REVIEW

Multivariate Pattern Recognition for Diagnosis and Prognosis in Clinical Neuroimaging: State of the Art, Current Challenges and Future Trends

Sven Haller · Karl-Olof Lovblad ·
Panteleimon Giannakopoulos · Dimitri Van De Ville

Abstract Many diseases are associated with systematic modifications in brain morphometry and function. These alterations may be subtle, in particular at early stages of the disease progress, and thus not evident by visual inspection alone. Group-level statistical comparisons have dominated neuroimaging studies for many years, proving fascinating insight into brain regions involved in various diseases. However, such group-level results do not warrant diagnostic value for individual patients. Recently, pattern recognition approaches have led to a fundamental shift in paradigm, bringing multivariate analysis and predictive results, notably for the early diagnosis of individual patients. We review the state-of-the-art fundamentals of pattern recognition including feature selection, cross-validation and classification techniques, as well as limitations including inter-individual variation in normal brain anatomy and neurocognitive reserve. We conclude with the discussion of future trends including multi-modal pattern recognition, multi-center approaches with data-sharing and cloud-computing.

Keywords MRI · Pattern recognition · Support vector machines SVM · Multivariate · Predictive modeling

Abbreviations
AD	Alzheimer disease
ADNI	Alzheimer disease neuroimaging initiative
DTI	Diffusion tensor imaging
fMRI	Functional magnetic resonance imaging
MCI	Mild cognitive impairment
MRI	Magnetic resonance imaging
SVM	Support vector machines
VBM	Voxel based morphometry

S. Haller (✉) · K.-O. Lovblad
Department of Imaging and Medical Informatics, University Hospitals of Geneva and Faculty of Medicine of the University of Geneva, Geneva, Switzerland
e-mail: sven.haller@hcuge.ch

P. Giannakopoulos
Department of Mental Health and Psychiatry, University Hospitals of Geneva and Faculty of Medicine of the University of Geneva, Geneva, Switzerland

D. Van De Ville
Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland

D. Van De Ville
Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Introduction

Many diseases cause systematic modifications in brain structure that can be imaged using MRI. These modifications may be subtle, in particular, at early stages of the disease, and, therefore, are not detectable by visual inspection alone. For many years, most neuroimaging studies have focused on group comparisons between healthy controls and patients. While such group-level results are fascinating from a research perspective and may reliably identify brain regions involved in a given disease, these findings do not automatically translate into useable diagnostic procedures at the individual level.

Therefore, the application of tools from pattern recognition to neuroimaging data contributed to a fundamental paradigm shift to develop novel and sensitive imaging-

based biomarkers, In principle, individual prediction is possible using univariate techniques, however, tools from pattern recognition—commonly referred to as multivariate pattern analysis (MVPA)—have the ability to integrate information across multiple variables. Exploiting multivariate data structure can significantly improve sensitivity, in particular when only subtle alterations occur, as it is typically the case in the context of early detection or diagnosis.

Another fundamental feature of pattern recognition is that it exploits in an elegant way high-dimensional data, even when conventional multivariate statistical methods would not work due to a limited number of data samples. This can be well illustrated by an example in the field of face perception based on images of female and male subjects (Haller et al. 2011). A (univariate) pixel-wise statistical test might for example find a significant difference of the nose region between female and male faces. However, it is impossible to reliably determine the gender based on the nose alone. Using a classifier, multivariate information is integrated using the pattern of multiple features such as ear, nose, eyebrows, chin etc. to recognize an individual face although each feature *per se* is not necessarily significantly different between groups.

The interest in applying machine learning techniques to neuroimaging data started by early work of Haxby and colleagues (Haxby et al. 2001) where they explicitly recognized the distributed nature of activation patterns from fMRI in the visual cortex. Until then, most fMRI analyses were performed using mass univariate techniques (i.e., voxel by voxel), which did not exploit inter-voxel dependencies. While, in principal, distributed patterns can be revealed by multivariate statistical techniques (e.g., partial least squares Krishnan et al. 2011), there was an almost immediate interest to employ these patterns using tools from machine learning that can then read out mental state from previously unseen data (Cox and Savoy 2003; Pereira et al. 2009). A multitude of interesting literature has demonstrated that it is possible to train data-driven models that can subsequently decode information from the subject's brain images; for example semantic meaning of words (Mitchell et al. 2008), emotional prosody pronounced by actors (Ethofer et al. 2009), or more recently visual imagery (Nishimoto et al. 2011) and even attempts to decode dreams (Horikawa et al. 2013).

These developments have also started a promising avenue for clinical neuroimaging. First, many neurological diseases and disorders are characterized by diffuse rather than focal changes (Seeley et al. 2009), and, therefore, multivariate methods should be more sensitive in picking up such changes. Second, predictive modeling at the level of the single subject is key to ultimately provide new neuroimaging markers with diagnostic value. Initially, whole-brain morphometry from structural MRI has been used to train models that can discriminate between healthy controls and patients, such as Alzheimer's disease and frontotemporal dementia (Kloppel et al. 2008; Fan et al., 2008a, c; Davatzikos et al. 2008), fragile-X syndrome (Hoeft et al. 2008), psychosis (Davatzikos et al. 2005; Fan et al. 2008b), depression (Costafreda et al. 2009), psychosis (Sun et al. 2009), multiple sclerosis (Weygandt et al. 2011) and so on. Advances in functional MRI, and more recently resting-state fMRI, have made it possible to study alterations in functional networks (Fox and Greicius 2010) without behavioral confounds (Bullmore 2012). Measures of functional connectivity between different brain regions can be used as features for pattern recognition methods, and has been applied, for example, to discriminate healthy from early AD (Wang et al. 2006), schizophrenia (Demirci et al. 2008), major depression (Craddock et al. 2009), pain perception (Marquand et al. 2010), and multiple sclerosis (Richiardi et al. 2012) or to predict brain maturity (Dosenbach et al. 2010). Pattern recognition approaches based on functional measures could potentially provide fMRI a more central role in the clinical field.

The application of these advanced pattern recognition data analysis techniques, in particular in a clinical context, requires understanding of the underlying principles and potential pitfalls, which will be discussed in the following sections.

State of the Art

Data Processing Pipeline

Conventional confirmatory analysis is based on a (predefined) generative model that is fitted to the data. Statistical hypothesis testing then provides forward inference on how well the model explains the data. Many group-studies use such schemes to identify significant differences between populations. Pattern recognition tools reverse the direction of inference; i.e., the model is learned *from the data* during a training phase in order to predict the explanatory variable. The model performance is then validated during a validation phase. The models depend on the type of classifier, but they are usually flexible and based on general assumptions about the data structure. The "task" of a classifier refers to the predicted variable; e.g., healthy versus patient. The data that is made available to the classifier, the definition of the task, and the performance of the classifiers are three essential ingredients for the interpretation of the results (see Fig. 1).

The pattern recognition pipeline (Lemm et al. 2011) has the following essential components. As the first step, feature extraction converts the raw data into the best possible form that maximizes the amount of information and
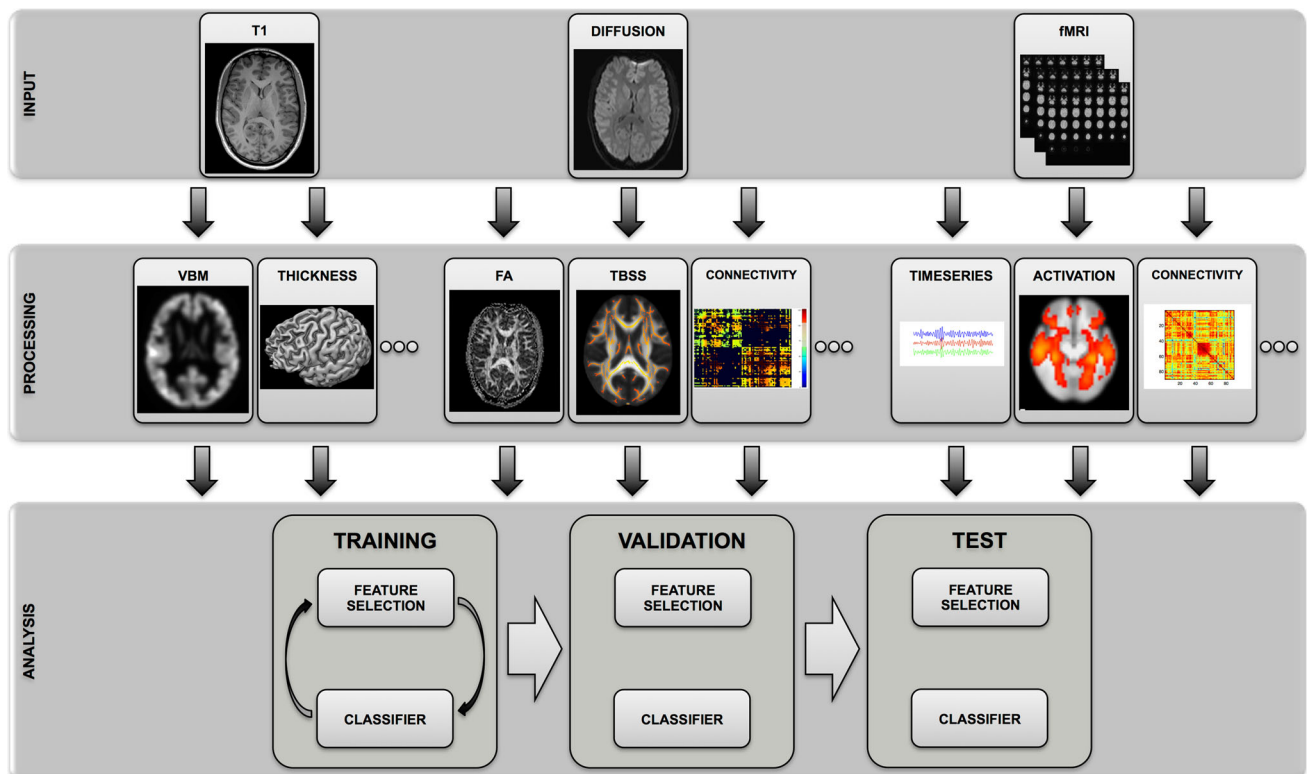
Fig. 1 Schematic illustration of the data processing pipeline. Several types of data input such as T1, diffusion weighted imaging or fMRI are in a first step pre-processed in multiples ways usually known from advanced image analysis techniques including depending on the specific demands spatial normalization into a standard space, field inhomogeneity correction, spatial (and temporal) smoothing, atlasing, independent component decomposition, structural or functional connectivity analyses and others. In the next step, the pre-processed data is used for feature selection and classification

minimizes the effect of confounds by various sources of noise. Structural or functional features can be extracted from imaging data, sometimes after a long procedure (e.g., spatial normalization to bring the data in the same "brain space"). For instance, a typical structural MRI dataset can easily contain more than $100'000$ voxels. In most cases, the features are related (similarity of adjacent voxels or voxels in homologous regions), and only a limited number of brain regions (and consequently features) will carry discriminative information. Other possibilities of features include those extracted from functional imaging (e.g., contrast of an experimental condition in a conventional activation study, or functional connectivity from resting-state fMRI), from diffusion-weighted imaging (e.g., measures of structural connectivity), neuropsychological and clinical measures. As discussed below in more detail, it is also possible to combine different types of features.

Feature Selection

In most cases, the number of features (e.g., voxels in the case of structural MRI) is much larger than the number of subjects available. The selection of the "best" features is a long-standing problem in machine learning that can be dealt with either explicitly (by a separate feature selection step) or implicitly (by regularization in the classification method). Conventional feature selection reduces the dimensionality of the feature space by ranking features according to their univariate statistical significance for distinguishing between two populations and the N-first features can be retained. Another approach is to apply multivariate techniques that project the feature space onto a (linear) subspace; e.g., principal components analysis can be used to keep those feature dimensions that explain most of the variance. In such case the new features will consists of linear combinations of the initial ones and it might be harder to interpret the final results. In general, correlated features are often not optimal for prediction (i.e., only the best predictive feature would be selected), however, from the neuroimaging point-of-view it is desirable to identify all correlated features to improve interpretability (Tolosi and Lengauer 2011). For instance, it is better to select a whole brain region as predictive for a certain task, instead of a few voxels that might not be very stable due to noise or slight variations in brain morphometry or the data.
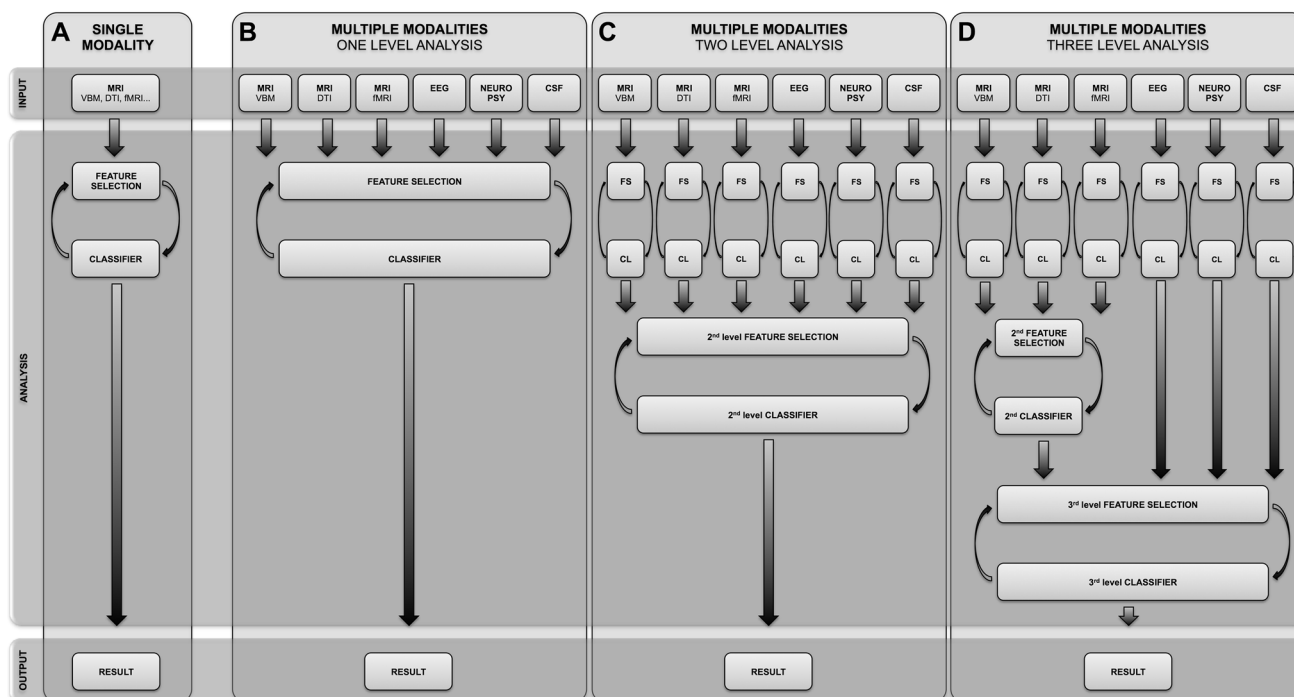
Fig. 2 Illustrates the simplified processing pipeline for a single domain classification (**a**) from data input to feature selection/classification to output. In multi-modal classification (**b**, **c**, **d**), all input data can directly enter one single feature selection/classification (**b**). Alternatively, it is possible to have one feature selection/classification per input data, followed by one single super- feature selection/classification at the second level (**c**). Moreover, it is possible to introduce an additional feature selection/classification levels, for example regrouping all imaging modalities, followed by a super feature selection/classification at the 3rd level (**d**)

Therefore, features are also sometimes transformed into another domain that is (approximately) decorrelating; e.g., the spatial wavelet transform leads to a representation that is more compact for piecewise smooth signals, a property that is often used to improve detectability in fMRI studies (Van De Ville et al. 2006). Finally, we also note that many classifiers have an implicit feature selection built-in termed as "regularization"; i.e., the classifier optimizes a criterion that favors fewer features.

Classification

Support-vector-machines (SVM) have been widely applied to neuroimaging data, mostly because of their robustness against outliers, but they are by far not the only choice available (Cristianini and Shawe-Taylor 2000). Typically, SVMs are used for binary classification such as discriminating patients from healthy controls. However, more advanced classifiers have been developed (e.g., decision forests (Criminisi and Shotton 2013) for a recent example) and should be considered for multi-class classification and regression as these tasks are increasingly important to build clinically relevant tools; e.g., for differential diagnosis. However, there is some trade-off between increasing the

complexity of the classifier and improving the feature extraction. Usually, good features allow for better classification with almost any classifier.

Another issue with classifiers arises when features are multimodal; i.e., combining different types of imaging data with other measures such as scores from neuropsychological tests. As the model of a single classifier will not allow the necessary flexibility, "ensemble learning" is a rich field in machine learning that deals with combining multiple classifiers (and thus multiple models) to integrate the richness of the data structure. One promising approach in the future is the use of multi-level classifier algorithms to aggregate information in a hierarchical way (see Fig. 2).

Cross-Validation

Classifiers are trained and evaluated on separate parts of the data according to a cross-validation methodology. Specifically, the classifier's model is trained using part of the data where both the features and the predicted variable are given to the algorithm. Then, the classifier is evaluated on the remaining part of the data by comparing its outcome against the ground-truth prediction. The procedure is repeated by removing different parts of the data and

summarizing the average performance across the validation folds (e.g., in terms of specificity and sensitivity). Ideally, the classifier should have learned the main data structure, ignoring the noise, and be able to generalize its model to unseen data.

Current Challenges and Future Trends

Compared to applications of pattern recognition in cognitive neuroimaging (Raizada and Kriegeskorte 2010), ameliorating the handling of various confounds is a remaining challenge before these techniques enter the clinical realm. In the following, we review the fundamental current challenges, potential pitfalls and limitations as well as future trends that deserve consideration, in particular with respect to clinical applications of pattern recognition analyses.

Normal Anatomical Inter-individual Variability

There is substantial normal inter-individual variation in brain morphology even in healthy volunteers, for example, approximately 15 % variability in cortical thickness (Haug 1987). In contrast, there is less variability in the inter-hemispheric difference between homologous brain regions. Correspondingly, a previous study demonstrated that due to this decreased variability in the within-subject cortical asymmetry, at-risk mental state subjects could be discriminated from volunteers only when considering the within-subject cortical asymmetry yet not based on direct assessment of cortical thickness between-subjects (Haller et al. 2009). Moreover, adjacent voxels are more likely to carry similar information than distant voxels in non-related areas. It is also notable that spatial information can be integrated at several scales; e.g. across neighboring voxels, subareas of predefined anatomical structures or even areas distributed over larger distances based on prior knowledge (e.g., Hackmack et al. 2012a).

Current classification analyses typically consider each voxel as an independent feature ignoring the highly ordered structure of the brain. In other words, SVMs are invariant to permutation of the feature dimensions. To acknowledge spatial neighborhood relationships, one can include spatial transformations in the feature extraction step (e.g., spatial averaging or wavelet transforms as mentioned above). This could potentially improve classification accuracy and robustness.

Incorporating Prior Knowledge

Numerous previous structural and functional group level studies have provided evidence of specific brain regions involved in a wide range of different diseases and conditions. This extensive prior knowledge is largely ignored in recent individual level pattern recognition analyses, although it might potentially improve classification accuracy and robustness.

Similar to the discussion about the anatomic structure of the brain above, it is likely beneficial to inject any available domain knowledge to improve the information content of the features despite the fact that feature selection and state-of-the-art classifiers are designed to deal with high-dimensional learning. Purely data-driven methods might miss important structures available in the data. In addition, the interpretation of the results could often become easier when features represent domain-relevant information. Advanced methods have recently been proposed to exploit spatial structure; e.g., based on hierarchical clustering to regroup similar voxels and reinforce the robustness (Cui et al. 2012). For some applications, features can also be extracted from specific regions-of-interest instead of the whole brain, or the "locality" of information can be probed by a "searchlight" approach where classification performance is reported in an information map for features extracted from (sliding) local neighborhoods (Kriegeskorte et al. 2006).

For functional data, there is often a significant amount of within-subject non-stationary activity that interferes with the signal of interest; in recent work, techniques for normalizing out the non-stationary component of within-subject activity show promising avenues for improving the robustness of subsequently extracted features (Samek et al. 2013). For resting-state fMRI data, interaction between timecourse can often be well summarized by functional connectivity measures such as Pearson correlation or more complicated measures of dependencies (Friston 2011). In principle, classifiers could learn the data structure that leads to considering features based on inter-hemispherical differences or functional connectivity, however, the available amount of data is often too limited and it often non-trivial to understand the properties of the learned model.

Normal Inter-individual Variation in Cognitive Reserve

Pattern recognition analyses at the individual level (as well as univariate group level analyses) are immanently based on the assumption that there is a direct relationship between brain pathology and symptomatology. In the example of cognitive decline, the assumption is that decreasing cognitive functions are paralleled by progressive brain atrophy. This assumption is, however, not necessarily true. Due to individual factors such as education and social integration, the same degree of clinically evident neurocognitive impairment can be caused by a variable

degree of brain pathology—or from the other perspective, the same degree of brain pathology can evoke variable degree of clinical neurocognitive impairment. This inter-individual variation in the neurocognitive reserve was described already in 1968 (Tomlinson et al. 1968). This means that for example of pattern recognition in the domain of cognitive decline, the same amount of structural brain alterations could be associated with clinically manifested cognitive decline if the neurocognitive reserve is exhausted, or the individual might still maintain intact cognitive functioning if there is sufficient neurocognitive reserve. This inter-individual variability in the neurocognitive reserve, or more generally variability between brain structural and functional alterations on the one hand, and clinically manifest symptoms on the other hand, represent a general and fundamental limitation for pattern recognition approaches especially in a clinical setting.

Dysbalance Between Number of Features and Number of Subjects, Data Reduction and Over-Fitting

It is worthwhile mentioning that a typical feature set extracted from MRI data can easily contain more than 100′000 features, which clearly exceeds the number of individuals in particular in single-center studies. In most cases, the features are related (similarity of adjacent or homologous voxels), and only a limited number will carry discriminative information.

While the cross-validation methodology is essential to train and evaluate classifiers, it still has the risk of over-fitting the data as the parameters can be tuned. Therefore, in practice, several nested levels of cross-validation should be used, as well as a separate independent test dataset that has not been used for any training/validation before. Often the sample size for single-center studies is insufficient to estimate such "real-world" performance, but this is one important objective for clinical neuroradiology that can benefit from ongoing multi-centric studies, data sharing, and cloud computing (see below).

Another important concept that is commonly used in machine learning is the "kernel trick" (Shawe-Taylor and Cristianini 2004). Instead of learning the structure in the high-dimensional feature space, which is rather empty as there are not enough data points, one transforms the data into a Euclidean space by a kernel function applying a kernel function to each pair of data points, as such constituting the kernel (Gram) matrix. The data structure can then be learned in a space that has as many dimensions as data points. Many different kernel functions have been proposed, including polynomial and Gaussian kernels (a.k.a. radial basis function kernel), or even kernels for graph structures. Many classifiers operate in this more

economical space and relate back the model to the full feature space.

In addition, parameter tuning for both feature selection and classifiers requires an additional inner cross-validation loop that decreases the available data for learning as well as increasing the risk of over-fitting.

In summary, selecting and determining the importance of features is an essential processing step in classification analyses, yet identifying the related parameters such as optimal number of features or regularization tuning is non-trivial in practice. This is still an active field of research in machine learning; e.g., stability selection, which refers to the consistency of features when subsampling the feature space (Meinshausen and Bühlmann 2010), including applications to neuroimaging (Langs et al. 2011).

Variability Related to Patient Selection, Inter-scanner Variability and Data Preprocessing

Additional potentially confounding factors, in particular with respect to clinical applications, include scanner heterogeneity (Abdulkadir et al. 2011; Kruggel et al. 2010), variability in data preprocessing and patient selection.

Multimodal Classification

One of the key challenges to obtain effective biomarkers for computer-aided diagnosis and prognosis is to incorporate information from different modalities. Next to imaging measures from functional and structural MRI, these should take into account additionally available parameters such as neuropsychological and clinical measures amongst others in order to improve classification accuracy and robustness, in particular for clinical applications.

Towards New Biomarkers

Classification of healthy subjects versus patients has limited importance for clinical practice. One clinically relevant application is *prediction* of individuals at risk for consequent cognitive decline in the domain of mild cognitive impairment (MCI), which represents a transition zone between normal aging and very early dementia (Petersen and Negash 2008). The definitions of MCI have substantially evolved and changed over the past years, which goes beyond the scope of this review. Depending on the MCI subtype, only about half of MCI subjects will progress to clinically overt dementia, whereas the other half may remain stable or evolve to other forms of dementia (Petersen 2004; Mariani et al. 2007; Forlenza

et al. 2009). Assuming that only about half of unselected MCI individuals will progress to clinically overt dementia, the *prediction* of individuals at risk for consecutive cognitive decline is of paramount importance for early individual treatment as well as for clinical trials. In a typical placebo-controlled pharmaceutical trial, 25 % of unselected MCI will remain stable despite being in the placebo group, while only 25 % of individuals will progress and obtain the active compound. Therefore, *pre-selection* of at risk individuals for future cognitive decline would substantially improve the design of clinical trials. As MRI is routinely performed in the clinical workup of cognitive decline, advanced data analysis techniques as those by pattern recognition tools make use of already existing data, which is thus cost effective and without additional discomfort for the patient. It is possible to *predict* future cognitive decline in MCI using baseline MRI based on grey matter voxel based morphometry (VBM) (Plant et al. 2010; Misra et al. 2009; Fan et al. 2008c), white matter DTI (Haller et al. 2010a, b; O'Dwyer et al. 2012) or iron deposition (Haller et al. 2010a, b). Another potential application for pattern recognition is in patient follow-up by a surrogate marker of patients' cognitive function based on imaging data; e.g., MVPA has been proposed as one way to overcome the clinico-radiological paradox in multiple sclerosis (Hackmack et al. 2012b), and resting-state fMRI appears as a promising candidate to provide relevant features of MCI to AD progression (Damoiseaux et al. 2012).

It is further possible to combine for example DTI and resting state fMRI to identify MCI individuals (Wee et al. 2012), to predict MCI to AD conversion using multimodal measures also in combination with neuropsychological scores or cerebrospinal fluid biomarkers (Cui et al. 2011, 2012) or by the combination of structural MRI and FDG-PET (Zhang and Shen 2012a, b; Zhang et al. 2011). Finally, it is possible to classify MCI subtypes, who have different risk of disease progression and who might benefit from different types of treatment, for example, based on DTI (Haller et al. 2013b).

The clinical application of such pattern approaches is not limited to dementia, and may also for example be applied to predict the development of psychosis in at risk mental state individuals (Mourao-Miranda et al. 2012; Orru et al. 2012; Pettersson-Yeo et al. 2013; Gothelf et al. 2011).

Another clinically relevant task is to discriminate between patients with typical Parkinson's disease from patients with atypical Parkinson's disease (Haller et al. 2012, 2013a), which significantly modifies prognosis, outcome and treatment. These studies represent promising advances in early clinical detection of individual patients to predict outcome or select specific at risk patients for clinical trials.

Conclusions

In summary, the application of techniques from the field of pattern recognition to neuroimaging data is an emerging field. These methods have a number of attractive features, including the use of multivariate information and the possibility to predict for previously unseen data. Ongoing research is still needed to overcome a number of limitations, including optimal feature selection that incorporates better domain knowledge, and integration of multimodal measurements. In addition, future methodological developments should be increasingly based on large datasets and multi-centric studies to increase both reproducibility and predictability. Recent data sharing initiatives such as ADNI (Mueller et al. 2005), in combination with cloud-computing power, will provide the necessary prerequisites for these developments. We will hopefully see new advances in individual-level classification analysis in order to provide earlier and more accurate diagnoses, better estimation of prognosis, and eventually improve patient care.

Conflict of interest No conflicts of interest.

References

Abdulkadir A, Mortamet B, Vemuri P, Jack CRJ, Krueger G, Kloppel S (2011) Effects of hardware heterogeneity on the performance of SVM Alzheimer's disease classifier. Neuroimage 58:785–792

Bullmore E (2012) The future of functional MRI in clinical medicine. Neuroimage 62:1267–1271

Costafreda SG, Chu C, Ashburner J, Fu CH (2009) Prognostic and diagnostic potential of the structural neuroanatomy of depression. PLoS ONE 4:e6353

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19:261–270

Craddock RC, Holtzheimer PE, Hu XP, Mayberg HS (2009) Disease state prediction from resting state functional connectivity. Magn Reson Med 62:1619–1628

Criminisi A, Shotton J (2013) Decision forests for computer vision and medical image analysis. Retrieved from http://dl.acm.org/citation.cfm?id=2462584

Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Retrieved from http://books.google.com/books?hl=en&lr=&id=_PXJn_cxv0AC&oi=fnd&pg=PR9&dq=An+introduction+to+support+vector+machines+and+other+kernel-based+learning+methods+*+N+Cristianini,+J+Shawe-Taylor+-+2000,+Cambridge+University+press.&ots=xQUe3C_oXd&sig=CmI_iC03SDnJrhIu08P73L-vKSg

Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, Zhu W, Park M, Jiang T, Jin JS (2011) Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. PLoS ONE 6:e21896

Cui Y, Sachdev PS, Lipnicki DM, Jin JS, Luo S, Zhu W, Kochan NA, Reppermund S, Liu T, Trollor JN, Brodaty H, Wen W (2012) Predicting the development of mild cognitive impairment: a new use of pattern recognition. Neuroimage 60:894–901

Damoiseaux JS, Prater KE, Miller BL, Greicius MD (2012) Functional connectivity tracks clinical deterioration in Alzheimer's disease. Neurobiol Aging 33:828.e19–828.e30

Davatzikos C, Shen D, Gur RC, Wu X, Liu D, Fan Y, Hughett P, Turetsky BI, Gur RE (2005) Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. Arch Gen Psychiatry 62:1218–1227

Davatzikos C, Resnick SM, Wu X, Parmpi P, Clark CM (2008) Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. Neuroimage 41:1220–1227

Demirci O, Clark VP, Magnotta VA, Andreasen NC, Lauriello J, Kiehl KA, Pearlson GD, Calhoun VD (2008) 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 2:147–226

Dosenbach NU, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, Nelson SM, Wig GS, Vogel AC, Lessov-Schlaggar CN, Barnes KA, Dubis JW, Feczko E, Coalson RS, Pruett JRJ, Barch DM, Petersen SE, Schlaggar BL (2010) Prediction of individual brain maturity using fMRI. Science 329:1358–1361

Ethofer T, Van De Ville D, Scherer K, Vuilleumier P (2009) Decoding of emotional information in voice-sensitive cortices. Curr Biol 19:1028–1033

Fan Y, Batmanghelich N, Clark CM, Davatzikos C (2008a) Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. Neuroimage 39:1731–1743

Fan Y, Gur RE, Gur RC, Wu X, Shen D, Calkins ME, Davatzikos C (2008b) Unaffected family members and schizophrenia patients share brain structure patterns: a high-dimensional pattern classification study. Biol Psychiatry 63:118–124

Fan Y, Resnick SM, Wu X, Davatzikos C (2008c) Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. Neuroimage 41:277–285

Forlenza OV, Diniz BS, Nunes PV, Memoria CM, Yassuda MS, Gattaz WF (2009) Diagnostic transitions in mild cognitive impairment subtypes. Int Psychogeriatr 21:1088–1095

Fox MD, Greicius M (2010) Clinical applications of resting state functional connectivity. Front Syst Neurosci 4:19

Friston KJ (2011) Functional and effective connectivity: a review. Brain Connect 1:13–36

Gothelf D, Hoeft F, Ueno T, Sugiura L, Lee AD, Thompson P, Reiss AL (2011) Developmental changes in multivariate neuroanatomical patterns that predict risk for psychosis in 22q11.2 deletion syndrome. J Psychiatr Res 45:322–331

Hackmack K, Paul F, Weygandt M, Allefeld C, Haynes JD (2012a) Multi-scale classification of disease using structural MRI and wavelet transform. Neuroimage 62:48–58

Hackmack K, Weygandt M, Wuerfel J, Pfueller CF, Bellmann-Strobl J, Paul F, Haynes JD (2012b) Can we overcome the 'clinico-radiological paradox' in multiple sclerosis? J Neurol 259:2151–2160

Haller S, Borgwardt SJ, Schindler C, Aston J, Radue EW, Riecher-Rossler A (2009) Can cortical thickness asymmetry analysis contribute to detection of at-risk mental state and first-episode psychosis? A pilot study. Radiology 250:212–221

Haller S, Bartsch A, Nguyen D, Rodriguez C, Emch J, Gold G, Lovblad KO, Giannakopoulos P (2010a) Cerebral microhemorrhage and iron deposition in mild cognitive impairment: susceptibility-weighted MR imaging assessment. Radiology 257:764–773

Haller S, Nguyen D, Rodriguez C, Emch J, Gold G, Bartsch A, Lovblad KO, Giannakopoulos P (2010b) Individual prediction of cognitive decline in mild cognitive impairment using support vector machine-based analysis of diffusion tensor imaging data. J Alzheimers Dis 22:315–327

Haller S, Lovblad KO, Giannakopoulos P (2011) Principles of classification analyses in mild cognitive impairment (MCI) and alzheimer disease. J Alzheimers Dis 26(Suppl 3):389–394

Haller S, Badoud S, Nguyen D, Garibotto V, Lovblad KO, Burkhard PR (2012) Individual detection of patients with Parkinson disease using support vector machine analysis of diffusion tensor imaging data: initial results. AJNR Am J Neuroradiol 33:2123–2128

Haller S, Badoud S, Nguyen D, Barnaure I, Montandon ML, Lovblad KO, Burkhard PR (2013a) Differentiation between Parkinson disease and other forms of Parkinsonism using support vector machine analysis of susceptibility-weighted imaging (SWI): initial results. Eur Radiol 23:12–19

Haller S, Missonnier P, Herrmann FR, Rodriguez C, Deiber MP, Nguyen D, Gold G, Lovblad KO, Giannakopoulos P (2013b) Individual classification of mild cognitive impairment subtypes by support vector machine analysis of white matter DTI. AJNR Am J Neuroradiol 34:283–291

Haug H (1987) Brain sizes, surfaces, and neuronal sizes of the cortex cerebri: a stereological investigation of man and his variability and a comparison with some mammals (primates, whales, marsupials, insectivores, and one elephant). Am J Anat 180:126–142

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430

Hoeft F, Lightbody AA, Hazlett HC, Patnaik S, Piven J, Reiss AL (2008) Morphometric spatial patterns differentiating boys with fragile X syndrome, typically developing boys, and developmentally delayed boys aged 1 to 3 years. Arch Gen Psychiatry 65:1087–1097

Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y (2013) Neural decoding of visual imagery during sleep. Science 340:639–642

Kloppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CRJ, Ashburner J, Frackowiak RS (2008) Automatic classification of MR scans in Alzheimer's disease. Brain 131:681–689

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

Krishnan A, Williams LJ, McIntosh AR, Abdi H (2011) Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. Neuroimage 56:455–475

Kruggel F, Turner J, Muftuler LT (2010) Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. Neuroimage 49:2123–2133

Langs G, Menze BH, Lashkari D, Golland P (2011) Detecting stable distributed patterns of brain activation using Gini contrast. Neuroimage 56:497–507

Lemm S, Blankertz B, Dickhaus T, Muller KR (2011) Introduction to machine learning for brain imaging. Neuroimage 56:387–399

Mariani E, Monastero R, Mecocci P (2007) Mild cognitive impairment: a systematic review. J Alzheimers Dis 12:23–35

Marquand A, Howard M, Brammer M, Chu C, Coen S, Mourao-Miranda J (2010) Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. Neuroimage 49:2178–2189

Meinshausen N, Bühlmann P (2010) Stability selection. J R Stat Soc 72:417–473

Misra C, Fan Y, Davatzikos C (2009) Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in

prediction of short-term conversion to AD: results from ADNI. Neuroimage 44:1415–1422

Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA (2008) Predicting human brain activity associated with the meanings of nouns. Science 320:1191–1195

Mourao-Miranda J, Reinders AA, Rocha-Rego V, Lappin J, Rondina J, Morgan C, Morgan KD, Fearon P, Jones PB, Doody GA, Murray RM, Kapur S, Dazzan P (2012) Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. Psychol Med 42:1037–1047

Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L (2005) Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimers Dement 1:55–66

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Curr Biol 21:1641–1646

O'Dwyer L, Lamberton F, Bokde AL, Ewers M, Faluyi YO, Tanner C, Mazoyer B, O'Neill D, Bartley M, Collins DR, Coughlan T, Prvulovic D, Hampel H (2012) Using support vector machines with multiple indices of diffusion for automated classification of mild cognitive impairment. PLoS ONE 7:e32441

Orru G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A (2012) Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev 36:1140–1152

Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45: S199–S209

Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. J Intern Med 256:183–194

Petersen RC, Negash S (2008) Mild cognitive impairment: an overview. CNS Spectr 13:45–53

Pettersson-Yeo W, Benetti S, Marquand AF, Dell'acqua F, Williams SC, Allen P, Prata D, McGuire P, Mechelli A (2013) Using genetic, cognitive and multi-modal neuroimaging data to identify ultra-high-risk and first-episode psychosis at the individual level. Psychol Med 43:2547–2562

Plant C, Teipel SJ, Oswald A, Bohm C, Meindl T, Mourao-Miranda J, Bokde AW, Hampel H, Ewers M (2010) Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. Neuroimage 50:162–174

Raizada RDS, Kriegeskorte N (2010) Pattern-information fMRI: new questions which it opens up and challenges which face it. Int J Imaging Syst Technol 1–19. Retrieved from http://www.dartmouth.edu/~raj/papers/raizada_kriegeskorte_IJIST_review_preprint.pdf

Richiardi J, Gschwind M, Simioni S, Annoni JM, Greco B, Hagmann P, Schluep M, Vuilleumier P, Van De Ville D (2012) Classifying minimally disabled multiple sclerosis patients from resting state functional connectivity. Neuroimage 62:2021–2033

Samek W, Meinecke FC, Muller KR (2013) Transferring subspaces between subjects in brain-computer interfacing. IEEE Trans Biomed Eng 60(8):2289–2298

Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD (2009) Neurodegenerative diseases target large-scale human brain networks. Neuron 62:42–52

Shawe-Taylor J, Cristianini N (2004) Kernel methods for pattern analysis. Retrieved from http://carbon.videolectures.net/v001/9d/twpzi4psdbakxe3lqofwohumyu7zv7wa.pdf

Sun D, van Erp TG, Thompson PM, Bearden CE, Daley M, Kushan L, Hardt ME, Nuechterlein KH, Toga AW, Cannon TD (2009) Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: classification analysis using probabilistic brain atlas and machine learning algorithms. Biol Psychiatry 66:1055–1060

Tolosi L, Lengauer T (2011) Classification with correlated features: unreliability of feature ranking and solutions. Bioinformatics 27:1986–1994

Tomlinson BE, Blessed G, Roth M (1968) Observations on the brains of non-demented old people. J Neurol Sci 7:331–356

Van De Ville D, Blu T, Unser M (2006) Surfing the brain. IEEE Eng Med Biol Mag 25:65–78

Wang K, Jiang T, Liang M, Wang L, Tian L, Zhang X, Li K, Liu Z (2006) Discriminative analysis of early Alzheimer's disease based on two intrinsically anti-correlated networks with resting-state fMRI. Med Image Comput Comput Assist Interv 9:340–347

Wee CY, Yap PT, Zhang D, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen D (2012) Identification of MCI individuals using structural and functional connectivity networks. Neuroimage 59:2045–2056

Weygandt M, Hackmack K, Pfuller C, Bellmann-Strobl J, Paul F, Zipp F, Haynes JD (2011) MRI pattern recognition in multiple sclerosis normal-appearing brain areas. PLoS ONE 6:e21138

Zhang D, Shen D (2012a) Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. PLoS ONE 7:e33182

Zhang D, Shen D (2012b) Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. Neuroimage 59:895–907

Zhang D, Wang Y, Zhou L, Yuan H, Shen D (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 55:856–867