysis of Clinical and Biomedical Data in an Interdisciplinary Approach (2009); Image Data Analysis and Processing in Neuroscience (2014).

The author of this thesis has been leading an independent small research group focused on neuroimaging and machine learning since 2012. The composition of the group was changing according to the students involved; nevertheless the main specialties included were always: Computational Biology (45 %), Neuroscience (33 %), Biomedical Engineering and Biophysics (22 %). The main pedagogical activity here is to keep explaining students that despite of novel biological insights brought by combination of neuroimaging and machine learning approaches, it also holds the danger of potential unintentional misuse, as was explained in Chapter 2.4.

Nowadays, the group is able to produce reasonable analytical outputs, including original papers in impacted journals covering fields of neuroscience and biomedical engineering – with substantial contributions from master and doctoral students.

6 CONCLUSIONS

In this habilitation thesis, structural magnetic resonance imaging is used to investigate brain changes in psychiatric disorders – with main focus on schizophrenia. This mental disorder affects about 21 million people worldwide (WHO, 2015), from which ~2,14 million in Europe. About a third of patients do not respond well during the treatment of the first episode of schizophrenia. Developing and exploiting novel neuroimaging methods, which improve the specificity of the treatment by 25 %, would mean that more than 2 million patients worldwide (>200,000 in Europe) would have a more suitable treatment regimen. Once a sufficient accuracy is achieved, the imaging methods will be brought closer to the clinical psychiatry practice. Further, it will allow an objective diagnosis in psychiatry, which previously was not possible, because the current practice is based on subjective assessment of clinically evident changes in mental function and behaviour.

The first part of this thesis provides an incisive review of image registration methods and their applications in neuroimaging including a survey and classification of the image registration techniques related to this area. A substantial space is devoted to design and the use of deformation-based morphometry – an approach in which the deformations that align anatomical images to a standard template are analyzed to detect significant structural differences between populations of patients and healthy controls. This algorithm is compared to voxel-based morphometry - another automated whole-brain morphometry method that is used by the neuroimaging community much more frequently, but that has also many questionable steps in the image processing pipeline, such as modulation or smoothing. The state-of-the-art part is completed with selected results of early and introductory work of machine learning groups in the neuropsychiatric research, aiming at computer-aided diagnostics based on individual brain-image data. This is more ambitious, provocative and difficult than uncovering brain regions with morphological differences between patient and healthy populations. On the other hand, many results which can be found in the related literature suffer from low statistical power, positive-outcome bias and often invalid or non-generalizable conclusions drawn out of high-dimensional data with a limited number of subjects. An important challenge is therefore the development of an interdisciplinary educated workforce that can lead future efforts in research and development for brain mapping through neuroimaging and machine learning methods.

The second part of this thesis describes design of three different methods for recognition of patients with first-episode schizophrenia from healthy volunteers, and demonstrates results which were obtained on the dataset underlying a prospective observational study of patients with schizophrenia during their first episode to assess the usefulness of the MRI morphometric examination for diagnostic and treatment outcome prediction purposes. The presented methods differ in the approach to image preprocessing and features extraction. Furthermore, these methods vary in the way of building classifier ensembles as well as in the use of particular classifiers. The generalizability of the involved models also varies – from a naive approach which leads to overestimated accuracy, to elaborated cross-validation strategies. The achieved classification accuracies $\in \langle 73.2\%; 87.5\% \rangle$ are comparable to other state-of-the-art works on MRI-based schizophrenia classification. Unfortunately, the achieved accuracy of the presented methods is still low to enable their clinical application.

Future efforts fueled by easier availability of powerful imaging methods will include simultaneous analysis of multimodal image data, by extracting e.g. local volume changes from structural MRI data, functional connectivity patterns from fMRI data and anatomical connectivity patterns from DTI, together with leveraging techniques from multiple-kernel and structural learning. In the specific area of mental disorders, the move from biomedical engineering research

towards clinical practice will have to be supported with other interdisciplinary collaborations, in order to obtain pathophysiology of the disorders at multiple levels from cell, brain, to behaviour.

BIBLIOGRAPHY

- Ardekani, B.A., Tabesh, A., Sevy, S., Robinson, D.G., Bilder, R.M., Szeszko, P.R., 2011. Diffusion tensor imaging reliably differentiates patients with schizophrenia from healthy volunteers. *Hum. Brain Mapp.* 32, pp. 1–9.
- Ashburner, J., Friston, K.J., 2005. Unified segmentation. *NeuroImage* 26, pp. 839–851.
- Ashburner, J., Friston, K.J., 2000. Voxel-Based Morphometry—The Methods. *NeuroImage* 11, pp. 805–821.
- Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K., 1998. Identifying global anatomical differences: Deformation-based morphometry. *Hum. Brain Mapp.* 6, pp. 348–357.
- Bishop, C.M., 2006. Pattern recognition and machine learning, Information science and statistics. Springer, New York.
- Bookstein, F.L., 2001. “Voxel-based morphometry” should not be used with imperfectly registered images. *NeuroImage* 14, pp. 1454–1462.
- Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, pp. 365–376.
- Caprihan, A., Pearlson, G.D., Calhoun, V.D., 2008. Application of principal component analysis to distinguish patients with schizophrenia from healthy controls based on fractional anisotropy measurements. *Neuroimage* 42, pp. 675–682.
- Castellani, U., Rossato, E., Murino, V., Bellani, M., Rambaldelli, G., Perlini, C., Tomelleri, L., Tansella, M., Brambilla, P., 2012. Classification of schizophrenia using feature-based morphometry. *J. Neural Transm.* 119, pp. 395–404.
- Colas, F., Brazdil, P., 2006. Comparison of SVM and Some Older Classification Algorithms in Text Classification Tasks, in: *Artificial Intelligence in Theory and Practice*, IFIP International Federation for Information Processing. Springer Boston, pp. 169–178.
- Culhane, A.C., Perrière, G., Considine, E.C., Cotter, T.G., Higgins, D.G., 2002. Between-group analysis of microarray data. *Bioinforma. Oxf. Engl.* 18, 1600–1608.
- Davatzikos, C., 2004. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage* 23, pp. 17–20.
- Davatzikos, C., Shen, D., Gur, R.C., Wu, X., Liu, D., Fan, Y., Hughett, P., Turetsky, B.I., Gur, R.E., 2005. Whole-Brain Morphometric Study of Schizophrenia Revealing a Spatially Complex Set of Focal Abnormalities. *Arch Gen Psychiatry* 62, pp. 1218–1227.
- Demirci, O., Clark, V., Magnotta, V., Andreasen, N., Lauriello, J., Kiehl, K., Pearlson, G., Calhoun, V., 2008a. A Review of Challenges in the Use of fMRI for Disease Classification / Characterization and A Projection Pursuit Application from A Multi-site fMRI Schizophrenia Study. *Brain Imaging Behav.* 2, pp. 207–226.
- Dietterich, T.G., 2000. Ensemble methods in machine learning, in: *Multiple Classifier Systems*. Springer, pp. 1–15.
- Dluhoš, P., Schwarz, D., Kašpárek, T., 2014. Wavelet features for recognition of first episode of schizophrenia from MRI brain images. *Radioengineering* 23, pp. 274–281.
- Ellison-Wright, I., Glahn, D.C., Laird, A.R., Thelen, S.M., Bullmore, E., 2008. The Anatomy of First-Episode and Chronic Schizophrenia: An Anatomical Likelihood Estimation Meta-Analysis. *Am. J. Psychiatry* 165, pp. 1015–1023.
- Fanelli, D., 2012. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90, pp. 891–904.

- Fan, Y., Gur, R.E., Gur, R.C., Wu, X., Shen, D., Calkins, M.E., Davatzikos, C., 2008. Unaffected Family Members and Schizophrenia Patients Share Brain Structure Patterns: A High-Dimensional Pattern Classification Study. *Biol. Psychiatry* 63, pp. 118–124.
- Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C., 2007. COMPARE: Classification of Morphological Patterns Using Adaptive Regional Elements. *Med. Imaging IEEE Trans. On* 26, pp. 93–105.
- Friston, K.J., Ashburner, J., 2004. Generative and recognition models for neuroanatomy. *NeuroImage* 23, pp. 21–24.
- Fujita, A., Gomes, L.R., Sato, J.R., Yamaguchi, R., Thomaz, C.E., Sogayar, M.C., Miyano, S., 2008. Multivariate gene expression analysis reveals functional connectivity changes between normal/tumoral prostates. *BMC Syst. Biol.* 2, 106.
- Fukunaga, K., 1990. Introduction to statistical pattern recognition. Academic press.
- Gaser, C., Nenadic, I., Buchsbaum, B.R., Hazlett, E.A., Buchsbaum, M.S., 2001. Deformation-based morphometry and its relation to conventional volumetry of brain lateral ventricles in MRI. *NeuroImage* 13, pp. 1140–1145.
- Gaser, C., Schmidt, S., Metzler, M., Herrmann, K.-H., Krumbein, I., Reichenbach, J.R., Witte, O.W., 2012. Deformation-based brain morphometry in rats. *NeuroImage* 63, pp. 47–53.
- Gholipour, A., Kehtarnavaz, N., Briggs, R., Devous, M., Gopinath, K., 2007. Brain Functional Localization: A Survey of Image Registration Techniques. *Med. Imaging IEEE Trans. On* 26, pp. 427–451.
- Giuliani, N.R., Calhoun, V.D., Pearlson, G.D., Francis, A., Buchanan, R.W., 2005. Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. *Schizophr. Res.* 74, pp. 135–147.
- Gong, Q.-Y., Sluming, V., Mayes, A., Keller, S., Barrick, T., Cezayirli, E., Roberts, N., 2005. Voxel-based morphometry and stereology provide convergent evidence of the importance of medial prefrontal cortex for fluid intelligence in healthy adults. *NeuroImage* 25, pp. 1175–1186.
- Honea, R., Crow, T.J., Passingham, D., Mackay, C.E., 2005. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am. J. Psychiatry* 162, pp. 2233–2245.
- Ingalhalikar, M., Kanterakis, S., Gur, R., Roberts, T.P., Verma, R., 2010. DTI based diagnostic prediction of a disease via pattern classification, in: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*. Springer, pp. 558–565.
- Ioannidis, J.P., 2005. Why most published research findings are false. *PLoS Med.* 2, p. e124.
- Iwabuchi, S.J., Liddle, P.F., Palaniyappan, L., 2013. Clinical utility of machine-learning approaches in schizophrenia: improving diagnostic confidence for translational neuroimaging. *Front. Psychiatry* 4:95.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning*, Springer Texts in Statistics. Springer New York, New York, NY.
- Janoušová, E., Schwarz, D., Kašpárek, T., 2015. Combining various types of classifiers and features extracted from magnetic resonance imaging data in schizophrenia recognition. *Psychiatry Res. Neuroimaging* submitted, in revision.
- Jolliffe, I., 2002. *Principal component analysis*. Wiley Online Library.
- Kailath, T., 1967. The Divergence and Bhattacharyya Distance Measures in Signal Selection. *IEEE Trans. Commun.* 15, pp. 52–60.
- Karageorgiou, E., Schulz, S.C., Gollub, R.L., Andreasen, N.C., Ho, B.-C., Lauriello, J., Calhoun, V.D., Bockholt, H.J., Sponheim, S.R., Georgopoulos, A.P., 2011. Neuropsychological testing and structural magnetic resonance imaging as diagnostic biomarkers early in the course of schizophrenia and related psychoses. *Neuroinformatics* 9, pp. 321–333.

- Kasperek, T., Prikryl, R., Mikl, M., Schwarz, D., Ceskova, E., Krupa, P., 2007. Prefrontal but not temporal grey matter changes in males with first-episode schizophrenia. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 31, pp. 151–157.
- Kasperek, T., Thomaz, C.E., Sato, J.R., Schwarz, D., Janousova, E., Marecek, R., Prikryl, R., Vanicek, J., Fujita, A., Ceskova, E., 2011. Maximum-uncertainty linear discrimination analysis of first-episode schizophrenia subjects. *Psychiatry Res. Neuroimaging* 191, pp. 174–181.
- Kawasaki, Y., Suzuki, M., Kherif, F., Takahashi, T., Zhou, S.-Y., Nakamura, K., Matsui, M., Sumiyoshi, T., Seto, H., Kurachi, M., 2007. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *NeuroImage* 34, pp. 235–242.
- Keller, S.S., Mackay, C.E., Barrick, T.R., Wiesmann, U.C., Howard, M.A., Roberts, N., 2002. Voxel-based morphometric comparison of hippocampal and extrahippocampal abnormalities in patients with left and right hippocampal atrophy. *NeuroImage* 16, pp. 23–31.
- Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *IJCAI*. pp. 1137–1145.
- Koutsouleris, N., Meisenzahl, E.M., Davatzikos, C., Bottlender, R., Frodl, T., Scheuerecker, J., Schmitt, G., Zetzsche, T., Decker, P., Reiser, M., Moller, H.-J., Gaser, C., 2009. Use of Neuroanatomical Pattern Classification to Identify Subjects in At-Risk Mental States of Psychosis and Predict Disease Transition. *Arch Gen Psychiatry* 66, pp. 700–712.
- Kumari, S., Vijay, R., 2012. Effect of symlet filter order on denoising of still images. *Adv. Comput. Int. J. ACIJ* 3, pp. 137–143.
- Kuncheva, L.I., 2004. *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons.
- Kuncheva, L.I., Rodriguez, J.J., Plumpton, C.O., Linden, D.E.J., Johnston, S.J., 2010. Random Subspace Ensembles for fMRI Classification. *IEEE Trans. Med. Imaging* 29, pp. 531–542.
- Legendre, P., Legendre, L., 1998. *Numerical Ecology, Volume 24, (Developments in Environmental Modelling)*.
- Lemm, S., Blankertz, B., Dickhaus, T., Müller, K.-R., 2011. Introduction to machine learning for brain imaging. *NeuroImage* 56, pp. 387–399.
- Leonard, C.M., Kuldau, J.M., Breier, J.I., Zuffante, P.A., Gautier, E.R., Heron, D.-C., Lavery, E.M., Packing, J., Williams, S.A., DeBose, C.A., 1999. Cumulative effect of anatomical risk factors for schizophrenia: an MRI study. *Biol. Psychiatry* 46, pp. 374–382.
- Leucht, S., Davis, J.M., Engel, R.R., Kane, J.M., Wagenpfeil, S., 2007. Defining Response in Antipsychotic Drug Trials: Recommendations for the Use of Scale-Derived Cutoffs. *Neuropsychopharmacology* 32, pp. 1903–1910.
- Liu, M., Wang, L., Shen, H., Liu, Z., Hu, D., 2011. A study of schizophrenia inheritance through pattern classification, in: *Intelligent Control and Information Processing (ICICIP), 2011 2nd International Conference on*. IEEE, pp. 152–156.
- Liu, M., Zhang, D., Shen, D., Alzheimer's Disease Neuroimaging Initiative, 2012. Ensemble sparse classification of Alzheimer's disease. *NeuroImage* 60, pp. 1106–1116.
- Liu, Y., Teverovskiy, L., Carmichael, O., Kikinis, R., Shenton, M., Carter, C.S., Stenger, V.A., Davis, S., Aizenstein, H., Becker, J.T., 2004. Discriminative MR image feature analysis for automatic schizophrenia and Alzheimer's disease classification. Springer.
- Maintz, J.B.A., Viergever, M.A., 1998. A survey of medical image registration. *Med. Image Anal.* 2, pp. 1–36.
- Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Anal. Mach. Intell. IEEE Trans. On* 11, pp. 674–693.
- Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J.-M., 2007. *Wavelets and their Applications*. Wiley Online Library.

- Mourão-Miranda, J., Bokde, A.L., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *NeuroImage* 28, pp. 980–995.
- Mourao-Miranda, J., Reinders, A.A.T.S., Rocha-Rego, V., Lappin, J., Rondina, J., Morgan, C., Morgan, K.D., Fearon, P., Jones, P.B., Doody, G.A., Murray, R.M., Kapur, S., Dazzan, P., 2012. Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. *Psychol. Med.* 42, pp. 1037–1047.
- Nakamura, K., Kawasaki, Y., Suzuki, M., Hagino, H., Kurokawa, K., Takahashi, T., Niu, L., Matsui, M., Seto, H., Kurachi, M., 2004. Multiple Structural Brain Measures Obtained by Three-Dimensional Magnetic Resonance Imaging To Distinguish Between Schizophrenia Patients and Normal Subjects. *Schizophr. Bull.* 30, pp. 393–404.
- Nenadic, I., Gaser, C., Sauer, H., 2012. Heterogeneity of Brain Structural Variation and the Structural Imaging Endophenotypes in Schizophrenia. *Neuropsychobiology* 66, pp. 44–49.
- Nieuwenhuis, M., van Haren, N.E., Hulshoff Pol, H.E., Cahn, W., Kahn, R.S., Schnack, H.G., 2012. Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *Neuroimage* 61, pp. 606–612.
- Niznikiewicz, M.A., Kubicki, M., Shenton, M.E., 2003. Recent structural and functional imaging findings in schizophrenia. *Curr. Opin. Psychiatry* 16, pp. 123–147.
- O’Bryant, S.E., Humphreys, J.D., Smith, G.E., Ivnik, R.J., Graff-Radford, N.R., Petersen, R.C., Lucas, J.A., 2008. Detecting dementia with the mini-mental state examination in highly educated individuals. *Arch. Neurol.* 65, pp. 963–967.
- Ota, M., Sato, N., Ishikawa, M., Hori, H., Sasayama, D., Hattori, K., Teraishi, T., Obu, S., Nakata, Y., Nemoto, K., 2012. Discrimination of female schizophrenia patients from healthy women using multiple structural brain measures obtained with voxel-based morphometry. *Psychiatry Clin. Neurosci.* 66, pp. 611–617.
- Pautasso, M., 2010. Worsening file-drawer problem in the abstracts of natural, medical and social science databases. *Scientometrics* 85, pp. 193–202.
- Perkins, D.O., Gu, H., Boteva, K., Lieberman, J.A., 2005. Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. *Am. J. Psychiatry* 162, pp. 1785–1804.
- Pohl, K., Sabuncu, M., 2009. A Unified Framework for MR Based Disease Classification, in: Prince, J., Pham, D., Myers, K. (Eds.), *Information Processing in Medical Imaging, Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, pp. 300–313.
- Radua, J., Canales-Rodríguez, E.J., Pomarol-Clotet, E., Salvador, R., 2014. Validity of modulation and optimal settings for advanced voxel-based morphometry. *NeuroImage* 86, pp. 81–90.
- Rohr, K., 2000. Elastic Registration of Multimodal Medical Images: A Survey. *Künstl. Intell.* 2000, pp. 11–17.
- Rueckert, D., Aljabar, P., 2010. Nonrigid Registration of Medical Images: Theory, Methods, and Applications. *Signal Process. Mag. IEEE* 27, pp. 113–119.
- Samartzis, L., Dima, D., Fusar-Poli, P., Kyriakopoulos, M., 2014. White matter alterations in early stages of schizophrenia: a systematic review of diffusion tensor imaging studies. *J. Neuroimaging Off. J. Am. Soc. Neuroimaging* 24, pp. 101–110.
- Sato, J.R., Fujita, A., Thomaz, C.E., Martin, M. da G.M., Mourão-Miranda, J., Brammer, M.J., Junior, E.A., 2009. Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction. *NeuroImage* 46, pp. 105–114.
- Sato, J.R., Thomaz, C.E., Cardoso, E.F., Fujita, A., Martin, M. da G.M., Amaro, E., 2008. Hyperplane navigation: A method to set individual scores in fMRI group datasets. *NeuroImage* 42, pp. 1473–1480.
- Schwarz, D., 2005. Automated morphometry of MRI brain images with the use of deformable registration, PhD thesis. ed. Brno University of Technology, Brno.

- Schwarz, D., Kasperek, T., 2014. Brain morphometry of MR images for automated classification of first-episode schizophrenia. *Inf. Fusion* 19, pp. 97–102.
- Schwarz, D., Kasperek, T., 2011. Comparison of two methods for automatic brain morphometry analysis. *Radioengineering* 20, pp. 996–1001.
- Schwarz, D., Kasperek, T., Provaznik, I., Jarkovsky, J., 2007. A Deformable Registration Method for Automated Morphometry of MRI Brain Images in Neuropsychiatric Research. *Med. Imaging IEEE Trans. On* 26, pp. 452–461.
- Shen, H., Wang, L., Liu, Y., Hu, D., 2010. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. *Neuroimage* 49, pp. 3110–3121.
- Shenton, M.E., Dickey, C.C., Frumin, M., McCarley, R.W., 2001. A review of MRI findings in schizophrenia. *Schizophr. Res.* 49, pp. 1–52.
- Sotiras, A., Christos, D., Paragios, N., 2012. Deformable Medical Image Registration: A Survey (No. RR-7919). INRIA.
- Starck, J.-L., Murtagh, F., Fadili, J.M., 2010. Sparse image and signal processing: wavelets, curvelets, morphological diversity. Cambridge University Press.
- Sun, D., van Erp, T.G.M., Thompson, P.M., Bearden, C.E., Daley, M., Kushan, L., Hardt, M.E., Nuechterlein, K.H., Toga, A.W., Cannon, T.D., 2009. Elucidating a Magnetic Resonance Imaging-Based Neuroanatomic Biomarker for Psychosis: Classification Analysis Using Probabilistic Brain Atlas and Machine Learning Algorithms. *Biol. Psychiatry* 66, pp. 1055–1060.
- Sun, J., Maller, J.J., Guo, L., Fitzgerald, P.B., 2009. Superior temporal gyrus volume change in schizophrenia: A review on Region of Interest volumetric studies. *Brain Res. Rev.* 61, pp. 14–32.
- Takayanagi, Y., Kawasaki, Y., Nakamura, K., Takahashi, T., Orikabe, L., Toyoda, E., Mozue, Y., Sato, Y., Itokawa, M., Yamasue, H., 2010. Differentiation of first-episode schizophrenia patients from healthy controls using ROI-based multiple structural brain variables. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 34, pp. 10–17.
- Thomaz, C.E., Boardman, J.P., Counsell, S., Hill, D.L., Hajnal, J.V., Edwards, A.D., Rutherford, M.A., Gillies, D.F., Rueckert, D., 2007a. A multivariate statistical analysis of the developing human brain in preterm infants. *Image Vis. Comput.* 25, pp. 981–994.
- Thomaz, C.E., Duran, F.L., Busatto, G.F., Gillies, D.F., Rueckert, D., 2007b. Multivariate Statistical Differences of MRI Samples of the Human Brain. *J Math Imaging Vis* 29, pp. 95–106.
- Van Haren, N., Cahn, W., Hulshoff Pol, H., Kahn, R., 2012. The course of brain abnormalities in schizophrenia: can we slow the progression? *J. Psychopharmacol. (Oxf.)* 26, pp. 8–14.
- Vapnik, V.N., 1999. An overview of statistical learning theory. *Neural Netw. IEEE Trans. On* 10, pp. 988–999.
- Wallisch, P., 2014. MATLAB for neuroscientists: an introduction to scientific computing in MATLAB, Second edition. ed. Academic Press, Amsterdam.
- Wang, S., Li, D., Song, X., Wei, Y., Li, H., 2011. A feature selection method based on improved fisher's discriminant ratio for text sentiment classification. *Expert Syst. Appl.* 38, pp. 8696–8702.
- Wang, Z., Wang, J., Calhoun, V., Rao, H., Detre, J.A., Childress, A.R., 2006. Strategies for reducing large fMRI data sets for independent component analysis. *Magn. Reson. Imaging* 24, pp. 591–596.
- WHO, 2015. Mental health - Schizophrenia. [online]. Available: http://www.who.int/mental_health/management/schizophrenia/en/ [Accessed 27 Feb 2015]

- Wright, I.C., Rabe-Hesketh, S., Woodruff, P.W., David, A.S., Murray, R.M., Bullmore, E.T., 2000. Meta-analysis of regional brain volumes in schizophrenia. *Am. J. Psychiatry* 157, pp. 16–25.
- Yoon, U., Lee, J.-M., Im, K., Shin, Y.-W., Cho, B.H., Kim, I.Y., Kwon, J.S., Kim, S.I., 2007. Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. *NeuroImage* 34, pp. 1405–1415.
- Zanetti, M.V., Schaufelberger, M.S., Doshi, J., Ou, Y., Ferreira, L.K., Menezes, P.R., Sczufca, M., Davatzikos, C., Busatto, G.F., 2013. Neuroanatomical pattern classification in a population-based sample of first-episode schizophrenia. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 43, pp. 116–125.
- Zarogianni, E., Moorhead, T.W., Lawrie, S.M., 2013. Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level. *NeuroImage Clin.* 3, pp. 279–289.
- Zitová, B., Flusser, J., 2003. Image registration methods: a survey. *Image Vis. Comput.* 21, pp. 977–1000.

ABSTRACT

This habilitation thesis contributes mainly to the field of biomedical engineering – one of many sciences active in brain research. The first part summarizes the state of the art of image registration and pattern recognition techniques applied in neuropsychiatric research. The second part describes original algorithms for recognition of first-episode schizophrenia patients from healthy controls. Various approaches to the classification or prediction tasks include automated whole-brain morphometry of magnetic resonance images and methods of supervised learning with ensemble strategies. Results obtained from a dataset, collected during a prospective observational study of first-episode schizophrenia patients, are presented and discussed in the light of the findings published in the recent literature.