International Graduate Program Medical Neurosciences

Master's Thesis

Spatially Constrained Clustering of Voxels Based on Cross-Individual Consistency

by Stephan F. Mauermann,
Matriculation Nr. 220686

Date of Submission:
October 31st, 2016

Visual Perception Laboratory, Charité - Universitätsmedizin Berlin
Direct Supervisor: Dr. Matthias Guggenmos

Responsible Group Leader:

Prof. Dr. Philipp Sterzer,
Charité - Universitätsmedizin Berlin,
Department of Psychiatry,
Charitéplatz 1, 10117 Berlin,
philipp.sterzer@charite.de

Suggested Second Reviewer:

Dr. Carsten Bogler,
Bernstein Center for Computational Neuroscience, Humboldt-Universität zu Berlin, Philippstraße 13, 10115, Berlin,
carsten.bogler@bccn-berlin.de

Contents

Acknowledgements

First and foremost, I want to thank Dr. Matthias Guggenmos for being my supervisor. Matthias was a great mentor, and provided me with huge amounts of help, answers and a lot of inspiration. I also want to thank Prof. Dr. Philipp Sterzer for his support and for giving me the opportunity to spend time in his lab to pursue my Master's project. And also a big thanks to everyone else in the lab for their advice, feedback, and good conversations. Finally, I want to thank my family for their endless support, and the other MedNeuros for being such good classmates.

Abstract

In the past two decades, magnetic resonance imaging (MRI) has become an indispensable tool for the study of mental disease in humans. Neuroimaging studies however often entail high-dimensional and complex data sets. This a critical challenge for researchers, especially for analyses that itself are relatively complex and computationally expensive, such as machine-learning-based analyses. It is thus essential to develop methods that can reduce the complexity of experimental data to facilitate further analyses. This is typically done via data compression (e.g. principal component analysis) or feature selection. In the case of MRI studies the latter often encompasses a manual definition of regions of interest (ROIs) or atlas-based ROIs. However, the definition of custom ROIs or the selection of atlas-based ROIs is often subjective or even arbitrary, making a data-driven approach desirable. Here we aimed at developing an unbiased data-driven algorithm for the definition of ROIs (clustering) based on cross-individual consistency of voxels. The idea of cross-individual consistency is that a correlated behavior of neighboring (hence *spatially constrained*) voxels across individuals is unlikely based on noise only, thus increasing the probability that these voxels generally contain meaningful signal. To this end we implemented two distinct algorithms: a) the PearsonMerger, which progressively merges voxels into clusters that show suprathreshold correlations; and b) spectral clustering, a method exploiting spectral graph theoretical methods to cluster data using eigenvectors of similarity matrix representations of the data. Both clustering methods were applied to a structural dataset from a study investigating alcohol use disorder (AUD), and classification based on these clusters was performed to predict the diagnosis of AUD (healthy control vs. AUD patient). Both clustering algorithms were able to deliver better classification scores when compared to using conventional atlas-derived ROIs. Specifically, PearsonMerger clustering resulted in slightly better performance than spectral clustering in terms of prediction accuracy. The PearsonMerger algorithm furthermore identified regions affected by AUD, that were

independently discovered by a previous study. In sum, spatially constrained clustering based on cross-individual consistency is a promising tool to discover interesting brain regions (PearsonMerger) or to parcel the brain (spectral clustering) in an unbiased data-driven fashion and to aid machine-based classification.

Abbreviations

(d)ACC	(dorsal) anterior cingulate cortex
ANN	artificial neural network
ADS	alcohol dependence scale
AUD	alcohol use disorder
CV	cross-validation
DSM	Diagnostic and Statistical Manual of Mental Disorders
FN	false negatives
FWHM	full-width at half-maximum
JHU	Johns Hopkins University
LeAD	learning and alcohol dependence
MDD	major depressive disorder
MNI	Montreal Neurological Institute
(f, s)MRI	(functional, structural) magnetic resonance imaging
MVPA	multivariate pattern analysis
N	negative predictions
P	positive predictions
PCA	principal component analysis
RF	random forest
ROI	region of interest
SPM	statistical parametric mapping
SVM	support vector machine
TP, TN	true positive, true negative predictions
VBM	voxel-based morphometry
WeiRD	Weighted Robust Distance

1 Introduction

Magnetic resonance imaging (MRI) studies of the human brain have become the key technology to investigate brain alterations that might underlie psychiatric disorder. But not only in disease this technique is fundamental to gaining new insights on the brain's structure and function. Thanks to the widespread adoption of MRI techniques and constant improvement of imaging and analysis procedures, important findings have surfaced and shed new light onto areas governing normal brain function, such as the functional topology in visual cortex and extrastriate visual areas. Moreover, MRI has been employed extensively for the investigation of pathological mechanisms related to psychiatric disorders such as major depressive disorder (MDD), schizophrenia or drug addictions. Different modalities of MRI allow neuroscientists and physicians, in an non-invasive manner, not only to reveal information about brain anatomy but also to indirectly track brain activity in conjunction with cognitive tasks — namely by structural and functional MRI (sMRI, fMRI), respectively. Many important findings on pathological brain disturbances have helped broadening our understanding of brain disorders, such as hippocampal atrophy in Alzheimer's disease (Johnson et al. 2012) or hypoactivity in reward-related areas (ventral striatum) in MDD patients (Admon & Pizzagalli 2015).

1.1 MRI Data Analysis

A challenging aspect of the analysis of brain imaging data is their high complexity and multidimensional nature. Recently, machine learning methods, e.g., support vector machines (SVM) or artificial neural networks (ANN), have gained substantial traction in the domain of neuroimaging data analysis. These multivariate methods - also termed multivariate pattern analysis (MVPA) - are able to detect even subtle correlations and patterns present in the data, overcoming limitations of the standard univariate approach to MRI analysis (Haynes 2015). Given a set of training images and corresponding labels (e.g., healthy control or MDD patient)

machine learning methods can be used to learn a mapping from brain patterns to labels in a supervised manner. In other words, the algorithm learns a model that connects patterns in brain images to categories (e.g., MDD). Ideally, after sufficient training the algorithm is capable of correctly classifying novel images by generalizing from distinctive patterns in the training data, according to the categories it has been trained on. This feature can be invaluable to clinical and research applications. MVPA in the clinical setting has the potential to become a standard diagnosis tool for neurological and psychiatric disorders. The ability to detect even subtle patterns of abnormal brain anatomy and activity can complement traditional symptom-based diagnosis, and may lead to more reliable decisions (Klöppel et al. 2012). A prime example would be the early detection and diagnosis of Alzheimer's disease from structural brain scans (Sabuncu et al. 2015). Applications of MVPA are helping to push the boundaries of neurosciences: for example researchers are now able to decode decisions from participant's brain activity even before they become aware of making that decision (Soon et al. 2008), a technique called mind reading.

1.2 Analytical Challenges

As stated in the previous paragraph, machine-learning based analysis approaches are highly capable, however pose a serious challenge to be applied correctly and efficiently. There is a series of issues that need to be considered and to be taken care of attentively when planning and executing MVPA (Haynes 2015). A central concern in the application of voxel-wise supervised classification algorithms is that the dimensionality of MRI images (the number of voxels in an image can easily reach hundreds of thousands) is far greater than the amount of available training samples (i.e., the number of subjects in a study, typically in the order of 10 to 100). High-dimensional data is problematic in several aspects: (1) time and memory required to train and test analysis algorithms scale exponentially in relation to the number of dimensions; and (2) the predictive power of classification algorithms is negatively correlated to the number of dimensions present in the data

it has been trained on (Hughes 1968). In the machine learning community this obstacle is called the *curse of dimensionality*. It is thus essential to obtain reduced but still meaningful representations of MRI data for further classification via *data compression* or *feature selection* (Kuhn et al. 2013). Feature selection procedures are designed to eliminate redundant or irrelevant features from the data and hence to improve analysis, resulting in shorter training times and reduced *overfitting*. The latter issue relates to the case when a classification algorithm has fitted a model to the data, which merely is dependent on the noise variance of the underlying data instead of the 'true' underlying relationship (Hawkins 2004). The generalization performance of such overfitted models is usually weak for novel samples (i.e., samples outside the scope of the training data) and thus predictions based on such models are not reliable and can lead to misinterpretation. In other words, feature selection promises not only to improve classification accuracy and to speed up the analysis process, but also to lead to better generalization.

One of the most popular methods for feature selection in a neuroimaging context is the application of predefined regions of interest (ROIs). Dimensionality reduction is achieved by reducing the number of considered brain regions and often also by averaging across voxels within the regions. In this way, the number of features can be reduced quite substantially. The challenge herein lies at how these regions are defined *a priori*. Typically, ROIs are defined based on previous studies (literature-based ROIs) (Saxe et al. 2006) or by using brain atlases (Mazziotta et al. 2001). However, such an approach is not without issues. First, using ROI coordinates from other studies or external atlases may lead to a poor definition of ROIs for the sample at hand. For instance, brain atlases are constructed from specific sample populations that may differ quite a bit from the population in the study to be analyzed. This may lead to cases where interesting information may be located just outside of these regions. Second, such a procedure is often associated with a certain degree of subjectivity or bias and may lead to less reproducible research outputs. It is thus desirable to develop tools for an unbiased data-driven creation of ROIs by means of parcellation or clustering algorithms, i.e. tools exploiting the inherent structures and characteristics of the data. Substantial effort

has been devoted to functional parcellation of brain images in fMRI studies (Thirion et al. 2014; Heller et al. 2006), however there is only a comparably small number of studies tackling this problem in a structural MRI setting (Hrdlicka et al. 2005).

1.3 Data-driven Clustering

The aim of the thesis project presented here was to develop a tool for data-driven clustering and parcellation of structural MRI images, independent of *a priori* hypotheses or anatomical brain atlases. Central to the clustering approach presented here is the idea to extract spatially connected clusters from whole-brain sMRI data that show consistent patterns across subjects. The underlying notion is that if pairs of voxels - which will be eventually merged into clusters - behave consistently across different subjects, there is a good chance that these clusters contain actual biologically meaningful information. Whether this information is relevant also for the actual contrast of interest can not be assumed — however, by applying such inter-individual clustering we can eliminate voxels that are likely noise-driven (i.e., that have no information content). In other words, given the low probability that large ensembles of noisy voxels would show consistent patterns across a sample group of subjects, we propose that our method is able to eliminate such noise-driven voxels. For the implementation of an unbiased, data-driven clustering method we developed a novel algorithm — the *PearsonMerger* (see 2.1 for detailed description). Briefly, voxel-wise Pearson correlations among spatially connected (adjacent) voxels across subjects are computed and appropriate thresholding of correlations is applied. In this way the algorithm progressively merges voxel pairs that show consistent correlations into clusters. Voxels that correlate with other voxels, which are already part of an existing cluster, will be assigned to that very cluster while voxels with a subthreshold correlation will not be considered part of this cluster.

In addition, we applied a second algorithm – *Spectral Clustering* (see 2.2 for more details), which is based on a similar idea, but serves not to select certain clusters, but rather to parcel the brain in a sensible way (Craddock et al. 2012). The idea here is that if one chooses to average across voxels within the parcels of a brain parcellation scheme, it would be desirable that the within-parcel variance is as small as possible, in order to lose as little information as possible through the averaging procedure. In short, spectral clustering is an established parcellation approach exploiting characteristics of sets of *eigenvalues* of matrices (the *spectrum* of a matrix) to group similar objects. To take into account the spatial connections of voxels, brain images are transformed into graph representation reflecting adjacency; put simply, a matrix containing spatial information about the brain image's voxels. By applying a distance function to all voxel pairs, each element of the aforementioned matrix is assigned a value reflecting the degree of similarity (i.e., the correlation between voxel pairs across subjects). Finally, after computing eigenvectors from the matrix holding information on relative similarity, a k-means clustering algorithm is applied to this eigenvector subspace. As a result, a clustering of the data is obtained, which is sensitive to the spatial properties and degrees of similarity (here, correlation) present in the data. Of note, spectral clustering differs from the earlier described PearsonMerger in that no thresholding is applied to between-voxel correlations. Consequently, each and every voxel in a brain image is assigned to a cluster based on their topological and correlational characteristics, while clusterings obtained via PearsonMerger only consider suprathreshold voxels. Applications and advantages of the here proposed data-driven clustering approaches can be summarized as follows:

(1) Unsupervised identification of coherent regions with consistent between-subject variance. Irrespective of any further application, the algorithm provides an unsupervised method to identify spatially connected clusters that show consistent variation across subjects, which may be biologically interesting in itself. At the very least it provides a way to explore the between-subject structure of a dataset without the need to put any additional information into the analysis.

(2) Feature selection. Noisy voxels are a primary concern in almost any neuroimaging setting. Since voxels that show entirely random variance and thus do not resemble any biologically meaningful information are very unlikely to show high correlations (i.e., correlations greater than expected by the spatial correlation structure of noise), the algorithm provides means to eliminate purely noise-driven voxels.

(3) Optimize ROIs / parcels for regional averaging. The curse of dimensionality is a major problem in neuroimaging classification problems and a common method is therefore to average across voxels within literature-informed ROIs or regions derived from anatomical brain atlases. A critical point is that a potential information loss through regional averaging is unknown. In addition, it is merely an assumption that the defined regions and region boundaries bear any relevance for the classification problem at hand. The present algorithm has two advantages in this regard: (a) it identifies clusters with a consistent pattern across subjects and it is thus natural to average across the voxels within a cluster; (b) the definition of regions and boundaries is data-driven and ensures that the selected regions show consistent variation across subjects (whether the variance is relevant to the classification problem at hand or not).

1.4 The Dataset

We will apply the here presented algorithms to a structural dataset from the Learning and Alcohol Dependence (LeAD) study (see 2.4). This dataset is comprised of sMRI images from patients (N = 119) suffering from alcohol use disorder (AUD, see next chapter) as well as from healthy controls (N = 97). Goal of the LeAD study was to gain a better understanding of the neurobiological basis of AUD. This dataset is particularly interesting because AUD is known to cause substantial structural damage to cortical as well as subcortical areas (Chanraud et al. 2007). The assumption for a machine-learning-based approach is that these brain alterations show a consistent pattern within the group of AUD patients.

1.5 Alcohol Use Disorder

Alcohol use disorders, entailing alcohol abuse and alcohol dependence, are one of the most common and undertreated psychopathologies in developed countries. Only less than 15 % of individuals affected by AUD receive therapy (Cohen et al. 2007). Patients suffering from AUD show a lack of control over their alcohol consumption and drinking is sustained despite exerting tremendous adverse effects on health, cognition and the personal environment. AUD is most prevalent in young adults, mild forms mostly resolve without therapy but more severe progressions require clinical intervention (Grant et al. 2015). Excessive alcohol consumption is accompanied by a number of comorbid diseases and other risks, e.g., accidents, liver damage, heart disease or cancer. Apart from these factors, alcohol intake can result in cognitive, emotional and motivational dysregulation. Heavy alcohol consumption causes cognitive and psychomotor deficits and elevated aggression levels. Initial pleasurable experiences with alcohol, repeated reward expectancy from consuming alcohol, genetic risk, social status and a person's belief to be able to abstain from drinking contribute to the development of AUD (Connor et al. 2011). Diagnosis of AUD typically arises as a side-effect of a treatment or hospitalization due to one of the various health risks associated with the disease. The desired treatment outcome for most AUD patients is sustained abstinence and entails cognitive-behavioral or pharmacological therapy, as well as combinations thereof.

2 Methods

2.1 PearsonMerger

The principle underlying this approach is rather straightforward: by correlating the values of a given voxel with the values of neighboring voxels across individuals, we aim to identify regions that behave consistently over a group of individuals. Moreover, voxels exhibiting large variance across individuals are selected. The key principle is schematized in Figure 1|A: in both scenarios (small and large cross-individual variance) the true values of a voxel pair show consistent behavior across individuals, these true values however are superimposed with noise. Noise is likely to mask the consistent behavior in case of low cross-individual variance (Figure 1|A, left diagram), whereas consistent behavior of voxel pairs in a large-variance environment is still detectable even under identical noise levels (Figure 1|A, right diagram). We thereby employ the Pearson correlation coefficient, which is defined by following equation as the covariance of both samples (vectors carrying voxel values) divided by the product of the two samples' standard deviations:

$$\rho_{x,y} = \frac{cov\,(X,Y)}{\sigma_x \sigma_y}$$

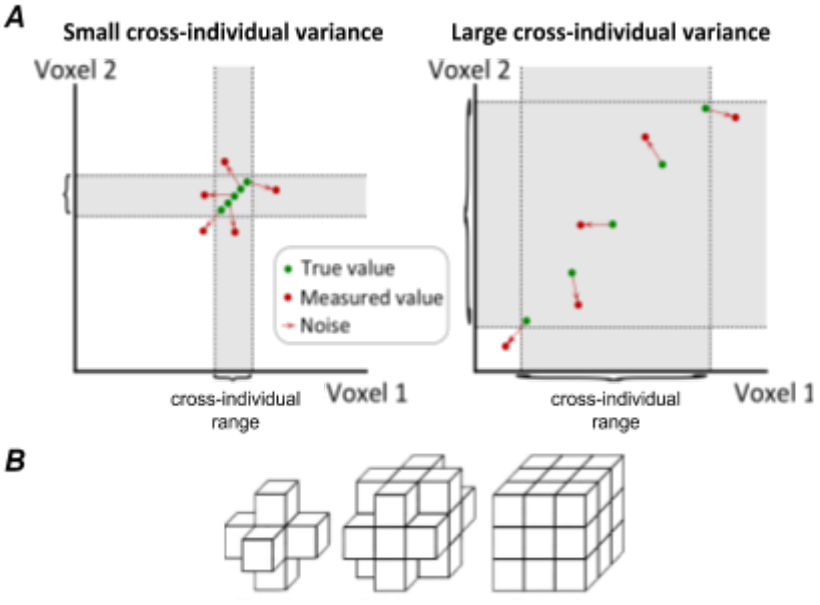

Figure 1| Central idea of Pearson-Merger algorithm and topological voxel neighborhoods. A) The PearsonMerger algorithm relies on large cross-individual variance to detect and select voxels with consistent behavior. Both voxel pairs (left vs. right diagram) have true values of consistent behavior across subjects and identical levels of noise. can this only. However, is consistency only detectable in groups of large variance. In sum, the algorithm was designed to pick up and cluster voxel pairs with consistent values and large cross-individual variance. Both characteristics are promising to enhance classification performance. **B)** Voxels associated adjacent to a center voxel in the 6- (left), 18- (middle), and 26-neighborhood (right). Adopted from Kinkingnéhun et al. (2007).

The algorithm iterates over all voxels of a structural MRI image and computes a set of correlations for all voxel pairs, that is, the current center voxel and all adjacent voxels. We decided to only take into account voxels of the 6-neighborhood to reduce computational complexity (see Figure 1|B). If any of these pairs exhibit a correlation that is larger than a predefined threshold, this pair will be merged and considered as one cluster or ROI. In this manner, spatially connected regions that strongly correlate will be merged into spatially contiguous clusters. As mentioned above, the algorithm accepts a threshold parameter for the correlation coefficient *r* and optionally also for *p*-values. However, in the present work we have focused on relatively high *r*-thresholds ($r > 0.8$), which inherently have high *p*-values, and thus did not investigate thresholding based on *p*-values. Furthermore, users can specify a minimum cluster size below which clusters are filtered out in a final step. An example of the PearsonMerger applied to clustering task can be seen in Figure 2.

Figure 2| Exemplary clustering result for PearsonMerger. Obtained at r-threshold of 0.8717, a total of 30 clusters was identified with a mean cluster size of 98 voxels. Clusters are plotted with unique color coding.

2.2 Spectral Clustering

Spectral clustering methods originate from spectral graph theory, where eigenvalues and eigenvectors of matrices are studied in relation to their graphs. In general, spectral clustering algorithms adhere to the following pattern: (1) transform data into a graph representation, the *similarity graph*:

$$G = (V, E)$$

with vertices v_i representing individual voxels of the dataset and edges e_{ij} indicating spatial connectedness to model local relationships between voxels. In

this way we obtain a matrix harboring every possible pair of voxels, with adjacent voxels assuming values of one while all other pairs are zero. (2) Assign weights to connections between adjacent voxels (the edges) by applying a similarity function to voxel pairs in order to obtain the *adjacency matrix* W. Weight values range from zero (not adjacent) to one (identical voxels). A popular choice for measuring similarity is the *Gaussian similarity function*:

$$s\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right)$$

Here the scale factor σ determines at which rate the euclidian distance (see the numerator in equation above) decays as a function of topological distance. The implementation presented here allows to choose between a slightly modified Gaussian similarity function and a Pearson correlation-based similarity function. (3) From adjacency matrix W the *graph Laplacian* is derived as follows:

$$L = D - W$$

with D being the *diagonal matrix* containing the sum of weights connected to each voxel:

$$D_{ii} = \sum_j A_{ij}$$

Of note, there are many more ways to construct the graph Laplacian, for example the normalized graph Laplacian. (4) Now the first k eigenvectors ($k = 100$ throughout the experiment) of the eigenproblem $Lu = \lambda Du$ are computed and finally a k*-means* algorithm is applied to the matrix containing the k eigenvectors to obtain clustering of the dataset into a predefined number of clusters. k-means clustering itself is a widespread and established clustering tool (Lloyd 1982). Given a set of features (i.e., voxels within an MRI image), k-means clustering divides the features into k sets by minimizing voxel-to-center distances (with μ_i being the centroid and S possible sets of clusters):

$$\arg\min_S \sum^k \sum^x \left|\left|x - \mu_i\right|\right|^2$$

Spectral clustering is a very robust and popular clustering algorithm but is mathematically quite abstract to non-mathematicians. Thus, the simplest version of the algorithm is presented here for simplicity. For a more detailed introduction see (von Luxburg 2007). For the project presented here a Python-based scikit-learn implementation of the spectral clustering algorithm was used. Results from an exemplary clustering can be taken from Figure 3.

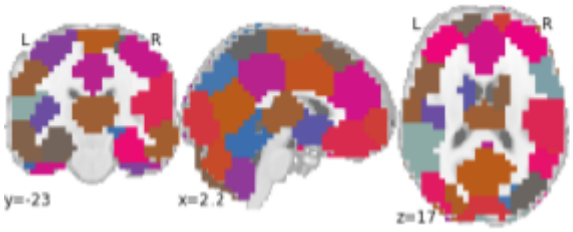

Figure 3| Exemplary clustering outcome using Spectral Clustering method. Predefined cluster number of 50. Clusters are plotted with unique color coding.

2.3 Classification

For classification we employed a novel algorithm that was developed in the Sterzer lab, the *WeiRD* (Weighted Robust Distance) classifier (Guggenmos et al. 2016). On simulated data, but also on real world data, the WeiRD algorithm performed better than the two frequently applied Support Vector Machine (SVM) or Random Forest (RF) classifiers. WeiRD belongs to the class of distance-to-centroid classifiers, with centroid referring to the arithmetic mean of all points within a given shape (i.e., a cluster). WeiRD operates by granting each feature (cluster) of the sample a vote regarding class membership. For each feature f_i of a test sample, the vote v_i is calculated by:

$$v_i = \left| f_i - m_i^A \right| - \left| f_i - m_i^B \right|$$

with $m_i^{A/B}$ being the mean of feature *i* for class A or B in the training data, respectively. The final prediction p of class membership is computed as the weighted sum over all votes:

$$p = sgn\left(\sum_i t_i v_i\right)$$

with t_i being the respective between-class *t*-value for feature *i* on the training data. The final prediction indicates whether a sample belongs to class A, if $p = -1$ while for class B values of $p = 1$ are assumed. To compare different classification outcomes we used the *balanced accuracy score* (Brodersen et al. 2010), defined as:

$$\text{balanced accuracy} = \frac{1}{2}\left(\frac{TP}{P} + \frac{TN}{N}\right)$$

with *TP* referring to true positives, *TN* to true negatives, *P* to positive predictions and *N* to negative predictions. This method of scoring is advantageous over conventional classification scores (e.g., recall scoring) as it is robust against an overestimation of the classification performance in case of imbalanced datasets.

2.4 The LeAD Study, Data Acquisition and Preprocessing

The structural data analyzed here were generated by the LeAD study conducted in Berlin and Dresden (www.lead-studie.de; clinical trial number: NCT01679145), a bicentric study of patients with AUD (N_{AUD} = 119) and healthy controls ($N_{controls}$ = 97). Exclusion criteria for subjects were left-handedness, previous or current substance use except for nicotine dependence in healthy controls and nicotine and alcohol dependence in the patient group. Additionally, major DSM-IV axis one psychiatric disorders and neurologic disorders also led to exclusion. Alcohol-dependent patients had been detoxified on a ward and also were recruited during that phase. Alcohol dependence had to be persistent for at least three years and alcohol dependence severity was assessed using the Alcohol Dependence Scale (ADS) (Skinner & Horn 1984); of note, several other alcoholism-related metrics were determined but are insignificant for the project presented here.

2.4.1 MRI Acquisition

High-resolution T1-weighted structural MRI scans were acquired on a 3-Tesla Siemens Trio scanner using a magnetization-prepared rapid gradient echo sequence (repetition time: 1900 ms; echo time: 5.25 ms; flip angle: 9°; field of view: 256 × 256 mm^2; 192 sagittal slices; voxel size: 1 mm isotropic).

2.4.2 Data Preprocessing

Data were preprocessed and analyzed via SPM12 and voxel-based morphometry (VBM) (Ashburner & Friston 2000). The goal of VBM is to detect local differences in brain tissue composition. This is achieved by spatial normalization of structural images to a common stereotactic space (MNI template) followed by a segmentation into grey matter, white matter, and cerebrospinal fluid based on local probability maps. For the subsequent analyses in the current project we used regional grey matter volume estimates corrected for brain size. To obtain grey matter volume estimates, normalized grey matter images were modulated through a nonlinear-only transformation on normalized images, resulting in relative grey-matter volume maps (Ashburner & Friston 2005). These maps were then smoothed with a Gaussian kernel of 8 mm (full-width at half-maximum, FWHM) and, for computational efficiency, resampled to a resolution of 5 x 5 x 5 mm. Furthermore, the data were linearly corrected for age, sex as well as for the site of image acquisition, collectively referred to as age-correction from here on (since age correction has by far the greatest effect). To this end, voxel-wise linear regression models were fitted to the data with three regressors for sex, age and scanning site. The obtained residual variance constituted the age-corrected data. Finally, the mean across all voxels was subtracted from voxel values, separately for each subject. This normalization procedure ensured that cross-individual consistency was not merely based on individual differences with respect to the overall mean.

3 Results

3.1 Brief Recapitulation of the Project

Aim of this thesis project was to develop a clustering method for structural neuroimaging data, with a focus on classification tasks. The novelty in the approach presented here was its data-driven nature. At the core of the implementation are two distinct clustering algorithms: (1) the PearsonMerger, a threshold-based method that progressively merges voxels into clusters that show consistent behavior between subjects; and (2) spectral clustering, which utilizes spectral graph theory and k-means clustering to group (all) voxels in an image based on a similarity measure. In the following chapters I will describe the results we obtained by applying these data-driven clustering tools to a structural dataset centered on alcohol use disorder.

3.2 PearsonMerger

3.2.1 Qualitative View on Clustering

In the current chapter I want to examine the cluster results obtained with the PearsonMerger algorithm from a qualitative perspective with regards to cluster characteristics. Figure 4 shows exemplary clusterings at varying correlation thresholds. Note that the clustering was based on all participants, that is healthy controls and AUD patients. Furthermore, the left panel shows age-corrected data (see 2.4.2) while the right panel shows clusters generated from non-corrected data . Starting from the top correlation thresholds increase, and with that the number of clusters increases simultaneously (e.g., one cluster in Figure 4|A and 29 clusters in 4|C). With increasing cluster numbers, the average size of clusters dropped. Once thresholds however reached values closer to 1.0 (which would reflect perfect correlation), the number of clusters decreased again (e.g., the drop in cluster number from 4|C to 4|E). In most cases no large differences were found between

clusters generated on age-corrected and not-corrected data, however with one notable exception. The dorsal anterior cingulate cortex, which has been found to be most affected by age (Guggenmos et al., manuscript in preparation), also showed the greatest reduction in cluster size (before age correction: 50 voxels; after age correction: 37 voxels at r = 0.88). Most other regions were consistent in both samples with only moderate deviations in size.

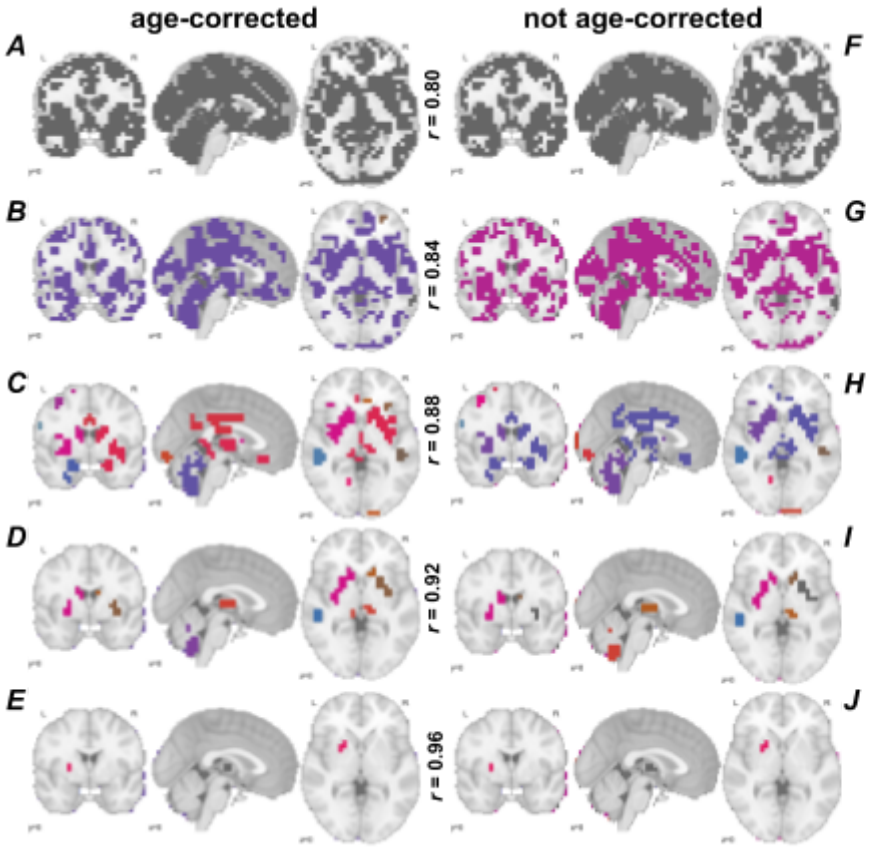

Figure 4| PearsonMerger clusterings at varying Pearson correlation thresholds based on all participants. Images were normalized to zero mean and age-corrected if indicated. **A)** Threshold of r = 0.80: one cluster was detected with size of 8429. **B)** Threshold of r = 0.84: nine clusters were detected, comprising 603 voxels on average. **C)** Threshold of r = 0.88: 29 clusters were created with average size of 83 voxels. **D)** Threshold of r = 0.92: eight clusters were detected with average size of 94 voxels. **E)** Threshold of r = 0.9062, 23 clusters with average size of 95 detected. **F)** Threshold of r = 0.80: one cluster was detected with size of 8397. **G)** Threshold of r = 0.84: six clusters were detected, comprising 911 voxels on average. **H)** Threshold of r = 0.88: 25 clusters were created with average size of 98 voxels. **I)** Threshold of r = 0.92: eight clusters were detected with average size of 92 voxels. **J)** Threshold of r = 0.96, two clusters with average size of 22 detected. Clusters are plotted in unique, arbitrary colors. Slice coordinates: x, y, z = (0, 0, 0).

A few regions are consistently included in clusters, even under strict correlation thresholding: cerebellum, putamen, globus pallidus, caudate nucleus, thalamus and amygdala (see Figure 4|C, D and Table 1). Moreover hippocampus, fusiform gyrus, medial temporal lobe, cingulate cortex and gyrus rectus are found to be covered by clusters even though at lower thresholds. These regions were found in the analysis using all participant data. Comparing clusterings generated on healthy controls and clusterings based on only AUD patients, the following regions show a higher prevalence in the AUD group: middle frontal gyrus, middle temporal

gyrus, superior frontal gyrus, dorsal anterior cingulate gyrus, insula and middle occipital gyrus (cluster images for AUD patients and healthy controls can be found in the supplemental Figures S1 & S2). Interestingly, these regions were also identified as group contrast by a previous study conducted at the Sterzer lab (Guggenmos et al., manuscript in preparation). Regions were identified by comparison of cluster positions with that of brain regions from an anatomical brain atlas (JHU atlas) (Faria et al. 2012).

Cerebellum
Fusiform gyrus
Thalamus
Precentral gyrus
Dorsal anterior cingulate gyrus
Posterior cingulate gyrus
Putamen, globus pallidus, caudate nucleus, nucleus accumbens
Posterior superior temporal gyrus
Superior and middle frontal gyrus
Superior and middle temporal gyrus
Insula

Table 1| Anatomical brain regions covered by clusters. Clusters generated by the PearsonMerger algorithm were compared to anatomical regions found in the JHU brain atlas.

3.2.2 Clustering and Classification performance

After presenting a qualitative view on the clustering results, we next explored as a proof-of-concept whether the data-driven PearsonMerger approach is able to improve the classification between patients and controls. To this end, we applied clustering (over a range of correlation thresholds) to brain scans from participants of the LeAD study, averaged across voxels within these clusters and fed the cluster-wise averages to the WeiRD classifier (see 2.3). Training and testing of the classifier was achieved by leave-one-out cross-validation (CV). Briefly, the algorithm is trained on *N* - 1 samples, and tested on the remaining sample that was excluded from training. Such procedures are common practice and necessary to assess the ability of a classification algorithm to generalize from known samples to unknown

samples and to detect and avoid overfitting. Figure 5 depicts classification scores (via balanced accuracy, see 2.3), mean cluster size and the number of detected clusters over a large range of correlation thresholds (i.e., from $r = 0.01$, up until the threshold where no more clusters were found at $r = 0.96$). We chose to focus on age-corrected data and normalized data, the overview plot for unnormalized data can be found in the supplement (Figure S3). At low thresholds, essentially no effective clustering took place recognizable by the mean cluster size plotted in Figure 5, mid panel. Clustering at such low thresholds effectively created clusters, which contain all or a major fraction of voxels contained in the brain image — the whole brain was treated as a single ROI. With advancing thresholds of $r = 0.7$ and greater, the mean cluster size dropped and at the same time the number of detected clusters increased. Interestingly, the maximum of detected clusters approximately fell together with the optimal classification accuracy obtained on the respective clusterings. Best classification and the peak in cluster number were observed in the range of $0.8 \leq r \leq 0.9$. Upon further increasing the correlation threshold, fewer and fewer clusters survived and concomitantly classification performance and mean cluster size decreased quickly, likely because the data was reduced too much to allow robust classification. Taken together, low thresholds led to single clusters that essentially contained (nearly) all voxels of the brain scan image, increasing thresholds resulted in more detected clusters and improved classification accuracy (with a global maximum). Further incrementing correlation thresholds lowered the number of clusters, and impaired classification, up to the point at which no more clusters were detectable. A closer inspection of classification performance revealed that the peak performance of 0.7418 of the classifier was reached at $r = 0.8677$ (see Figure 6). Briefly on how to interpret these scores: a balanced accuracy score of 0.5 would indicate that the classifier randomly assigns labels to samples, i.e., the classifier is only ‘guessing’, larger values indicate better classification and that the algorithm is able to generalize from training samples to unfamiliar test samples. To benchmark our approach, we used ROIs derived from the JHU brain atlas to compete against ROIs determined via the PearsonMerger algorithm. Using these brain-atlas based ROIs we had reached a classification score of 0.7021, while

classification on clusters generated with our approach yielded classification performances superior to that.

A limitation at present is that the optimal threshold is of course unknown. Future applications will have to show how general the obtained optimal threshold is or whether otherwise the threshold has to be estimated in a nested procedure. Interestingly, we found that the optimal threshold matched the threshold where the largest number of clusters was generated. This is depicted in Figure 7, which shows the obtained number of clusters for different correlation thresholds. Again, future applications will have to show whether this can serve as a heuristic to determine the optimal correlation threshold. Of note, clusterings presented in this chapter, and all results building up on these (e.g., classification scores), were obtained from age-corrected and normalized brain data (see 2.4.2). The supplement chapter contains graphs showing the results for unnormalized data (Figures S3, S4 & S5).

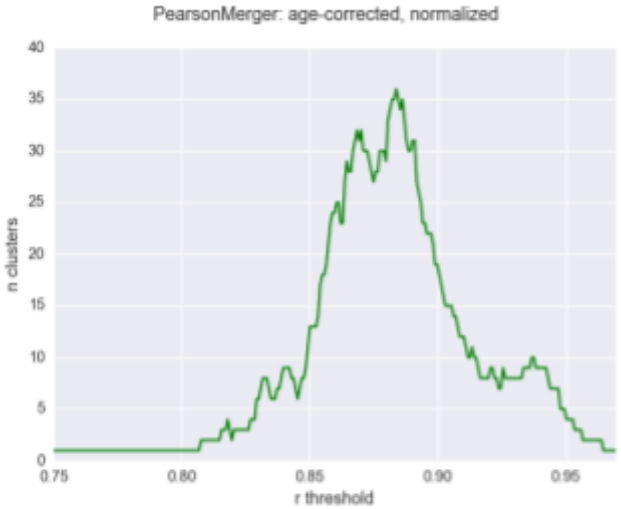

Figure 7| Number of clusters across a fine-resolution range of r thresholds. A maximum of 36 clusters was detected at the correlation threshold of r = 0.8838. Data was age-corrected and normalized to zero mean.

3.3 Spectral Clustering

3.3.1 Qualitative View on Clustering

The idea of our second analytic approach was to generate a data-driven parcellation of the entire brain, similar to the parcellation provided by an anatomical atlas. For this approach, we opted to implement a clustering technique termed Spectral Clustering (see 2.2). Spectral clustering is a method widely used in image segmentation and is based on graph theoretical models taking into account similarity between neighboring image elements (pixels, or voxels in 3D images). Several studies have used spectral clustering in the setting of fMRI, however was it applied only rarely to structural brain scans and never based on cross-individual consistency (see 1.2). A fundamental difference to the PearsonMerger algorithm is that spectral clustering always clusters all voxels of the image into contiguous regions, which are spatially coherent and highly similar. PearsonMerger in contrast, only clusters voxels displaying supra-threshold correlations with its neighbors. It was this elemental difference that made it appealing to implement spectral clustering as a complementing method. As indicated in Figure 8, spectral clustering requires as a parameter the desired number of clusters. Exemplary clusterings resulting from our current dataset are shown in Figure 8, ranging from 10 clusters to 1000 clusters. One can see that, as described above, all voxels of the image are assigned to clusters. With higher cluster numbers, the size of these clusters naturally decreased. Between age-corrected and not-corrected data no apparent divergence could be observed (see Figure 8, left vs. right panel). It would have been beyond the scope of this thesis to quantify to what degree ROIs created via spectral clustering corresponded to regions found in an anatomical brain atlas. However, one feature that is quite apparent is that the clustering, just like an anatomical atlas, has a large degree of hemispherical symmetry (see especially Figure 8|A).

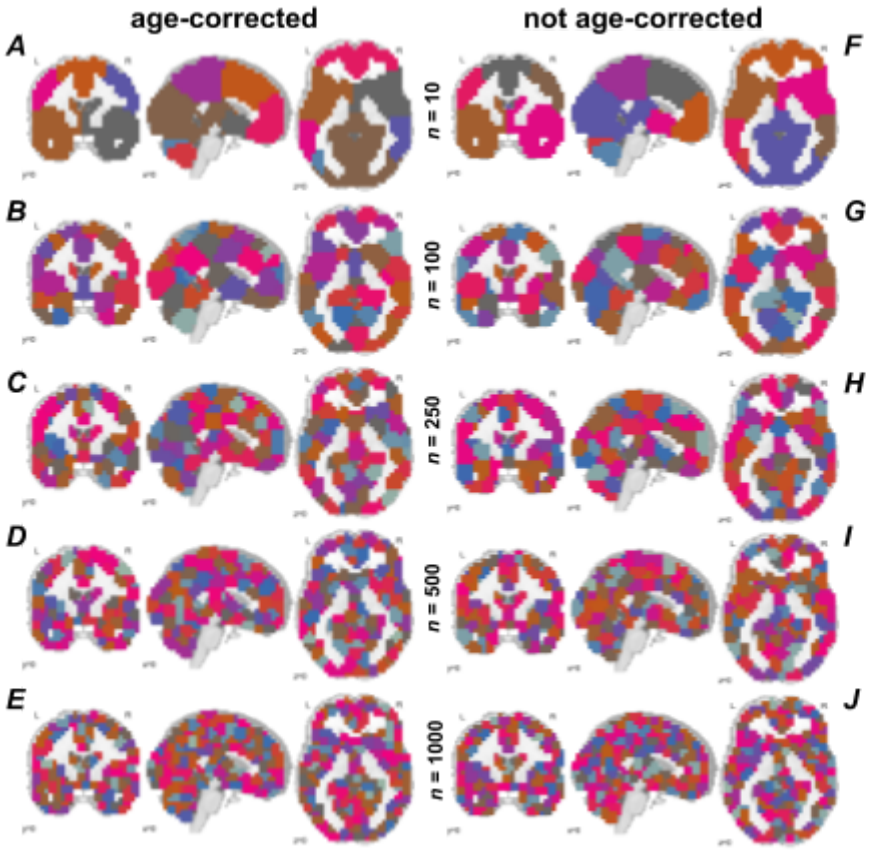

Figure 8| Results of Spectral Clustering at varying cluster numbers based on all participants. Images were normalized to zero mean and age-corrected if indicated. **A, F)** n clusters = 10, mean size of 1117 voxels. **B, G)** n clusters = 100, mean size of 112 voxels. **C, H)** n clusters = 250, mean size of 45 voxels. **D, I)** n clusters = 500, mean size of 22 voxels. **E, J)** n clusters = 1000, mean size of 11 voxels. Clusters are plotted in unique, arbitrary colors. Slice coordinates: x, y, z = (0, 0, 0).

3.3.2 Clustering and Classification performance

Just as we did with PearsonMerger clusterings, we subjected clustered brain images to classification using WeiRD. Again, target of the classifier was to predict whether a brain image belongs to the group of AUD patients or healthy controls. Figure 9 shows balanced accuracy scores for classifications performed on images clustered by spectral clustering. Especially at high cluster numbers this method becomes computationally very expensive, thus regions of lower cluster numbers were computed at higher density while regions of higher cluster numbers were computed more sparsely (note the density of data points across the range of cluster numbers). In contrast to the PearsonMerger-based classifications, an optimal range of classification is less obvious. However, by applying change point analysis to the data (Basseville & Nikiforov 1993), we could obtain estimates for the number of clusters at which classification operated at best (dark curves in the graph below). Briefly, change point detection serves the purpose to detect sudden changes in the

data at hand. In our case, we fitted three linear curves to the data by minimizing the mean squared error. Next we iterated over a range of possible locations at which the three curves intersect. Finally, three curves emerged that indicate the flow of the data: a steep initial increase, followed by a moderate growth phase containing the optimal classification performance (0.7056) and eventually a small but consistent decline in classification scores. The estimated maximum was at 311 clusters. This maximum classification accuracy was below the maximum accuracy level obtained via PearsonMerger clustering; however, it outperformed classification based on the JHU-atlas (see red dashed line in Figure 9) over almost the whole range of cluster numbers. Normalization of the data did not affect classification scores as prominently as in the PearsonMerger approach. Classification scores on clusterings from unnormalized data can be seen in Figure S6.

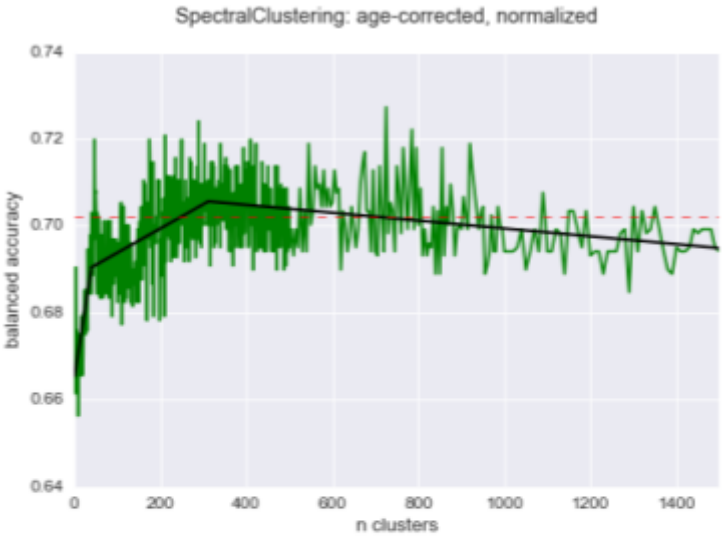

Figure 9| Balanced Accuracy Scores over a range of clusters obtained via Spectral Clustering. Green line represents balanced accuracy scores, red dashed line marks classification performance using the JHU atlas (0.7021), black lines are the result of a change point analysis. Slopes for first line m = 0.00064071 (p = 0), y-intersect n = 0.6648; slope of the second fitted line m = 0.00005587 (p = 0), y-intersect n = 0.6882; and slope of the third line m = -0.00000905 (p = 0), y-intersect n = 0.7085. The first changepoint lies at 39 clusters and the second change point was found at 311 clusters, classification score at second change point is 0.7056. Data was corrected for age and normalized to zero mean prior to clustering.

4 Discussion

4.1 Clustering vs. conventional ROI Creation

Benchmarking classification based on data processed according to the two clustering algorithms against classification based on ROIs derived from the JHU atlas, has shown that our data-driven approach may be superior to a generic anatomical brain atlas in terms of classification performance. Both implemented clustering techniques (i.e., PearsonMerger and Spectral Clustering) yielded higher classification scores than JHU-based ROIs (0.7021). Importantly, PearsonMerger-based clusters achieved better maximal scores (0.7418, see Figure 6) as compared to ROIs created with spectral clustering (0.7056, see Figure 9, via change point analysis). Notably, the range for optimal classification using PearsonMerger is very narrow (around $r = 0.87$) while spectral clustering yielded classification scores that were less sensitive to the number of clusters albeit generally lower. Of note, for PearsonMerger-based classification we observed that the best classification performance was achieved in a range where also the number of detected clusters was largest, although it is unclear at present whether this could provide the basis for a threshold determination heuristic.

In the context of this comparison one may wonder whether the PearsonMerger approach qualifies as a form of double-dipping. Atlas-based ROIs are independent from the brainscan data to be analyzed, while our data-driven approach naturally relies on these data. However, the crucial point is that the PearsonMerger procedure did not make use of the class labels, that is, the ROI generation approach was not informed by the final contrast of interest. In this sense the procedure is connatural to other commonly applied preprocessing methods, e.g. the removal of voxels that show little variance across individuals (in the case of a between-subject design) or across measurements within an individual (e.g. functional MRI).

We furthermore observed that age-correction and normalization of data led to better classification performance on clustered images. With improved classification over conventional atlas-based ROIs, we demonstrated that the here proposed methods for sMRI data clustering may be beneficial for classification tasks. Such classifications are valuable not only for research questions (e.g., to identify regions characteristic for mental disorders) but also in the clinical setting where objective and reliable algorithmic diagnosis is promising to aid physicians as a diagnostic tool. Furthermore, applying a data-driven approach to clustering is likely to alleviate some of the aforementioned issues with conventional ROI-creation.

4.2 PearsonMerger Clustering, Anatomical Regions, and AUD

The PearsonMerger approach generated not only clusters that enhanced classification performance, but these clusters also matched regions that were previously identified as relevant in a AUD versus control group comparison: unpublished work from the Sterzer lab on premature brain aging in AUD has revealed regions that were affected by excessive alcohol consumption most severely in terms of grey matter loss (Guggenmos et al., manuscript in preparation). Among the regions detected by both studies were: dorsal anterior cingulate gyrus, superior frontal gyrus, middle frontal gyrus, superior temporal gyrus, thalamus, insular, fusiform gyrus, and cerebellum. The combined outcome of both studies suggests that regions detected by the PearsonMerger algorithm as having consistent cross-individual variance corresponded to those regions that showed consistent variance between groups in the actual contrast of interest. Although the focus of the project presented here was on the methodical side and to develop the algorithms for data-driven clustering, it is nevertheless encouraging to gain support for the plausibility of detected clusters by another study, which employed a different set of methods.

4.3 Unbiased Creation of Tailor-made Brain Atlases

We developed two distinct methods to cluster sMRI data into coherent, spatially contiguous regions based on consistent behavior across subjects. Consistent behavior was assessed by computing Pearson correlation coefficients of adjacent voxels over all subjects. Both PearsonMerger and Spectral Clustering were able to outperform JHU-based ROIs in terms of classification, with PearsonMerger showing somewhat better maximal classification scores. It is thus conceivable that both methods are able to reduce the complexity of the underlying brain scan data via feature selection (see 1.2) in a way that enhances classification. As such, it provides researchers with a tool to almost automatically create clusterings of their data without requiring to apply a host of *a priori* information. In other words, the use of the techniques presented here makes it possible to create tailor-made ROIs or atlas-like parcellation schemes for individual studies. Both methods require just one critical input parameter, that is the threshold for Pearson correlation for PearsonMerger and the number of clusters in the case of spectral clustering.

Apart from their application as a feature selection step in machine-learning analysis pipelines, they may also serve as an exploratory tool to scrutinize the brain for clusters showing consistent variation across individuals. An application of this exploratory approach could be the minimization of within-group variance of brain data. Minimizing within-group variance is a desirable goal in a machine learning context, as the classification is most accurate in cases of minimal within-group variance and maximal between-group variance. Consider the fact that the PearsonMerger identified the dorsal anterior cingulate cortex (dACC) as a primary region of consistent variation across participants. From previous studies it is known that this brain area is among the ones most affected by aging (Pardo et al. 2007). And indeed, we found that the extent of consistent variation in the dACC was reduced when age-correction was applied, effectively reducing the between-subject variance for this variable of no interest. By identifying brain areas with consistent

variation within a group, the PearsonMerger could thus help to minimizes sources of disadvantageous within-group variance.

4.4 PearsonMerger vs. Spectral Clustering

Finally, I want to highlight the difference between the two clustering methods that we have implemented. Spectral clustering on the one hand is a mathematically complex method using eigenvectors of similarity matrix representations of the data to cluster all elements of the image, and is computationally expensive. In contrast, the PearsonMerger method is less complex and computationally cheaper method that progressively merges neighboring voxels into clusters when predefined correlation threshold are exceeded. Fundamental difference is here that clustering via spectral clustering always includes all voxels, while PearsonMerger clustering only contains those voxels that exhibit suprathreshold correlations with neighboring voxels. PearsonMerger clusters performed punctually better than clusters generated via spectral clustering. As described earlier, the range for optimal classification with PearsonMerger clusters is very narrow while spectral-clustering based clusters perform well over a wide range of input parameters (number of clusters).

4.5 Summary and Limitations

Taken together, with this exploratory proof-of-principle study, we present an unbiased approach to programatically cluster structural neuroimaging data in coherent clusters based on cross-individual consistency; and consistency being assessed via Pearson correlation coefficients among spatially connected voxels. We have shown that our approach can contribute to improvement of classification results over a conventional method of ROI creation, namely brain-atlas based ROIs. The biggest advantage of such data-driven clustering techniques over classic methods is that the obtained clusters fit the data very well and that the danger of discarding informative voxels by manual selection of ROIs is eliminated. To further validate our approach however, the methods presented here need to be tested on

more structural (or even functional) MRI datasets. The intention of the current project was to establish a method for dimensionality reduction (via feature selection) to improve classification. Thus it is favorable to also test for putative enhancement of classification accuracy when using other common classification algorithms such as support vector machines or random forest classifiers to further underpin the integrity of the clustering approaches proposed here.

5 Supplemental Figures

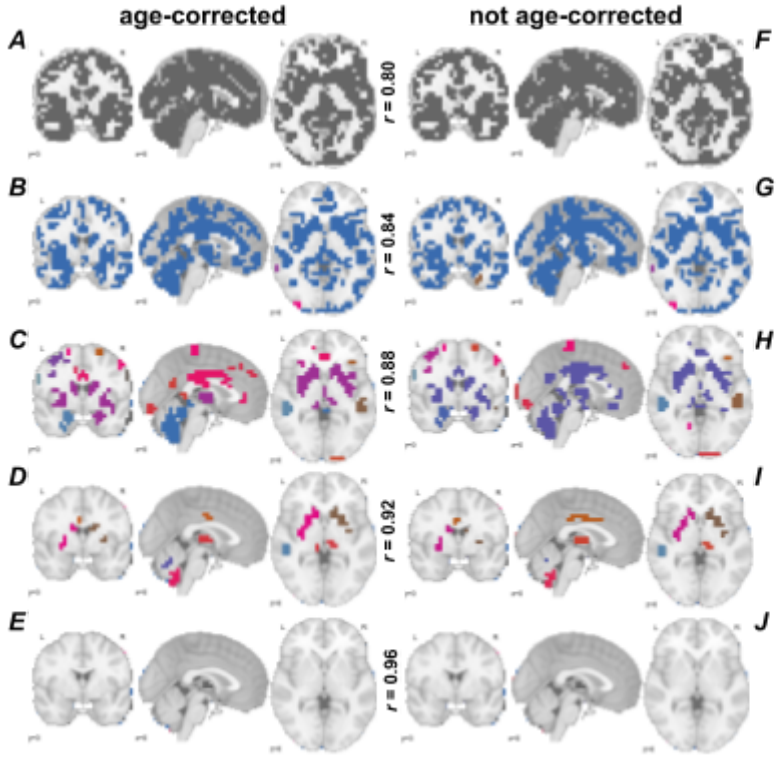

Figure S1| PearsonMerger clusterings at varying Pearson correlation thresholds based on AUD patients. Images were normalized to zero mean and age-corrected if indicated. **A)** Threshold of r = 0.80: one cluster was detected with size of 8647. **B)** Threshold of r = 0.84: seven clusters were detected, comprising 832 voxels on average. **C)** Threshold of r = 0.88: 37 clusters were created with average size of 72 voxels. **D)** Threshold of r = 0.92: ten clusters were detected with average size of 75 voxels. **E)** Threshold of r = 0.9062, one clusters with average size of 40 detected. **F)** Threshold of r = 0.80: one cluster was detected with size of 8551. **G)** Threshold of r = 0.84: seven clusters were detected, comprising 837 voxels on average. **H)** Threshold of r = 0.88: 30 clusters were created with average size of 89 voxels. **I)** Threshold of r = 0.92: 11 clusters were detected with average size of 72 voxels. **J)** Threshold of r = 0.96, one clusters with average size of 30 detected. Clusters are plotted in unique, arbitrary colors. Slice coordinates: x, y, z = (0, 0, 0).

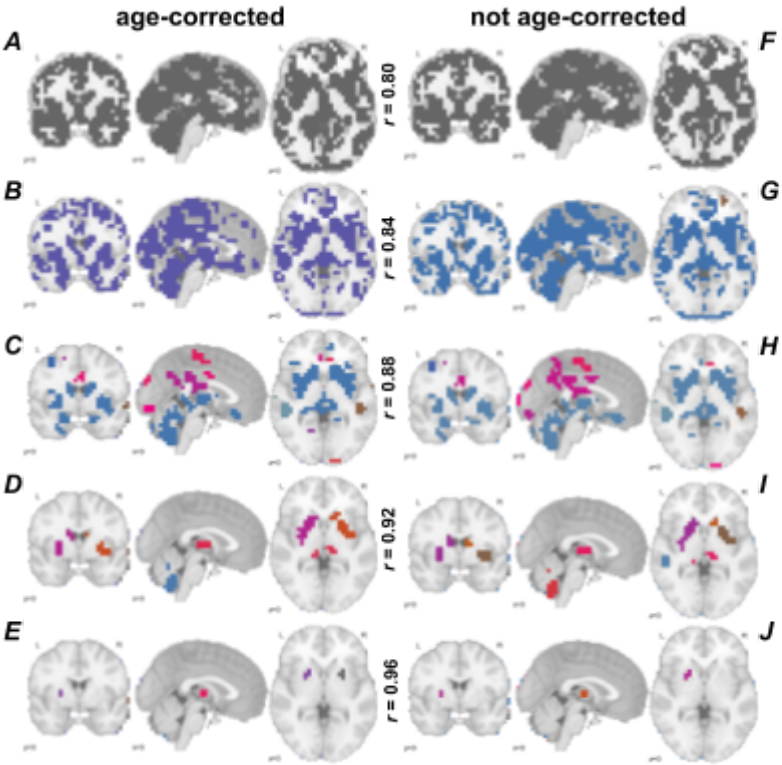

Figure S2| PearsonMerger clusterings at varying Pearson correlation thresholds based on healthy controls. Images were normalized to zero mean and age-corrected if indicated. **A)** Threshold of r = 0.80: one cluster was detected with size of 8685. **B)** Threshold of r = 0.84: five clusters were detected, comprising 1205 voxels on average. **C)** Threshold of r = 0.88: 33 clusters were created with average size of 87 voxels. **D)** Threshold of r = 0.92: nine clusters were detected with average size of 97 voxels. **E)** Threshold of r = 0.9062, four clusters with average size of 20 detected. **F)** Threshold of r = 0.80: one cluster was detected with size of 8397. **G)** Threshold of r = 0.84: eight clusters were detected, comprising 755 voxels on average. **H)** Threshold of r = 0.88: 30 clusters were created with average size of 96 voxels. **I)** Threshold of r = 0.92: ten clusters were detected with average size of 84 voxels. **J)** Threshold of r = 0.96, three clusters with average size of 18 detected. Clusters are plotted in unique, arbitrary colors. Slice coordinates: x, y, z = (0, 0, 0).

Figure S3| Overview plot for clustering obtained by PearsonMerger. A range of thresholds for Pearson correlation coefficients *r* was tested. As a function of *r*: top graph depicts classification performance (assessed by balanced accuracy scores); middle graph shows the mean size of generated clusters; and bottom graph indicates the number of clusters the algorithm was able to detect. MRI data was not normalized but corrected for age.

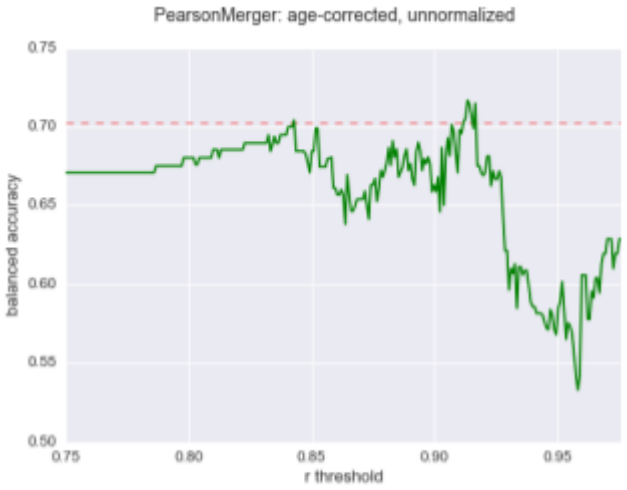

Figure S4| Classification Performance across a fine-resolution Range of Correlation Thresholds. The red dashed line represents classification accuracy on JHU atlas-derived clusters (0.7021). PearsonMerger-based clusters and their classifications scores are drawn in green, optimal performance of 0.7166 is reached at a threshold of *r* = 0.9134. Data was corrected for age but not normalized.

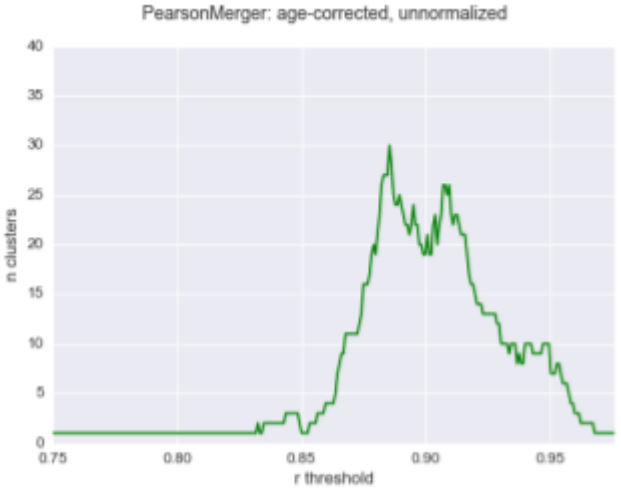

Figure S5| Number of clusters across a fine-resolution range of r thresholds. A maximum of 30 clusters was detected at the correlation threshold of *r* = 0.8854. Data was age-corrected but not normalized.

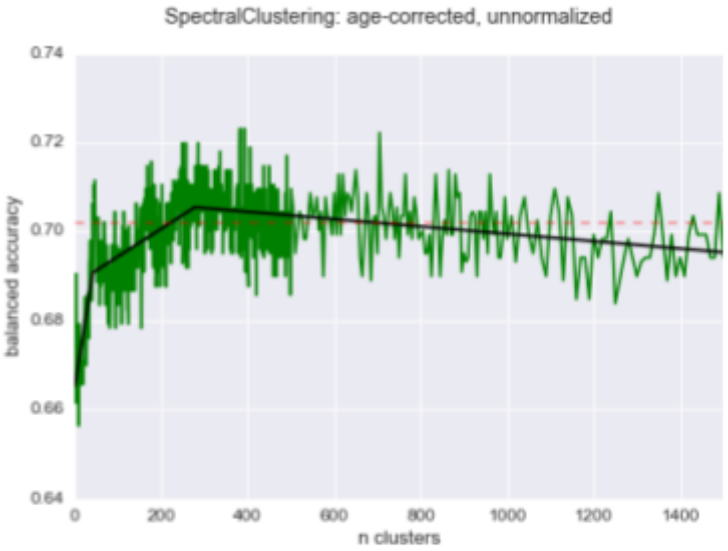

Figure S6| Balanced Accuracy Scores over a range of clusters obtained via Spectral Clustering. Green line represents balanced accuracy scores, red dashed line marks classification performance using the JHU atlas as ROI template (0.7021). Black lines are the result of a change point analysis: slopes for first line m = 0.00061875 (p = 0), y-intersect n = 0.6647; slope of the second fitted line m = 0.00006230 (p = 0), y-intersect n = 0.6881; and slope of the third line m = -0.00000828 (p = 0), y-intersect n = 0.7077. The first changepoint lies at 41 clusters and the second change point was found at 277 clusters. Change point analysis give optimal score of 0.7054 at the second change point. Data was corrected for age but not normalized prior to clustering.

6 References

Admon, R. & Pizzagalli, D.A., 2015. Dysfunctional Reward Processing in Depression. *Current opinion in psychology*, 4, pp.114–118.

Ashburner, J. & Friston, K.J., 2000. Voxel-based morphometry--the methods. *NeuroImage*, 11(6 Pt 1), pp.805–821.

Ashburner, J. & Friston, K.J., 2005. Unified segmentation. *NeuroImage*, 26(3), pp.839–851.

Basseville, M. & Nikiforov, I.V., 1993. *Detection of abrupt changes: theory and application*, Prentice-Hall, Inc.

Brodersen, K.H. et al., 2010. The Balanced Accuracy and Its Posterior Distribution. In *2010 20th International Conference on Pattern Recognition*. Available at: http://dx.doi.org/10.1109/icpr.2010.764.

Chanraud, S. et al., 2007. Brain morphometry and cognitive performance in detoxified alcohol-dependents with preserved psychosocial functioning. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, 32(2), pp.429–438.

Cohen, E. et al., 2007. Alcohol treatment utilization: Findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *Drug and alcohol dependence*, 86(2-3), pp.214–221.

Connor, J.P. et al., 2011. A prospective study of alcohol expectancies and self-efficacy as predictors of young adolescent alcohol misuse. *Alcohol and alcoholism* , 46(2), pp.161–169.

Craddock, R.C. et al., 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8), pp.1914–1928.

Faria, A.V. et al., 2012. Atlas-based analysis of resting-state functional connectivity: evaluation for reproducibility and multi-modal anatomy-function correlation studies. *NeuroImage*, 61(3), pp.613–621.

Grant, B.F. et al., 2015. Epidemiology of DSM-5 Alcohol Use Disorder: Results From the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA psychiatry* , 72(8), pp.757–766.

Guggenmos, M., Accelerated brain aging in alcohol use disorder. *manuscript in preparation*.

Guggenmos, M. et al., 2016. WeiRD - a fast and performant multivoxel pattern classifier. In *2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*. Available at: http://dx.doi.org/10.1109/prni.2016.7552349.

Hawkins, D.M., 2004. The Problem of Overfitting. *ChemInform*, 35(19). Available at: http://dx.doi.org/10.1002/chin.200419274.

Haynes, J.-D., 2015. A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron*, 87(2), pp.257–270.

Heller, R. et al., 2006. Cluster-based analysis of FMRI data. *NeuroImage*, 33(2), pp.599–608.

Hrdlicka, M. et al., 2005. Subtypes of autism by cluster analysis based on structural MRI data. *European child & adolescent psychiatry*, 14(3), pp.138–144.

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory / Professional Technical Group on Information Theory*, 14(1), pp.55–63.

Johnson, K.A. et al., 2012. Brain Imaging in Alzheimer Disease. *Cold Spring Harbor perspectives in medicine*, 2(4), pp.a006213–a006213.

Klöppel, S. et al., 2012. Diagnostic neuroimaging across diseases. *NeuroImage*, 61(2), pp.457–463.

Kuhn, M., Max, K. & Kjell, J., 2013. An Introduction to Feature Selection. In *Applied Predictive Modeling*. pp. 487–519.

Lloyd, S., 1982. Least squares quantization in PCM. *IEEE transactions on information theory / Professional Technical Group on Information Theory*, 28(2), pp.129–137.

von Luxburg, U., 2007. A tutorial on spectral clustering. *Statistics and computing*, 17(4), pp.395–416.

Mazziotta, J. et al., 2001. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 356(1412), pp.1293–1322.

Pardo, J.V. et al., 2007. Where the brain grows old: decline in anterior cingulate and medial prefrontal function with normal aging. *NeuroImage*, 35(3), pp.1231–1237.

Sabuncu, M.R., Konukoglu, E. & Alzheimer's Disease Neuroimaging Initiative, 2015. Clinical prediction from structural brain MRI scans: a large-scale empirical study. *Neuroinformatics*, 13(1), pp.31–46.

Saxe, R. et al., 2006. Divide and conquer: A defense of functional localizers. *NeuroImage*, 30(4), pp.1088–1096.

Skinner, H.A. & Horn, J.L., 1984. *Alcohol Dependence Scale: User's Guide*, Toronto, Canada: Addiction Research Foundation.

Soon, C.S. et al., 2008. Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, 11(5), pp.543–545.

Thirion, B. et al., 2014. Which fMRI clustering gives good brain parcellations? *Frontiers in neuroscience*, 8. Available at: http://dx.doi.org/10.3389/fnins.2014.00167.

7 Declaration of authorship

I hereby declare that I wrote this Master thesis completely on my own and without the use of any other sources or tools apart from those cited. All explanations or citations that I copied directly or in their sense are marked as such.

The thesis or its content has not been handed in to any other official commission.

_____

Stephan F. Mauermann

Berlin, 31st of October 2016