Computer-based diagnostic and prognostic approaches in medical research using brain MRI

Kumulative Habilitationsschrift

zur Erlangung der Lehrbefähigung für das Fach Psychologie

vorgelegt dem Rat der Lebenswissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

von

Dr. rer. nat. Martin Weygandt

Prof. Dr. Jan-Hendrik Olbertz

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Richard Lucius

Dekan der Lebenswissenschaftlichen Fakultät

Berlin, 29. April 2016

Content

1 Introduction

With the advent of powerful computers and magnetic resonance imaging (MRI) scanners, a novel diagnostic and prognostic approach in medicine developed recently that uses computer-based regression algorithms evaluating brain MRI signals for disease prediction. In this approach, a regression algorithm typically first extracts certain regularities related to a criterion of interest (e.g. the presence or severity of a disease) from exemplary MRI data and then evaluates MRI data of unknown subjects with regard to this criterion based on the extracted regularities. The approach is currently very popular as it promises to enable automated diagnosis of disease and to aid in the early detection of disorders and in prognosis of treatment responses. Consistent with the expected utility of the approach, it yielded hundreds of publications and is used for neurologic, psychiatric and metabolic disorders (see Sundermann et al., 2014; Cohen et al., 2011; Bray et al., 2009; Demicri et al., 2008 for an overview). In this synopsis, I will describe the basic concept of the approach, its development, and research aims pursued with it in the Introduction. In the second part, I will describe the articles that I submit for habilitation in terms of the concepts and methods delineated in the Introduction.

1.1 Basic concept

Although a variety of slightly different methods are used in the framework of the proposed approach, the majority of these methods are regression techniques used in a model-validation framework. In the literature, this combination of regression and model-validation is frequently referred to as *multivariate pattern analysis* (MVPA; see Sundermann et al., 2014; Cohen et al., 2011; Bray et al., 2009; Pereira et al. 2009; Demicri et al., 2008; Haynes et al., 2006; Norman et al., 2006 for an overview). Algorithms used in MVPA can be subdivided into algorithms for nominal regression (usually referred to as ‘classifiers’) and algorithms for continuous regression. Please note that I use the term ‘regression’ synonymously for binary and continuous regression in the following due to this technical relationship of 'classification' and 'regression'. Furthermore, the use of MVPA techniques for data analysis can be described with a three-step scheme comprising *Feature determination*, *Training* and *Test*. In short, a set of features relevant for an investigated disorder is selected from or computed based on all available MRI features in the feature determination stage. After the features have been determined, the corresponding signals are sampled across all available subjects and are combined in subject-specific feature patterns. In the training stage, patterns of a subset of subjects (i.e. the training subjects) are used to determine a regression function characterizing the regularities in the training patterns with regard to a criterion of interest. In the test stage, MRI feature patterns of novel subjects (i.e. the test subjects) are evaluated with regard to this criterion based on the regularities extracted in the training stage. In the following, the three stages are described in more detail.

1.1.1 Feature determination

In this stage, explanatory MRI features (e.g. blood flow or cortical thickness at specific coordinates) must be selected from or computed based on all available raw MRI variables (typically MRI signals measured in small cuboid-shaped spatial measurement units called voxels) that are considered relevant for a given research question. After the features have been determined, the corresponding signals from these features are sampled across all subjects and combined in subject-specific feature patterns. In the literature, various feature determination methods are used.

The simplest way is to directly use the MRI signals contained in all available voxels as features (i.e. to use a 'whole brain' approach as e.g. in Fu et al., 2008; Zhang et al., 2005). However, although the method is easy to implement it has also two important drawbacks. First, it is typically associated with a poor multivariate signal-to-noise ratio (SNR) as dysregulation of functional processes in psychiatric or neuropathology in neurologic disease is typically not uniformly distributed across the brain and the approach thus tends to include a large number of disorder-irrelevant features. Second, redundancy or collinearity among features can become a problem in the whole brain approach. Specifically, since the number of unknowns (i.e. MRI features; usually several thousand) exceeds the number of equations in this approach (i.e. training subjects; usually several dozens), determination of a regression function can (depending on the regression algorithm) correspond to solving an underdetermined set of equations from a mathematical perspective. In underdetermined equation systems, any unknown can be expressed as a linear combination of a subset of other unknowns from the same system as soon as the number of unknowns in the subset equals the number of equations. Correspondingly, underdetermined equation systems are highly redundant what can in turn lead to unstable estimates for regression coefficients in linear multiple regression meaning that already tiny variations in the features can induce large changes in the regression coefficient estimates. Thus, especially the choice of the regression algorithm has to be handled with care in a whole brain approach (see Dormann et al., 2013 for an overview).

To address these issues, 'region of interest' (ROI) approaches are frequently applied for feature selection. In this approach, MRI signals are only drawn from voxels located in functionally relevant networks in psychiatric research (e.g., Weygandt et al., 2012) or located in regions with a relevant tissue type (grey or white matter) in neurologic research (e.g., Weygandt et al., 2015a; Lee et al., 2014; Bendfeldt et al. 2012; Klöppel et al., 2008a, b). Alternatively, ROI are used for pattern generation by computing single aggregate measures across a whole ROI such as the hippocampal volume in Alzheimer's Disorder (AD) that are then integrated into patterns consisting of several of such features (Ritter et al., 2015; Cuingnet et al., 2011). Finally, in a specific variant of the ROI approach (that might also be considered as a mixture of the whole-brain and the ROI approach) known as searchlight method (Haynes et al., 2007; Kriegeskorte et al., 2006), disease-related MRI signals are selected from voxels in small spherical areas

surrounding a center voxel (Hackmack et al., 2012a; Weygandt et al., 2012) and regression models and prediction accuracy measures are then computed for these local patterns. This local regression procedure is repeated until each voxel included in a given analysis has been used once as center of a searchlight. Consequently, the method allows searching for informative areas regarding a given disorder across the whole brain by exploiting local multivariate imprints of functional dysregulation or neuropathology while circumventing the poor signal-to-noise problem of the whole-brain approach.

Component-based methods such as independent (ICA) or principal component analysis (PCA) correspond to another group of feature determination methods that can be applied to address the problems of the whole-brain approach. In particular, these methods identify directions (i.e., components) in voxel space that are either mutually orthogonal (PCA) or stochastically independent (ICA). Please see Stone (2001) for a comparative overview on these methods. After these directions have been identified, voxel information is projected onto them and the resulting component loadings are used as feature patterns for a subject (Wilette et al. 2014; Kawasaki et al. 2007; Ford et al., 2003).

Finally, methods frequently used in computer vision such as such as Scale Invariant Feature Transformation (Chen et al., 2014), Wavelet Analysis (Hackmack et al., 2012b), or Texture Analysis (Zhang et al., 2008) can be considered as another larger (but heterogeneous) group of feature determination methods. Consistently, these methods compute higher-level image (here: volume) statistics such as edges and spatial intensity contrasts that can then be combined in feature patterns handed over to a regression algorithm for training and testing.

1.1.2 Training

In this stage, a regression algorithm determines a mathematical regression function that maps the feature patterns onto a criterion vector following a statistical optimization procedure specific for the algorithm. Importantly, only patterns of a subset of subjects (i.e. of the training subjects) are used in this mapping procedure. Depending on the application (i.e. diagnosis of medical status or prediction of disease severity), the criterion vector can either code the group membership (e.g., diseased or healthy; Weygandt et al., 2015a; 2012) or symptom scores of training subjects as e.g., assessed by the Expanded Disability Status Score (EDSS; Kurtzke, 1983) in multiple sclerosis (MS; Hackmack et al., 2012a) or the Mini-Mental Status Examination (Folstein et al., 1975) in AD (Stonnington et al., 2010). To achieve the mapping, a large variety of regression algorithms has been used in the computerized MRI disease prediction framework. For example Canonical Correlation Analysis (CCA; Hackmack et al., 2012a), Linear Discriminant Analysis or Multivariate Analysis of Variance (MANOVA; Allefeld & Haynes, 2014) respectively, Logistic Regression (Weygandt et al., 2015a), Partial Least Square Regression (Chen et al., 2010), Support Vector Classifiers (SVC; Weygandt et al., 2012; Fu et al., 2008; Klöppel et al., 2008a, b)

and Support Vector Regression (Stonnington et al., 2010) to name just a few. Each of these regression algorithms is characterized by a specific optimization method. For example, SVC try to find a direction in feature space that maximizes the distance between patterns located in the transition area between classes, i.e. to find a large-margin that separates the classes (c.f., Müller et al., 2001). Alternatively, MANOVA identifies the direction(s) in space with the maximal ratio between inter-group and intra-group variation (Allefeld & Haynes, 2014). For a detailed overview on methodological aspects of a variety of MVPA algorithms, please see Duda et al. (2000).

1.1.3 Test

In the test stage, the mathematical regression model determined during training is used to evaluate patterns of novel subjects not used for model determination (i.e. of the test subjects) with regard to the criterion of interest. After this has been done for each available test pattern, accuracy measures are computed such as the percentage of correctly classified patterns, the mean of sensitivity and specificity, or the multiple correlation coefficient to name a few. Then, measures of inferential statistics are computed for these accuracy parameters e.g., by permutation testing (Good, 2005). In short, in permutation testing the true criterion vector is permuted several times and the classification procedure (i.e. training and testing including accuracy determination) is conducted for each of the permutations separately which finally yields a distribution of permutation accuracies. After this distribution has been determined, it is used to assess the probability of the empirically observed accuracy in terms of the relative proportion of permutation accuracies having minimally the extent of the empirical accuracy. Please see e.g., Weygandt et al. (2015a) for details of permutation testing in an MRI disease prediction framework.

1.2 Development of the method

The development of the approach from its beginning to now can probably be described best in terms of three epochs characterized by the discovery of diagnostic information in MR signals, the first use of statistical regression techniques to analyze this information, and finally the mass use of the approach.

1.2.1 Discovery of diagnostically relevant information in MR signals

In particular, the pivotal study of Raymond Damadian in the early 1970s *'Tumor Detection by Nuclear Magnetic Resonance'* (Damadian, 1971) might be considered as the starting point of this

development. In this study, Damadian used nuclear magnetic resonance (NMR) measurements to discriminate between healthy and cancerous tissue in rats based on differences of ^{1}H NMR relaxation times and thus showed for the first time that MR properties of biological entities can contain diagnostically exploitable information.

1.2.2 First analyses of diagnostic information with computer-based regression

After this first conceptual step, a small series of neurological studies evolved in the 1990s that started to analyze brain MRI scans with MVPA techniques for disease classification in a second stage. In particular, Namer and colleagues (Namer et al., 1993) used MVPA techniques and MRI-derived biomarkers to separate MS patients from patients showing high-risk factors of cerebrovascular diseases but no focal neurological symptoms. The study revealed high sensitivity of the computer-based diagnostic approach being approximately 20% superior to that achieved by human experts using MRI criteria existing at that time (Fazekas et al., 1988). Importantly, high diagnostic accuracy of the computer-based approach in this early work was already achieved in a model-validation framework, i.e. by applying a regression model derived from a set of training subjects to an independent set of test subjects. Inspired by this work, Barkhof and colleagues (Barkhof et al., 1997) used multivariate regression techniques to compare existing MRI criteria for future conversion from clinically-isolated syndrome (CIS) to clinically definite MS or more precisely they used the method to select an optimal set of MRI criteria for diagnosis performed by neurologists. Similar to the results obtained by Namer et al. (1993), results of this study (Barkhof et al., 1997) showed that the optimal set of features identified by MVPA was superior to feature sets suggested in diagnostic criteria existing at that time (Fazekas et al., 1988; Paty et al., 1988). Finally, the first study that used functional MRI (fMRI) for classification of neurologic patients originated in the early 2000s. In particular, Ford et al. (2003) evaluated brain activity acquired with fMRI during a word generation task to separate between elderly patients diagnosed with early AD and healthy control (HC) subjects of comparable age with MVPA including model-validation techniques.

In psychiatric research, computer-based patient classification using MRI signals was first performed in the early 2000s. In particular, Ford et al. (2002) used MRI-based MVPA including model-validation techniques to separate 15 schizophrenic patients from 8 HC subjects based on hippocampal MRI signals. The authors achieved good diagnostic accuracy although it is unclear whether the unequal group sizes might have biased the results because the authors exclusively reported the percentage of correctly classified patterns instead of measures correcting for unequal group sizes (e.g., the mean of sensitivity and specificity). Furthermore, it has to be mentioned that an MVPA neuroimaging approach for classification of schizophrenic patients (versus HC subjects) was already applied in 2001 by Meyer-Lindenberg and colleagues (Meyer-Lindenberg et al., 2001) in a Positron Emission Tomography (PET) study. In particular, the authors evaluated PET connectivity patterns during a working memory task to separate between patients and

controls. Importantly, they also used model-validation techniques as described for Namer et al. (1993). Unfortunately, the results of the model-validation analysis are hard to interpret in terms of a diagnostic approach since multiple patterns of individual subjects entered the analysis and the authors only report overall pattern separability and not separability of patterns specific for individual subjects.

1.2.3 Mass use of the computerized MRI disease prediction approach

Inferred from the distribution of publication dates reported for computer-based MRI disease prediction studies in review papers (e.g., Sundermann et al., 2014; Cohen et al., 2011; Bray et al., 2009; Demicri et al., 2008), a third epoch in the development of the approach characterized by its mass use started approximately in 2008 and lasts until now. This time point is also consistent with the publication of the most cited research article from the field by Klöppel et al. (2008a) with currently more than 500 citations (according to Google Scholar). During this third stage, the approach has been applied to a vast number of neurological disorders, e.g., in AD (Ritter et al., 2015; Dukart et al., 2013; McEvoy et al., 2009), MS (Weygandt et al., 2015a; Bendfeldt et al., 2012), Parkinson's Disease (Long et al. 2012), and stroke (Saur et al., 2010). In psychiatric research, the approach has been used in attention deficit hyperactivity disorder (Bohland et al., 2012), autism (Anderson et al., 2011), depression (Hahn et al., 2011), eating disorders (Weygandt et al. 2012), obsessive-compulsive disorder (OCD; e.g., Hoexter et al., 2013), pedophilia (Ponseti et al., 2012) and schizophrenia (Zarogianni et al., 2013). Please note, that these studies are mentioned here exclusively to give a small impression on the range of disorders the approach has been applied to and is far from being complete.

1.3 Research aims

Besides characterizing studies in the field based on different developmental stages or investigated disease domains, they can also be categorized based on the research aims pursued with the computer- and MRI-based disease prediction approach. Reviewing the literature shows that computer- and MRI-based disease prediction is (considered to be) applied in three frameworks. First, in an automated diagnosis setting. Second, to refine existing diagnostic criteria for diagnosis performed by human medical experts. Finally, the approach is used to identify novel diagnostic biomarkers. Please note that these goals are not mutually exclusive, i.e. a given study can address more than one of these objectives at a time.

1.3.1 Automated diagnosis

When a computerized patient classification system based on MRI signals is (intended to be) used in an automated diagnosis framework, it is expected to diagnose the unknown clinical status of a subject blindly i.e., without further human assistance exclusively based on a given set of MRI signals. Furthermore, such a system should reach very high diagnostic accuracy that is optimally superior to that obtained by human diagnosticians. Finally, such an approach should achieve constant diagnostic accuracy independent of e.g., differences in radiological expertise across medical experts. That this is possible could at least in part be shown in a study by Klöppel et al. (2008b). In this study, the authors compared the performance of a computer-based system in diagnosing AD with that of physicians having two different levels of expertise in this diagnostic task, i.e. general radiologists who evaluated brain scans during less than 5% of their daily work time and neuroradiologists who performed this task during more than 40%. The results showed that the computer-based approach and neuroradiologists both achieved very high (and comparable) accuracies in separating patients and HC subjects whereas overall radiologists only achieved moderate diagnostic accuracies.

Despite the promising character of such results suggesting a potential role of computer- and MRI-based diagnosis in clinical practice, several difficulties have to be coped with in a real-world application that do not apply in a research setting like in Klöppel et al. (2008b). In particular, (data of) Alzheimer's patients fulfilled a variety of well-defined research criteria e.g. the AD diagnoses were confirmed histopathologically (c.f. McKhann et al., 1984) in this study. Furthermore, patients were controlled for covariates such as having no family history of AD. Consequently, MVPA algorithms could extract / apply classification rules from / to very well-defined and prototypical data. In a real-world setting however, the majority of test data would presumably be less prototypically as e.g. Schneider et al., (2009) could show that patients suffering from AD frequently have a mixed neuropathology, i.e. they have features characteristic for AD but also features characteristic for vascular dementia such as cerebral infarcts. Due to such overlaps, diagnostic accuracies obtained by the computer-based approach in a research setting as in the study of Klöppel et al. (2008b) could overestimate accuracies obtainable in a real-world clinical setting to an unknown extent. Thus, in addition to substantial infrastructural demands, handling of missing data, and further aspects this problem might be an important reason for the fact that the approach is only used in research and not in an applied clinical routine framework until now. However, despite these issues there are currently several research projects such as the *'iDSS - Integrative Decision Support System for diagnosis of dementia in hospitalized elderly patients'* from the Bernstein Center for Computational Neuroscience Berlin and other groups (e.g., Ritter et al., 2015) or *'PredictAD'*[1] that aim to solve these problems and to establish the approach in a clinical real-world setting.

[1] See http://geriatrie.charite.de/en/research/research_projects_of_the_working_group_age_technology/idss/ and www.predictad.eu for further information

1.3.2 Refinement of diagnostic guidelines

Furthermore, the approach is also used to refine existing diagnostic guidelines for human diagnosticians. For example, Barkhof et al., (1997) used it to select an optimal set of MRI criteria for prognosis of future conversion from CIS to clinically definite MS. Specifically, they first pooled continuously scaled MRI features included in various diagnostic guidelines existing for this task at the time of the study (e.g. the number of periventricular hyperintense lesions in T2-weighted MR images; Fazekas et al., 1988; Paty et al., 1988) and then used varying cut-off scores to dichotomize the continuous feature scores into present and absent. After dichotomization, all cut-off dependent variants of all features were handed over to a Logistic Regression algorithm. Specifically, they used a stepwise regression approach to identify a set of non-redundant MRI criteria and corresponding cut-off scores that was superior to the (sets included in the) initial guidelines (Fazekas et al., 1988; Paty et al., 1988). They found that a four-parameter model including periventricular, infratentorial, and juxtacortical lesions together with gadolinium-enhancement best predicts conversion from CIS to MS. Thus, by using this approach, they provided easy to use diagnostic guidelines for daily routine in clinical hospitals.

In another attempt to use the approach for refinement of diagnostic guidelines, I used it to clarify whether it is possible to separate between subjects suffering from Bulimia Nervosa (BN) and subjects suffering from so called Binge-Eating Disorder (BED; Weygandt et al., 2012; see 2.1 for details). The question is important, since it was unclear for a long time whether BED should be considered as an autonomous diagnostic category or as a variant of BN because BN and BED share important commonalities on the phenomenological level (e.g., bingeing, comparable levels of body dissatisfaction and fear of weight gain). Consequently, BED was included as a provisional diagnostic category in the Diagnostic and Statistical Manual of Mental Disorders (DSM), edition IV (American Psychiatric Association, 1994) only.

1.3.3 Identification of novel diagnostic and prognostic biomarkers

Finally, MRI and computer-based regression are used to search for novel diagnostic biomarkers in a wide range of medical disorders. For example, I used the method to search for non-lesion related markers of MS in MRI images measured with standard MRI-sequences (Weygandt et al., 2015a). Van den Bogaard et al. (2012) / Kidwell et al. (2013) give an overview on the use of these techniques for identification of biomarkers in Huntington's Disease / stroke. In psychiatric research, I used the technique to identify novel diagnostic biomarkers for Eating Disorders (Weygandt et al., 2012). Hahn et al. (2011), McGrath et al. (2013), Pizzagalli (2011) and Fu et al. (2008) describe neuroimaging-derived biomarkers for depression. Furthermore, Zarogianni et al. (2013) give an overview on the use of SVC for identification of MRI-derived biomarkers in Schizophrenia. Finally, I used MRI and computer-based regression in the domain of metabolic

disorders, i.e. to search for prognostic biomarkers of short- (Weygandt et al., 2013) and long-term dietary success in obesity (Weygandt et al., 2015b).

At this point, it is important to mention that although it might appear that only classification techniques used in a model-validation framework can be used to search for diagnostic biomarkers due to their subject- or exemplar-oriented perspective also traditional methods such as t-tests can be used for this purpose. In order to understand this, one might consider a single hypothetical MRI feature measured for an arbitrary patient and a control group characterized by a high (absolute) t-statistic. Given the definition of the t-statistic in a two-sample t-test for independent samples with unequal variance:

$$t = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

with $m_{1/2}$ being the group means, $s_{1/2}$ the standard deviations, and $n_{1/2}$ the group sizes, this feature would also enable high diagnostic separability of subjects because the group means would be localized fairly apart from each other and the signal value of each specific group member would be quite close to the mean signal of its group. Under these circumstances, a simple hypothetical classifier not using model-validation that classifies subjects into groups based on the minimal Euclidean distance to the respective group means would also obtain high diagnostic accuracy. Consequently, unless no model-validation I used it is quite obvious that accuracy obtained by classifiers on one hand and t-statistics computed by t-tests on the other are highly related.

However, given that classification procedures typically involve a model-validation step, one might ask what the relation of classification accuracy obtained in a validation framework and t-scores not computed in such a framework is. To understand this, techniques of model-validation frequently used in classification such as the leave-one-out (LOO) cross-validation (CV) strategy have to be considered (for an overview on CV see e.g., Duda et al. [2000]). In LOO CV, only one pattern is removed from the set of all available patterns in a single CV iteration and is used for testing, the remaining patterns are used as training patterns. This procedure is repeated until all patterns were once treated as test pattern and a measure of accuracy is computed that summarize diagnostic separability of test patterns / individual subjects belonging to the different groups. When keeping this procedure in mind, it becomes clear that LOO CV does not change that relation dramatically since when only one pattern is removed from all available patterns during division into training and test, the statistical properties relevant for classification in each LOO iteration (here: group means and intra-group signal variations) would be more or less identical to these properties obtained for the full set of patterns. Consequently, one can conclude that the test statistics obtained by classifiers using certain common model-validation techniques (i.e. 'leave-few-samples-out' CV) and by more traditional procedures such as t-tests not using model-validation are highly related because also the relation

of underlying statistical properties is strong. Importantly, the same applies for the link of multivariate classification with CV and traditional multivariate statistics without CV (e.g. Hotellings T^2 for independent samples) as well as for uni- and multivariate continuous regression with and without CV. This can be seen in Figure 1 that summarizes the results of corresponding simulation analyses.

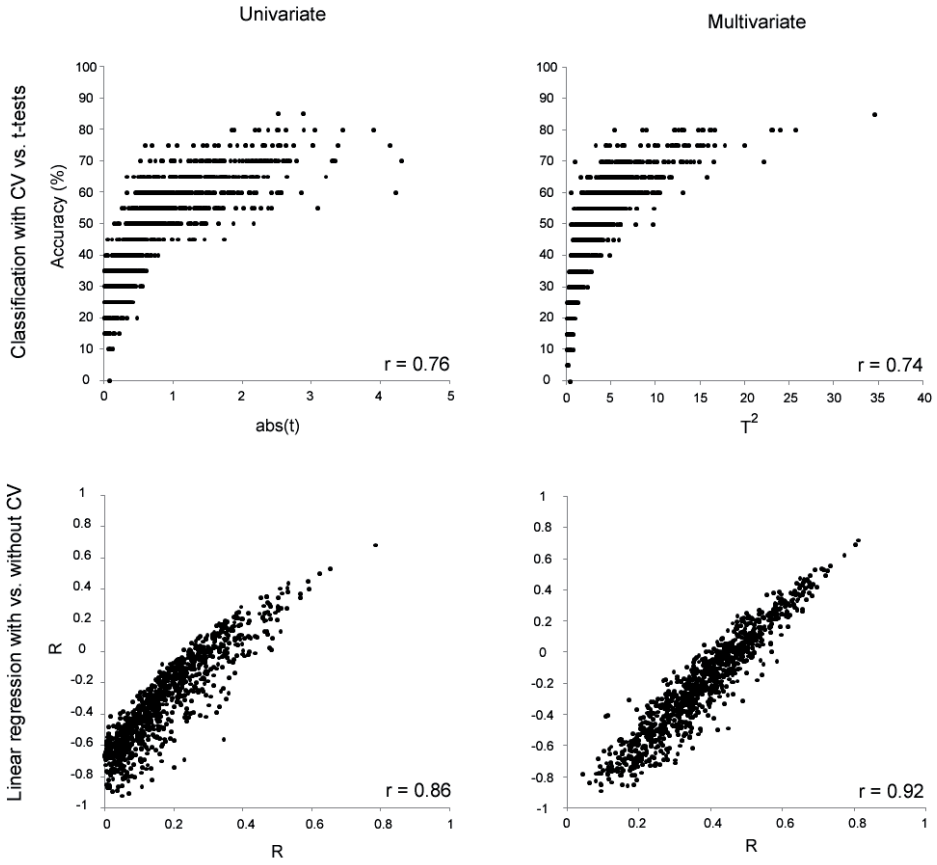

Figure 1. Relation between statistical parameters obtained by combining regression and model-validation techniques on one hand and parameters used in traditional analyses not using model-validation on the other for simulated data. In particular, the upper panel depicts this relation for binary regression (i.e. classification) with cross-validation (CV) and different t-tests without CV. Specifically, the left graph shows the link between the absolute Student t-statistic computed for 1000 single random variables (n = 20) drawn from a standard normal distribution and classification accuracy obtained for the same data with a univariate Logistic Regression classifier in a CV framework. In each of the 1000 variables, ten data points (i.e., 'subjects' or 'exemplars') corresponded to a hypothetical group A and ten corresponded to a hypothetical group B. In order to compute 'diagnostic' accuracy, a balanced leave-two-out CV approach was used, i.e. one exemplar from group A and one from group B in each iteration during the CV procedure was excluded from training and used for testing instead. To determine a simple measure of the association between absolute Student t-statistics and diagnostic accuracies, I computed the Pearson correlation coefficient that assesses the linear link between two variables. Across all 1000 variables this revealed a correlation of r = 0.76 and thus a strong (Cohen, 1988) linear dependency. The upper right

graph depicts the link of Hotellings T^2 for independent samples computed for 1000 multivariate feature sets consisting of 3 random variables drawn form a standard normal distribution on one hand and the 'diagnostic' accuracy obtained for the same data with a multivariate Logistic Regression classifier in a cross-validation framework on the other. As for the univariate case, ten data points in each of the feature sets correspond to a hypothetical group A and ten correspond to a hypothetical group B. Hotellings T^2 for independent samples can be considered as the multivariate counterpart of the Student t-statistic and is computed following $T_3^2 = \Delta M \cdot \left(\frac{Cov_A}{n_A} + \frac{Cov_B}{n_B} \right)^{-1} \cdot \Delta M^T$. In this equation, ΔM corresponds to a row vector containing the mean difference of each feature between the two groups, Cov_A and Cov_B correspond to the covariance matrices of the three features in group A and B, and n_A and n_B correspond to the group size of each of the groups. The CV method used was the same as described for the univariate case. The Pearson correlation coefficient computed between Hotellings T^2 and diagnostic accuracies across all 1000 features sets was r = 0.74 and was thus almost identical as for the univariate data set. In the lower panel, the relation between uni- (left graph) and multivariate (righ graph) continuous regression with and without model-validation is depicted. In particular, the left graph depicts the relation between correlation coefficients between true and predicted criterion vectors computed with linear regression based on a single predictor (plus constant) and a criterion vector in a leave-two-out CV framework and correlation coefficients between true and predicted criterion vectors computed without CV. Please note, that the latter technique (one predictor of interest, one constant predictor, one criterion, no CV) is sometimes also referred to as 'simple regression'. Again, this relation was computed for 1000 random predictor variables and a single random criterion (n = 20) that were all drawn from a standard normal distribution. The Pearson correlation between (Fisher-Z transformed) correlations computed with and without CV was r = 0.86, indicating a quite strong linear dependency. The right graph depicts the same association except for the fact that the random criterion was now predicted based on multivariate feature sets containing three random predictors (plus constant). For the multivariate case, the relation between (Fisher-Z transformed) correlations computed with and without CV was very similar to that computed with a univariate model, i.e. r = 0.92.

Thus, when integrating these considerations and results one can conclude that just as MVPA and CV, traditional mass univariate analysis techniques (c.f., Kiebel & Friston, 2004) without CV (e.g., t-tests [Hölzel et al., 2008] and simple regression [Weygandt et al., 2015b; 2013]) can be used to search for diagnostic and prognostic biomarkers. However, despite a similar utility of these methods in this specific aspect, MVPA and CV are obviously better suited for applied automated diagnosis and for refinement of diagnostic guidelines as they are directly addressing the separability of individual subjects and not the (dis-) similarity of group-wise parameter distributions.

2 Description of submitted articles

2.1 Diagnosing different binge-eating disorders based on reward-related brain activity patterns

In this fMRI picture perception study (Weygandt et al., 2012), we investigated how visual food cues are encoded in areas of the reward and incentive salience systems in eating disordered patients (bulimia nervosa [BN; n = 14] and binge-eating disorder [BED; n = 17]) and control subjects (overweight to obese, n = 17; normal weight, n = 19). Furthermore, we tested whether classifiers can utilize differences in food cue-elicited brain activity patterns in reward- and motivation-related areas to separate between individual BN and BED patients. This question is important, since it was long unclear whether BED is an autonomous diagnostic category or a variant of BN because BN and BED share important commonalities on the level clinical symptoms (e.g., bingeing, comparable levels of body dissatisfaction and fear of weight gain). Consistently, BED was included in the DSM-IV (American Psychiatric Association, 1994) as a provisional diagnostic category only. Inspired by this ambiguity, the study aimed to clarify whether existing diagnostic criteria of eating disorders can be refined by the application of MRI- and computer-based patient classification.

To address both questions, we used an SVC algorithm in two slightly different searchlight decoding frameworks and conducted two pattern recognition analyses. In the first analysis, we searched for brain regions in a large ROI comprising areas of the reward and the incentive salience systems (i.e. the amygdala, anterior cingulate cortex, insula, lateral and medial OFC, and the ventral striatum) that separate between high calorie food and neutral control pictures based on local searchlight fMRI activation patterns within groups. For this aim, a leave-two-out CV approach was used. In particular, in each CV iteration a SVC classifier was trained with the searchlight patterns of voxel regression coefficients reflecting the local responsivity to high-calorie pictures and with patterns of voxel regression coefficients for control pictures of all but one subject of a given group. These voxel-wise regression coefficients for both picture categories were computed in a traditional first-level within-subject analysis using linear regression / a general linear model in advance to the pattern recognition analysis. In the test stage, the given searchlight classifier was used to predict the picture categories (food vs. neutral) of both searchlight regression coefficient patterns of the left out subject.

In the second pattern recognition analysis, we used a searchlight ensemble approach and a nested CV procedure to test whether eating disordered patients (BED and BN) can be separated based on searchlight voxel contrast parameters assessing the difference between voxel regression coefficients for high-calorie minus neutral pictures. In the nested CV approach, a second-level LOO CV loop was conducted based on the training data of (i.e. ‘inside’) the first-level LOO CV loop as multivariate features selection method. In particular, after diagnostic accuracy for patient separation of both groups was determined for all searchlights in the second-level loop (i.e. based

on the patterns of all but one subject), the five best performing searchlight classifiers were selected as members for a searchlight ensemble classifier. These five local classifiers were then retrained with all training data of the first-level LOO CV loop and used to predict the class of the patient left out as test subject in the first level LOO CV loop. Finally, the average decision across classifiers in the ensemble was used for diagnosis. This ensemble decoding approach was performed for each of the subregions included in our reward and salience ROI separately (i.e., for amygdala, anterior cingulate cortex, etc.).

In the first pattern recognition analysis investigating the encoding of visual food cues, significant accuracy for separation of food and control stimuli was possible in the insula in all groups. This finding is in line with the fact that the insula contains the primary gustatory cortex (Augustine, 1996) and could thus be understood as a basic brain response pattern to food cues reflecting gustatory properties of nourishment. Importantly however, stimulus separation in areas belonging to the dopaminergic reward system (ventral striatum) and the incentive salience network (anterior cingulate cortex and amygdala) that are related to addiction-like phenomena such as craving (Pelchat et a, 2009) was only possible in eating disordered patients. In the second analysis investigating the differential diagnosis of BED vs. BN, significant classification accuracy was obtained in the anterior cingulate cortex, insular cortex and ventral striatum. Interestingly, patterns separating between individuals of the different eating disorders were consistently characterized by slightly higher activation in BN what might indicate a stronger cue-elicited urge to consume food or stronger food craving respectively in BN.

Taken together, our finding of BED and BN separability based on reward/motivation-related fMRI activation patterns fits well to the inclusion of BED as an autonomous eating disorder in the current edition of the DSM-V (American Psychiatric Association, 2013). Furthermore, the results suggest that the computer- and MRI-based approach for patient classification can contribute to the refinement of existing diagnostic guidelines for psychiatric disorders.

2.2 Role of neural impulse control mechanisms for dietary success in obesity

Here (Weygandt et al., 2013), we investigated whether neurobehavioral correlates of control in reward-related decision-making, a concept closely connected to a key criterion in substance dependence (considered as psychiatric disorder in the ICD), i.e. loss of control over begin and end of substance consumption, contains prognostic information for dietary success in obesity. Importantly, obesity is considered as a metabolic disorder in the ICD.

Influential neuroscientific studies on decision-making (e.g., Hare et al., 2011; 2009) propose that behavioural control in reward-related decision-making strongly depends on a modulatory influence of the dorsolatelral prefrontal cortex (DLPFC) on ventromedial prefrontal cortex (VMPFC). In particular, in Hare et al. (2009) self-reported dieters had to select repeatedly

between two food items, a constant reference item with average taste- and health-related properties determined for each subject individually and another food item that varied in terms of taste- and health-related aspects. The study revealed several interesting findings. First, it showed that VMPFC activity reflected the subjective value of / relative preference for food items independent of taste and healthiness of a given item (i.e. independent of how much control was necessary to prefer the item). Second, activity in the DLPFC was positively linked to successful self-control (i.e. showed higher activity when subjects chose an item that was healthier and less tasty than the reference item). Finally, VMPFC and DLPFC activity showed significant correlation during successful control trials only. From these findings, the authors concluded that the VMPFC computes a value-signal (which might also be considered as a "liking" signal in terms of the motivation/addiction literature; e.g. Volkow et al., 2002) that integrates control-related aspects of cued reinforcers computed in the DPLFC (health) and rewarding properties of these stimuli (taste).

To investigate the relevance of such processes for success of obese patients (n = 16) in a 12-week diet, we conducted an fMRI study measuring food-specific delay discounting at baseline and used neural and behavioral parameters derived thereof to predict dietary success across the 12-week interval. Dietary success was assessed as body-mass index [BMI] at the beginning of the diet minus BMI at the end of the diet, corrected for several covariates-of-no-interest (see Weygandt et al., 2013 for details regarding the covariates correction procedure). In the food-specific delay discounting task conducted directly before the beginning of the diet, subjects had to choose between an immediately gratified but small portion of a preferred meal and a larger portion of the meal gratified with a certain time delay.

We found a) a negative correlation of food-related impulsivity and dietary success, and b) value-based VMPFC and DLPFC activity with a significant positive correlation to dietary success. c) Diet-predictive DLPFC activity was negatively linked to impulsivity of food choices. Furthermore, d) connectivity between VMPFC and DLPFC was positively linked to dietary success and e) simultaneously to impulse-control in food choices. See Figure 2 for imaging results.

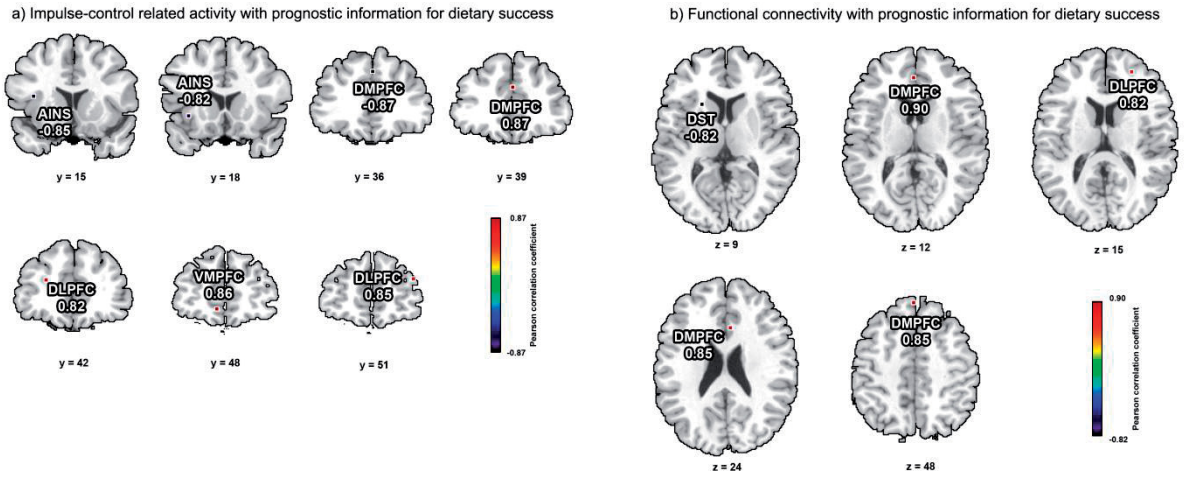

Figure 2. Neural impulse control mechanisms and their relation to dietary success in obesity. a) Control-related brain activity with prognostic information for dietary weight loss. Brain area abbreviations highlight regions with above chance correlations between control-related activity and weight loss as determined by permutation testing. Coordinates reported were significant on a threshold corrected for multiple testing with the False-Discovery-Rate method ($\alpha_{FDR} = 0.025$). Numbers beneath abbreviations correspond to Pearson correlation coefficients obtained in these areas. Indices beneath each coronar brain slice report the y-coordinate of this slice in the standard space of the Montreal Neurological Institute (MNI) brain template (Tzourio-Mazoyer et al., 2002). b) Functional connectivity of the VMPFC with prognostic information for dietary weight loss. Prognostic information was determined based on voxel parameters characterizing the connectivity of a given voxel with the VMPFC voxel depicted in (2a, i.e. for MNI: -6, 48, -9). Brain area abbreviations highlight regions with above chance ($\alpha_{FDR} = 0.025$) correlations between connectivity and weight loss as determined by permutation testing. Numbers beneath abbreviations correspond to Pearson correlation coefficients obtained in these areas. Indices beneath each transversal brain slice report the z-coordinate of this slice in the standard space of the MNI brain template. Slices in a and b are displayed in neurological orientation. Brain area abbreviations: AINS, anterior insula; DST, dorsal striatum; DLPFC, dorsolateral prefrontal cortex; DMPFC, dorsomedial prefrontal cortex; VMPFC, ventromedial prefrontal cortex. Graph and caption of the figure are adapted from Weygandt et al. (2013).

Taken together, the results of our study nicely complement the findings of Hare et al. (2009) by showing that DLPFC brain activity and VMPFC - DLPFC connectivity is not only linked to food-related control in a purely experimental framework but is also a prognostic biomarker for a real-world outcome parameter of control – dietary success in obesity. Furthermore, our findings are well in line with the assumptions of the influential 'impaired response inhibition and salience attribution' theory proposed for addiction by e.g. Volkow et al. (2010) and Goldstein & Volkow (2002). This theory holds that inhibitory influences of the (dorsal) prefrontal cortex on structures computing a value signal (such as the ventral striatum or the VMPFC) are reduced in drug addiction due to an insensitivity induced by overstimulation.

Following this theory, reduced top-down influence covaries with disinhibition of stimulus driven behaviours such as drug craving and thus consumption.

2.3 Impulse control in the dorsolateral prefrontal cortex counteracts post-diet weight regain in obesity

In this study (Weygandt et al., 2015b), we investigated whether control-related brain activity measured with fMRI directly after the end of a 12-week very low calorie diet can be considered as prognostic biomarker for real-world success in sustained post-diet weight maintenance.

As we and others have shown, a variety of methods exist to induce short-time dietary success in obese patients (e.g., Weygandt et al., 2013; Shaw et al., 2006; Sheperd, 2003). Unfortunately however, a large number of subjects start regaining weight when weight-loss treatment ends - independent of the type of treatment applied e.g. psychological as well as pharmacological (e.g., Sjöström et al., 1998; Wadden et al., 1998). Consequently, we continued our longitudinal line of research started in Weygandt et al. (2013) and now searched for neurobehavioral biomarkers of sustained weight maintenance across a one-year post-diet interval. To achieve this goal, we applied the same experimental fMRI delay-discounting paradigm in the same pool of subjects as in Weygandt et al. (2013). In particular, we measured food-related neurobehavioral impulse control at two time points, i.e. the beginning ('T0') and the end ('T12') of a one-year weight maintenance interval directly following a 12-week diet (addressed in Weygandt et al., 2013). At T0, fMRI data of 23 subjects were assessed and fMRI data of 19 subjects were assessed at T12. We tested whether activity in DLPFC at T0 and whether activity changes in this area across the weight maintenance period (T0 - T12) were linked to the degree of success in maintenance (i.e. BMI at T0 - BMI at T12, corrected for several covariates-of-no-interest). The same was done for behavioral impulsivity parameters Figure 3 depicts the results for behavioral impulse control parameters and their link to the degree of success in one-year post diet weight maintenance. Figure 4 depicts the results for their neural counterparts.

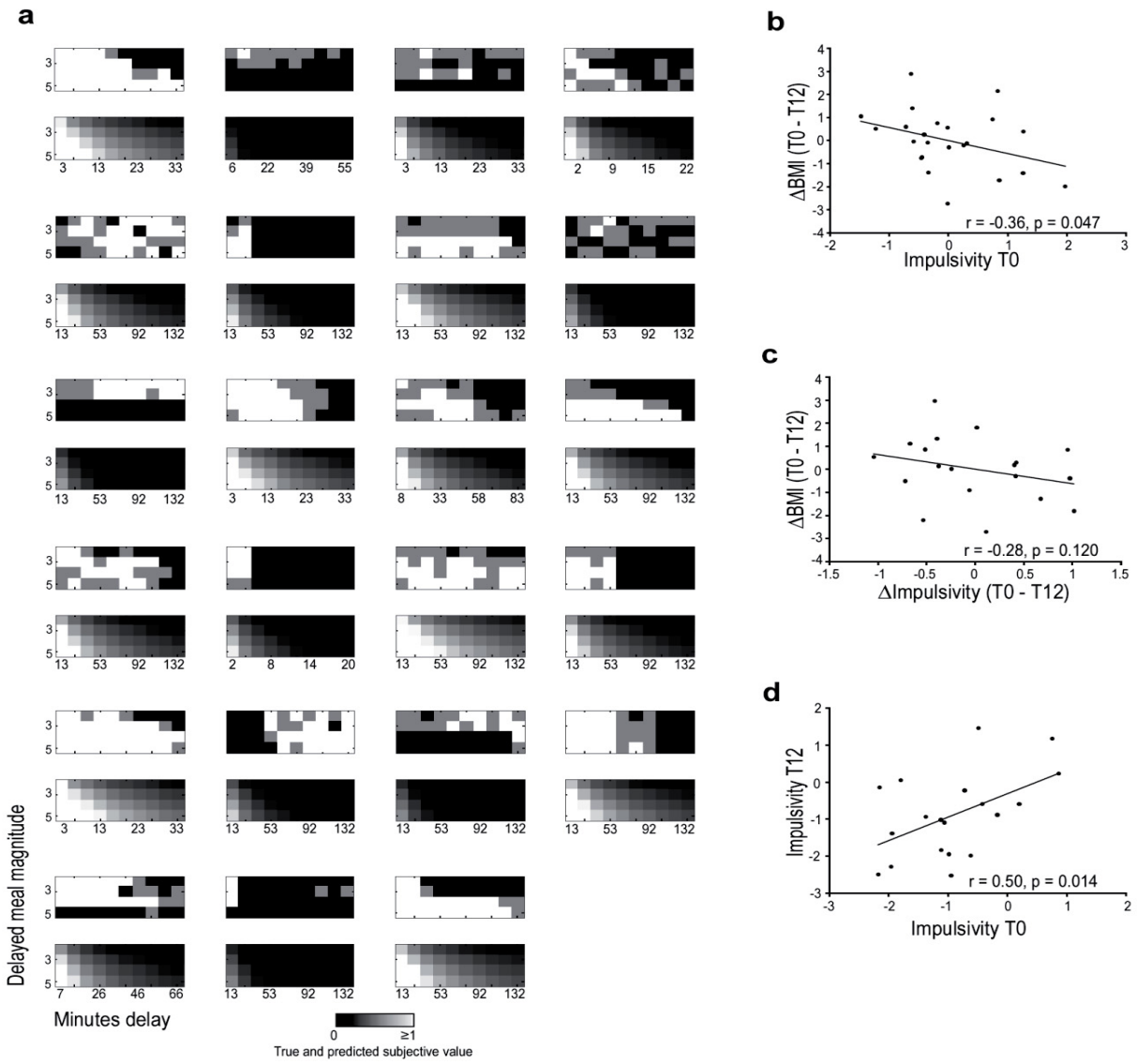

Figure 3. Delay discounting, behavioral impulsivity and success in weight maintenance. a) Value of delayed meals for each subject at T0 as function of meal magnitude and delay. For each subject, a pair of vertically aligned graphs depicts the true value of delayed meals in the upper graph and the value predicted by a hyperbolic discounting function (c.f., Mazur, 1987; see below) in the lower graph. Numbers at the bottom of a pair depict subject-specific delay times. Left to each row of graphs the magnitude of delayed meal options is depicted. The true subjective value in the upper graph was computed as the average of run-wise choices (0 = immediate, 1 = delayed) for each combination of meal magnitude and delay. Consequently, the true subjective value corresponds to the empirical choice probability of the delayed meal option for a given pair of meal magnitude and delay. The predicted subjective value for each combination of meal magnitude and delay in the lower graph was computed following a hyperbolic discounting function, i.e. subjective value = meal magnitude / (1 + behavioral impulsivity k_{T0} · meal delay). The pairs of graphs are sorted by success in weight maintenance.

Specifically, the pair of graphs for the least successful subject is depicted in the upper left, the pair for the most successful subject in the lower right. b) Correlation of impulsivity at T0 and weight maintenance. c) Correlation of longitudinal impulsivity changes and weight maintenance. d) Stability of food-related impulsivity. The figure and its caption are adapted from Weygandt et al. (2015b).

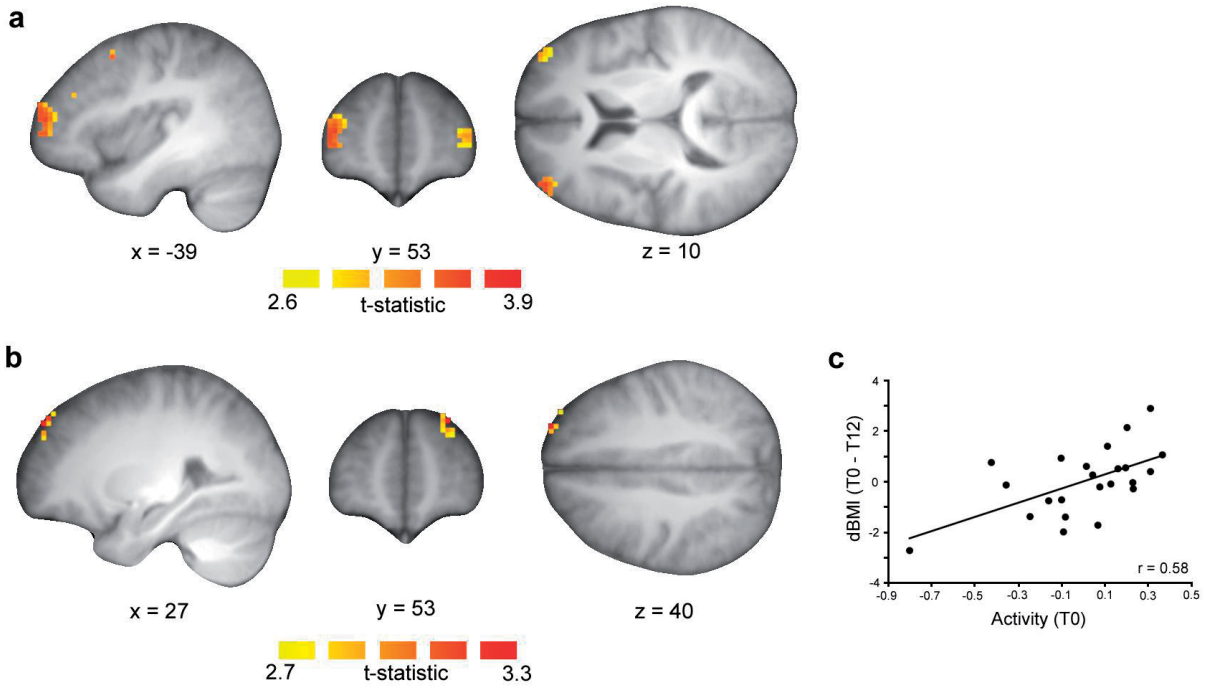

Figure 4. DLPFC activity reflecting impulse control and its relation to the degree of weight maintenance. a) DLPFC activity in the anatomical standard space of the Montreal Neurological Institute (MNI; Tzourio-Mazoyer et al., 2002) brain reflecting impulse control at T0, centered on the peak voxel (MNI: -39, 53, 10). b) Activity at T0 with information for weight maintenance, centered on the voxel with peak information (MNI: 27, 53, 40). c) Correlation of activity at T0 for MNI: 27, 53, 40 and weight maintenance. Coronal slices are shown in neurological orientation. The figure and its caption are adapted from Weygandt et al. (2015b).

Taken together, the analyses conducted show that control-related DLPFC activity at T0 is coupled to success in weight maintenance. Thus, control-related brain activity measured with fMRI directly after the end of a 12-week very low calorie diet can be considered as prognostic biomarker for real-world success in sustained post-diet weight maintenance in obesity.

2.4 Investigation of mindfulness meditation practitioners with voxel-based morphometry

In this study, we investigated brain morphological biomarkers of mindfulness meditation (Hölzel et al., 2008). Mindfulness meditation is a technique that is characterized by non-evaluative observations of the stream of interoceptive sensations and has been shown to improve overall well-being in healthy subjects (Brown & Ryan, 2003) but also symptoms in various disorders such as MS (Grossmann et al., 2010), pain, depression and anxiety (Grossmann et al., 2003). Consequently, identification of brain-derived biomarkers of mindfulness meditation can be important for the treatment of neurologic as well as psychiatric disorders.

To investigate markers of mindfulness meditation, we conducted two analyses. In particular, we analyzed differences in local gray matter concentration between a group of mindfulness meditation practitioners (n = 20) and non-meditating control subjects (n = 20; matched with regard to gender, age, education and handedness) on a voxel-level using two-sample t-tests in the first analysis. Furthermore, we tested whether we could predict the mean gray matter concentration in meditation practitioners in several ROI known to be involved in mindfulness meditation on a functional level (Cahn & Polich, 2006) based on the overall meditation practice in a second analysis using linear multiple regression.

The first analysis revealed that meditation practitioners were characterized by greater grey matter concentration than non-meditating controls in the right anterior insula, the left inferior temporal gyrus and the right hippocampus. Furthermore, the second analysis showed that average grey matter concentration in the left inferior temporal cortex could be predicted by the hours of meditation practice. Importantly, identification of anterior insula is consistent with findings in a similar study investigating differences of cortical thickness between meditation practitioners and control subjects (Lazar et al., 2005) and with its involvement in interoceptive perception in normal healthy subjects (Critchley et al., 2004). Furthermore, identification of hippocampus is consistent with findings showing that this area regulates cortical arousal and responsiveness (Newberg & Iversen, 2003).

To summarize, in this study we were able to identify MRI biomarkers of mindfulness meditation, a mental technique that has been proven to improve symptoms in neurologic (Grossmann et al., 2010) as well as psychiatric disorders (Grossmann et al., 2003).

2.5 Can we overcome the 'clinico-radiological paradox' in multiple sclerosis?

Inspired by the fact that established lesion-related biomarkers for MS such as the overall number and volume are only weakly associated with clinical disability in MS (a fact that is known as 'clinico-radiological paradox'; Zivadinov & Leist, 2005; Miki et al., 1999; Filippi et al., 1995),

we conducted a study that should clarify the link between locally specific MRI voxel intensity alterations and symptom severity in relapsing-remitting MS patients (RRMS; n = 40) assessed in several functional domains (Hackmack et al., 2012a). Specifically, we used CCA to determine the link between local searchlight voxel intensity patterns and cognitive dysfunction, motor and overall disability, and finally disease duration in a LOO CV framework. Cognitive dysfunction was assessed by the Paced Auditory Serial Addition Test (PASAT; Cutter et al., 1999), motor disability was measured by the Timed 25-Foot Walk Test [TWT; Cutter et al., 1999] and the 9-Hole Peg Test [9-HPT; Cutter et al., 1999]), and overall disability by the EDSS (Kurtzke, 1983). In this study, the searchlight CCA approach was performed based on voxel intensity signals sampled from T1- (i.e. magnetization prepared rapid gradient echo) and T2-weighted (i.e. Turbo Inversion Recovery Magnitude) MRI images.

As expected, searchlight CCA revealed local MRI voxel intensity variations with information for cognitive dysfunction in areas involved in working memory such as posterior parietal cortex. Furthermore, also areas in middle cingulate gyrus, inferior temporal lobe, and fusiform gyrus contained predictive information for cognitive impairment. In addition, motor disability could be predicted in subregions of the cerebellum, thalamus, and primary motor cortex as well as in posterior parietal cortex and middle frontal gyrus. For overall disability a cluster in the inferior frontal gyrus contained maximal predictive information. Finally, the analysis revealed areas in somatosensory cortex and posterior parietal cortex as being predictive of disease duration.

Taken together, contrary to the traditional procedure linking lesion-related parameters aggregated across the whole brain to clinically disability in MS, the searchlight approach revealed a variety of regionally specific MRI biomarkers for disability in RRMS that are in good accordance with existing knowledge on the functional specialization of these regions (see e.g., Dehaene et al., 2004). Together with the predictive accuracy obtained in the LOO CV framework, these results suggest that the proposed method might be able to overcome the so-called clinico-radiological paradox in MS.

2.6 MRI-based diagnostic biomarkers for early onset pediatric multiple sclerosis

A growing body of research addresses neurological (e.g. Bigi & Banwell, 2012; Vargas-Lowy et al., 2012), epidemiological (e.g., Renoux et al., 2007; Ruggieri et al., 2004) and diagnostic (e.g., Polman et al., 2011; Sadaka et al., 2012) features of pediatric MS (PMS) which is typically defined as a variant of MS that occurs with an onset of prior or up to an age of 16 (e.g. Sadaka et al., 2012; Renoux et al., 2007). Within this age-dependent definition, a prevalence of 2.2% to 5% has been reported for PMS among all MS patients (e.g., Renoux et al., 2007; Ruggieri et al., 2004). However, despite these efforts it is yet unclear whether the term PMS denotes a single,

homogenous disease. For example, a study evaluating the validity of the revised McDonald criteria for the diagnosis of PMS (Sadaka et al., 2012) found that these criteria were not well suited to predict the development of clinically definite MS after acute demyelination for early onset pediatric MS (EOPMS) patients with onset prior to the age of eleven. Moreover, a study that compared the characteristics of hyperintense lesions in T2-weighted images between EOPMS and late onset pediatric MS (LOPMS) patients found that EOPMS patients had less well-defined ovoid lesions (Chabas et al., 2008). Furthermore, EOPMS seems to be associated with a different pattern of symptoms on a clinical level. Specifically, Ruggieri et al. (1999) found that the predominant pattern for EOPMS patients involved consciousness disturbances, seizures, and ataxia as compared to predominant optic nerve involvement in LOPMS and adult onset MS. Based on such findings, it was recently questioned whether it might be useful to complement the MRI-based diagnostic criteria for EOPMS (Chabas et al., 2008).

Consequently, we combined model-validation and univariate logistic regression techniques based on tissue probability parameters of single voxels to search for novel MRI-based diagnostic biomarkers for EOPMS in this study (Weygandt et al., 2015a). EOPMS was defined as MS with an onset occurring prior to the age of twelve. In particular, we investigated the separability of EOPMS patients (n = 16) vs. LOPMS patients (onset < 16 yrs; n = 17) based on tissue probability parameters extracted from areas located in grey and white matter separately. Moreover, we also investigated the separability of EOPMS patients vs. HC subjects (n = 15) based on these parameters and finally the separability of LOPMS and HC. Patients in the LOPMS group and HC subjects were matched to the patients of the EOPMS group with regard to gender and the two PMS groups were additionally matched in terms of disease duration, age and lesion load (please see Weygandt et al. [2015a] for details regarding the matching procedure).

Results obtained by the logistic regression analyses searching for diagnostic information confirmed existing assumptions on the key role of hyperintense periventricular WM lesions for the diagnosis of (P)MS. However, by using the combined regression and model-validation approach, we also found novel diagnostic biomarkers for EOPMS in frontal gyri and WM areas adjacent to these gyri. Consequently, we were able to differentiate EOPMS from LOPMS as well as both PMS groups from HC based on local tissue probability information extracted from GM and WM areas in datasets acquired with clinical routine brain MRI sequences in this study. Importantly, our results validate existing assumptions on the key role of hyperintense WM lesions for MS diagnosis but we also found novel non-lesion related biomarkers for PMS. The latter finding suggests that conventional MRI contains a richer set of diagnostically relevant features related to MS than is typically assumed. Furthermore, diagnostic classification was independent of group differences in gender, disease duration, age, and lesion load, as these parameters were balanced. Thus, the MRI biomarkers identified reflect neuropathological processes specific for MS with very early disease onset.

2.7 Multi-scale classification of disease using structural MRI and wavelet transform

Consistent with the prompt to search for novel MRI biomarkers in MS (Charil et al., 2006) due to the only moderate MS-specificity of brain lesions and their only moderate link to clinical disability in MS (Zivadinov & Leist, 2005; Miki et al., 1999; Filippi et al., 1995), we tested the diagnostic utility of wavelet transformations of raw spatial MRI voxel intensity signals as features for an MVPA approach in this study (Hackmack et al., 2012b). In short, similar to Fourier Analysis, Wavelet Analysis is a complex mathematical technique for analyzing systematic signal alterations of varying frequency (see e.g. Graps [1995] for an overview). The latter ability makes both methods a potential candidate for analyzing systematic spatial signal variations in structural MRI images on different spatial scales simultaneously. However, because Wavelet Analysis is better suited to decompose discontinuous or non-stationary signals than Fourier Analysis due the finite character of basis functions evaluated (i.e. the Wavelets), we decided to use the former method to analyze spatial MRI signal alterations on multiple spatial resolutions. In particular, the spatial extent of these scales ranged from cuboid-shaped areas of few mm^3 volume on the smallest scale to cuboid-shaped areas covering 12.5% of the field of view on the largest. Furthermore, the method was applied separately to a dataset consisting of structural MRI data of RRMS patients (n = 41) and HC subjects (n = 26), and a dataset consisting of structural MRI data of AD patients (n = 20) and HC subjects (n = 20) taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

For each of the two data sets, two different types of features were derived from raw spatial MRI voxel intensity signals and thus also two different analyses were conducted for each data set. First, Wavelets were used to characterize signal variations across three large ('global') tissue-dependent ROI (areas containing lesions, areas containing normal-appearing brain tissue, and the whole brain) on six different spatial scales for each subject included in each of the two data sets separately. On each spatial scale, three-dimensional Wavelet basis functions with 28 different spatial orientations were used to characterize signal variations across the whole ROI. After the fit of each orientation-specific Wavelet basis function was determined, these fit parameters (which might roughly be understood as indicators of explained variance), were combined to yield a pattern for each ROI, scale, subject, and data set containing 28 features. Then, a LOO CV MVPA approach using an SVC algorithm was applied to compute diagnostic separability of groups based on these global indicators of signal variation specifically for each scale and dataset. In the second analysis, local Wavelet coefficients were determined as indicators of local signal variations. As in the global analysis, 28 orientation-specific Wavelet basis-functions were used to characterize systematic local signal alterations for each of six specific spatial scales in each of the ROI. In the next step, a LOO CV MVPA approach using a SVC algorithm was applied to compute diagnostic separability of groups based on local Wavelet indicators of signal variation.

For the separation of RRMS and HC based on global signal variations, Wavelet coefficients derived from lesioned tissue separated best. Specifically, consistent with lesion load only in RRMS and not HC, patients were characterized by more/stronger systematic signal variations than HC subjects. For the separation of RRMS and HC based on local wavelet coefficients, similar results were obtained. Interestingly, especially for lesioned tissue, separability increased together with the spatial scale of Wavelet basis functions across all positions in the brain in the local analyses. For the separation of AD and HC, the results were less conclusive, i.e. accuracy obtained was lower and separability was less specific for selected (i.e. the larger) scales than in separation of RRMS and HC.

Taken together, this was the first study that characterized the utility of diagnostic MRI biomarkers for MS and AD in terms of the spatial extent of features. Importantly, it showed that diagnostically relevant neuropathology induces large-scale structural changes in RRMS e.g., such as ventricle widening due to thalamic atrophy (Houtchens et al., 2007).

2.8 Multimodal prediction of conversion to Alzheimer based on incomplete biomarkers

In this study (Ritter et al., 2015), we predicted conversion from normal cognition (NC) or mild cognitive impairment (MCI) to AD in individual subjects in a three-year follow-up period after baseline using extensive multimodal data taken from the ADNI database. In particular, a set of 288 features from 10 different domains or modalities respectively was included in the analysis taken from 237 subjects that were diagnosed as NC or MCI patients at baseline. For 86 of these subjects, the diagnosis changed to AD during the three-year interval, for 151 it did not. Features were derived from these modalities: MRI (lesion load, volume and voxel information), Positron Emission Tomography, laboratory measurements, medical history, medical symptoms at baseline, neurological and physical exams, neuropsychological testing and demography.

Importantly, we addressed several problems in this study that have to be solved if the computer- and MRI-based diagnostic approach proposed should be established as a useful tool in a clinical real-world setting for automated diagnosis. Specifically, we first investigated the problem of how to deal with missing data. This aspect might be especially important in a highly multimodal clinical real-world setting since most likely parameters acquired will vary across hospitals, attending physicians, and patients. To address this aspect, we compared the results of three different interpolation methods for missing data handling on prognostic accuracy (i.e. a method using measures of central tendency, a method using an Expectation-Maximization [EM] approach, and a combination). Furthermore, we tested procedures to reduce the impact of unbalanced training data (i.e. unequal group sizes of converters and non-converters) on prognostic accuracy. Finally, we compared the utility of different feature selection approaches

(selection of features by human experts and different automated feature selection methods) on accuracy.

Maximal accuracy in predicting conversion from NC or MCI to AD in a three-year interval after baseline across all methods compared was 73% (mean of sensitivity and specificity). Furthermore, maximal modality-specific accuracies were obtained based on neuropsychological markers and markers derived from both imaging modalities. The best individual feature was the Functional Activities Questionnaire (Pfeffer et al., 1982). In addition, the results show that methods used for handling of unbalanced training data had a significant effect only when using an SVC algorithm for prognosis, i.e. the impact of such methods was classifier-specific. Presumably due to the presence of features measured on a continuous but also an ordinal level together with the unsuitability of the EM-method to handle ordinal data, the combined interpolation method was best suited for missing data handling. Finally, although feature selection by human experts achieved highest accuracy among all methods compared, also automated features selection methods significantly improved prognostic accuracy compared to no feature selection.

Taken together, the results of the study suggest that a combination of automated missing data substitution, automated feature selection and computer-based classification can be used in a clinical real-world setting for automated diagnostics or as a decision support system in the closer future.

3 Conclusion

This habilitation synopsis gives an overview on methodological aspects of the computer-based disease prediction approach, its development, the research aims pursued with it, and it describes the research articles from this field that I submit for habilitation. In particular, the synopsis starts by outlining the basic concept of the approach and by describing the three fundamental procedural stages characterizing it, i.e. the feature determination, training and test stages. Then, I continued by delineating the development of the approach in terms of three epochs that are characterized by the discovery of diagnostic information in MR signals, the first use of statistical regression techniques to analyze this information, and the mass use of the approach. Finally, I outlined research aims pursued with the approach, i.e. automated diagnosis, refinement of diagnostic guidelines, and identification of novel diagnostic biomarkers. In the second part of the synopsis, I described the peer-reviewed research articles submitted for habilitation. Across these articles or studies respectively, all of the three research aims pursued with the approach were addressed. Furthermore, technical challenges connected to the approach were addressed in various different fashions. Thus, these studies provide a substantial overview on the methodological diversity of the field.

4 References*

Allefeld C, Haynes JD. Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *NeuroImage*. 2014; 89: 345–357.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th ed. 2013.Washington, DC: American Psychiatric Association.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. 1994.Washington, DC: American Psychiatric Association.

Anderson JS, Nielsen JA, Froehlich AL, DuBray MB, Druzgal TJ, et al. Functional connectivity magnetic resonance imaging classification of autism. *Brain*. 2011; 134: 3742–3754.

Augustine JR. Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Res Rev*. 1996; 22: 229–244.

Barkhof F, Filippi M, Miller DH, Scheltens P, Campi A, et al. Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain*. 1997; 120: 2059 - 2069.

Bendfeldt K, Klöppel S, Nichols TE, Smieskova R, Kuster P, et al. Multivariate pattern classification of gray matter pathology in multiple sclerosis. *NeuroImage*. 2012; 60: 400–408.

Bigi S, Banwell B. Pediatric multiple sclerosis. *J Child Neurol*. 2012; 27:1378-1383.

Bohland JW, Saperstein S, Pereira F, Rapin J, Grady L. Network, anatomical, and non-imaging measures for the prediction of ADHD diagnosis in individual subjects. *Front Syst Neurosci*. 2012; 6: 78. doi: 10.3389/fnsys.2012.00078. eCollection 2012.

Bray S, Chang C, Hoeft F. Applications of multivariate pattern classification analyses in developmental neuroimaging of healthy and clinical populations. *Front Hum Neurosci*. 2009; 3:32. doi: 10.3389/neuro.09.032.2009. eCollection 2009.

Brown KW, Ryan RM. The benefits of being present: Mindfulness and its role in psychological well-being. *J Pers Soc Psychol*. 2003; 84: 822 - 848.

Cahn BR, Polich J. Meditation states and traits: EEG, ERP, and neuroimaging studies. *Psychol Bull*. 2006; 132: 180 – 211.

Chabas D, Castillo-Trivino T, Mowry EM, Strober JB, Glenn OA, Waubant E. Vanishing MS T2-bright lesions before puberty. A distinct MRI phenotype? *Neurology*. 2008; 71: 1090 - 1093.

Charil A, Yousry TA, Rovaris M, Barkhof F, De Stefano N, et al. MRI and the diagnosis of multiple sclerosis: expanding the concept of ’no better explanation’. *Lancet Neurol*. 2006; 5: 841-852.

* Articles submitted for habilitation are cited in bold face.

Chen Y, Storrs J, Tan L, Mazlack LJ, Lee JH, et al. Detecting brain structural changes as biomarker from magnetic resonance images using a local feature based SVM approach. *J Neurosci Methods*. 2014; 221: 22 - 31.

Chen J, Phan TG, Reutens DC. Ridge penalized logistical and ordinal partial least squares regression for predicting stroke deficit from infarct topography. *J Biomed Sci Eng*. 2010; 3: 568 – 575.

Cohen JR, Asarnow RF, Sabb FW, Bilder RM, Bookheimer SY, et al. Decoding continuous variables from neuroimaging data: basic and clinical applications. *Front Neurosci*. 2011; 5: 75. doi: 10.3389/fnins.2011.00075. eCollection 2011.

Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd ed., 1988. Hillsdale: Lawrence Erlbaum Associates.

Critchley HD, Wiens S, Rotshtein P, et al. Neural systems supporting interoceptive awareness. Nature Neurosci. 2004; 7: 189–195.

Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage*. 2011; 56:766 - 781.

Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain*. 1999; 122: 871 – 882.

Damadian R. Tumor detection by nuclear magnetic resonance. *Science*. 1971; 171: 1151 – 1153.

Dehaene S, Molko N, Cohen L, Wilson AJ. Arithmetic and the brain. *Curr Opin Neurobiol*. 2004; 14: 218 – 224.

Demirci O, Clark VP, Magnotta VA, Andreasen NC, Lauriello J, et al. A Review of Challenges in the Use of fMRI for Disease Classification / Characterization and A Projection Pursuit Application from Multi-site fMRI Schizophrenia Study. *Brain Imaging Behav*. 2008; 2: 147 - 226.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G., et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. 2013; 36: 27 – 46, 2013.

Duda RO, Hart P, Stork DG, 2000. Pattern Classification. Wiley-Interscience; 2nd ed.

Dukart J, Mueller K, Barthel H, Villringer A, Sabri O, et al. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. *Psychiatry Res*. 2013; 212: 230 - 236.

Fazekas F, Offenbacher H, Fuchs S, Schmidt R, Niederkorn K, et al. Criteria for an increased specificity of MRI interpretation in elderly subjects with suspected multiple sclerosis. *Neurology*. 1988; 38: 1822–1825.

Filippi M, Paty DW, Kappos L, Barkhof F, Compston DA, et al. Correlations between changes in disability and T2-weighted brain MRI activity in multiple sclerosis: a follow-up study. *Neurology*. 1995; 45: 255-260.

Folstein MF, Folstein SE, McHugh PR. Mini-Mental State (a practical method for grading the state of patients for the clinician). *J Psychiat Res*. 1975; 12: 189–198.

Ford J, Farid H, Makedon F, Flashman LA, McAllister TW, et al. Patient Classification of fMRI Activation Maps. Lecture Notes in Computer Science Volume. 2003; 2879: 58-65.

Ford J, Shen L, Makedon F, Flashman LA, Saykin AJ. A combined structural-functional classification of schizophrenia using hippocampal volume plus fMRI activation. In Second joint EMBS/BMES conference. 2002. Houston, TX.

Fu CH, Williams SC, Cleare AJ, Scott J, Mitterschiffthaler MT, et al. Neural responses to sad facial expressions in major depression following cognitive behavioral therapy. *Biol Psychiatry*. 2008; 64: 505-512.

Goldstein RZ, Volkow ND. Drug addiction and its underlying neurobiological basis: neuroimaging evidence for the involvement of the frontal cortex. *Am J Psychiatry*. 2002; 159:1642–1652.

Good P. Permutation, Parametric and Bootstrap Tests of Hypotheses. 3rd ed. 2005. New York: Springer.

Graps A. An introduction to wavelets. *Comput Sci Eng*. 1995; 2: 50–61.

Grossman P, Kappos L, Gensicke H, D'Souza M, Mohr DC, et al. MS quality of life, depression, and fatigue improve after mindfulness training: a randomized trial. *Neurology*. 2010; 75: 1141 - 1149.

Grossman P, Niemann L, Schmidt S, Walach H. Mindfulness-based stress reduction and health benefits: a meta-analysis. *J Psysom Res*. 2003; 57: 35–43.

Hackmack K, Weygandt M, Würfel J, Pfueller CF, Bellmann-Strobl J, et al. Can we overcome the 'clinico-radiological paradox' in multiple sclerosis? *J Neurol*. 2012a; 259: 2151 - 2160.

Hackmack K, Paul F, Weygandt M, Allefeld C, Haynes JD, et al. Multi-scale classification of disease using structural MRI and wavelet transform. *NeuroImage*. 2012b; 62: 48 – 58.

Hahn T, Marquand AF, Ehlis AC, Dresler T, Kittel-Schneider S, et al. Integrating neurobiological markers of depression. Arch Gen Psychiatry. 2011; 68: 361 – 68.

Hare TA, Malmaud J, Rangel A. Focusing attention on the health aspects of food changes value signals in vmpfc and improves dietary choice. *J Neurosci*. 2011; 31: 11077 – 11087.

Hare TA, Camerer CF, Rangel A. Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System. *Science.* 2009; 324, 646–648.

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, et al. Reading hidden intentions in the human brain. *Curr Biol*. 2007; 17: 323 – 328.

Haynes JD, Rees G. Decoding mental states from brain activity in humans. *Nat Rev Neurosci*. 2006; 7: 523 – 534.

Hoexter MQ, Miguel EC, Diniz JB, Shavitt RG, Busatto GF, et al Predicting obsessive-compulsive disorder severity combining neuroimaging and machine learning methods. *J Affect Disord*. 2013; 150: 1213 – 1216.

Hölzel BK, Ott U, Gard T, Hempel H, Weygandt M, et al. Investigation of mindfulness meditation practitioners with voxel-based morphometry. *Social Cognitive and Affective Neuroscience*. 2008; 3: 55 - 61.

Houtchens MK, Benedict RH, Killiany R, Sharma J, Jaisani Z, et al. Thalamic atrophy and cognition in multiple sclerosis. *Neurology*. 2007; 69: 1213 - 1223.

Kawasaki Y, Suzuki M, Kherif F, Takahashi T, Zhou SY, et al. Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. *NeuroImage*. 2007; 34: 235 - 242.

Kidwell CS. MRI biomarkers in acute ischemic stroke: a conceptual framework and historical analysis. *Stroke*. 2013; 44: 570 – 578.

Kiebel SJ, Friston KJ. Statistical parametric mapping for event-related potentials: I. Generic considerations. *NeuroImage.* 2004; 22: 492 - 502.

Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, et al. Automatic classification of MR scans in Alzheimer's disease. *Brain.* 2008a; 131: 681 – 689.

Klöppel S, Stonnington CM, Barnes J, Chen F, Chu C, et al. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain.* 2008b; 131: 2969 - 2974.

Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. *Proc Natl Acad Sci USA*. 2006; 103: 3863–3868.

Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983; 33: 1444 – 1452.

Lazar SW, Kerr CE, Wasserman RH, Gray JR, Greve DN, et al. Meditation experience is associated with increased cortical thickness. *Neuroreport.* 2005; 16: 1893–1897.

Lee GY, Kim J, Kim JH, Kim K, Seong JK. Online Learning for Classification of Alzheimer Disease based on Cortical Thickness and Hippocampal Shape Analysis. *Healthc Inform Res*. 2014; 20: 61 - 68.

Long D, Wang J, Xuan M, Gu Q, Xu X, et al. Automatic classification of early Parkinson's disease with multi-modal MR imaging. *PLoS One.* 2012; 7:e47714. doi: 10.1371/journal.pone.0047714. Epub 2012 Nov 9.

Mazur JE. An adjusting procedure for studying delayed reinforcement. In: Commons ML, Mazur JE, Nevin JA, Rachlin H, editors. Quantitative Analysis of Behavior: The Effect of Delay and of Intervening Events of Reinforcement Value, 1987. vol. 5. Erlbaum, Hillsdale, NJ, pp. 55–73.

McEvoy LK, Fennema-Notestine C, Roddey JC, Hagler DJ Jr, Holland D, et al. Alzheimer disease: Quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. *Radiology*. 2009; 251: 195–205.

McGrath CL, Kelley ME, Holtzheimer PE, Dunlop BW, Craighead WE, et al. Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry*. 2013; 70: 821 - 829.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDSADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology*. 1984; 34: 939 – 944.

Meyer-Lindenberg A, Poline JB, Kohn PD, Holt JL, Egan MF, et al. Evidence for abnormal cortical functional connectivity during working memory in schizophrenia. *Am J Psychiatry*. 2001; 158: 1809 – 1817.

Miki Y, Grossman RI, Udupa JK, Wei L, Polansky M, et al. Relapsing-remitting multiple sclerosis: longitudinal analysis of MR images-lack of correlation between changes in T2 lesion volume and clinical findings. *Radiology*. 1999; 213: 395-399.

Müller KR, Mika S, Rätsch G, Tsuda K, Schölkopf B. An introduction to kernel-based learning algorithms. *IEEE Trans Neural Netw*. 2001; 12: 181 - 201.

Namer IJ, Yu O, Mauss Y, Dumitresco BE, Chambron J. An evaluation of the significance of areas of intense signal in the MR brain images of patients with multiple sclerosis. *Magn Reson Imaging*. 1993; 11: 311 - 317.

Newberg AB, Iversen J. The neural basis of the complex mental task of meditation: neurotransmitter and neurochemical considerations. *Medical Hypotheses*. 2003; 61: 282 – 291.

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mindreading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci*. 2006; 10: 424 – 430.

Paty DW, Oger JJF, Kastrukoff LF, Hashimoto SA, Hooge JP, et al. MRI in the diagnosis of MS: a prospective study with comparison of clinical evaluation, evoked potentials, oligoclonal banding, and CT. *Neurology*. 1988; 38: 180 – 185.

Pelchat ML. Food Addiction in Humans. *J Nutrition*. 2009; 139: 620 – 622.

Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage.* 2009; 45:S199 – S209.

Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. *J Gerontol.* 1982; 37: 323 - 329.

Pizzagalli DA. Frontocingulate dysfunction in depression: toward biomarkers of treatment response. *Neuropsychopharmacology*. 2011; 36: 183 - 206.

Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol*. 2001; 69: 292 – 302.

Ponseti J, Granert O, Jansen O, Wolff S, Beier K, et al. Assessment of pedophilia using hemodynamic brain response to sexual stimuli. *Arch Gen Psychiatry*. 2012; 69: 187 - 194.

Renoux C, Vukusic S, Mikaeloff Y, Edan G, Clanet M, et al. Natural history of multiple sclerosis with childhood onset. *New Engl J Med*. 2007; 356: 2603 - 2613.

Ritter K, Schumacher J, Weygandt M, Buchert R, Allefeld C, et al. Multimodal prediction of conversion to Alzheimer based on incomplete biomarkers. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2015. Accepted.

Ruggieri M, Iannetti P, Polizzi A, Pavone L, Grimaldi LM, 2004. Multiple sclerosis in children under 10 years of age. *Neurol Sci*. 2004; 25(Suppl. 4): S326 – 335.

Ruggieri M, Polizzi A, Pavone L, Grimaldi LM. Multiple sclerosis in children under 6 years of age. *Neurology*. 1999; 53: 478 - 484.

Sadaka Y, Verhey LH, Shroff MM, Branson HM, Arnold DL, et al.Canadian Pediatric Demyelinating Disease Network 2010 McDonald Criteria for Diagnosing Pediatric Multiple Sclerosis. *Ann Neurol.* 2012; 72: 211–223.

Saur D, Ronneberger O, Kümmerer D, Mader I, Weiller C, et al. Early functional magnetic resonance imaging activations predict language outcome after stroke. *Brain.* 2010; 133: 1252 - 1264.

Schneider JA, Aggarwal NT, Barnes L, Boyle P, Bennett DA. The neuropathology of older persons with and without dementia from community versus clinic cohorts. *J Alzheimers Dis*. 2009;18: 691 - 701.

Shaw KA, Gennat HC, O'Rourke P, Del Mar C. Exercise for overweight or obesity. Cochrane Database Syst Rev CD003817. 2006.

Shepherd TM. Effective management of obesity. *J Fam Pract.* 2003; 52, 34-42.

Sjöström, L., Rissanen, A., Andersen, T., Boldrin, M., Golay, A., et al, Randomised placebo-controlled trial of orlistat for weight loss and prevention of weight regain in obese patients. *Lancet.* 1998; 352, 167 – 172.

Stone JV. Independent component analysis: an introduction. *Trends Cogn Sci*. 2002; 6: 59 - 64.

Stonnington CM, Chu C, Klöppel S, Jack CR Jr, Ashburner J, et al. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *NeuroImage*. 2010; 51: 1405 - 1413.

Sundermann B, Herr D, Schwindt W, Pfleiderer B.Multivariate classification of blood oxygen level-dependent FMRI data with diagnostic intention: a clinical perspective. *AJNR Am J Neuroradiol*. 2014; 35: 848 - 855.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*. 2002; 15: 273 – 289.

Van den Bogaard S, Dumas E, van der Grond J, van Buchem M, Roos R. MRI biomarkers in Huntington's disease. *Front Biosci*. 2012; 4: 1910 - 1925.

Vargas-Lowy D, Chitnis T. Pathogenesis of Pediatric Multiple Sclerosis. *J Child Neurol*. 2012; 27: 1394 - 1407.

Volkow ND, Wang GJ, Fowler JS, Tomasi D, Telang F, et al. Addiction: Decreased reward sensitivity and increased expectation sensitivity conspire to overwhelm the brain's control circuit. *Bioessays*. 2010; 32: 748–755.

Volkow ND, Fowler JS, Wang GJ, Goldstein RZ. Role of dopamine, the frontal cortex and memory circuits in drug addiction: insight from imaging studies. *Neurobiol Learn Mem*. 2002; 78, 610–624.

Wadden TA, Sternberg JA, Letizia KA, Stunkard AJ, Foster GD, 1998. Treatment of obesity by very low calorie diet, behaviour therapy, and their combination: a five year perspective. *Int J Obes,* 1998; 13(Suppl. 2), 39–46.

Weygandt M, Hummel HM, Schregel K, Ritter K, Allefeld C, et al. MRI-based diagnostic biomarkers for early onset pediatric multiple sclerosis. *NeuroImage: Clinical*. 2015a; 7: 400 – 408.

Weygandt M, Mai K, Dommes E, Ritter K, Leupelt V, et al. DLPFC impulse control counteracts a yo-yo effect in post-diet weight maintenance. NeuroImage. 2015b; 109: 318 – 327.

Weygandt M, Mai K, Dommes E, Leupelt V, Hackmack K, et al. Role of neural impulse control mechanisms for dietary success in obesity. NeuroImage. 2013; 83: 669 - 678.

Weygandt M, Schäfer A, Schienle A, Haynes JD. Diagnosing different binge-eating disorders based on reward-related brain activity patterns. *Hum Brain Mapp*. 2012; 33: 2135 – 2146.

Wilette AA, Calhoun VD, Egan JM, Kapogiannis D; Alzheimer's Disease Neuroimaging Initiative. Prognostic classification of mild cognitive impairment and Alzheimer's disease: MRI independent component analysis. *Psychiat Res*. 2014; 224: 81 - 88.

Zarogianni E, Moorhead TW, Lawrie SM. Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level. *NeuroImage: Clinical*. 2013; 3: 279 - 289. Doi

Zivadinov R, Leist TP. Clinical-magnetic resonance imaging correlations in multiple sclerosis. *J Neuroimaging*. 2005; 15(suppl 1): 10S-21S.

Zhang J, Tong L, Wang L, Li N. Texture analysis of multiple sclerosis: a comparative study. *Magn Reson Imaging*. 2008; 26: 1160 - 1166.

Zhang L, Samaras D, Tomasi D, Volkow N, Goldstein R: Machine learning for clinical diagnosis from functional magnetic resonance imaging. *Comp Vis Pat Rec. IEEE Proc of CVPR*. 2005; 1:1211 – 1217.

5 Statements of authorship and originality

This habilitation synopsis is entirely my own work. To the best of my knowledge and belief, it does not contain any material previously published or written by another person, except where due reference is made in this synopsis itself. Neither the habilitation synopsis, nor any section thereof, has been previously submitted towards a degree or diploma in any university or other higher education institution.

_____ _____

Place and date signed Signature

6 Acknowledgement

Bei der Durchführung dieser Arbeit habe ich viele Hinweise und freundliche Unterstützung erhalten.

Zuallererst möchte ich mich bei John-Dylan Haynes bedanken, der zentral zum grundlegenden Zustandekommen der Projekte beigetragen hat, mich bei allen Studien mit konstruktivem Feedback unterstützt hat, und mir einen weitreichenden Einblick in die vielen Facetten des Wissenschaftsbetriebs ermöglicht hat.

Weiterhin gilt mein Dank Kerstin Ritter und Carsten Allefeld, die in vielen produktiven Gesprächen zur methodischen Realisierung der Studien beigetragen haben.

Darüber hinaus möchte ich mich bei Knut Mai, Jens Würfel, Britta Hölzel, Anne Schienle, Joachim Spranger und Friedemann Paul bedanken, die wesentlich zum grundlegenden Zustandekommen der Projekte beigetragen haben, mich konstruktiv bei deren Organisation unterstützt haben, und mir in vielen interessanten Gesprächen Hinweise zu klinischen Aspekten gegeben haben.

Schließlich möchte ich mich bei Freunden und Familie bedanken, die mich zu allen Zeiten mit Rat und Tat, Freude und Zuneigung unterstützen.