

Diese Arbeit wurde vorgelegt am Lehrstuhl für Informatik 5 (Informationssysteme)
The present work was submitted to the Chair of Computer Science 5 - Information
Systems

**Analyse heterogener Strukturen innerhalb der
Parkinson-Krankheit basierend auf Maschinellern Lernen**
**Machine Learning based analysis of
heterogeneity in the Parkinson's disease**

Bachelorarbeit
Computational Engineering Science

März 2018

Vorgelegt von Presented by	Philippe Drießen Lochnerstraße 55, 52064 Aachen Matrikelnummer 334068 philippe.driessen@rwth-aachen.de
Erstprüfer First Examiner	Prof. Dr. Thomas Berlage Lehrstuhl Informatik 5 RWTH Aachen University
Zweitprüfer Second Examiner	PD Dr. Cord Spreckelsen Institut für Medizinische Informatik RWTH Aachen University
Externer Betreuer External Supervisor	Ph.D. M.Sc. Kaustubh R. Patil Institute of Neuroscience and Medicine (INM-7) Research Centre Juelich

Aachen, 27.03.2018

Eidesstattliche Versicherung

Statutory Declaration in Lieu of an Oath

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)

Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/
Masterarbeit* mit dem Titel

I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

independently and without illegitimate assistance from third parties. I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:

Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt

Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

Para. 156 StGB (German Criminal Code): False Statutory Declarations

Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.

§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt

(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.

(2) Strafflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.

Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence

(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not exceeding one year or a fine.

(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:

I have read and understood the above official notification:

Ort, Datum/City, Date

Unterschrift/Signature

Bachelorthesis

Mr. Cand.-Ing.: Philippe Drießen, 334068

Title: **Machine Learning based analysis of heterogeneity in the Parkinson's disease**

Medical imaging is becoming the main framework for studying the anatomy and function of the human brain. In neuroimaging, multivariate pattern analysis techniques and group analyses are carried out to reveal differences between populations and to derive highly sensitive and specific biomarkers of diseases. The driving assumption behind the majority of the currently existing methods is that a pattern differentiates between two subgroups. This assumption ignores the ample characteristics of heterogeneity in brain diseases.

Disentangling this heterogeneous nature of Parkinson's could contribute substantially to our understanding of this brain disease and may lead to a more accurate diagnosis, prognosis and eventually to a personalized treatment.

Within the scope of this thesis the possibility of revealing heterogeneity in the Parkinson's disease is tackled. Thereby, CHIMERA, a novel probabilistic clustering approach that has been shown to perform well on the Alzheimer's disease, was taken into account. In order to identify certain pathological patterns within the patient population, CHIMERA was applied to different kinds of preprocessed structural, voxel based Parkinson's MRI data. The clustering was carried out with distinct parameters and was afterwards compared to a commonly used k-means clustering algorithm.

The goal of this thesis was to reveal differences between CHIMERA and the classical k-means clustering, according to different parameterization and preprocessing. By showing disparity within the outcome of both algorithms I hope to lay the foundation of a novel strategy in heterogeneity discovery in the Parkinson's disease.

The scope of this thesis covered the following aspects:

- Literature review of the state-of-the-art methodology in pattern recognition, used in neuroimaging.
- Investigation of the potential of the current methods to reveal heterogeneous patterns within the Parkinson's disease.
- Application of state-of-the-art multivariate pattern analysis.
- Evaluation of the selected approaches.

Acknowledgements

I would first like to express my gratitude to my supervisor Ph.D. Kaustubh R. Patil for the useful comments, remarks and engagement throughout the learning process of this thesis. Mr. Patil consistently allowed this paper to be my own work, but steered me in the right direction whenever he thought I needed it.

As this work was created in close collaboration with the *Research Centre Juelich* and especially with the *Institute of Neuroscience and Medicine* (INM-7) I would like to thank Prof. Dr. Simon B. Eickhoff as head of INM7 for making this thesis possible and guiding me whenever I had a question about my research.

Furthermore I would like to thank PD Dr. Cord Spreckelsen from the *Institut für Medizinische Informatik* of the RWTH Aachen University as second supervising professor of this thesis for his very valuable comments on this work and his great support in making this thesis possible, despite the sometimes complicated formalities. Moreover I would like to thank the managing director of the *Institut für Medizinische Informatik* Prof. Dr. Dr. Klaus Kabano for his involvement in finding a suitable supervisor for this thesis.

Also I would like to thank Prof. Dr. Thomas Berlage from the RWTH Aachen *Institute for Information Systems* and head of *Life Science Informatics* at *Fraunhofer Institute for Applied Information Technology* (FIT) for acting as primary supervising professor of this thesis.

Last but definitely not least, I would like to express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Danke!



Philippe Drießen

Contents

1	Introduction	8
1.1	Motivation	8
1.2	Objective	9
1.3	Structure of the thesis	10
2	State-Of-The-Art Methodologies	11
2.1	Neuroimaging techniques	11
2.1.1	Magnetic Resonance Imaging	11
2.2	Pattern analysis in neuroimaging	13
2.2.1	Voxel-Based Analysis	13
2.2.2	Multivariate pattern analysis	14
2.3	Clustering of heterogeneous patterns within disease	14
2.3.1	The a priori approach	14
2.3.2	Clustering of the diseased	15
2.3.3	Introduction to CHIMERA	15
2.4	Summary	16
3	Methodology of clustering heterogeneous disease effects in the Parkinson's disease	17
3.1	Parkinson's Disease	17
3.1.1	Overview	17
3.1.2	Heterogeneity in the Parkinson's disease	18
3.2	Pipeline	19
3.3	Dimensionality reduction	21

3.3.1	Non-Negative Matrix Factorization	22
3.3.2	Principle Component Analysis	24
3.4	Confound Filtering	26
3.4.1	Identifying Confounds	26
3.4.2	Analysis of Variance	26
3.5	Clustering Methods	27
3.5.1	CHIMERA Algorithm	27
3.5.2	K-means Algorithm	31
3.6	Postprocessing	33
3.6.1	Silhouette	33
3.6.2	Adjusted Rand Index	34
3.7	Visualization	36
3.7.1	Multidimensional scaling	36
3.8	Summary	37
4	Results	38
4.1	The Data	38
4.2	Preprocessing	41
4.2.1	Analysis of Variance	41
4.2.2	Principal Component Analysis	42
4.2.3	Non-negative Matrix Factorization	43
4.3	Clustering	45
4.3.1	Parameters	45
4.3.2	Visualization	46
4.3.3	Adjusted Rand Index	48
4.3.4	Silhouette	50
4.4	Cluster Analysis	51
4.4.1	Age	51
4.4.2	Disease Duration	52

4.4.3 Gender	53
4.4.4 Scanner	54
5 Discussion and Prospects	55
5.1 Evaluation	55
5.1.1 Data Set	55
5.1.2 Clustering Analysis	56
5.1.3 Cluster Analysis	57
5.2 Limitations and Outlook	59
5.2.1 Limitation	59
5.2.2 Outlook	60
5.3 Conclusion	61
6 Appendix	62
6.1 Scree Plot	62
6.2 Clustering	63
6.3 Rand Index	64
6.4 Silhouettes	65
6.5 Age	66
6.6 Disease Duration distribution	67
6.6.1 Gender	68
6.6.2 Scanner	69
List of Figures	70
List of Tables	73
Bibliography	74

Chapter 1

Introduction

1.1 Motivation

Several disorders, especially brain diseases, are characterized by heterogeneous symptoms. Diseases that are specified by a clinical heterogeneous presentation include neurodevelopmental and neurodegenerative disorders. For instance in the case of the Alzheimer's disease (AD) researchers showed that it is possible to separate the diseased into different subtypes based upon their distribution of neurofibrillary tangles [1]. In the autism spectrum disorder (ASD) neurodevelopmental disorders are characterized by deficits in repetitive behaviors and social communication [2][3]. For the Parkinson's- and Schizophrenia disease it is possible to obtain certain subtypes by separating symptomatology to discrete symptom domains [4][5][6][7][8][9].

Disentangling heterogeneity within diseases may significantly improve our understanding which can then be used for a more accurate diagnosis, prognosis and eventually for a personalized treatment.

To reveal differences among population in neuroimaging, group analyses are performed. These analyses are carried out to describe for instance disease effects comparing patient and control populations [10] or seeking to find characteristics of brain development by comparing different subjects according to their age [11]. In order to disentangle heterogeneity among populations a wide range of image types are used throughout studies to carry out statistical group analysis. These types of images include functional MRI [12], structural MRI [13][14], and diffusion tensor imaging [15].

A common assumption in the group analysis is that there is a single image pattern that distinguishes the groups [16]. To put it drastically, most of the commonly used computational neuroimaging group analyses only assume one single unifying pattern in the pathophysiological process. For instance they presume that there is a specific disease effect which can be found by comparing a control with a diseased population. However these approaches set value on ignoring the heterogeneous nature of disease phenotypes and thereby lack crucial information when modeling the disease effects. Hence, state of the art analyses may fail to identify the pathological subtypes.

In order to address this problem Aoyan Dong et. al. propose a novel probabilistic clustering approach, *CHIMERA*, in 2016 for modeling the pathological process. They demonstrate that

CHIMERA produces better clustering results compared to two common clustering approaches and also that it is able to discover two main reproducible Alzheimer's disease subtypes that display significant differences in cognitive performance [17].

Utilizing these promising results as foundation, this thesis focused on applying *CHIMERA* to Parkinson's patients, aiming to evaluate clustering results with different kinds of preprocessing methods and comparing them to a standard clustering approach.

1.2 Objective

Ayong Dong et. al. have already evaluated their probabilistic cluster approach on a synthetic dementia and an Alzheimer's disease (AD) data set, demonstrating the superiority in comparison to classical clustering approaches and the discovery of two main and reproducible AD subtypes.

Based on these encouraging results the main objective of this thesis was to evaluate *CHIMERA*'s performance on the Parkinson's disease in comparison to a classical *K-means* clustering approach. The novel method is able to account for possible confounding for example due to gender, age or the locality where the data has been acquired. Correcting confounding effects is crucial in order to prevent falsified results that may not reflect the actual relationship anymore. Thus, as *K-means* is not able to account for confounders itself, *Analysis of Variance* [18] was used to filter out confounding effects before applying *K-means* to achieve a fairer comparison between these two algorithms. Moreover different kinds of preprocessing methodologies were applied to reduce the dimensionality. Having data of high dimensionality can be problematic because computational cost is potentially increased and the learning model might not generalize well. Hence, here namely *Non-Negative Matrix Factorization* (NMF) [19] and *Principal Component Analysis* (PCA) [20] were performed to transform the initial data into different lower dimensional representations.

Although it is hard to analyze clustering results in cases where the ground truth is not known, this work also investigated the obtained results throughout different data representations and clustering methods.

As only a few researchers have taken the effort to reveal the inherent heterogeneity in disease, there is still a wide range of possible directions to investigate. This thesis may only be a small fraction in the process of disentangling heterogeneity amongst diseases, but it could be a start in the right direction.

1.3 Structure of the thesis

This thesis is organized as follows:

In chapter 2 modern state of the art methods are investigated under the desired characteristics of being able to model the pathological process among diseases. Moreover two kinds of image acquisition techniques are explored.

In chapter 3 the algorithms underlying *K-means* and *CHIMERA* are explained in detail. Furthermore different kinds of preprocessing methods are outlined. Besides the technical aspects the Parkinson's disease is reviewed regarding its heterogeneous nature.

In chapter 4 the clinical data set which was used for discovering the patients' subtypes is presented. Moreover clusterings generated by the different algorithms and granularities (different representation of the initial data in lower dimensionality) are illustrated and compared. Furthermore a deeper look into the actual clusters is taken. That is picturing the obtained partitions regarding their reflection of patients' age, sex, scanning site or disease duration.

Chapter 5 is focusing on discussing and evaluating the results of the previous chapter, outlining prospects as well as limitations of this work.

Chapter 2

State-Of-The-Art Methodologies

Medical images and their automated analysis have become the main framework for studying anatomy and function of the human brain. Typically this is done by either applying *Voxel Based Analysis* (VBA) or *Multivariate Pattern Analysis* (MVPA) techniques.

This chapter focuses on these two methodologies outlining why they have possible disadvantages in order to identify pathological subgroups. Also different approaches which were developed to overcome some of the limitations are considered. Moreover a short introduction into medical imaging, specifically *magnetic resonance imaging* is given.

2.1 Neuroimaging techniques

In general neuroimaging can be classified into two broad categories:

1. **structural imaging** deals with the visualization of the anatomical structure of the human brain. In the neurological field structural imaging is for example used to localize pathological structures such as tumors or injuries [21].
2. **functional imaging** is a technique of detecting changes in the metabolism, blood flow or change of chemical composition in specific regions of interests. In contrary to structural imaging the focus lies on revealing physiological activities. For instance *functional magnetic resonance imaging* (fMRI) measures brain activity utilizing the magnetic property of oxygenated blood [22].

Considering that there are a lot of different techniques applied in the field of neuroimaging and that most of the underlying physics of these methods would exceed the limited scope of this thesis by far, only *magnetic resonance imaging* is briefly explained in the following, aiming to explain the fundamentals.

2.1.1 Magnetic Resonance Imaging

In order to obtain images from an MRI scanner the patient is positioned within the machine so that it forms a strong magnetic field around the area of interest. Hydrogen atoms in tissues

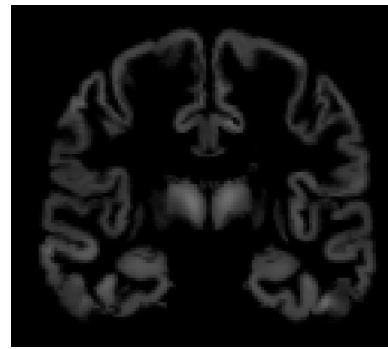
containing water molecules emit a signal that can be processed to form an image of the region of interest (ROI). As energy from an oscillating magnetic field is applied temporarily at a specific frequency the excited protons (hydrogen atoms) beam a radio frequency signal which is received by a coil. The contrast between different kinds of tissues is measured by the rate at which excited atoms return back to their equilibrium state [23].

The deterioration (relaxation) of a textitnuclear magnetic resonance (NMR) can be analyzed using two separate methodologies each with different time constraints. The first one accounts for the loss of signal intensity. To be more precise: it focuses on the time constant associated with the physical process, responsible for the relaxation vector which is in parallel to the externally applied magnetic field. This time constraint is called *spin-lattice relaxation time* T_1 . The second methodology is subjected to the broadening of the signal, which is the relaxation transverse to the static magnetic field named *spin-spin relaxation time* T_2 [24].

T_1 -weighted images usually have an excellent contrast: fluids appear dark, water-based tissues can be seen as mid-gray and fat based tissues are bright. T_1 scans are known as 'anatomy scans' as they show most clearly the boundaries between two different tissues. Examples of T_1 -weighted images are shown in figure 2.1. In T_2 -weighted images, fluids appear with the highest intensity. T_2 images are often considered as 'pathology' scans due to their ability of highlighting abnormal fluid in contrast to darker normal tissue [25]. Figure 2.2 displays two T_2 -weighted images.

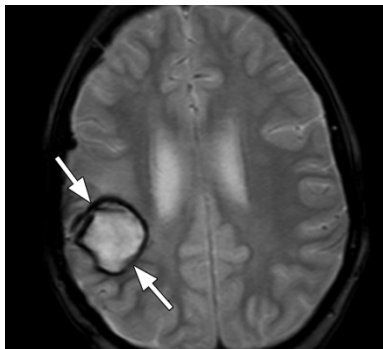


(a) Axial T_1 -MR image illustrating healthy anatomy.

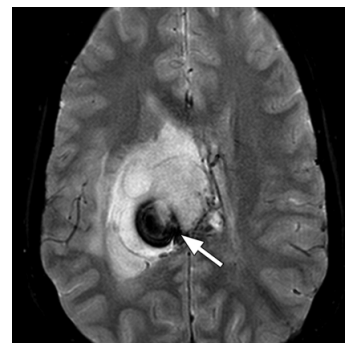


(b) Coronal T_1 -MR image illustrating healthy anatomy.

Figure 2.1: Illustration of T_1 weighted physiological brain images. Images were obtained from the clinical data set provided for this work



(a) Axial T_2 -MR image of a cerebral hematoma (arrows).



(b) Axial T_2 -MR image of a tumoral hemorrhage (arrow) .

Figure 2.2: Illustration of T_2 weighted pathological brain images [26].

2.2 Pattern analysis in neuroimaging

Automated analysis of medical images that are spatially aligned with respect to each other evolved to be the main framework for studying the anatomy and function of the human brain. Usually this is employed by either *Voxel Based analysis* (VBA) or *multivariate pattern analysis* (MVPA) [17]. The aligning of the images is crucial as each human brain differs in its anatomy. In order to be comparable each brain has to be registered to the same normalized template.

In the following sections both MVPA and VBA were investigated and their limitations were outlined in order to identify pathological patterns.

2.2.1 Voxel-Based Analysis

The increase of resolution of anatomical scans especially of the human brain and the sophistication of processing images put emphasis on characterizing differences of neuroanatomical shapes throughout human brains.

One simple example for *voxel-based analysis* is a voxel-wise comparison of the local concentration of gray matter (major component of the central nervous system, consisting mostly of neuronal cells) between two different groups of subjects, e.g. patient and healthy control groups. A voxel is a volume element or region in a tissue slice (for instance obtained by an MRI scan), intuitively a voxel can be seen as a three-dimensional version of a pixel. VBA involves spatial normalization of high-resolution images into the same stereotactic space. After normalization the gray matter segments are averaged with their neighbours (smoothing). After these preprocessing steps, voxel-wise parametric tests are performed comparing the smoothed gray matter images from the two populations [13] .

To put this more generally, *voxel-based analysis* techniques are complementing different regions of interest volumetrically and thereby providing a comprehensive assessment of anatomical differences throughout the brain. Usually *voxel-based analysis* performs mass-univariate statistical tests for instance on tissue components intending to reveal differences in anatomy or shape [27][28][29][30][31][32].

Problems that might occur using these methodologies are twofold, on one hand voxel-wise methods suffer from low statistical power and on the other hand - and more crucial since most of them are only performing mass-univariate statistical tests - they are more likely to ignore multivariate relationships in the data. As shown before there is ample evidence of the heterogeneous nature of diseases especially in case of dementia including the Parkinson's disease.

Considering these aspects, voxel-wise methods might lack potential in revealing different subtypes of disease.

2.2.2 Multivariate pattern analysis

As voxel-wise methodologies are limited by their univariate statistical tests, *multivariate pattern analysis* (MVPA) has gained significant attention due to the capability to capture complex relationships of image signals among brain regions.

Being able to capture multiple relationships in the data these types of methods allow to better characterize group differences among populations and thus have the potential for improved diagnosis and targeted treatment [16].

Based on these encouraging outlooks, machine learning approaches for classification or clustering like *support vector machines* or *linear discriminants* have been used with increased success in order to derive individual, sensitive and specific biomarkers of disease [33][34][35][36][37].

Although *multivariate pattern analysis* methods lead to promising potential in order to identify specific biomarkers of disease, their capability to capture pathological heterogeneous patterns is limited. Reason for this is that typically both *multivariate pattern analysis* and *voxel-based analysis* share a common assumption in which there is only a single pattern that distinguishes between two contrasted groups. Putting it drastically, most of the commonly used computational neuroimaging group analyses only assume a single unifying pattern in the pathophysiological process [16].

2.3 Clustering of heterogeneous patterns within disease

As outlined in subsections 2.2.1 and 2.2.2 both *voxel-based analysis* and *multivariate pattern analysis* lag potential in disentangling the heterogeneous nature of disease. The following sections are focusing on different ways to tackle this problem. The first two segments are covering two different classical approaches while the last one is giving an introduction to the novel clustering method used in this work, *CHIMERA*.

2.3.1 The a priori approach

In order to reveal the inherent disease heterogeneity the first class of methods uses a priori knowledge to predefine the patient population into coherent groups. The splitting criteria for this subdivision might be based on independent criteria (e.g. clinical) and opts to identify group-level anatomical or functional diversity by using univariate statistical methodologies [4][5][6][7][9].

As already mentioned in section 2.2.1 univariate statistical methods ignore multivariate relationships in the data. Furthermore, and more importantly, these kinds of methodologies highly depend on an a priori disease subtype definition. Obtaining these properly can cause several problems as for instance cognitive or clinical evaluation might be noisy, non specific or are in general difficult to obtain [16].

2.3.2 Clustering of the diseased

The second category of methodologies typically performs multivariate clustering that is usually applied directly to the patient population.

Although this group of methods has the potential to capture multivariate patterns it is more likely to put emphasis on clusters that reflect brain anatomies instead of the pathological subtypes. Due to this possibility there is a risk of obtaining clusters which are derived based on normal inter-individual variability caused by age, sex or other confounding variables. [9][4][38][39] In other words clustering applied directly to the diseased population might reflect more confound variability instead of highlighting the desired disease heterogeneity.

2.3.3 Introduction to CHIMERA

As outlined in 2.3 and 2.3.2 commonly used state of the art methodologies fail to identify heterogeneity within the diseased.

In 2016 Aoyan Dong et. al. proposed a novel probabilistic clustering method called *CHIMERA*. In order to capture the pathological process *CHIMERA* uses multiple regularized transformation from a normal control population (e.g cognitive stable, normal brain development etc.) to a patient distribution caused by heterogeneous disease effects.

Considering both populations as point distributions in *CHIMERA* the pathological subtypes are identified by matching the normal and the patient distribution while taking covariates such as age, gender, sex or scanner variability into account. Hereby, each transformation that is being obtained during the mapping process corresponds to one pathological subtype.

Intuitively *CHIMERA* is seeking to find the so called disease effect. That is, given an 80-years old female who is suffering from the Parkinson's disease would have been an 80-years old healthy female had she been spared from the disease; the transition between these two states is considered to be the disease effect [17].

Figure 2.3 shows a visualization of the *CHIMERA* mapping process:

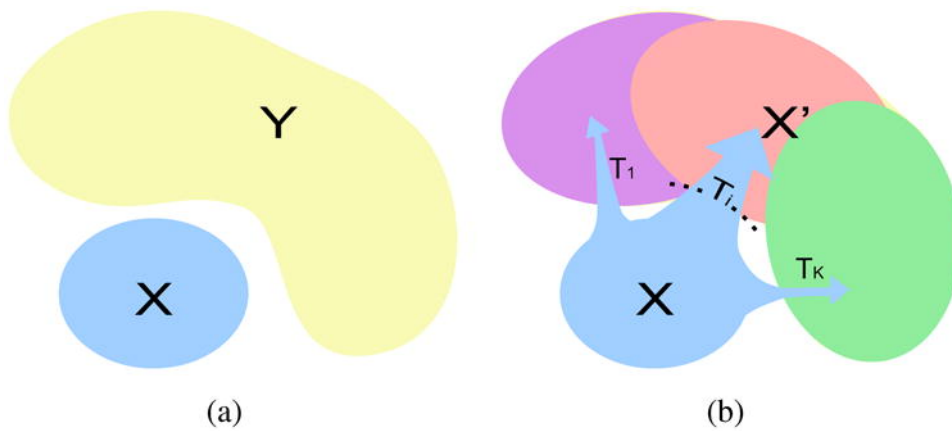


Figure 2.3: (a) the problem setting is displayed where \mathbf{X} denotes the control distribution and \mathbf{Y} the patient population respectively. (b) the model assumption: \mathbf{X} is transformed into a distribution \mathbf{X}' covering the distribution \mathbf{Y} by a set of K different transformations [17].

2.4 Summary

The main framework for studying anatomy and function of the human brain is the automated analysis of spatially aligned medical images. This is typically performed by either *voxel-based analysis* or *multivariate pattern analysis*. As outlined previously a common assumption behind both methodologies is that there is a single pattern which distinguishes between contrasted groups and thus ignores the heterogeneous nature of disease.

Furthermore two different kinds of state of the art methods have been investigated in order to tackle this problem: the first approach uses predetermined subdivisions of the diseased samples obtained by a priori knowledge and tried to identify group-level anatomical or functional differences using univariate statistical methods. Hence, multivariate relationships in the data are ignored. Moreover these types of methods depend on a priori definition of the subtypes, which can be noisy or hard to obtain. The second method applies clustering directly to the patient images. This approach puts emphasis on clusters that are more likely to reflect the normal inter-individual variability of a human being instead of highlighting disease heterogeneity.

Lastly the probabilistic *CHIMERA* clustering approach has been examined. This approach models the heterogeneity in diseases as a set of transformations from a control to a patient population, each transformation corresponding to a pathological subtype. Hence, this clustering approach constitutes a method that is neither depending on a priori knowledge nor only considering the diseased. Thus *CHIMERA* introduces a novel strategy to tackle disease heterogeneity.

Chapter 3

Methodology of clustering heterogeneous disease effects in the Parkinson's disease

In this chapter the Parkinson's disease is briefly outlined regarding its symptoms, treatment and heterogeneity (3.1). Moreover the process of obtaining clusterings from medical image data that was used for this work is outlined in the pipeline section (3.2). Furthermore the fundamentals of the methodologies used in this thesis are investigated.

This includes the clustering algorithms themselves and also pre- and postprocessing approaches. Specifically, the methods of *CHIMERA* (3.5.1) and *K-means* (3.5.2) are explained in detail. As reduction tools *principle component analysis* (3.3.2) and *non-negative matrix factorization* (3.3.1) are considered. Also the correction for possible confounds is presented (3.4). For postprocessing (3.6) or cluster evaluation the *silhouette* and the *adjusted rand index* are outlined. Finally *multidimensional scaling* (3.7.1) is introduced as a visualization technique for high dimensional data.

3.1 Parkinson's Disease

In this section the Parkinson's disease is briefly explained. Also the heterogeneity of the diseased is tackled.

3.1.1 Overview

The Parkinson's disease (PD) is a progressive multi-system neurodegenerative disease that mainly strikes patients in their later years of life. It is considered the second most common neurodegenerative disease worldwide. The prevalence is on the rise along with the change in population demographics [40].

Patients suffering from PD undergo distinctive neuropathological brain changes. For instance the formation of abnormal proteinaceous spherical formations, named Lewy bodies, takes place. The noradrenergic and adrenergic degeneration of dopaminergic nigrostriatal neurons

with Lewy bodies is considered to be the primary neuropathological correlate of motor impairment in Parkinson's disease. Nevertheless there are other kinds of nerve cells which may show similar damage in their cytoskeleton [41].

The presence of symptoms in the PD are generally categorized into motor and non-motor symptoms. Motor symptoms for instance include tremor, slowness of movement, rigidity or postural instability. Non-motor manifestations that can occur are disorders in cognition, mood or behavior and thought [41].

The aetiology of the Parkinson's disease is (probably) multicultural and at the moment there is no available treatment that can either halt or stop the progression of the disease. Current treatments are symptomatic and only aim to correct motor disturbances [42].

3.1.2 Heterogeneity in the Parkinson's disease

Parkinson's disease is a heterogeneous disease. In other words instead of considering only one single disease type there can be several manifestations of the condition derived.

For instance a study focusing on early clinical stages revealed four subtypes of the PD. The first group contains patients with a younger disease onset. The second class includes diseased with a dominant tremor. Another group is characterized by a non-tremor dominant subgroup of patients along with significant levels of cognitive impairment and mild depression. The last subtype identified consists of patients with a rapid disease progression but no cognitive impairment [4].

Roughly speaking Parkinson's disease's heterogeneity can be addressed by separating its symptomatology to discrete symptom domains [4] [9].

Disentangling the heterogeneous nature of disorders is especially important for those without a suitable treatment. Being able to correctly identify a patient's subtype may significantly contribute to understanding the aetiology of the disease. Eventually this knowledge can potentially lead to a more accurate diagnose, prognosis and some day maybe even to a personalized treatment.

3.2 Pipeline

In this thesis several steps were taken before obtaining the actual results outlined in chapter 4. This section is explaining each step that was taken along the process. Figure 3.1 shows an illustration of each of those steps.

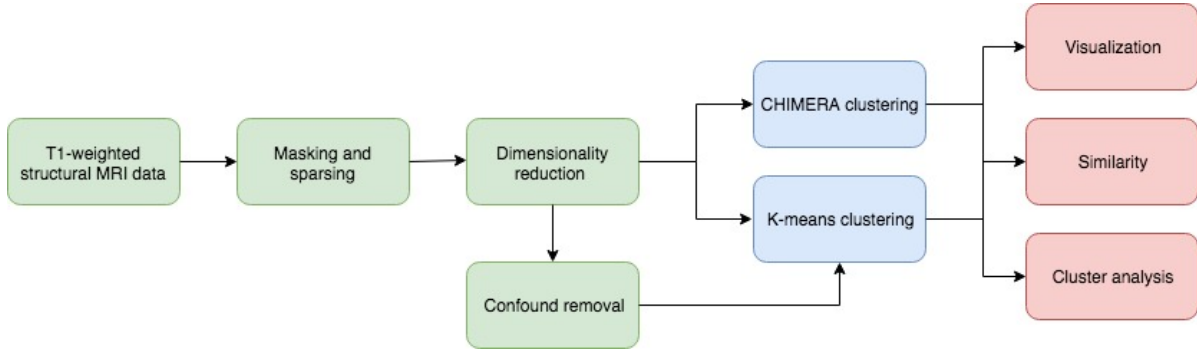


Figure 3.1: illustrates the step-wise procedure of this work. The different process steps are colored as follows: green indicating preprocessing, blue corresponding to the actual clustering and red reflecting the postprocessing.

Preprocessing:

For this work T_1 -weighted images were provided for each subject on which *voxel based morphometry* (VBM) was performed. VBM provides the voxel-wise element estimation of the local amount or volume of a specific tissue compartment [43]. One can think of a voxel as of a three-dimensional pixel within an MR-image. Here, the interest lied in investigating the local distribution of gray matter as research reveals that for instance atrophy in gray matter can be an indicator for different motor subtypes in the Parkinson's disease [44]. Gray matter is a major component of the human brain and comprises most of its neuronal cells. T_1 -weighted images are used due to the fact that they represent the anatomy and therefore the gray matter segments in the best possible fashion. The concept of VBM incorporates different preprocessing steps.

As every individual human brain is different in its shape or size each T_1 -weighted brain image is spatially aligned. That is registering each individual brain to a reference brain template. Moreover after removing any nonbrain parts the tissues are classified (segmentation) into gray and white matter as well as cerebrospinal fluid based on intensity values, also bias corrections for non-uniformities is performed. Lastly, the resulting segmentations are modulated by scaling with the amount of volume changes due to spatial registration so that the total amount of gray matter in the modulated image remains the same as it would be in the original image. An optional step in VBM is spatial smoothing which is typically applied before statistical analysis. After smoothing each voxel represents a weighted mean of its own and its neighbors' values. One reason is that the data is more normally distributed after smoothing and thus makes it more suitable for parametric tests. Another reason is that spatial normalization is not always perfect and small interindividual differences remain. Smoothing accounts for these residual small differences in local anatomy [13].

Starting from the given spatially normalized T_1 -weighted images the partial volumes were segmented according to the *computational anatomy toolbox* (CAT) [45] and the gray matter values for each voxel were extracted.

Instead of extracting the whole brain in this work a mask was used in order to consider only specific volumetric elements. By using only specific regions of interest (ROI) the number of extracted voxels was reduced, leading to faster computational time. Moreover, as only regions of the brain that are known to be involved in the pathological process of the disease were included, noise due to not contributing brain areas was excluded. The mask used for this work was well established and yielded good gray matter coverage. It has been provided by the *Institute of Neuroscience and Medicine (INM-7)* at *Research Centre Juelich*.

Figure 3.2 displays a subject's brain where the mask was overlayed. Eventually only gray matter values in the white colored areas were extracted.

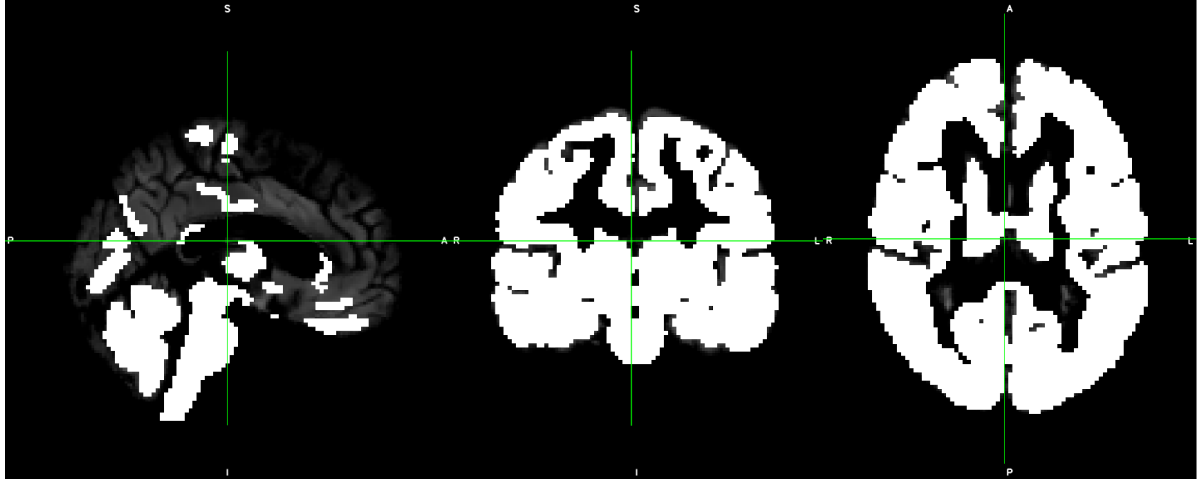


Figure 3.2: sagittal (left), coronal (middle) and axial (right) illustrations of a single subjects brain. Images are T1-weighted brain MR-images where the mask used in this work was overlayed.

After overlaying the mask the obtained voxel number was above 100.000. Before applying any further reduction technique the voxel data was sparsed. Only every second volumetric element was used. The medical interpretation of doing this is that brain tissues close to each other will provide the same sort of information. This way, the dimensionality can further be reduced while still retaining sufficient amount of information. Result of this sparsing process were 20792 voxels.

After applying the masking and sparsing process to all individual subjects all voxels were saved into a feature matrix. Equation 3.1 illustrates the composition of the feature matrix M_f where v denotes the number of extracted voxels and s the number of subjects.

$$M_f = \begin{bmatrix} gm_{1,1} & \dots & gm_{1,v} \\ \dots & \ddots & \dots \\ gm_{s,1} & \dots & gm_{s,v} \end{bmatrix} \quad (3.1)$$

As all images were aligned (normalized) to the same template each column of the matrix corresponds to a gray matter ($gm_{i,j}$) value for the j^{th} voxel of the brain and each row is representing the i^{th} subject respectively. Intuitively, each row of M_f can be seen as a subject represented in an v -dimensional feature space. Here, the features are the gray matter values for a specific voxel.

For faster computational power the dimensionality was drastically diminished. To accomplish this, two different kinds of methodologies were used, namely *principle component analysis* (3.3.2) and *non-negative matrix factorization* (3.3.1).

As *CHIMERA* is able to account for confounds, for a fairer comparison to the *K-means* algorithm, that by itself does not correct for the possible influences of covariates, the data had to be adjusted before applying *K-means*. In this thesis two different kinds of confound removal strategies were applied as reflected in section 3.4.

Clustering:

After preprocessing all samples the feature matrix (3.1) was used to cluster the diseased using both *CHIMERA* and *K-means* algorithm.

Postprocessing:

After clusters from both methods had been collected, different analyzing steps were taken. In order to see if clusterings differed between the *K-means* and the *CHIMERA* approach three different techniques were used to tackle this:

1. *Visualization*: this is reducing the feature space to a two-dimensional space. Samples and their clusterings can thus be investigated in an intuitive fashion. In this work *multi-dimensional scaling* was used [46] as reduction method.
2. *Adjusted Rand index*: the similarity between two clusterings can be mathematically expressed by the *adjusted Rand index*: a value of 1 indicating that each sample is labeled identically and a value of 0 stating that no single sample label matches [47].
3. *Silhouette*: is a method to interpret and validate the consistency within clusters of data. The technique provides a graphical representation of how well each object lies within its assigned cluster [48].

Besides investigating cluster differences the samples that were assigned to the same cluster were analyzed. Specifically, characteristics like gender, age or scanning location were considered.

3.3 Dimensionality reduction

In terms of performance, having data of high dimensionality is problematic because it can lead to high computational cost. Furthermore the trained model might be prone to overfit, which means that the model will perform well on the training data but poorly on test data.

To tackle this, two kinds of dimensionality reduction techniques were used in this work. In this chapter both methods are elucidated, namely *non-negative matrix factorization* (NMF) and *principle component analysis* (PCA).

3.3.1 Non-Negative Matrix Factorization

Non-negative matrix factorization (NMF) describes a group of algorithms where a matrix V is factorized into two significant lower rank matrices W and H . Hereby all three matrices have the property of being non-negative. The satisfaction of these constraints is what differentiates NMF from the rest of the methodologies (e.g. PCA) and leads to components with unique properties. Latter is the attribute that is of significant interest in medical imaging. By multiplying the factorized matrices W and H , V can then be approximated as:

$$V \approx W * H \quad (3.2)$$

NMF is used in multiple applications with either clustering or dimensionality reduction purposes. For instance in the medical field NMF is used for the analysis of structural neuroimaging data, using the property that the measured anatomic values, such as gray matter are always positive [49].

NMF has an inherent clustering property, where all columns of input matrix $V = (v_1, \dots, v_n)$ are clustered. Specifically, given a feature matrix $V_{samples}^{v,n}$ where v denotes the number of voxels which were extracted for each sample and n is the number of samples. $V_{samples}^{v,n}$ can then be approximated as:

$$V_{samples}^{v,n} = W^{v,k} * H^{k,n} \quad (3.3)$$

where k denotes the number of components or hidden features one would like to obtain. As a hyperparameter k is chosen by the operator. Each column in Matrix W can be seen as a hidden feature vector where each cell value defines the rank of each voxel in the feature. The higher a voxel's cell value the higher is the contribution to the corresponding feature. Thus W represents the cluster centroid (i.e. the k^{th} column corresponds to the cluster centroid of the k^{th} cluster). In the scenario above each cell value of the k^{th} column of H can be interpreted as the contribution of each subject to the k^{th} feature or cluster. Hence H yields the cluster indicators, assigning samples with higher impact to the corresponding cluster [50].

In this thesis NMF was mainly used as dimensionality reduction tool. Equation 3.2 can be written as:

$$v \approx W * h \quad (3.4)$$

where v and h are the corresponding columns of v and h . Thus each data vector v is approximated by a linear combination of the columns of W , weighted by the components of h . Hence, W can be regarded as containing a basis that is optimized for the linear approximation of the data in V . As relatively few basis vectors were used to represent significantly more data vectors, good approximation could only be achieved if the basis vectors discovered a structure that is latent (hidden features) in the data. In order to achieve a lower rank representation of $V_{samples}^{v,n}$ the voxel vectors were projected onto the new basis vectors [51].

As shown in equation 3.5 the reduced feature matrix $V_{new}^{n,k}$ can be calculated as the product of $V_{samples}^{v,n}$ transposed times W , with each row corresponding to a sample and each column to a latent voxel structure.

$$V_{new}^{n,k} = V_{samples}^{v,nT} * W^{v,k} \quad (3.5)$$

Typically in order to obtain the factorization matrices W and H the problem is stated as an energy minimization problem utilizing the *Frobeniusnorm* $\|\cdot\|_F^2$:

$$\min_{W,H} \|V * WH\|_F^2, \text{ subjected to } W \geq 0, H \geq 0 \quad (3.6)$$

Although there are several methods in which W and H may be found, a popular method due to its simplicity is the multiplicative update rule: W and H are initialized non negative. Next, the updated values are computed as follows, where n is the index of iteration.

$$H_{[i,j]}^{n+1} \leftarrow H_{[i,j]}^n \frac{((W^n)^T V)_{[i,j]}}{((W^n)^T W^n H^n)_{[i,j]}} \quad (3.7)$$

$$W_{[i,j]}^{n+1} \leftarrow W_{[i,j]}^n \frac{(V(H^{n+1})^T)_{[i,j]}}{(WH^{n+1}(H^{n+1})^T)_{[i,j]}} \quad (3.8)$$

Equations 3.7 and 3.8 are repeated until a predefined residual is achieved [52].

As in medical studies the sample size is usually considerably low in contrast to the feature size (e.g. images), a slightly different version of *NMF* was performed in this work. The tall data matrix V was organized in such a way that each data sample (brain voxel) was aligned column wise ($V = [v_1, \dots, v_N]$, $v_i \in \mathbb{R}^D$, a data sample refers to a vectorized image, also see 3.1). Here, an adopted version of the *projective non-negative matrix factorization* (PNNMF) [53] focusing on its orthonormal extension (*OPNNMF*) [49] was performed. *OPNNMF* factorizes the data matrix by minimizing the following energy:

$$\min_W \|V - WW^T X\|_F^2, \text{ subjected to } W \geq 0, W^T W = I \quad (3.9)$$

where $H = W^T V$ and I denotes the identity matrix. Hence, the solution could be approximated by iteratively applying a multiplicative update rule presented in equation 3.10.

$$W'_{ij} = W_{ij} \frac{(VV^T W)_{ij}}{(WW^T VV^T W)_{ij}} \quad (3.10)$$

In order to apply the scheme, an appropriate initialization is needed. Therefore, the performed initialization strategy proposed by *Boutsidis* and *Gallopoulos* was employed which generates sparse initial components that among other benefits improved the sparsity of the final non-negative decomposition [54].

The implications of estimating the loading coefficients of W are twofold. First, the overlap between the estimated components is greatly decreased, leading to components exhibiting high sparsity, which is a significant attribute in settings like brain modeling [55]. Secondly, the components now provide a grouping of the data variables [49].

3.3.2 Principle Component Analysis

Principle component analysis (PCA) or also known as *Karhunen–Loève* transform is a technique that is commonly used in the fields of dimensionality reduction, feature extraction and data visualization [56].

In general two different formulations of *PCA* are available that lead to the same algorithm: the first definition defines PCA as the mean squared distance between the data points and their projection [57]. Equivalently, it can be formulated as the orthogonal projection of the data onto a lower dimensional linear space (*principle space*), in such a way that the maximum of variance is preserved [58]. In figure 3.3 the orthogonal projection is pictured.

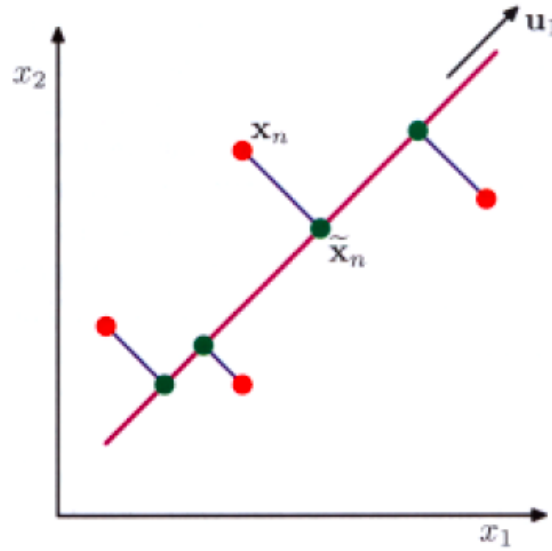


Figure 3.3: the magenta colored line denotes the lower dimensional *principal subspace* that *principle component analysis* seeks to find. The subspace is chosen in such a way that the orthogonal projection of the red colored data points onto the new subspace maximizes the variance of the projected green points. The alternative formulation of *PCA* that is based on minimizing the sum-of-squares of the projection errors is illustrated by the blue lines [59].

Due to the scope of this thesis only the maximum variance formulation is covered in the following.

Maximum variance formulation:

A data set of observations x_n is considered where $n = 1, 2, \dots, N$ and x_n is a Euclidean variable $x_n \in \mathbb{R}^D$. The aim is to obtain a lower dimensional space $M < D$ while maximizing the variance of the projected data points. For simplicity a projection onto a one-dimension space ($M = 1$) is considered. The direction in the data is defined as a unit vector ($u_1 \in \mathbb{R}^D$). Next, each data point x_n is projected onto a scalar value $u_1^T x_n$ where $u_1^T \bar{x}_n$ yields the mean of this projection. The sample set mean \bar{x}_n is given by

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad (3.11)$$

and the variance of the projected data is given by

$$\frac{1}{N} \sum_{n=1}^N \{u_1^T x_n - u_1^T \bar{x}\}^2 = u_1^T S u_1 \quad (3.12)$$

where S is defined as the covariance matrix:

$$S = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})(x_n - \bar{x})^T \quad (3.13)$$

The variance is maximized under the condition of $u_1^T u_1 = 1$ preventing $\|u_1\| \rightarrow \infty$. In order to enforce this constraint the *Lagrange multiplier* λ_1 is introduced which results in the following maximization function:

$$f(u_1, \lambda_1, S) = u_1^T S u_1 + \lambda_1(1 - u_1^T u_1) \quad (3.14)$$

leading to

$$\frac{\delta f}{\delta u_1} = S u_1 - \lambda_1 u_1 \stackrel{!}{=} 0 \Leftrightarrow \quad (3.15)$$

$$S u_1 = \lambda_1 u_1 \quad (3.16)$$

As can be seen from equation 3.16 u_1 must be an eigenvector of S . By multiplying the left hand side with u_1^T and making use of the unit property the variance is given as

$$u_1^T S u_1 = \lambda_1 \quad (3.17)$$

This variance is maximized when u_1 is equal to the eigenvector corresponding to the largest eigenvalue.

Additional principle components can be defined in an incremental fashion by choosing each new direction to be that one which maximizes the projected variance amongst all possible directions with the constraint to be orthogonal to the already existing components [20].

3.4 Confound Filtering

A confounder is a variable whose presence affects the variables being studied so that the results are falsified and do not reflect the actual relationship anymore.

In this section the confounding variables used in this work are specified. Furthermore the method of *analysis of variance* (ANOVA) is explained regarding how it was utilized in order to correct for confounds.

3.4.1 Identifying Confounds

In the field of neuro-imaging confounding variables are ubiquitous. Although there are many influencing covariates in the human brain, in this work the focus lied on correcting three of these possible factors.

1. *Sex*: gender studies of the human brain aiming to reveal differences in its connectivity and anatomy are widely prevalent. For instance research shows that the brain of a human male tends to have a larger total brain volume compared to a female [60]. As this work contained male and female subjects correction for possible differences in brain anatomies due to sex was necessary.
2. *Age*: the aging of the human brain is a whole research field on its own. A pattern that is reported by several MRI studies is that there is a consistent age-related volumetric reduction of gray matter (GM) in the human neocortex, involving mainly prefrontal regions as well as the parietal and temporal association cortices [61]. Hence, a correction for subjects' age was performed.
3. *Scanning site*: MRI data is inherently noisy. Effects such as patient positioning and field of view selection are not constant between scans. Furthermore different scanners might use different imaging protocols or scanner specific software versions for data processing [62]. As in this work the clinical data was provided by two hospitals with different scanners, controlling for these possible effects was crucial.

3.4.2 Analysis of Variance

In order to filter effects which are caused by gender, sex or scanner locality in this work the statistical method of *analysis of variance* (ANOVA) [63] was applied.

ANOVA, is a powerful statistical technique that involves partitioning the observed variance into different components. In this thesis the variance in the gray matter values was partitioned into components, accounting for different sources of variation (confounds). Once the variance due to the confounders in the dependent variable was explained, the residuals (the part of the total variability that still remained unexplained) yielded the confound corrected data.

3.5 Clustering Methods

Intuitively one can think of a clustering as a grouping of objects where objects within the same group (cluster) are more similar to each other than to those in other groupings.

In this section the two types of clustering methods applied in this work are explained in more detail, namely *CHIMERA* and *K-means*.

3.5.1 CHIMERA Algorithm

As previously explained in section 2.3.3 *CHIMERA* models the pathological effect as a set of transformations from a healthy to a patient distribution.

Method [17]:

Assuming a data set containing M normal control (NC) samples $X = \{x_1, \dots, x_m\}$ and N patients samples $Y = \{y_1, \dots, y_n\}$. For simplicity it is assumed that the samples can be described by two sets of features: a set of D_1 - dimensional image features: $x_m^v, y_n^v \in \mathbb{R}^{D_1}$; and a set of D_2 -dimensional covariate (e.g age, sex, scanner) features: $x_m^c, y_n^c \in \mathbb{R}^{D_2}$. Utilizing the more compact notation leads to: $x_m = (x_m^v, x_m^c), y_n = (y_n^v, y_n^c)$.

Given that there are enough NC points in the data set in order to describe the NC population, *CHIMERA* assumes that the estimated anatomy of patients, had they been spared from the disease, can be covered by the normal control distribution. Under the assumption that all the patients can be associated with NCs and that, contrarily, the transformed NC points cover the entire set of patients, the transformation T is retrieved by matching patient and NC distributions.

The maximum a posteriori optimization of the model leads to the following energy minimization:

$$\epsilon(X, Y, \theta) = -\mathcal{L}(X, Y, \theta) + \mathcal{R}(\theta) \quad (3.18)$$

where θ denotes the parameters of the model, \mathcal{L} the log-likelihood of the distribution X and Y given the parameters θ and \mathcal{R} denoting a regularization term for improving stability.

To cover the heterogeneous process a transition T consists of multiple possible transformations, each representing a pathological direction of image change:

$$x'_m = T(x_m) = (T(x_m^v), x_m^c) \quad (3.19)$$

where $X' = [x'_1, \dots, x'_M]$ are the transformed NC samples.

The transformation only applies to x_m^v covariate image features, x_m^c remaining the same.

The matching of distributions Y and X' is found by an adopted version of the *coherent point drift* algorithm [64]. Each point x'_m is a centroid of a spherical Gaussian cluster where each of those clusters is considered to have the same variance σ^2 . Points y_n are treated as i.i.d. data generated by a *Gaussian mixture model* (GMM) [65] having an equal weight $P(x'_m) = \frac{1}{M}$ for

each cluster. The similarity between the two distributions can thus be measured by the data likelihood of this mixture model.

$$\begin{aligned}
 P(X, Y) &= \prod_{n=1}^N \sum_{m=1}^M P(x'_m) P(y_n | x'_m) \\
 &= \prod_{n=1}^N \sum_{m=1}^M \frac{1}{M} \frac{r^{\frac{D_2}{2}}}{(\sqrt{2\pi}\sigma)^{D_1+D_2}} * \\
 &\quad \exp \left\{ \frac{\|y_n^v - T(x_m^v)\|^2 + r\|y_n^c - x_m^c\|^2}{-2\sigma^2} \right\}
 \end{aligned} \tag{3.20}$$

Equation 3.20 describes the likelihood of data Y generated by centroids X' where the distance between two points is measured by *RBF* kernels. The size of covariate features is chosen to be r times larger than the kernel size of the image feature, where r is determined by the ratio of total variance of the features. This is to correct for influence due to possible confounds.

The transformation from one NC point to the patient space can be denoted as

$$T(x_m^v) = \sum_{k=1}^K \zeta_{km} T_k(x_m^v) \tag{3.21}$$

where K is the number of pathological directions T_1, \dots, T_k one intends to find.

In the ideal case, if the disease subtypes would be distinct, ζ_{km} should take the value of 1 for the transformation responsible for this specific subtype that effects x_m and 0 otherwise. As different methodologies might correspond to the same point in the normal control space, *CHIMERA* uses a soft assignment in which the variable ζ_{km} is relaxed in such a way that it sums up to 1 for each x_m .

In order to map the two populations *CHIMERA* performs linear transformations. Hence, each T_k can be described as a set of parameters $(A_k, b_k) \in (\mathbb{R}^{D_1 \times D_1}, \mathbb{R}_1^D)$:

$$T(x_m^v) = \sum_{k=1}^K \zeta_{km} (A_k x_m^v + b_k) \tag{3.22}$$

where $\sum_k \zeta_{km} = 1$ and $\zeta_{km} \geq 0$.

Depending on how the matrix A_k is chosen, different kinds of transformation are possible:

1. *Full matrices*: affine transformation. This is rotating, scaling and translation.
2. *Diagonal matrices*: only translation and scaling is possible.
3. *Identity matrices*: only the translation b_k is considered.

Putting it all together, the obtained log-likelihood of the data is given by introducing equation 3.22 into formula 3.20:

$$\begin{aligned}\mathcal{L}(X, Y, \Theta) &= \prod_{n=1}^N \log \sum_{m=1}^M P(x'_m) P(y_n | x'_m) \\ &= \prod_{n=1}^N \log \sum_{m=1}^M \frac{1}{M} \frac{r^{\frac{D_2}{2}}}{(\sqrt{2\pi}\sigma)^{D_1+D_2}} * \exp\left\{\frac{r\|y_n^c - x_m^c\|^2}{-2\sigma^2}\right\} * \\ &\quad \exp\left\{\frac{\|y_n^v - \sum_{k=1}^K \zeta_{km}(A_k x_m^v + b_k)\|^2}{-2\sigma^2}\right\}\end{aligned}\quad (3.23)$$

Typically the sample size is quite low in medical settings (especially in image driven studies) in comparison to the large image scale. In order to prevent ill posed problems *CHIMERA* introduces a regularizing or penalty term into the energy function 3.18:

$$\mathcal{R}(\Theta) = \frac{\lambda_1}{2\sigma^2} \sum_k \|b_k\|_2^2 + \frac{\lambda_2}{2\sigma^2} \sum_k \|A_k - I\|_F^2 \quad (3.24)$$

To derive an analytical solution the Frobenius norm of $A_k - I$ and the ℓ^2 norm of b_k is penalized, where I denotes the identity matrix.

Optimization [17]:

In order to optimize the parameters $\Theta = (A, b, \zeta, \sigma^2)$ of the model, where $A = \{A_1, \dots, A_K\}$ and $b = \{b_1, \dots, b_K\}$ *CHIMERA* uses an *Expectation-Maximization* framework [66][65].

The algorithm introduces latent variables z indicating the posterior probability of data point for each mixture component, $q_{nm} = q(z_n = x'_m | y_n)$. Utilizing this a lower bound of the log-likelihood is achieved [65].

$$\mathcal{F}_0 = \sum_{n,m} q_{nm} \log\left(\frac{P(y_n, x'_m)}{q_{nm}}\right) \quad (3.25)$$

The energy ϵ is minimized in an iterative scheme, calculating alternately the expected value of q (E-step) with respect to the parameters obtained previously $\Theta^{(t-1)}$ and updating $\Theta^{(t)}$ by minimizing the objective function illustrated in equation 3.27 (M-step) in the current iteration t .

Specifically, *CHIMERA* calculates each step as followed:

E-step: using parameters $\Theta^{(t-1)}$ evaluated in previous M-step, equation 3.25 is optimized at $q_{nm} = P(z_n = x'_m | y_n)$:

$$q_{nm} = \frac{\exp\left(\frac{\|y_n^v - \sum_k \zeta_{km}(A_k x_m^v + b_k)\|_2^2 + r\|y_n^c - x_m^c\|_2^2}{-2\sigma^2}\right)}{\sum_{i=1}^M \exp\left(\frac{\|y_n^v - \sum_k \zeta_{ki}(A_k x_i^v + b_k)\|_2^2 + r\|y_n^c - x_i^c\|_2^2}{-2\sigma^2}\right)} \quad (3.26)$$

M-step: the objective function $\mathcal{F}(\Theta)$ is constructed as an upper bound of the energy function

ϵ . Minimizing $\mathcal{F}(\Theta)$ is equivalent to the minimization of ϵ [67].

$$\begin{aligned} \mathcal{F}(\Theta) = & \frac{1}{2\sigma^2} \sum_{m,n} q_{nm} \left(\|y_n^v - \sum_k \zeta_{km} (A_k x_m^v + b_k)\|^2 + r \|y_n^c - x_m^c\|_2^2 \right) \\ & + \frac{N(D_1 + D_2)}{2} \log \sigma^2 + \frac{\lambda_1}{2\sigma^2} \sum_k \|b_k\|_2^2 + \frac{\lambda_2}{2\sigma^2} \sum_k \|A_k - I\|_F^2 \end{aligned} \quad (3.27)$$

subjected to: $\sum_{k=1}^K \zeta_{km} = 1$ for $m = 1, \dots, M, 0 \leq \zeta_{km} \leq 1$

The objective function is not globally convex but jointly convex in each parameter. Hence *CHIMERA* performs an iterative procedure by minimizing the objective sequentially with respect to σ^2, ζ, A and b . The closed form solution for σ^2 , A and b is derived by setting the derivative of the objective function to zero. ζ is optimized using an advanced projected gradient descent algorithm.

The optimization is terminated when the residual between two iteration steps is small enough. Furthermore the *Expectation Maximization* only guarantees a local minimum solution, thus *CHIMERA* is re-run multiple times. Depending on the operator the solution that yields the smallest energy during multiple runs or the most reproducible model can be saved.

CHIMERA initializes the parameters σ^2 as mean distance between X and Y , ζ to be uniformly distributed for each x_m , each A_k to be the identity matrix I , the translation term b_k was sampled from a normal distribution $\mathcal{N}(0, 1)$.

Clustering [17]:

The probability for a normal control sample x_m to undergo the transformation T_k is considered in the coefficients ζ_{km} . The likelihood of a patient sample y_n to be associated with an NC x_m can be denoted as $P(y_n|x_m)$. Then, the probability of a given patient y_n being generated by the transformation of T_k is estimated by:

$$P_k(y_n) = \sum_m P(y_n|x_m) \zeta_{km} \quad (3.28)$$

As the posteriors q_{nm} are proportional to $P(y_n|x_m)$ and include a common denominator (equation 3.26) they can be used to partition the patient samples according to their main transformation. For each patient y_n the label l_n corresponds to the largest likelihood:

$$l_n = \underset{k}{\operatorname{argmax}} P_k(y_n) = \underset{k}{\operatorname{argmax}} \sum_m q_{nm} \zeta_{km} \quad (3.29)$$

As long as the ζ_{km} coefficients are stored, the labeling for a novel data s can be estimated by:

1. computing the likelihood $P(s|x_m)$ based on the distances between the novel sample s and the transformed controls X' .
2. calculating $P_k(s)$ and thereby obtaining the label $l_s = \underset{k}{\operatorname{argmax}} P_k(s)$.

3.5.2 K-means Algorithm

K-means is a popular used method for clustering. This method aims to partition N observations into k clusters where each observation is assigned to the nearest cluster center. Intuitively, one can think of *K-means* clusters as comprising a group of data points whose inter-point distances are small compared to observations outside of the cluster.

Method: [68]

Assuming a given data set $\{x_1, \dots, x_N\}$ consisting of N observations of a random D -dimensional Euclidean variable x . Let μ_k be the prototype of the k^{th} cluster, this is a D -dimensional vector which is representing the centers of the corresponding cluster. The cluster belongings are encoded in a one-hot notation. Therefore, a binary indicator $r_{nk} \in \{0, 1\}$ is introduced, where $k = 1, \dots, K$ describes which of the k clusters is assigned to data point x_n . If point x_n belongs to a specific cluster k then $r_{nk} = 1$ and $r_{ni} = 0$ for $i \neq k$ hold.

The objective function is defined as:

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2 \quad (3.30)$$

To minimize J , an iterative procedure is used in which each iteration involves two successive steps. First μ_k is initialized, afterwards J is alternately minimized with respect to r_{nk} and μ_{nk} taking the partial derivative respectively. It can be noticed that these two stages of updating correspond to the *E-step* and *M-step* of the *expectation maximization* algorithm.

In order to minimize J , r_{nk} has to be chosen such that:

$$r_{nk} = \begin{cases} 1, & \text{if } k = \operatorname{argmin}_j \|x_n - \mu_j\|^2 \\ 0, & \text{otherwise} \end{cases} \quad (3.31)$$

Which basically is the assignment of the n^{th} data point to the closest cluster center.

Considering the optimization of μ_k where r_{nk} is held fix (this is setting the partial derivative of J to zero):

$$\frac{\delta J}{\delta \mu_k} = 2 \sum_{n=1}^N r_{nk} (x_n - \mu_k) \stackrel{!}{=} 0 \quad (3.32)$$

solving for μ_k leads to:

$$\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}} \quad (3.33)$$

The denominator in this expression is equal to the number of points assigned to cluster k , thus μ_k is set to be equal to the mean of all data points x_n assigned to cluster k . For this reason this procedure is called *K-means* algorithm.

The two phases of re-assigning data points to clusters and re-computing the cluster means are repeated alternately until there is no further change in the assignments or until a maximum number of iterations is exceeded. Because each iteration reduces the value of the objective function J , convergence of the algorithm is assured.

As *K-means* is a heuristic algorithm, there is no guarantee that it converges to a global minimum, and the result may depend on the initial clusters. As the algorithm is usually very fast, it is common to run it multiple times with different parameters. Figure 3.4 illustrates the clustering process of *K-means*:

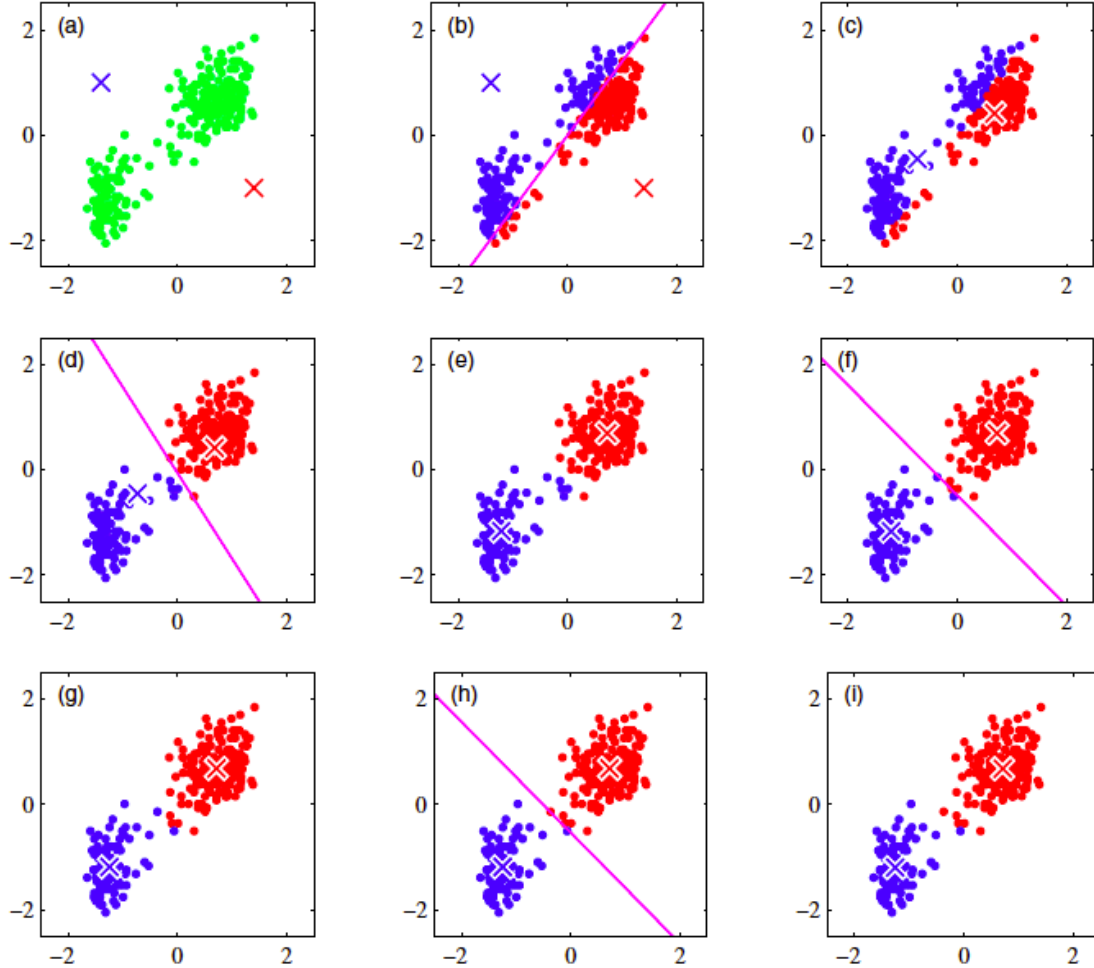


Figure 3.4: Visualization of K-means:

- (a) two-dimensional data in the Euclidean space (green points). Not yet clustered. The initial centroids or prototypes μ_1 and μ_2 are denoted by the red and blue crosses respectively.
- (b) the initial E-step assigns each data point to the red or the blue cluster according to the closest centroid. The perpendicular bisector of the two cluster centers, shown by the magenta denotes the decision boundary.
- (c) each cluster center is recomputed (M-step) to be the mean of the points assigned to the corresponding cluster
- (d) - (i) alternately E and M steps until final convergence of *K-means* is reached [69].

3.6 Postprocessing

This section tackles the analysis of the clusterings. Here, the interest lied in particular on characterizing the similarity between the two clustering methods. Hence, the *adjusted Rand index* and the *silhouette* that were applied are described in the following. Furthermore in order to visualize the high dimensional data in a lower dimensional space, the method of *multidimensional scaling* is outlined.

3.6.1 Silhouette

Intuitively one can think of the *silhouette* as a graphical representation of how well each object is located within its cluster. In this work, the *silhouette* was used to identify differences in similarity or consistency within a clustering method.

Specifically, the *silhouette* measures how similar an object is regarding its assigned cluster (cohesion) in comparison to other clusters (separation). Hence, it provides a method for interpretation and validation of a clustering.

The *silhouette* has a range from minus-one to plus-one where a high positive value indicates that the considered object fits well into its assigned cluster and poorly into the neighboring clusters. If the *silhouette* yields a small or negative value the object might be more similar to other clusters in comparison to its own. It is important to note that the *silhouette* defines similarity based on distances. This is an object with a small distance to its assigned cluster and a large gap to its neighbor clusters will score a high *silhouette* value [48]. Thus one can assume high silhouettes for objects in clusters produced by *K-means* as its algorithm is designed to assign observations to the closest cluster centers. As *CHIMERA* uses a different cluster technique different silhouettes were expected.

Method [48]:

Given an object o where $o \in A$ where A is the cluster it is assigned to. The silhouette is defined as:

$$s(o) = \begin{cases} 0, & \text{if } a(o) = b(o) = 0 \\ \frac{b(o) - a(o)}{\max\{a(o), b(o)\}}, & \text{otherwise} \end{cases} \quad (3.34)$$

where:

- $a(o)$ = average dissimilarity of o to all other objects of A . In figure 3.5 this is the average length of all lines within cluster A . Mathematically expressed $a(o)$ can be denoted as:

$$a(o) = \text{dist}(A, o) = \frac{1}{n_A} \sum_{a \in A} \text{dist}(a, o) \quad (3.35)$$

where $\text{dist}(i, j)$ defines the distance between i and j and n_a is the number of elements in cluster A .

- Let $d(o, D)$ be the average dissimilarity of objects o to all objects of other groupings for which $D \neq A$ holds. Specifically this is the average length of all lines going from o to C or B in figure 3.5. After computing $d(o, D)$ for all clusters the smallest value is selected

and denoted by:

$$\begin{aligned} b(o) &= \min_{D \neq A} d(o, D) \\ &= \min_{D \neq A} \left(\frac{1}{n_D} \sum_{d \in D} \text{dist}(d, o) \right) \end{aligned} \quad (3.36)$$

where n_D denotes the number of objects in the clustering. The nearest cluster to A is then called the neighboring cluster.

The average over all silhouettes of a cluster is called the *silhouette coefficient* and is defined as:

$$\frac{1}{n_A} \sum_{o \in A} s(o) \quad (3.37)$$

The *silhouette coefficient* can either be calculated for a specific cluster or for the whole set of clusters where n_A is the number of all *silhouettes* in cluster A or the whole data set A respectively. Hence, the *silhouette coefficient* provides a measure of all *silhouettes* in a cluster or data set.

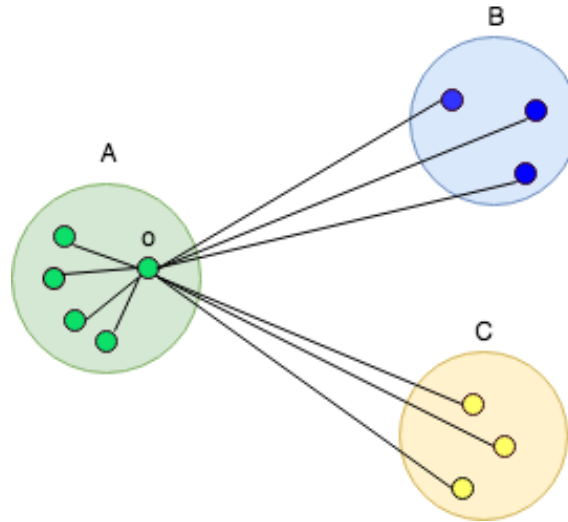


Figure 3.5: Illustration of elements involved in the computation of the silhouette $s(o)$. A, B and C are clusters, o is the observation that is considered.

3.6.2 Adjusted Rand Index

The *Rand index* [70] measures how similar two clusterings are to each other. The adjusted Rand index is extending the classical *Rand index* by correcting for chance in the sense that the index would take on some constant value (e.g. zero) under an appropriate null model of how the partitions have been chosen [47].

Assuming a set of partitions of n elements $E = \{e_1, \dots, e_n\}$ and two partitions of E , $X = \{x_1, \dots, x_r\}$ and $Y = \{y_1, \dots, y_s\}$ that group the elements of E in r and s clusters respectively. The *Rand index* R is then given by:

$$R = \frac{a + b}{\binom{n}{2}} \quad (3.38)$$

where

- a: denotes the number of pairs of elements in E that are in the same subset in X and in the same subset in Y.
- b: defines the number of pairs of elements in E that are in different subsets in X and in different subsets in Y.

The *Rand index* has a value in the range of 0 to 1. A value of 0 indicating that the two data clusterings do not agree on any pair of points and a value of 1 corresponding to identical groupings [70][47].

In contrary the *adjusted Rand index* can take values from -1 to +1. Where positive values close to one indicates that clusters are quite similar (a value of one implies that the clusterings are identical). A value close to zero indicates the two clustering assignments match in a similar fashion as to be expected if they were both randomly generated. Negative values indicate that there is less agreement between clusterings than expected from random partitions [70][47].

Method [70][47]:

Assuming a set of M of n elements and two clusterings $X = \{X_1, \dots, X_m\}$ and $Y = \{Y_1, \dots, Y_s\}$. The contingency table denoting the overlap between Y and X is then given by:

X/Y	Y_1	\dots	Y_s	\sum
X_1	n_{11}	\dots	n_{1s}	x_1
\vdots	\vdots	\ddots	\dots	\vdots
X_m	n_{m1}	\dots	n_{ms}	x_m
\sum	y_1	\dots	y_s	

Table 3.1: contingency table for clusterings X and Y

where each entry n_{ij} yields the number of objects for which $n_{ij} = |X_i \cap Y_j|$ holds.

The *adjusted Rand index* can then be defined as:

$$\begin{aligned}
 ARI &= \frac{Index - Expected Index}{Maximum Index - Expected Index} \\
 &= \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\left(\sum_i \binom{x_i}{2} \sum_j \binom{y_j}{2} \right) / \binom{n}{2} \right]}{\frac{1}{2} \left[\sum_i \binom{x_i}{2} + \sum_j \binom{y_j}{2} \right] - \left[\left(\sum_i \binom{x_i}{2} \sum_j \binom{y_j}{2} \right) / \binom{n}{2} \right]} \quad (3.39)
 \end{aligned}$$

where n_{ij} , x_i and y_j are collected from the contingency table 3.1.

3.7 Visualization

Visualizing high dimensional features in typically 2-dimensional space can give an intuition about the data. For instance points that strongly separate themselves from the other observation might be outliers that potentially disturb the data. In this work visualization was mainly used to illustrate subjects according to their brain voxels in 2-dimensional space.

Hence in this section *multidimensional scaling* is introduced which served as visualization technique.

3.7.1 Multidimensional scaling

Multidimensional scaling (MDS) is a technique for visualizing similarities of individual objects in a data set. An *MDS* algorithm's target is to reduce each object down to an N -dimensional space in such a way that the distances between objects are preserved as best as possible. Hereby, each object is represented in each of the N -dimensions.

Method [46]:

Given a distance matrix Δ of K objects:

$$\Delta = \begin{bmatrix} \delta_{1,1} & \cdots & \delta_{1,K} \\ \vdots & \ddots & \vdots \\ \delta_{K,1} & \cdots & \delta_{K,K} \end{bmatrix} \quad (3.40)$$

where $\delta_{i,j}$ denotes the distance between the i^{th} and the j^{th} object.

MDS aims to find D vectors $x_1, \dots, x_D \in \mathbb{R}^N$ such that the distances $\delta_{i,j}$ are preserved. Hence, for all $i, j \in 1, \dots, D$, $\|x_i - x_j\| \approx \delta_{i,j}$ holds where $\|\cdot\|$ denotes a vector norm (e.g. classical *MDS* uses the Euclidean distance).

In order to find the embedding of the D objects into the \mathbb{R}^N space *multidimensional scaling* applies different approaches to determine the vectors x_i . Typically *MDS* is formulated as an optimization problem, the loss function in this context is called *stress*, which is usually minimized by performing *stress majorization*.

Given the stress function S :

$$S(X) = \sum_{i < j \leq N} w_{i,j} (\delta_{i,j}(X) - \sigma_{i,j})^2 \quad (3.41)$$

where $w_{i,j}$ yields a weighting factor for the measurements between two object pairs i and j , $\delta_{i,j}$ denoting the distance matrix between the i^{th} and j^{th} objects and $\sigma_{i,j}$ representing the ideal distances between two points (this is the separation in their original space).

To reduce a given set of N M -dimensional data items to a configuration $X \in \mathbb{R}^r$ of K points where $r \ll M$ equation. This minimization will lead to a configuration of X in which objects that were close together in their original space will correspond to points that are close together in the lower space.

There are various approaches to minimize $S(X)$. Here, the *Kruskal*-method [71] that recommends an iterative steepest decent approach was performed.

3.8 Summary

In the beginning of this chapter the Parkinson's disease was briefly explained (section 3.1) showing that there is no available treatment that can stop the progression of the disease. Moreover it was outlined that Parkinson's is a heterogeneous disease and why it can be beneficial to disentangle this heterogeneous nature.

Secondly the pipeline of this thesis was introduced (section 3.2). Showing each step from the raw image data to the obtained clusterings. Based on this step wise procedure all necessary methodologies were described. Specifically, dimensionality reduction techniques (3.3), the two clustering methods (3.5), *CHIMERA* and *K-means* and *ANOVA* as a method that allows to correct for possible confounds influence (3.4) were explained. Lastly different kinds of postprocessing methods (3.6) were outlined: the *silhouette* as a technique to get an intuition of how well an object is assigned to its cluster and the *adjusted Rand index* as a method to compare two clusterings. Furthermore *multidimensional scaling* was described as a tool for visualizing high dimensional feature spaces in a lower dimensional space.

Chapter 4

Results

This chapter focuses on presenting the results obtained by applying both *CHIMERA* and *K-means* on different granularities of preprocessed data.

At first the data set provided for this work is presented. Afterwards, the application of confound filtering and the two different dimensionality reduction techniques are outlined.

The main part of this chapter is based on the obtained clusterings. Therefore the parameters that were utilized when performing each clustering method are briefly presented. Next, the outcome of both methodologies is visualized in 2D-space followed up by a similarity check, the *adjusted Rand index*. Moreover the *silhouette* of each clustering will be illustrated.

Lastly the collected clusters are presented. Specifically, the groupings are reviewed for patterns in age, gender, scanner protocol and disease duration.

4.1 The Data

The clinical data set which was investigated was acquired at *Universitätsklinikum Aachen* (UKA) and *Universitätsklinikum Düsseldorf* (UKD).

Both studies were ethically approved from the ethics board of Heinrich Heine and RWTH-Aachen University. Furthermore the usage of this data for scientific purposes was granted by both institutions.

Originally the combined data set contained 192 subjects, after removing outliers it reduced down to 180 samples. For instance subjects with an age below 35 were excluded.

The data comprises 91 Parkinson's patients and 89 healthy controls. Overall there were 98 males and 82 females. 102 samples were provided by UKD and 78 from UKA. Figure 4.1 illustrates the grouping of control and patients as well as the sex distribution regarding the two scanning sites.

In figure 4.2 boxplots for age and disease duration (that is the time patients suffer from the Parkinson's disease) are displayed. Regarding age the partitions reflect healthy controls and patients, the disease duration was partitioned based on the scanning site.

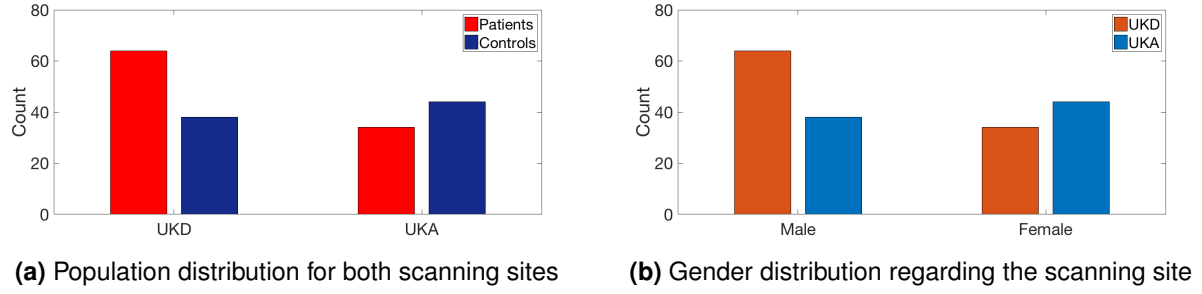


Figure 4.1: Scanning site comparison of subjects

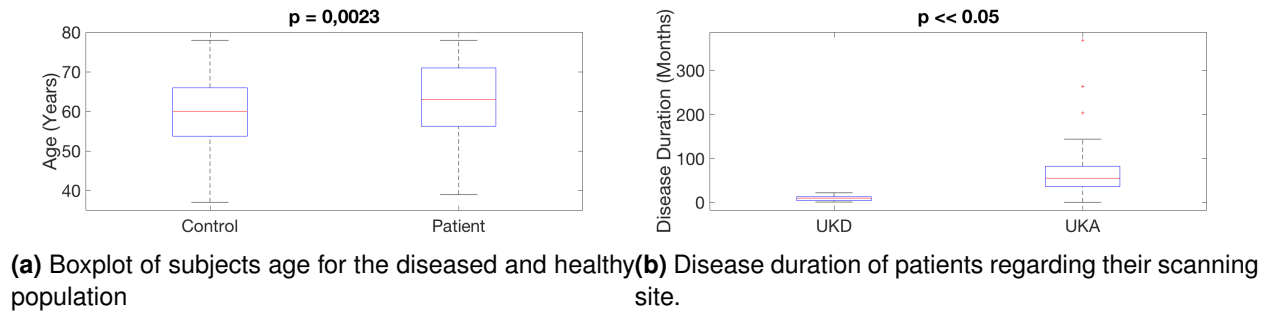


Figure 4.2: Boxplots: Age and disease duration. The p-value corresponds to a two-sample t-test of the respective distributions.

As one can observe from figure 4.1 patients were significantly older ($p < 0.05$) than the control population which is not optimal as distributions that are equally distributed are preferred.

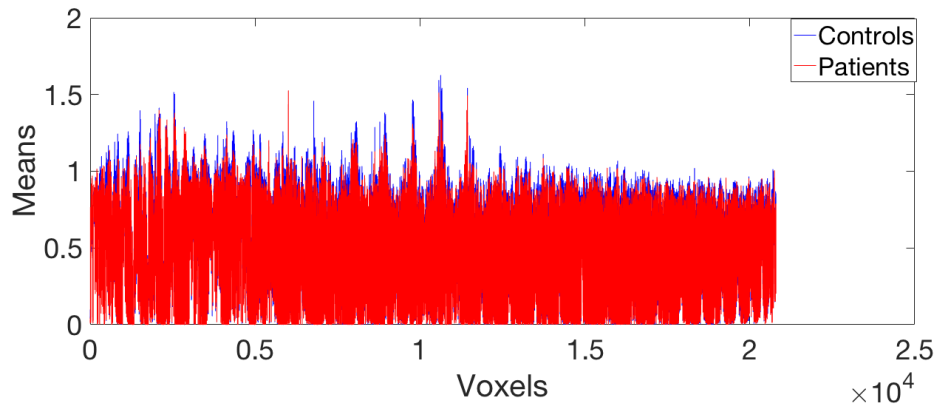
Even more diverse ($p \ll 0.5$) was the distribution of patients suffering from the disease illustrated in figure 4.2b: Parkinson's patients scanned at the *RWTH Aachen Uniklinik* were suffering from the Parkinson's disease significantly longer than patients from the *Düsseldorf Uniklinik* which had more or less been recently diagnosed.

In this work T_1 -weighted MRI-scans were used to try to identify the gray matter distribution for each individual subject. Based on these, groupings that reflected different subtypes of the disease (also see preprocessing in section 3.2) were identified. Specifically, the features that were the basis for the clustering methods were voxel wise measured gray matter values. In other words the goal was to find patterns based on volumetric differences or changes in the human brain.

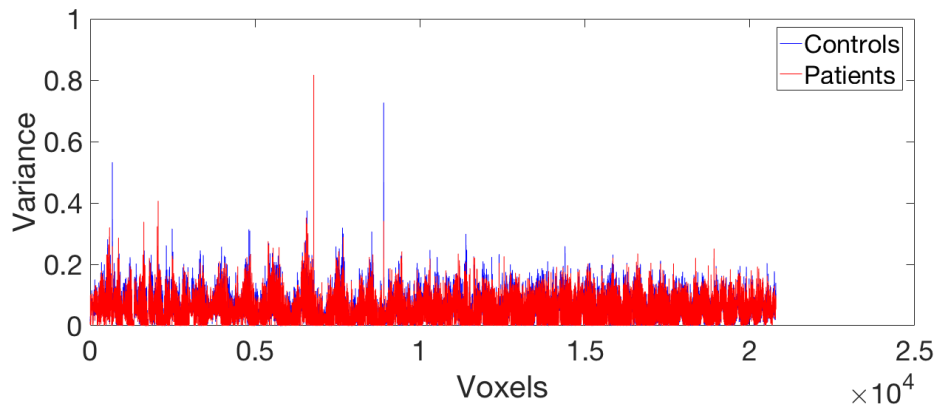
Research shows that gray matter can be a subtype disease indicator. For instance *Rosenberg-Katz et. al.* [44] showcase that in the tremor dominant (TD) subtype and the postural instability gait difficulty (PIGD) subtype, the characteristics in gray matter density of the human brain provide an indication for different manifestations (subtypes) of the Parkinson's disease.

After masking the T_1 -weighted gray matter MRI-scans (preprocessing in section 3.2) the feature matrix $M_f^{s,v}$ (3.1) contained 20792 voxels (v) extracted from 180 patients (s .)

Figure 4.3 gives an intuition of the mean and variance per voxel:



(a) Gray matter mean of every subjects evaluated at each voxel



(b) Gray matter variance value of every subject evaluated at each voxel

Figure 4.3: Gray matter mean and variance of every subject per voxel

As visualized above the variance and the mean between the normal control and patient population seemed to be quite similar. Except two volumetric elements that had conspicuously more variance (but stable means) there were no outliers detected. In order to preserve more information those two voxels were kept.

Figure 4.4 illustrates the samples (according to their voxel data) in a 2D-feature space which was obtained by applying *Multidimensional scaling* (3.7.1) to the original feature-space (20792-dimensions):

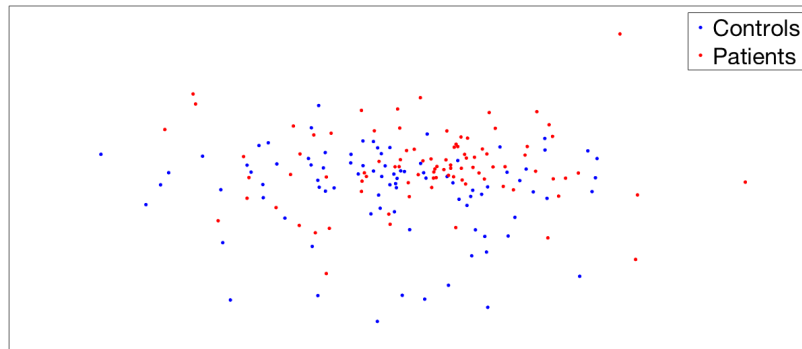


Figure 4.4: 180 samples in the 2D feature space. Each sample in the original space was represented by 20792 (number of voxels) grey matter values. Here, every sample is characterized by two values that were obtained by *multidimensional scaling*. Points close together in the original space are also close in the 2D space.

4.2 Preprocessing

This chapter outlines the different schemes that were used to filter for confound influence. Furthermore a closer look into the dimensional reduction methods is taken. Here the focus lied on how to choose a comparable number of representations of the data for both reduction methods while still preserving sufficient information to explaining the uncompressed data.

4.2.1 Analysis of Variance

Accounting for confounding (section 3.4) is crucial when searching for patterns which might be disguised by noise due to covariates.

K-means is not able to account for confounds, while *CHIMERA* is correcting for possible confounds in a way that distances of covariate features are larger than the imaging features.

For a fairer comparison the data accessed by *K-means* was manually filtered for confounds, specifically for age, sex and scanner location (*CHIMERA* accounts for the same confounders). Therefore an *analysis of variance* (ANOVA) was used in three different ways:

1. ANOVA based on the total population, that is patients and controls combined.

As in this work the interest lied on clustering the disease into different subtypes, *K-means* had only knowledge about the patient population. While regressing on both, diseased and healthy population, the idea was to create more similarity to the *CHIMERA* approach by preserving some of the healthy controls' knowledge.

Specifically, ANOVA was performed on the complete feature matrix (3.1) leading to a confound corrected representation of the samples. Afterwards *K-means* was applied to the resulting matrix where all healthy controls were excluded.

2. *ANOVA* based on the patient population only.

This is applying first the *ANOVA* and then *K-means* to the feature matrix (3.1) where all healthy controls were excluded beforehand.

3. No confound correction at all.

That is applying *K-means* directly to feature matrix (3.1) that had all healthy subjects removed. As already presented in section 3.4, not correcting for possible confounds might lead to falsified results. Here, the interest was to compare all three adjusting methods regarding their similarity to the *CHIMERA* approach.

4.2.2 Principal Component Analysis

Principle Component Analysis aims to persevere as much variance of the data as possible. Each eigenvalue of the normalized covariance matrix is representing the amount of variance retained from the original data. The percentage of variance that is retained in the i^{th} component is given by:

$$p_i = \frac{e_i}{\sum_i e_i} \quad (4.1)$$

where e_i denotes the i^{th} eigenvalue.

There are dozens of methods of determining how many *principle components* (granularities of the data) to select in order to represent the original data as best as possible. One would like to obtain the fewest number of *principle components* that persevere most of the variance of the original data.

The three most common selection methods (rules of thumb) are:

1. *Kaisers Rule*: keeping only components with eigenvalues for which $e_i \geq 1$ hold.
2. *Scree Plot*: this is plotting the number of eigenvalues in descending order (biggest to smallest). In some cases the obtained figure illustrates an *elbow* structure. The components that are selected are the ones prior to the *elbow* (see 6.1 for a screen plot).
3. *Proportion of variance*: selecting the components in such a way that at least two-third of the variance is explained.

In this work the selected components were chosen based on the proportion of variance retained. Figure 4.5 displays the variance retained of the feature matrix (3.1) for each component (in descending order):

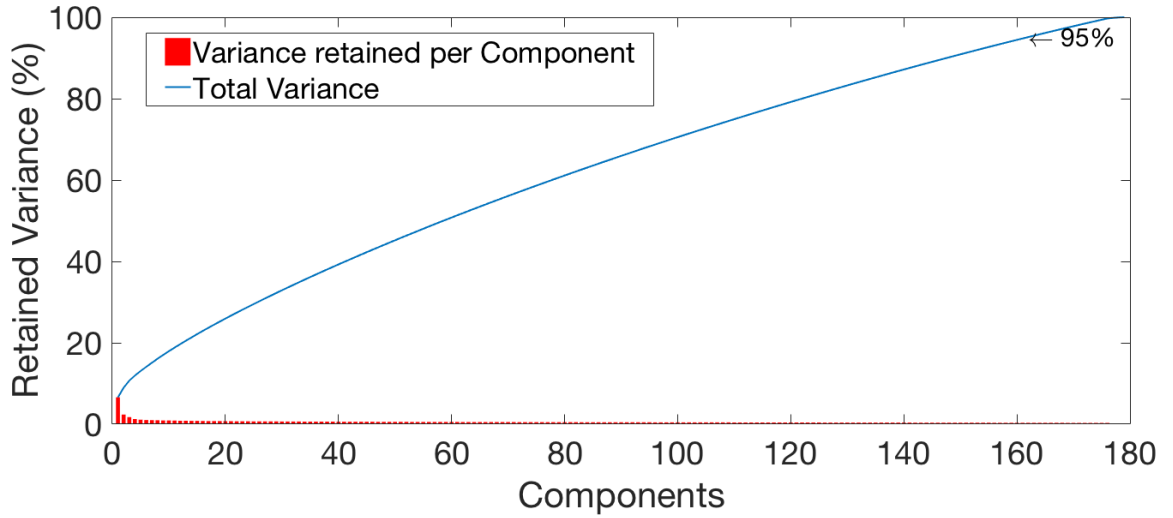


Figure 4.5: Variance of the feature matrix retained for each principal component

In this work multiple components had been chosen to compare their effects on the clustering. To be more precise, 162 components (95% of variance retained) were selected as a benchmark of orientation. In the end five different sets of components were used, two above and two below the threshold of 162, specifically: 140, 150, 162, 170, 179-components.

4.2.3 Non-negative Matrix Factorization

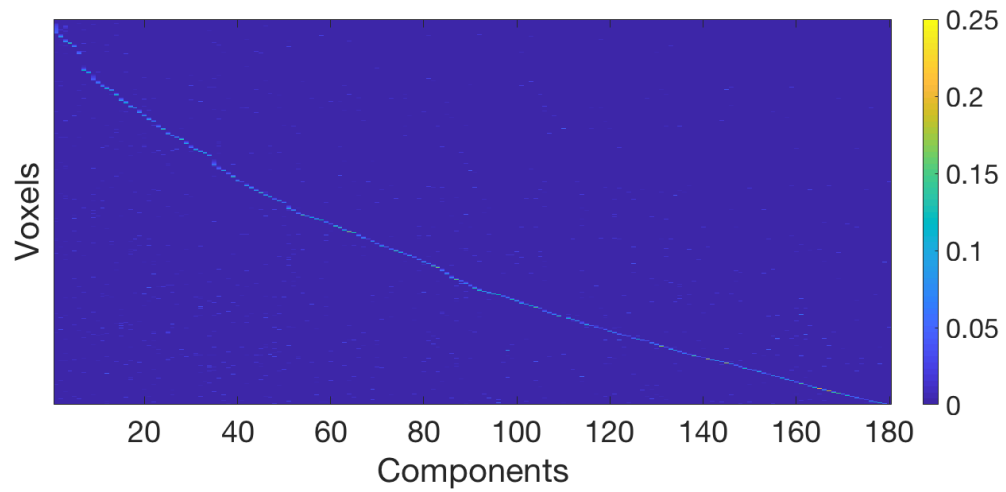
Non-negative Matrix Factorization requires an a priori definition of the maximum number of components (that is the number of columns in the W - and the number of rows in the H matrix respectively) one likes to obtain. Hence, for *non-negative Matrix Factorization* (NMF) it is not possible to extract the retained variance for each component as easily as in *principal component analysis* (PCA) (for instance a plot like for figure 4.5 cannot be obtained as the components are chosen a priori to construct the two factorization matrices).

Thus, in order to be comparable to PCA, a range of granularities was chosen so that they could be compared to multiple principal components.

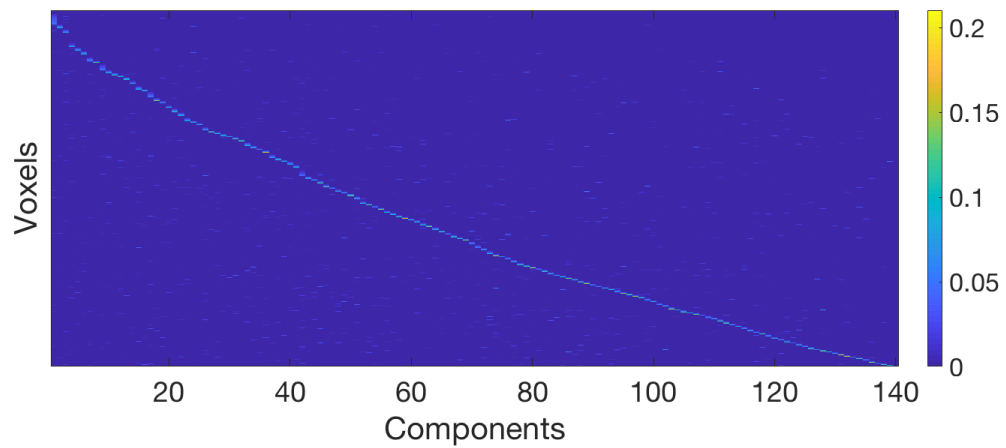
The selected number of components for NMF were 140, 150, 160, 170 and 180. The reason for choosing specifically these components was based on the assumption that NMF would need at least the same number of components as PCA to preserve the same amount of variance, as NMF is limited by its constraints to only positive values, only additive, not subtractive, combinations are allowed.

Specifically, for each component the transposed feature matrix M_f^T was partitioned into two lower rank matrices W and H . The dimensional reduced representation of the data was then obtained by multiplying the feature matrix times the obtained W matrix for each individual component.

Figure 4.6 illustrates the contribution of each voxel to the k^{th} component for the 140 and 180 component representation:



(a) contribution of each voxel to the 180-component NMF representation



(b) contribution of each voxel to the 140-component NMF representation

Figure 4.6: Contribution of each voxel to the components in different NMF granularities

4.3 Clustering

In this chapter the clusterings obtained by *CHIMERA* and *K-means* are visualized and compared.

First the parameters used to perform both methods are outlined, followed up by a 2D-visualization as well as similarity and consistency tests. Specifically, the *adjusted Rand index* between *CHIMERA* groupings and the three different confound adjusted *K-means* clusters is visualized. The consistency is addressed by the *silhouette* for different clusterings.

4.3.1 Parameters

CHIMERA:

CHIMERA comes with several hyper parameters which were originally [17] determined based on cross-validation [72]. Here, due to the limit of the scope, most specifications were chosen by default.

CHIMERA was performed with the following parameters:

- the number of clusters was set to 2. This is the minimal amount of clusters that can be obtained. Although there might be several more subtypes one can find that the minimal number of cluster was a reasonable point to start from.
- covariates (confounding factors) that were introduced to *CHIMERA* were: age, sex and scanner location.
- as stopping criterion a residual value of 0.001 was set (default).
- for λ_1 a value of 10 and for λ_2 a value of 100 was chosen (default).
- as transformation matrix a full matrix was set (affine matrix, also default).
- the maximum number of iterations was set to 10000 (default 1000) to ensure a minimal solution.
- number of runs with different initialization (replicates) was set to 50 (default).
- the data was normalized to be in a range of zero to one (default).
- the result that was saved during multiple re-initializations was the one which yielded the minimal energy (default).

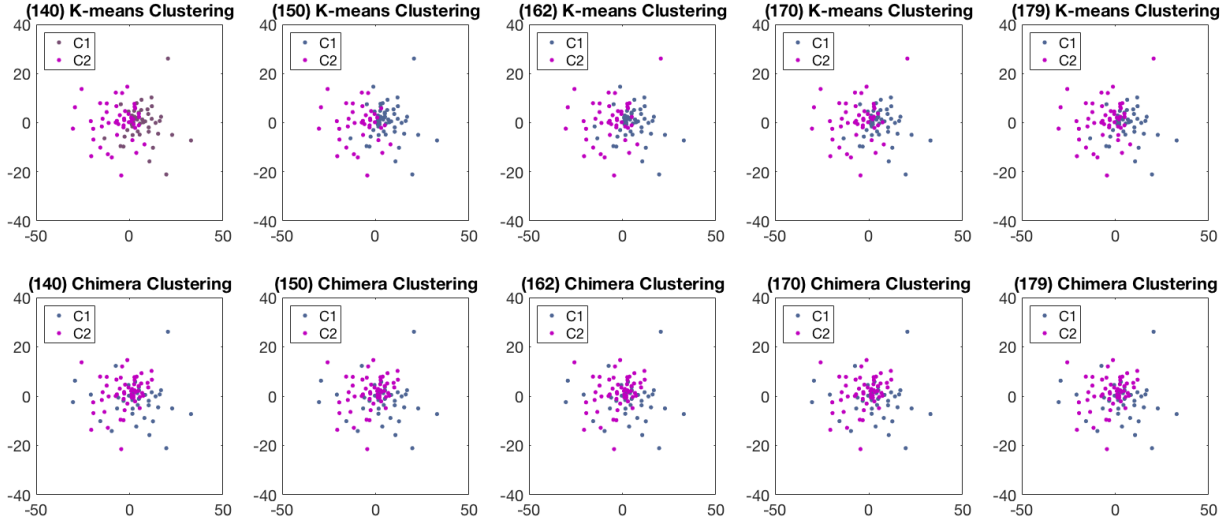
K-means:

K-means was executed with 150 replicates (this is the number of times to repeat the clustering, each with a new set of initial centroids), as it depends on its initialization and hence might only find a local minimum. Furthermore the number of maximum iterations was set to 10000 to ensure a minimal solution. The parameter k was set to 2 in order to be comparable to *CHIMERA*.

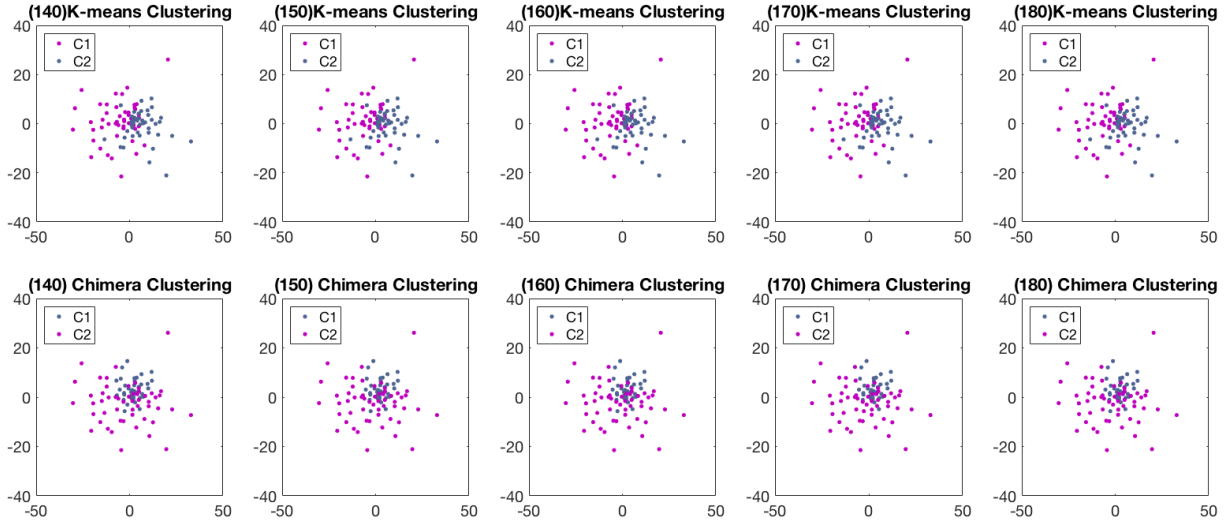
4.3.2 Visualization

Figure 4.7 and 4.8 are illustrating the clusterings of multiple representations of the original data. The groupings for *K-means* were based on data that was corrected for possible confounds with the two different schemes (3.4).

Note that for better comparison the feature spaces visualized here (for both 4.7 and 4.8) are corresponding to raw 20792 space which was reduced to 2D using *multidimensional scaling*. The actual feature space depends on both methodology and confound adjustment.

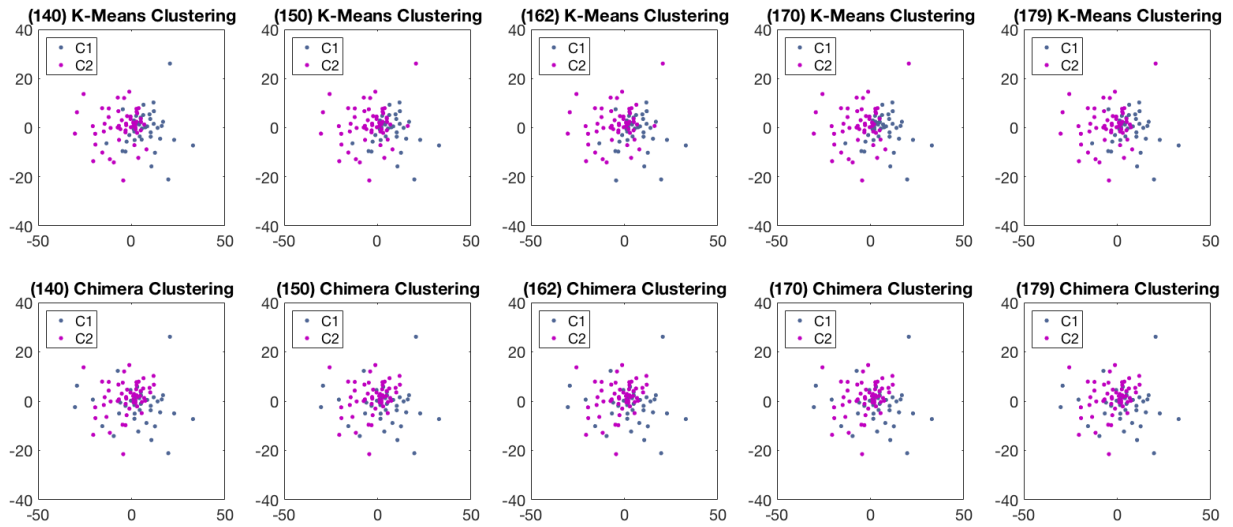


(a) Clusterings for different PCA components. The top row refers to *K-means* clusterings, the bottom to *CHIMERA* groupings.

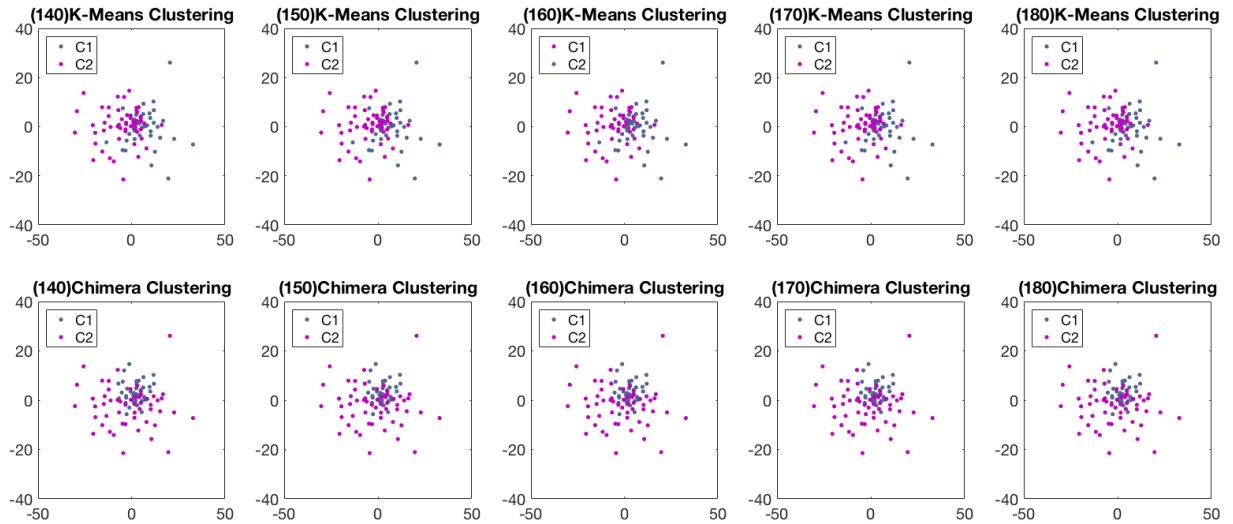


(b) Clusterings for different NMF components. The top row refers to *K-means* clusterings, the bottom to *CHIMERA* groupings.

Figure 4.7: 2D visualization of *CHIMERA* and *K-means* clusterings based on different data representations. For *K-means* possible confounds were corrected according to the patient population. The number of components for each reduction method is denoted in (·).



(a) Clusterings for different PCA components. The top row refers to *K-means* clusterings, the bottom to *CHIMERA* groupings.



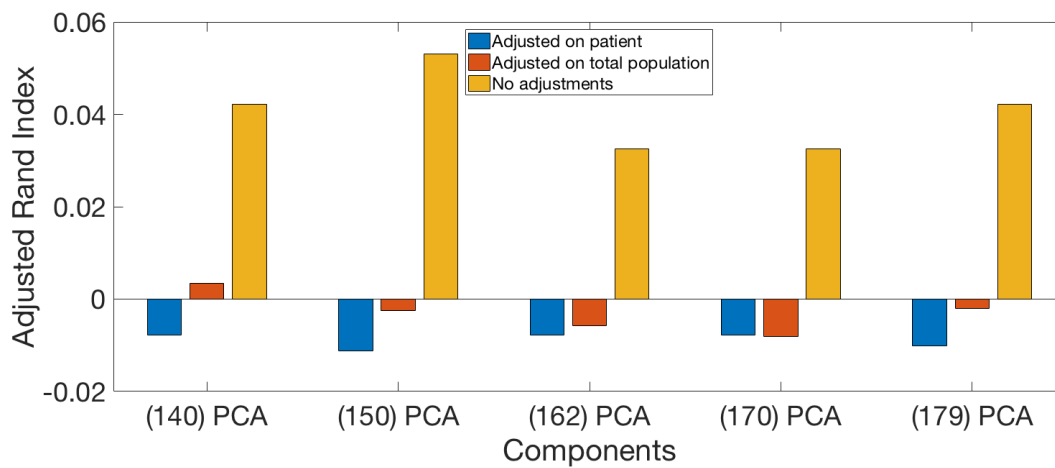
(b) Clusterings for different NMF components. The top row refers to *K-means* clusterings, the bottom to *CHIMERA* groupings.

Figure 4.8: 2D visualization of *CHIMERA* and *K-means* clusterings based different data representations. For *K-means* possible confounds were corrected based on the total population. The number of components for each reduction method is denoted in (·).

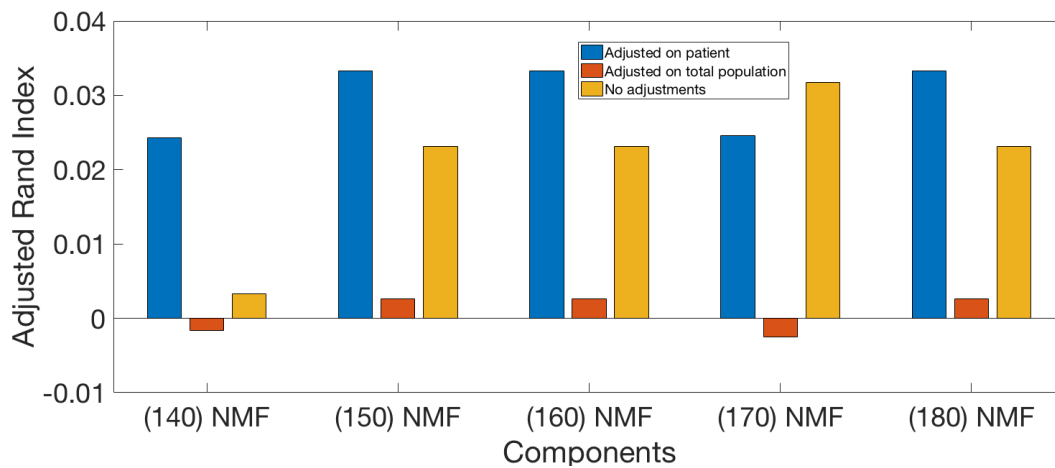
Clusterings obtained where no a priori confound removal was performed are visualized in appendix 6.2.

4.3.3 Adjusted Rand Index

Illustration 6.3 displays the *adjusted Rand index* comparing *CHIMERA* clustering to groupings of *K-means* that are adjusted for possible confounds in different schemes. Both clusterings were applied to a lower representation of the data using NMF and PCA respectively as reduction method.



(a) ARI for multiple *principal components*, comparing *K-means* with *CHIMERA*



(b) ARI for multiple NMF components, comparing *K-means* with *CHIMERA*

Figure 4.9: Adjusted Rand index between *CHIMERA* and different schemes of *K-means*. The number of components is denoted in (·).

In appendix 6.3 the classical *Rand index* is illustrated.

Table 4.1 illustrates the consistency (similarity) of *CHIMERA* clusters within a reduction method and throughout the respective approach.

Each cell represents an *adjusted Rand index* value. Blue and red colored cells denote cluster similarity within the given reduction method, magenta cells yielding the similarity comparing both reduction techniques against each other.

	(140) NMF	(150) NMF	(160) NMF	(170) NMF	(180) NMF	(140) PCA	(150) PCA	(162)PCA	(170) PCA	(179) PCA
(140) NMF	1									
(150) NMF	0.871	1								
(160) NMF	0.871	1	1							
(170) NMF	0.829	0.955	0.955	1						
(180) NMF	0.871	1	1	0.955	1					
(140) PCA	0.413	0.331	0.331	0.305	0.331	1				
(150) PCA	0.413	0.331	0.331	0.305	0.331	1	1			
(162) PCA	0.413	0.331	0.331	0.305	0.331	1	1	1		
(170) PCA	0.413	0.331	0.331	0.305	0.331	1	1	1	1	
(179) PCA	0.413	0.331	0.331	0.305	0.331	1	1	1	1	1

Table 4.1: Cluster consistency in *CHIMERA* clusters. The number of components for each reduction method is denoted in (·).

Table 4.2 and 4.3 are illustrating the consistency of *K-means* clusterings with different confound correcting schemes for both NMF and PCA reduced data:

	(140) NMF	(150) NMF	(160) NMF	(170) NMF	(180) NMF	(140) PCA	(150) PCA	(162)PCA	(170) PCA	(179) PCA
(140) NMF	1									
(150) NMF	1	1								
(160) NMF	1	1	1							
(170) NMF	1	1	1	1						
(180) NMF	1	1	1	1	1					
(140) PCA	0.790	0.790	0.790	0.790	0.790	1				
(150) PCA	0.639	0.639	0.639	0.639	0.639	0.675	1			
(162) PCA	0.750	0.750	0.750	0.750	0.750	0.570	0.639	1		
(170) PCA	0.712	0.712	0.712	0.712	0.712	0.750	0.675	0.570	1	
(179) PCA	0.790	0.790	0.790	0.790	0.790	0.675	0.537	0.712	0.750	1

Table 4.2: Cluster consistency in *K-means* clusterings, possible confounds were corrected based on patients. The number of components for each reduction method is denoted in (·).

	(140) NMF	(150) NMF	(160) NMF	(170) NMF	(180) NMF	(140) PCA	(150) PCA	(162)PCA	(170) PCA	(179) PCA
(140) NMF	1									
(150) NMF	1	1								
(160) NMF	1	1	1							
(170) NMF	1	1	1	1						
(180) NMF	1	1	1	1	1					
(140) PCA	0.830	0.830	0.830	0.830	0.830	1				
(150) PCA	0.870	0.870	0.870	0.870	0.870	0.717	1			
(162) PCA	0.829	0.829	0.829	0.829	0.829	0.750	0.789	1		
(170) PCA	0.639	0.693	0.639	0.639	0.639	0.639	0.750	0.639	1	
(179) PCA	0.789	0.789	0.789	0.789	0.789	0.789	0.829	0.789	0.750	1

Table 4.3: Cluster consistency in *K-means* clusterings, possible confounds were corrected based on patients and healthy controls. The number of components for each reduction method is denoted in (·).

4.3.4 Silhouette

Figure 4.10 displays the silhouette for two different components (here the widest difference in components are chosen for each method) of NMF for both the *CHIMERA* clusters and the *K-means* groupings. Correction for possible confounds was based on the patient population. Figure 4.11 illustrates the silhouettes for *PCA* respectively. Silhouettes for data adjusted for possible confounds based on the total population are provided in appendix 6.4.

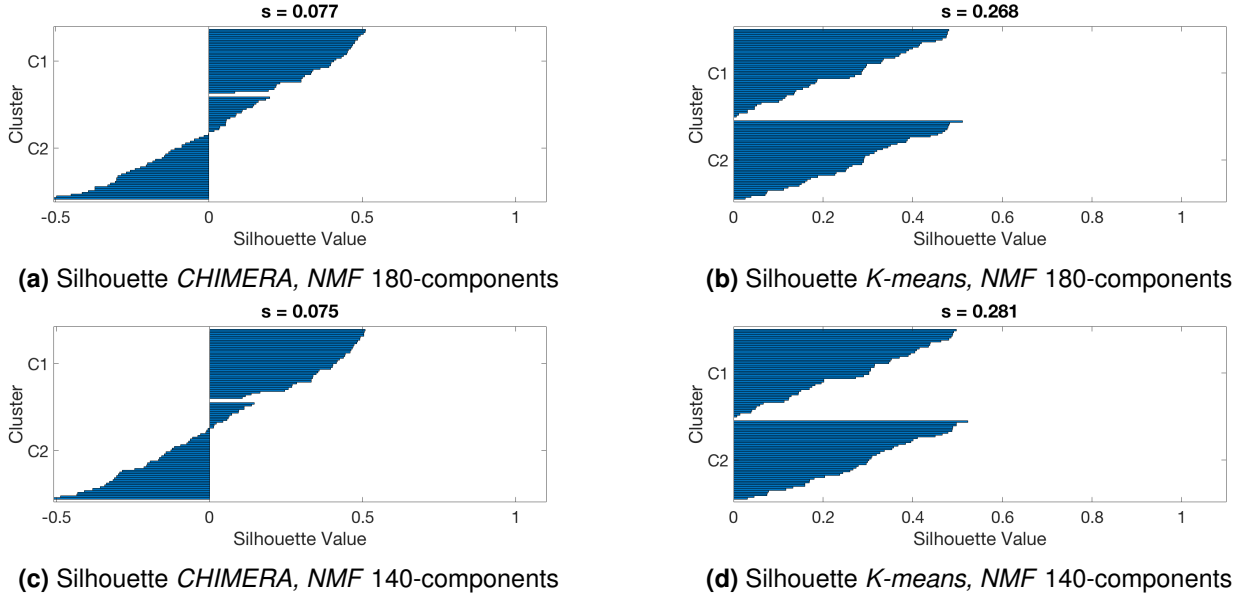


Figure 4.10: Silhouette for *CHIMERA* and *K-means* for two different NMF granularities. For *K-means* possible confounds were corrected based on the patients distribution, s denotes the *silhouette coefficient*.

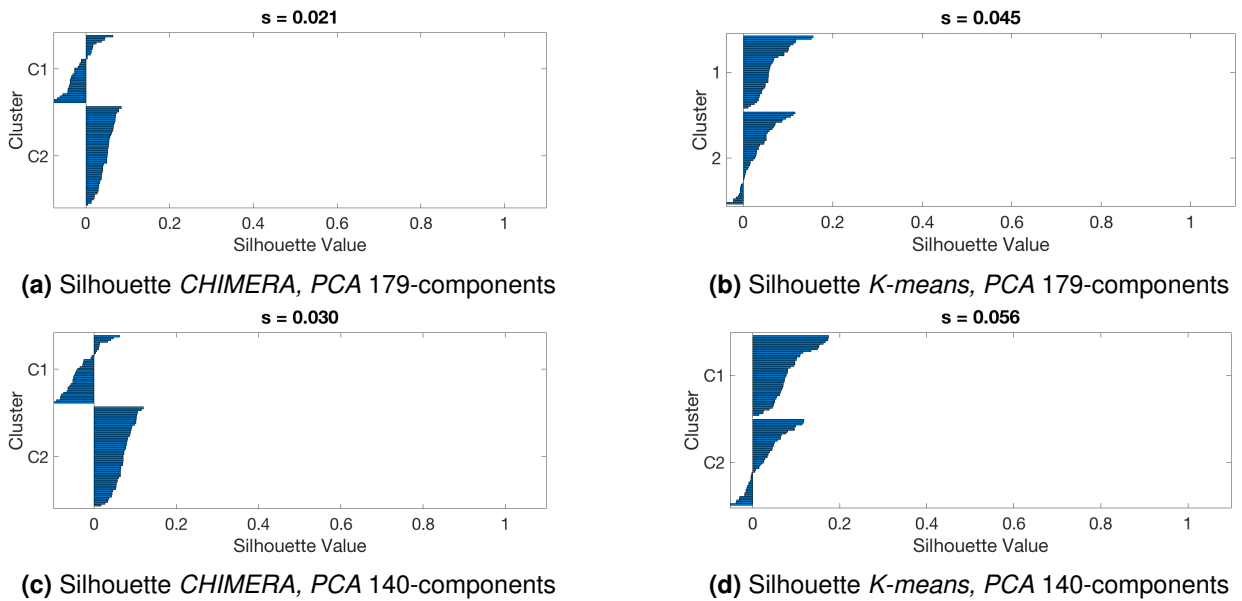


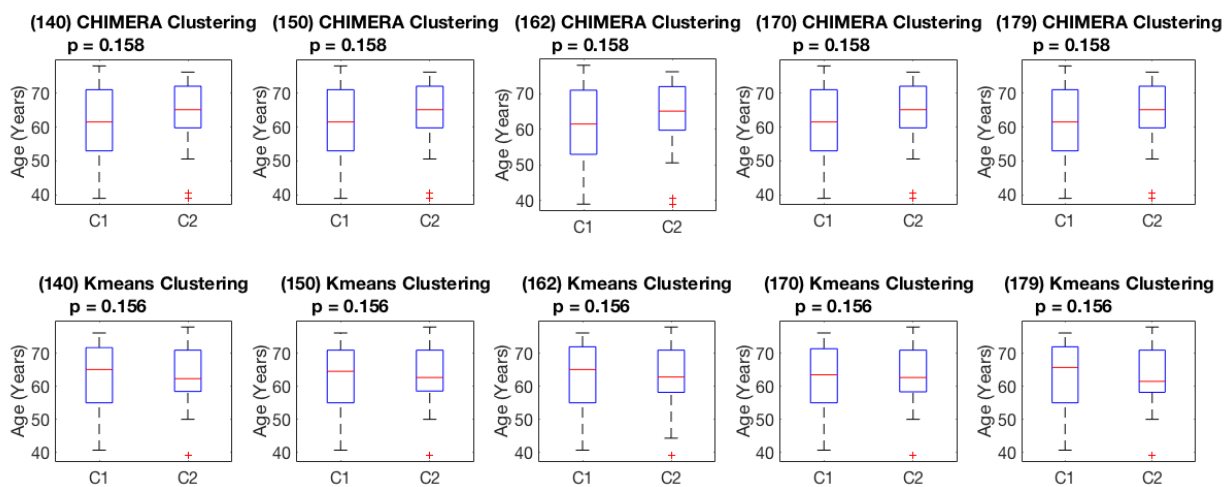
Figure 4.11: Silhouette for *CHIMERA* and *K-means* for two different NMF granularities. For *K-means* possible confounds were corrected based on the patients and healthy distribution, s denotes the *silhouette coefficient*.

4.4 Cluster Analysis

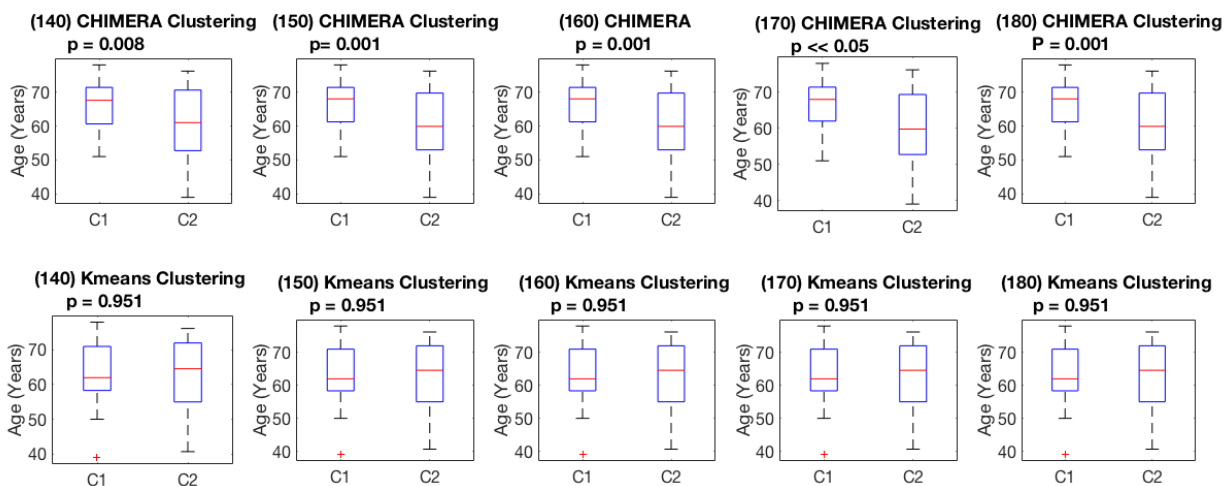
This section will focus on the samples within a cluster. Specifically, sex, age, scanner protocol and disease duration are investigated.

Note that due to the scope of this work in the following chapter only *K-means* clusterings that were corrected for possible confounds based on the patient samples are illustrated. See appendix 6 for illustrations that were corrected based on patients and healthy controls. Sections are named respectively.

4.4.1 Age



(a) Boxplot regarding age for different PCA components



(b) Boxplot regarding age for different NMF components

Figure 4.12: Boxplots of the age distribution in clusterings (C1,C2) obtained from both, *CHIMERA* and *K-means* for multiple NMF and PCA components. P-value was obtained by a two sample t-test. Correction for possible confounds was based on the patient samples. The number of components is denoted in (-).

4.4.2 Disease Duration

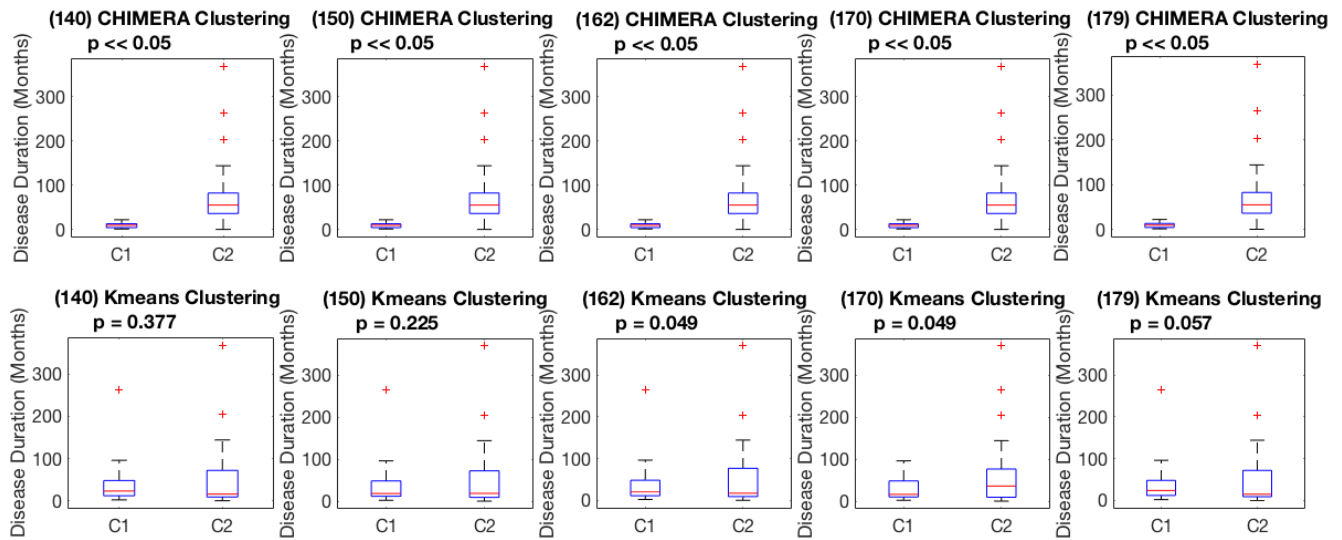
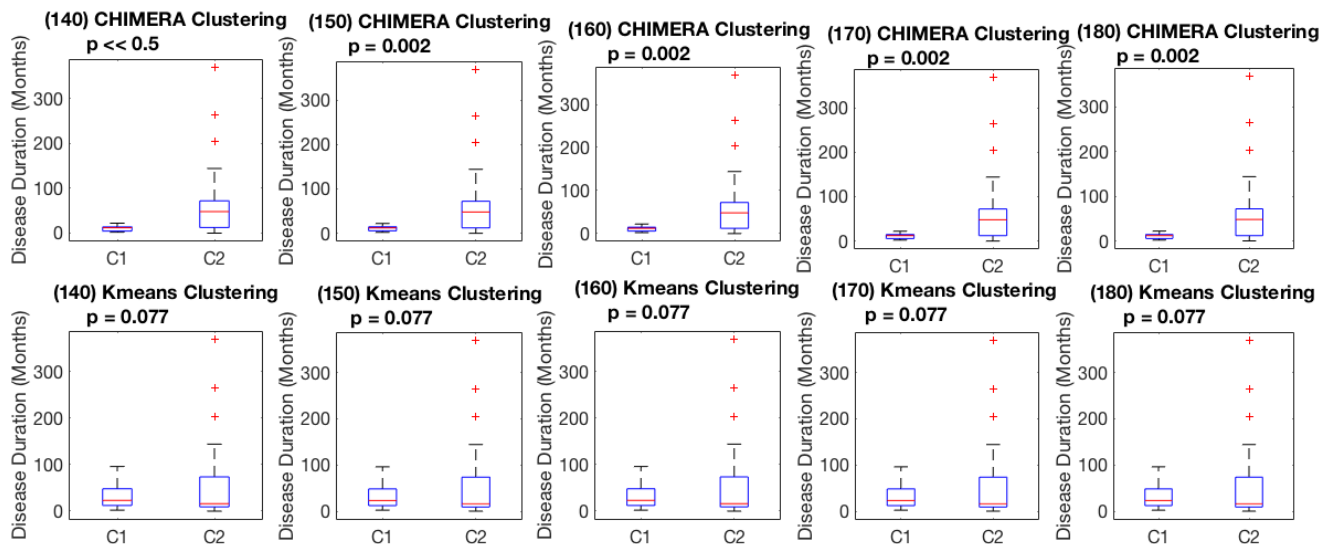
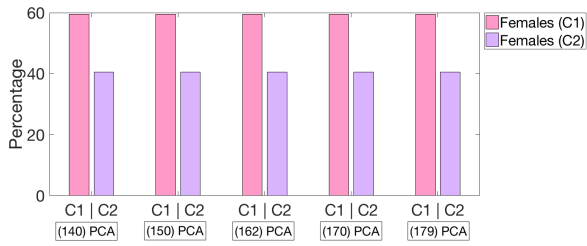
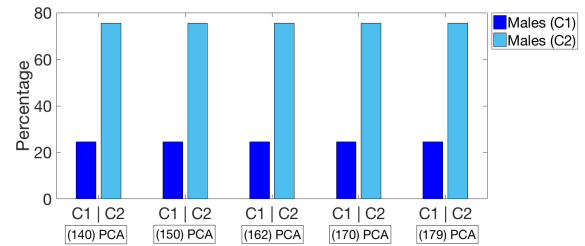
(a) Boxplot regarding patients disease duration for different *PCA* components(b) Boxplot regarding patients disease duration for different *NMF* components

Figure 4.13: Illustration of the disease duration of patients in clusterings (C1,C2) obtained from both *CHIMERA* and *K-means* for multiple NMF and PCA components. The correction for possible confound was based on the patient population. P-values were calculated by performing a two sampled t-test for the specific clustering. The number of components is denoted in (.).

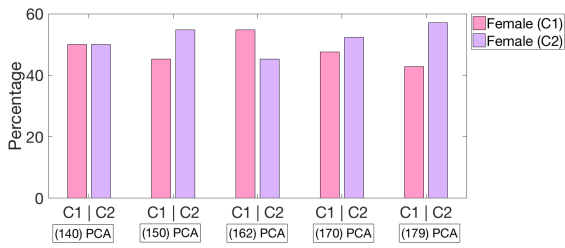
4.4.3 Gender



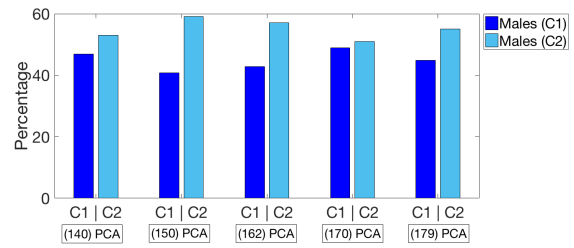
(a) Female distribution for *CHIMERA* clusters based on PCA-reduced data



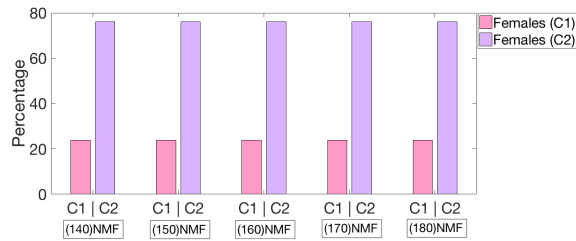
(b) Male distribution for *CHIMERA* clusters based on PCA-reduced data



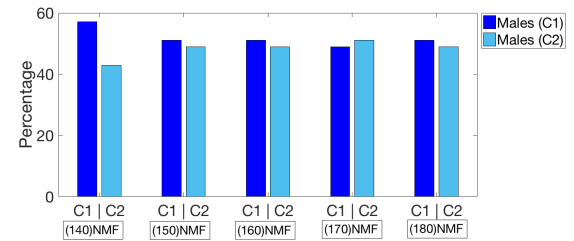
(c) Female distribution for *K-means* clusters based on PCA-reduced data



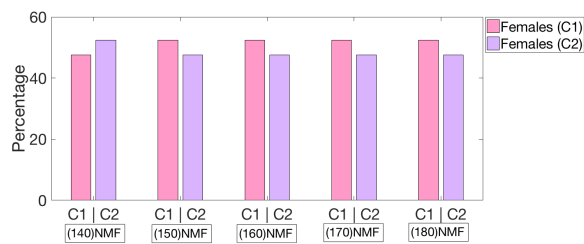
(d) Male distribution for *K-means* clusters based on PCA-reduced data



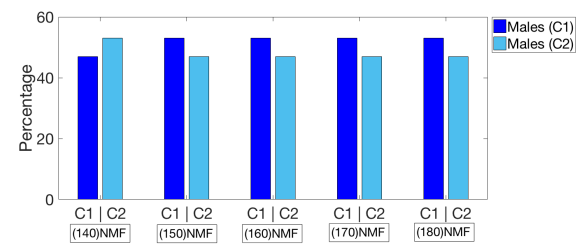
(e) Female distribution for *CHIMERA* clusters based on NMF-reduced data



(f) Male distribution for *CHIMERA* clusters based on NMF-reduced data



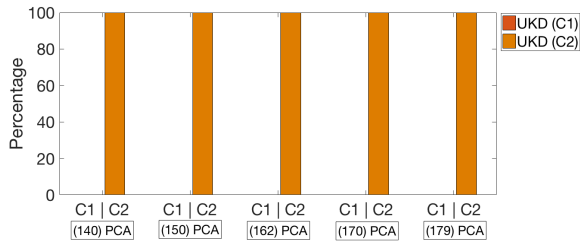
(g) Female distribution for *K-means* clusters based on NMF-reduced data



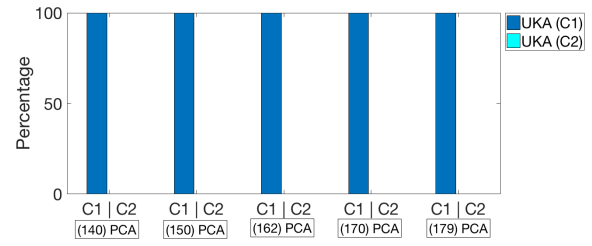
(h) Male distribution for *K-means* clusters based on NMF-reduced data

Figure 4.14: Illustration of the percentage of males and females in clusters C1 and C2 obtained from both *CHIMERA* and *K-means* based on multiple NMF and PCA components. Correction for possible confounds was based on patient samples. The number of components is denoted in (·).

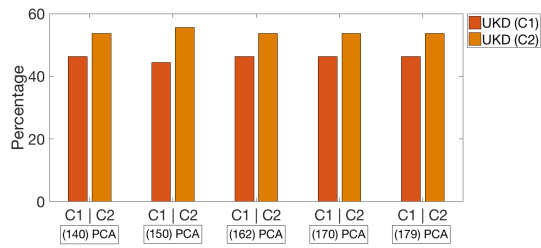
4.4.4 Scanner



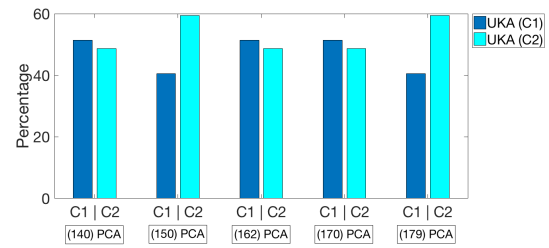
(a) UKD distribution for *CHIMERA* clusters based on PCA re-reduced data



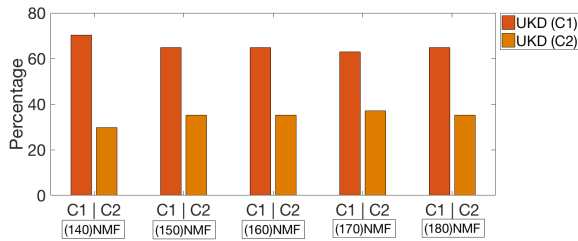
(b) UKA distribution for *CHIMERA* clusters based on PCA re-reduced data



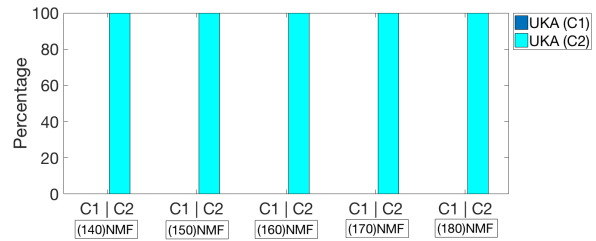
(c) UKD distribution for *K-means* clusters based on PCA re-reduced data



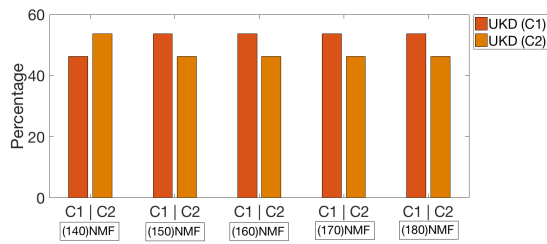
(d) UKA distribution for *K-means* clusters based on PCA re-reduced data



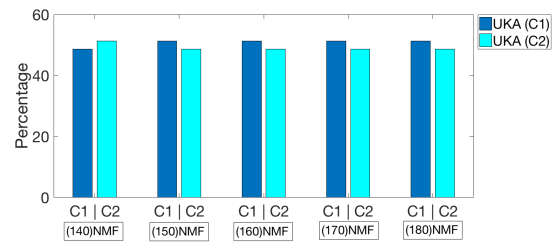
(e) UKD distribution for *CHIMERA* clusters based on NMF re-reduced data



(f) UKA distribution for *CHIMERA* clusters based on NMF re-reduced data



(g) UKD distribution for *K-means* clusters based on NMF re-reduced data



(h) UKA distribution for *K-means* clusters based on NMF re-reduced data

Figure 4.15: Illustration of the percentage of patients scanned at *Universitätsklinikum Aachen* (UKA) and *Universitätsklinikum Düsseldorf* (UKD) in clusters C1 and C2 obtained from both *CHIMERA* and *K-means* for multiple NMF and PCA components. Correction for possible confounds were based on patient samples. The number of components is denoted in (·).

Chapter 5

Discussion and Prospects

This section is evaluating the results gathered in chapter 4. Firstly the data set provided for this work is examined (5.1.1). Secondly the results obtained by both clustering methods are discussed. The latter is divided into the comparison of the clusters based on their similarity in the feature space (5.1.2) and the characteristic of samples that they comprise (5.1.3). Lastly the limitations of this work are reviewed and possible outlooks outlined (5.2).

5.1 Evaluation

This section discusses the outcome of chapter 4.

5.1.1 Data Set

Before directly considering the clustering results a closer look at the clinical data set is required.

First of all the data set was considerably small compared to other data driven approaches as for instance the image classification of cars where (image) data is ubiquitous. The fact that data is rare in medical settings is a significant limitation as more data usually provides more information and thus better results.

Secondly, equally distributed data is preferred. Here, the data set consisted of a larger number of male (98) than female (82) subjects as illustrated in figure 4.1b. Furthermore study participants were scanned at multiple sites where more subjects were studied in *Uniklinik Düsseldorf* (UKD) compared to *Uniklinikum Aachen* (UKA), figure 4.1a. As figure 4.2 illustrates there was a significant difference in age ($p < 0.05$) and disease duration ($p \ll 0.05$) regarding the healthy and the patient distribution as well as the scanning site. In other words patients were significantly older than the healthy controls and diseased who were scanned at UKD were more recently diagnosed with Parkinson's whereas patients studied at UKA were suffering significantly longer from the disease.

As outlined above, the data set was not optimal which is not unusual in clinical settings as data is rare.

Looking at figure 4.5 one can notice that relatively many components were needed to retain a high amount of variance. Here, the first 162 components retained 95% of the variance of the original (not reduced) data. The maximum number of components one can reduce to is bounded by the sample size (180 subjects). Usually, when applying PCA the first few components are preserving the most variance of the data and thus comprising the directions (trend or patterns) in the data. Hence, having more or less an equal distribution of variance throughout all components, emphasized that there was no significant trend in the data. In other words every human brain seemed to be unique.

The diversity of each individual subject was also reflected by components of NMF. Looking at figure 4.6 it becomes clear that there was only very small contribution of voxels to an NMF component (indicated by the dark blue).

5.1.2 Clustering Analysis

One of the objectives of this thesis was to find out whether *CHIMERA* and *K-means* clusters differ and if so how significant this discrepancy is.

Considering the two 2D-visualization of the two different cluster approaches for both confound filtering techniques (figure 4.7 and figure 4.8) one can easily observe that the clusterings obtained by *CHIMERA* differed from those found by *K-means*.

The significance of this difference is illustrated in barplots 6.3. For both reduction methodologies as well as for both confound filtering techniques the *adjusted Rand Index* did not exceed a value of 0.05. Considering that an *adjusted Rand index* value of 1 yields identical clusterings whereas a value of near to 0 indicates that the two clustering assignments match about as might be expected if they were both randomly generated, the obtained results proofed that there are significant differences between the two clustering methods. *K-means* cluster based on data that was not adjusted for possible covariate interference sometimes even leads to more similar results. Moreover the diversity between *CHIMERA* and *K-means* was consistent throughout reduction and confound correction methods.

Investigating the consistency between different components of NMF and PCA it becomes obvious that *CHIMERA* was sensitive to the reduction methods. In other words clusterings varied depending on which reduction method was performed. Table 4.1 comprises different *adjusted Rand index* values obtained comparing different clusterings. Notably, values among reduction methods were high which indicates high (or for PCA complete) similarity between the clusterings. That is the blue marked cells for NMF and the red marked cells for PCA reduction. Although there was consistency among methods, there was a difference when comparing clusters of both methodologies. This is indicated by the magenta cells that were comparing clusters obtained from different representations of the data (components) against the two reduction methods, leading to significant lower values and thus poor similarity.

Considering the same structured table for *K-means* clustering where either possible confounds were corrected based on the patient samples (table 4.2) or filtered by combining the patients with healthy controls (table 4.3), it can be seen that clusters were similar within reduction techniques (blue and red cells) and also comparing the methods (magenta cells) lead to

akin results for both confound correction schemes. In other words in comparison to *CHIMERA*, *K-means* was not sensitive to either of the reduction methods performed in this work.

By taking a closer look at the silhouettes of the clusterings the differences of the clusters in their corresponding feature spaces and their underlying principles are revealed. As can be seen in figure 4.10 for NMF and also in illustration 4.11 for PCA, clusters obtained with *K-means* yielded in most cases a high *silhouette* for observations whereas subjects in clusters produced by *CHIMERA* scored a negative *silhouette* for one grouping.

By measuring the distances between neighboring clusterings, the silhouette determines how consistent a cluster is. Hence, high values for *K-means* clustering were expected. The reason for this behaviour is that *K-means* is based on assigning objects to the nearest cluster center. Thus, the object wise distance within a cluster was small in comparison to large distances to observations in other clusters. In *CHIMERA* clusterings most of the objects in one cluster yielded to a negative silhouette. In other words, the observation was closer to the nearest neighboring cluster than to the cluster it was assigned to.

This again illustrated the fundamental difference that was outlined earlier in the methodology. *K-means* aimed to find clusterings based on high dimensional voxel distances whereas *CHIMERA* sought to find groupings via mapping and matching of distributions.

5.1.3 Cluster Analysis

While in the previous section 5.1.2 the clusterings were investigated regarding their similarity and structure in high dimensional space this section is outlining the differences within a cluster and possible patterns that might be reflected.

A qualitative assessment of the clusterings without any further knowledge and due to the limited scope of this thesis is very difficult. Thus, here the different cluster analysis steps that were taken in section 4.4 were consequently analyzed focusing on obvious patterns. Furthermore it was already shown that *K-means* and the *CHIMERA* groupings were fundamentally different for both confound correction schemes, hence in the following for simplicity only clusterings that were adjusted for possible confounding influence based on the patient population were considered.

Age:

Figure 4.12 illustrates the age distribution in the clusters for multiple granularities for PCA and NMF.

An observation for *CHIMERA* clustering based on NMF reduced data was that there were significant differences ($p < 0.05$) between the two resulting groupings (for all granularities) whereas for PCA reduced data none of the components yielded a remarkable age contrast in the clusters. This dissimilarity illustrated again that *CHIMERA* clusterings are sensitive to a priori reduction methods. Every cluster obtained by *K-means* did not have any significant difference in the age distribution in the clusters, emphasizing the dissimilarity of both clustering methods.

Although there was a difference in *CHIMERA* clusters regarding based on NMF reduced data for PCA preprocessed data this did not hold. Hence, summarizing the above, there seemed to be no constant obvious pattern in patients' age reflected up by either clustering methods.

Gender:

The percentage of males or females in each specific cluster is illustrated in figure 4.14 where the percentage was derived based on all male and female subjects in the data set.

Comparing the resulting clusters based on multiple *principal components* *CHIMERA* yielded more males (around 78%) in cluster C2 than in cluster C1 (22%) whereas females were more equally spread (max. difference of 20 % between clusters). In contrast, *K-means* structured clusters in a more balanced way (max. difference of 20 % between clusters) for both female and male patients.

A similar pattern can be seen for clustering based multiple representations (components) of the data obtained by NMF whereas in contrary to groupings obtained based on *principle components* cluster C2 contained more female patients (46% difference) than in C1. This again illustrated the sensitivity of *CHIMERA* to the applied reduction methods. *K-means* as also previously observed seemed to spread genders more or less equally over clusters.

Although there were differences in diseased subjects gender for clusters obtained by *CHIMERA* the pattern was not consistent for both reduction techniques. In contrary *K-means* lead to a balanced distribution of males and females in the clusterings for both dimensionality reduction techniques. Hence, a consistent obvious pattern for either clustering method was not obtained.

Disease Duration:

Figure 4.13 illustrates the duration of patients suffering from the Parkinson's disease.

It can be observed that there was a significant difference ($p < 0.05$) in terms of patients disease duration in clusterings for both PCA and NMF reduced data. In other words, cluster C1 reflected diseased subjects that were more or less recently diagnosed whereas cluster C2 contained patients in later stages of the disease. In contrast *K-means* clusterings did not reflect any significant difference ($p > 0.5$) or pattern in disease duration.

These results lied emphasis on *CHIMERA* reflecting the duration of illness of patients. As the disease duration was not introduced as a covariate to the algorithm, capturing this pattern just based on anatomical differences in human brains gray matter structure would be a significant indication that *CHIMERA* clusters reflected the pathological process of patients. Researchers already demonstrated that this disease duration has an influence on gray matter structure of Parkinson's patients. For instance accelerated cortical atrophy is found in PD subjects with a duration of illness of 1-5 years whereas striatal atrophy occurred in PD subjects that were suffering less than 1 year [73]. In contrast the *K-means* approach did not reflect a significant difference ($p > 0.05$) regarding patients' disease duration in clusters indicating the superiority of *CHIMERA* to identify the pathological process in diseased.

Scanning location:

Figure 4.15 illustrates the clusterings regarding where patients were scanned.

CHIMERA clusters which were derived based on different NMF and PCA preprocessed data reflected the scanning site. For clusters based on PCA reduced data one contained all patients from Aachen the other cluster comprised all diseased subjects from Düsseldorf. A similar pattern was found by clustering different NMF components, where cluster C2 contained 100% of UKA patients and 30% of UKD diseased subjects, cluster C1 contained the rest of the UKD patients. This pattern was reflected even though *CHIMERA* accounted for the scanner location via confound correction for site (4.3.1).

Considering that the site was introduced as a possible confound to *CHIMERA* and that there was a significant difference in disease duration regarding where patients were scanned (patients scanned at UKD were more recently diagnosed and in contrary subjects scanned at UKA already had suffered significantly ($p \ll 0.5$) longer from the disease (figure 4.2)). This emphasized that *CHIMERA* clusters reflected the duration of illness.

Furthermore, observing the clusterings obtained by *K-means*, it becomes obvious that there was no pattern captured by this method. Patients scanned at different locations were rather equally partitioned throughout the clusters.

5.2 Limitations and Outlook

This section outlines the limitation of this work. Furthermore an outlook is given on how one could proceed based on the results that this thesis yielded.

5.2.1 Limitation

Data set:

As previously mentioned in 5.1.1 the provided data set was not optimal. Especially the fact that patients from different sites had suffered for a different amount of time from the Parkinson's disease was very unfortunate. As for now further investigation steps will be required to identify if *CHIMERA* reflected different conditions at the two scanner locations or identified the pathological progression according to the duration of illness.

Furthermore the sample size (180) was small compared to the relatively large feature size (20792). There are two problems which might have occurred:

- Firstly, as the the sample size yielded an upper bound for the dimensionality reduction methods applied in this work the maximum number of components of the reduced data was limited by the number of subjects, here 180. In settings where the feature space is significantly higher than the sample size, the intrinsic feature space (that is the minimal number of hidden features that explain the data) might be higher than what can be obtained by the components of the reduction methods. If this holds, both PCA and NMF

were not capable to explain the data even if all components would have been kept. In age prediction researchers found the best performance within a 300 to 500 range for NMF components [74] which is almost twice as many components as provided for this work.

- Secondly, one fundamental assumption of *CHIMERA* is that enough healthy patients have to be collected in order to describe the normal control population. If so, one can assume that the estimated anatomy of patients, had they been spared of the disease, is covered by the healthy distribution [17]. Given that *CHIMERA* was originally evaluated on a data set twice as large as the size of this work and also included a higher ratio of normal controls in comparison to the diseased, the number of normal controls in this work might not have been sufficient to capture the healthy anatomy.

Parameters

Most of the parameters in this thesis were chosen by default, that is taking the authors predefined parameter settings. Most factors however were originally obtained via *cross-validation* [44] leading to problem specific characteristics [17]. As parameters were kept fixed for the entire work different parameterizations would probably have influenced the clusterings.

5.2.2 Outlook

There are a few approaches in order to tackle some of the limitations outlined in 5.2:

1. in order to find out whether *CHIMERA* reflected the disease duration one could either use a different data set (where the disease duration is more equally spread) or rerun the algorithm while introducing duration of illness as a covariate. If *CHIMERA* identified the pathological subtypes according to patients' duration of illness the results are expected to be similar whereas different results must lead to further investigations.
2. As reduction methods are bound to the sample size including all the limitations that come along with it, instead of applying these methods an a priori reduction method, also called atlas, could be used. An atlas is a predefined parcellation of the brain where multiple voxels are combined for a specific region. Another advantage of using an atlas is that, once the clusters are calculated, region or voxel wise tests (e.g. a two-sample t-test) can be performed comparing different clusters of patients as well as the diseased with the healthy population. This could lead directly to areas that are effecting the clustering (for instance atrophy in a specific brain region might correspond to a specific subtype and hence to a specific cluster).
3. In order to identify more suitable parameter settings, *cross-validation* [72] is a commonly used approach to tackle this issue. Being able to generate problem specific parameterization might enhance the capability of *CHIMERA* to identify subtypes among the diseased.

5.3 Conclusion

One of the objectives of this thesis was to find whether it makes a difference to apply *CHIMERA* or *K-means* to Parkinson's patients and if so, how significant this dissimilarity is. This thesis could outline that *CHIMERA* led to fundamentally diverse groupings compared to *K-means*. Furthermore it has been shown that this diversity was consistent over multiple components of dimensionality reduction methods, namely *principle component analysis* and *non-negative matrix factorization*. In this work also two different schemes of confound filtering techniques were performed. Neither has shown to produce clusterings which were more similar to the ones *CHIMERA* yielded. Moreover this thesis was able to outline that based on the used parameter setting *CHIMERA* was sensitive to reduction methods. Clusterings that were obtained based on data that was reduced in dimensionality using PCA were different from those that utilized data reduced by NMF. Although the clusters were different when comparing the reduction techniques, clusters were consistent over multiple components within the methods.

While looking at characteristics of subjects in each specific cluster *CHIMERA* seemed to reflect the disease duration of patients. Unfortunately as the patients from Düsseldorf were relatively recently diagnosed with the disease in contrast to the subjects scanned in Aachen which were suffering significantly longer from the disease, this could also be a side effect. As *CHIMERA* found clusters that reflected the duration of illness and scanning locality, *K-means* in contrary had neither a significant pattern of scanning site nor disease duration in its clusters. Specifically, *K-means* did not capture any obvious trend within clusterings whereas *CHIMERA* clusters lied emphasis on a pattern of pathological progression. Under the assumptions that the results yielded by *CHIMERA* indeed reflected the advancement of the Parkinson's disease this would be a significant indication of the superiority of *CHIMERA* to capture subtypes of the disease in comparison to *K-means*.

However there were several limitations of this thesis. Most of them regarding to the provided data set or to the parameters used in this work. Hence, all acquired results are provided with caution.

A qualitative statement about which of the investigated clustering methods performed better or was even capable to capture the pathological process in the Parkinson's disease was - without any further investigation - impossible to make. Nevertheless, this work could show that the methodologies differed and that there was a chance of *CHIMERA* clusters reflecting the pathological progression of the diseased.

Certainly, further investigation is required to overcome the limitations of this work but potentially the results of this thesis can be a good point to start from.

Chapter 6

Appendix

6.1 Scree Plot

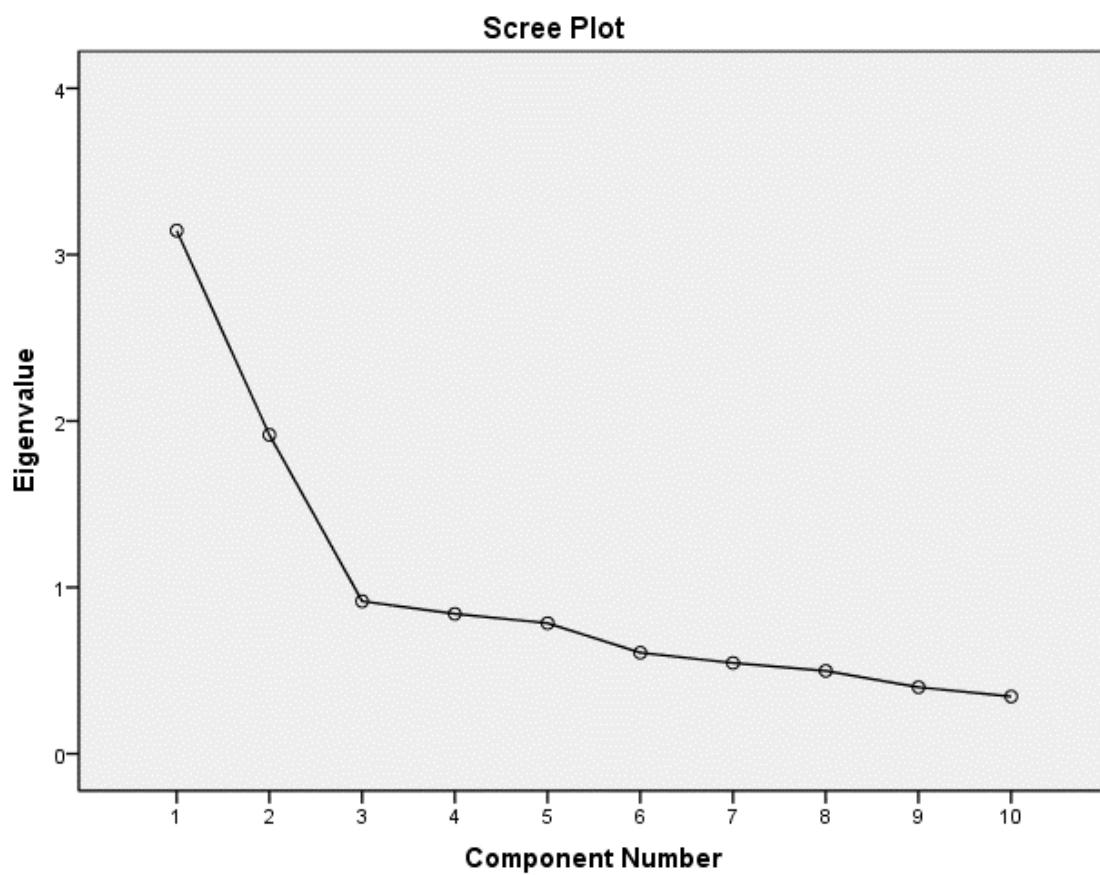
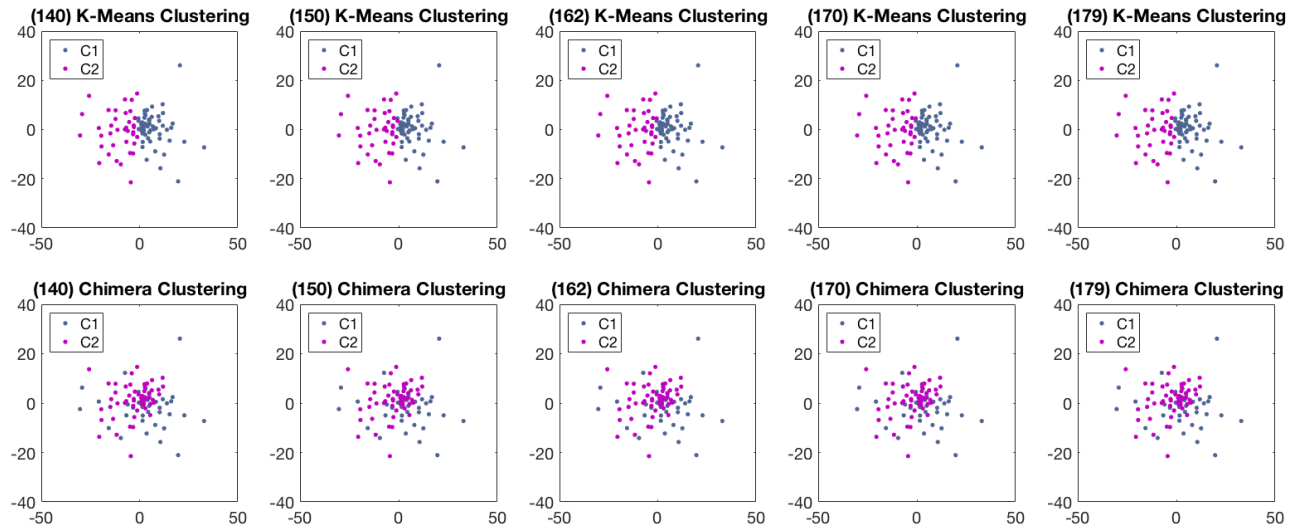


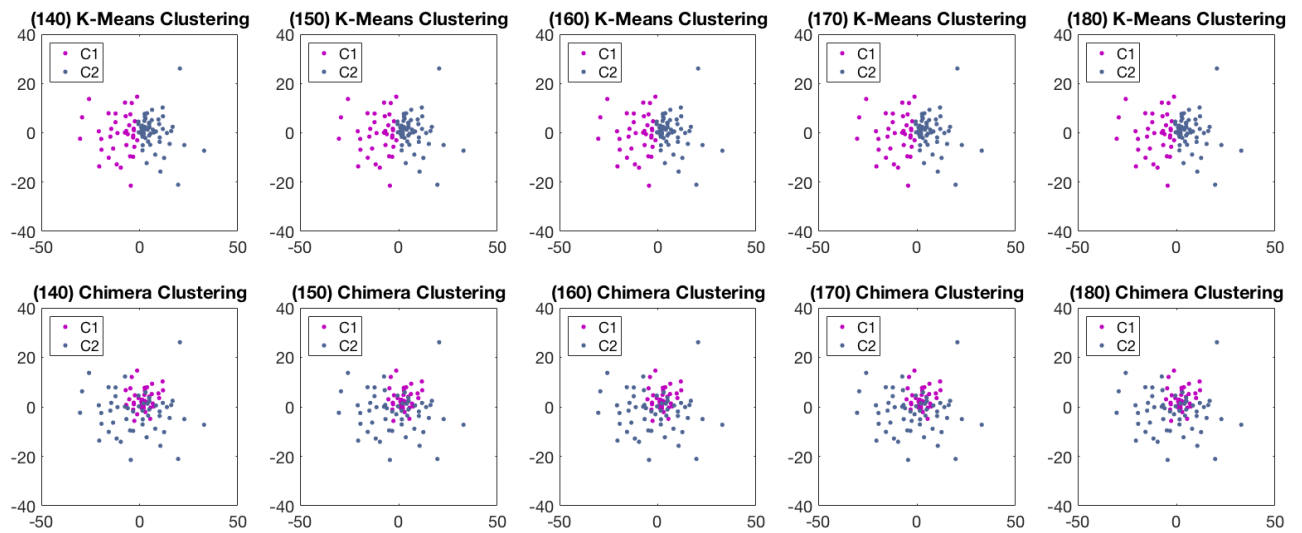
Figure 6.1: Scree Plot

In this example the first two components would be selected.

6.2 Clustering



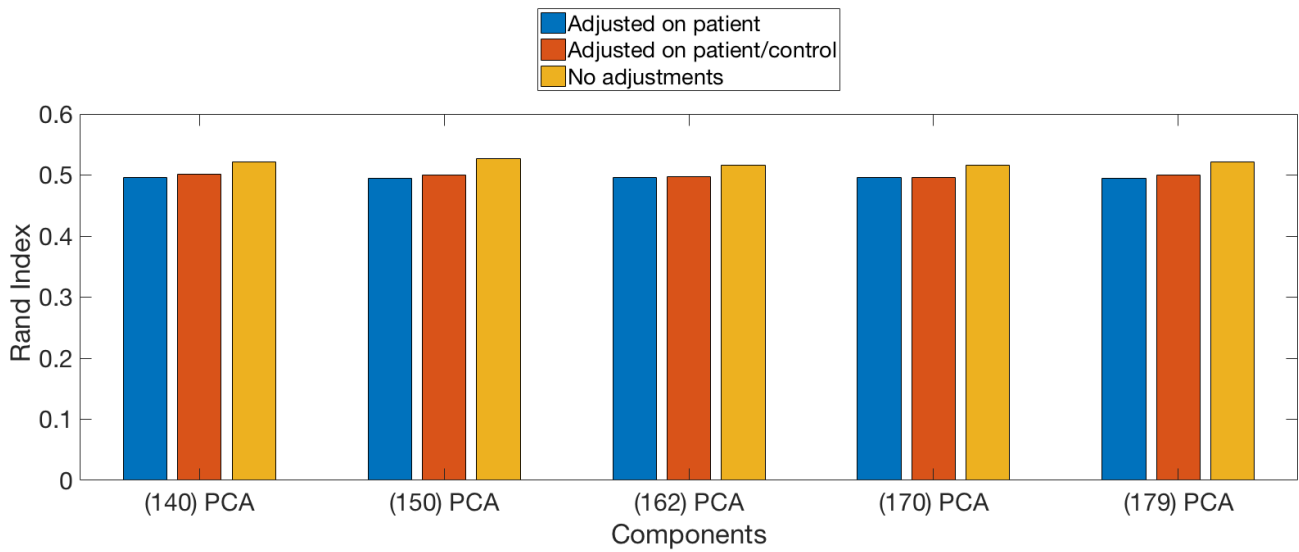
(a) Clusterings for different *PCA* components



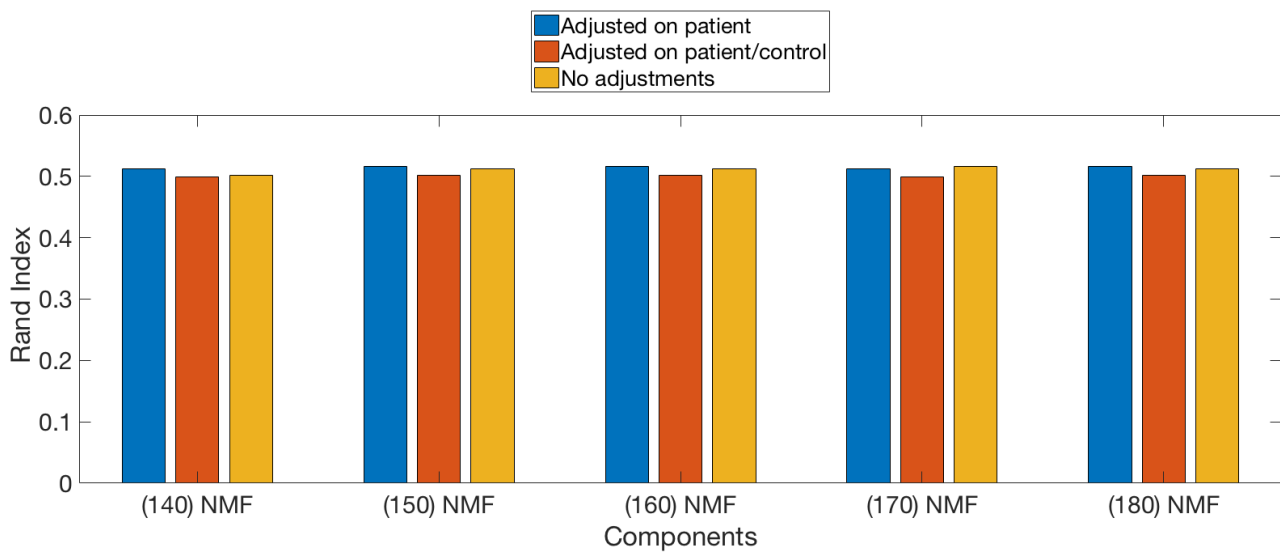
(b) Clusterings for different *NMF* components

Figure 6.2: 2D visualization of *CHIMERA* vs. *K-means* clusterings with different granularities. For *K-means* without any correction for confounds. The number of components is denoted in (\cdot) .

6.3 Rand Index



(a) Rand Index for multiple PCA components



(b) Rand index for multiple NMF components

Figure 6.3: Rand index between *CHIMERA* and different confound correction schemes of *K-means*. The number of components is denoted in (\cdot) .

6.4 Silhouettes

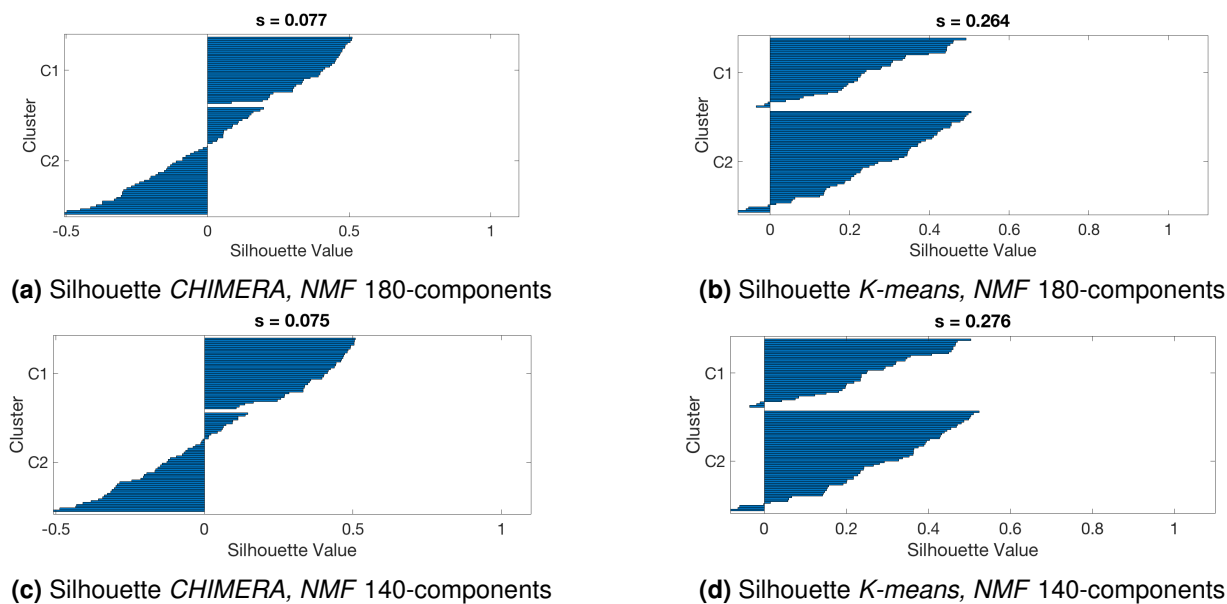


Figure 6.4: Silhouette for CHIMERA and K-means for two different NMF granularities. For K-means confounds were removed based on the total population.

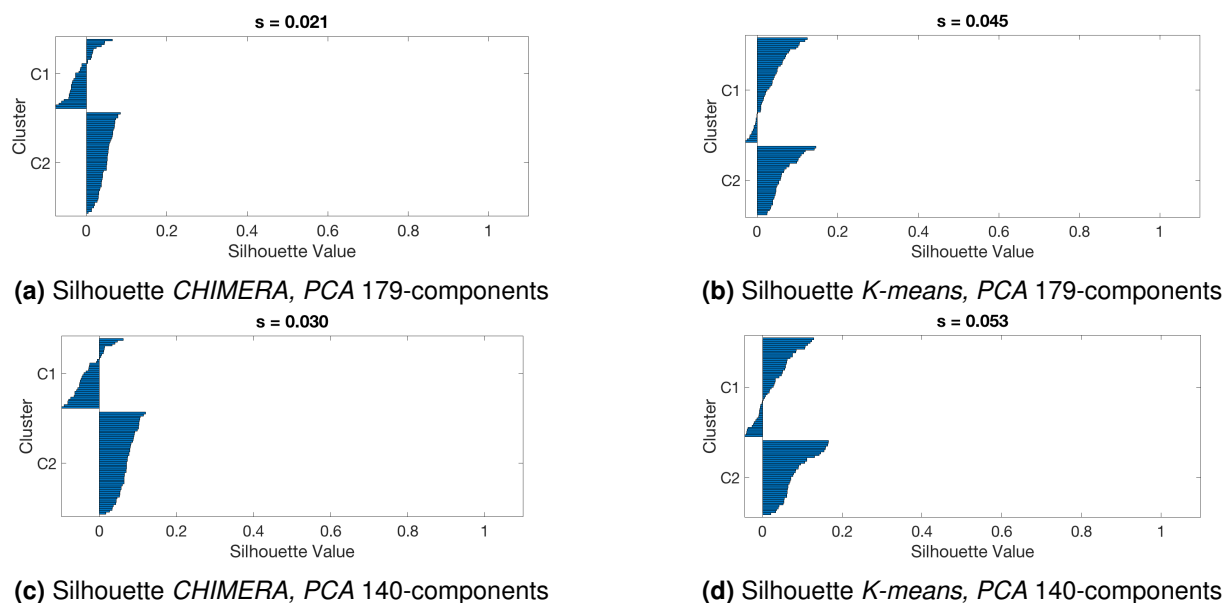
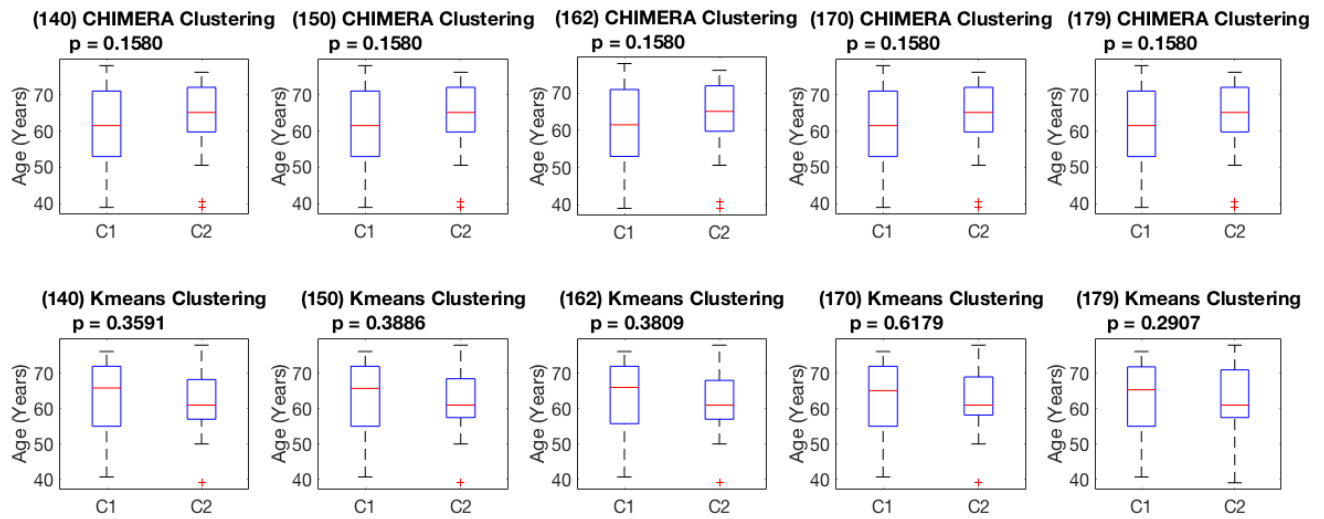
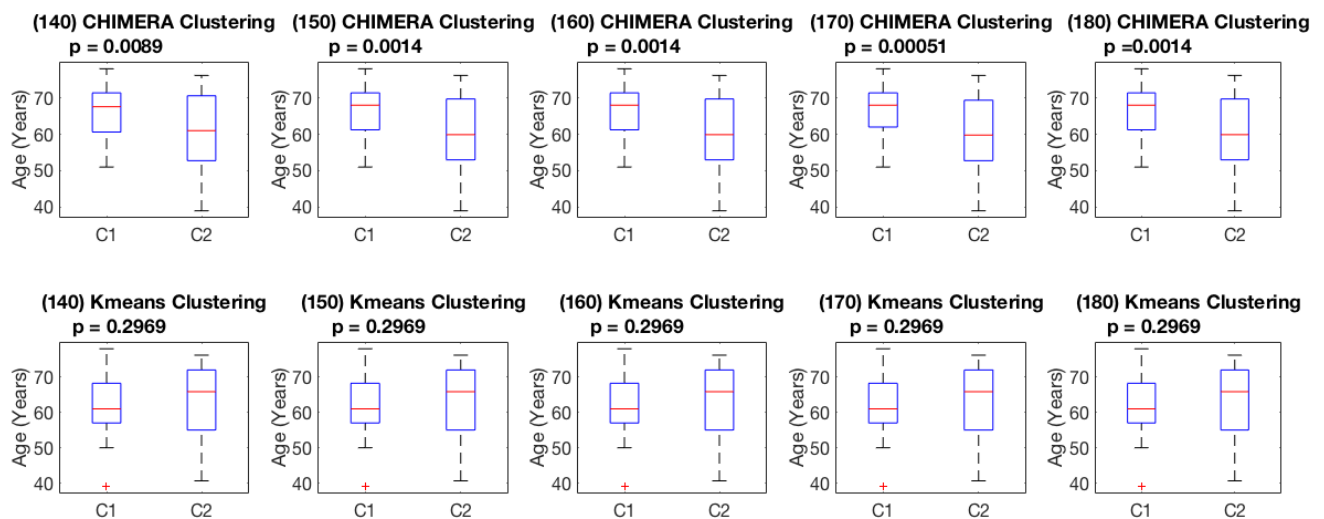


Figure 6.5: Silhouette for CHIMERA and K-means for two different PCA granularities. For K-means confounds were removed based on the total population.

6.5 Age



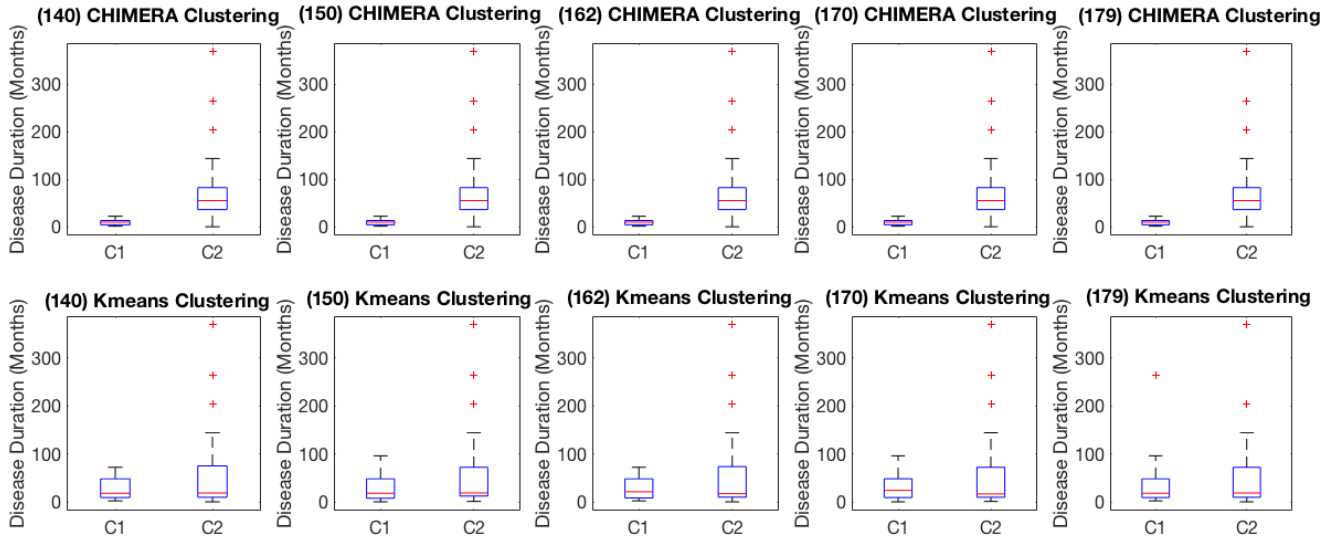
(a) Boxplot regarding age for different *PCA* components



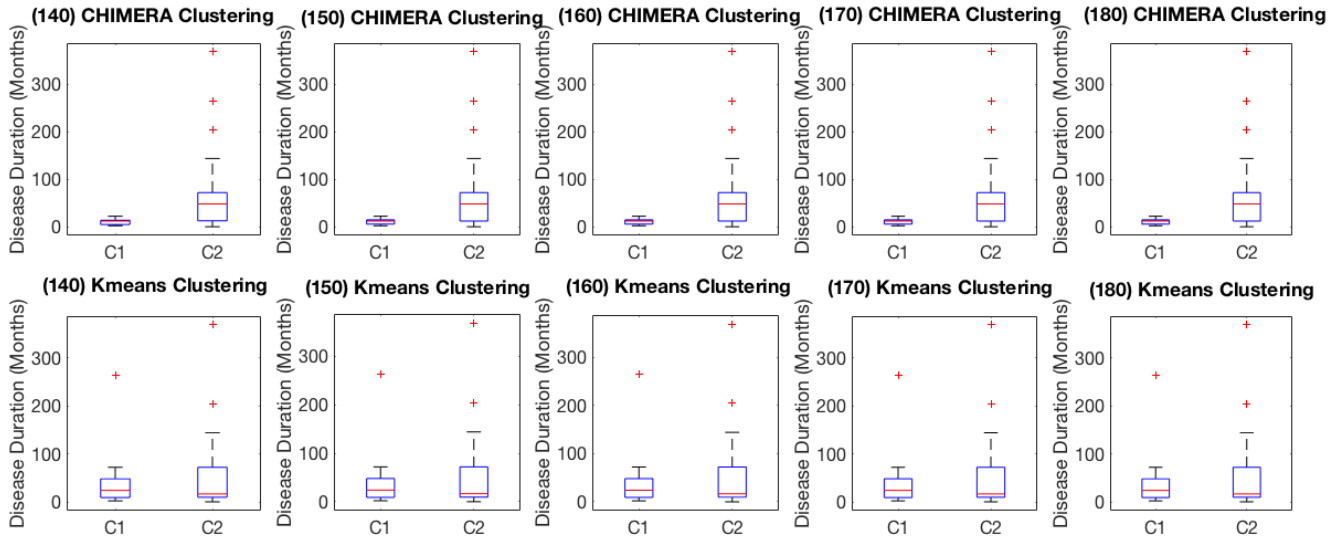
(b) Boxplot regarding age for different *NMF* components

Figure 6.6: Boxplots regarding age in the clusters for multiple *NMF* and *PCA* granularities. Confound correction based on the total population. P value refers to a two-sample t-test between both clusterings. The number of components is denoted in (·).

6.6 Disease Duration distribution



(a) Boxplot regarding patients disease duration for different *PCA* components



(b) Boxplot regarding patients disease duration for different *NMF* components

Figure 6.7: Boxplots regarding patients disease duration for both clustering methods and multiple *NMF* and *PCA* granularities. For *K-means* confounds correction was based on the total population. P value refers to a two-sample t-test between both clusterings. The number of components is denoted in (·).

6.6.1 Gender

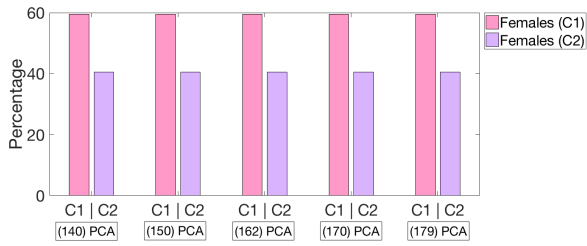
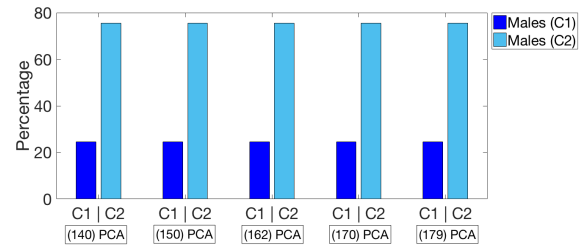
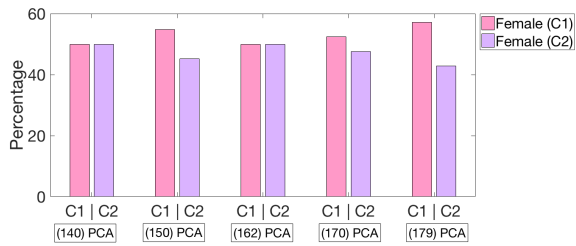
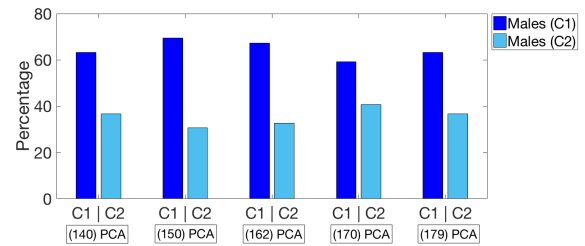
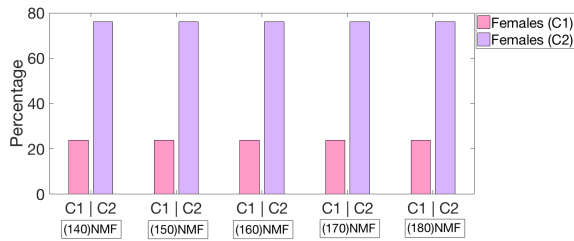
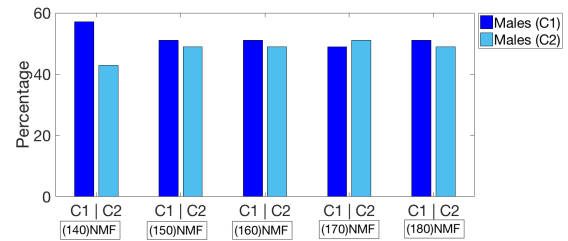
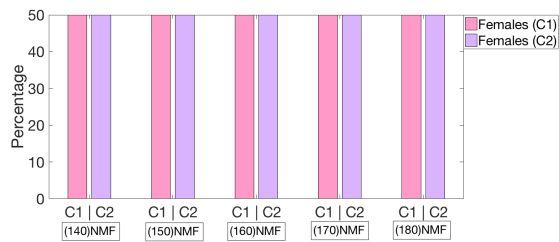
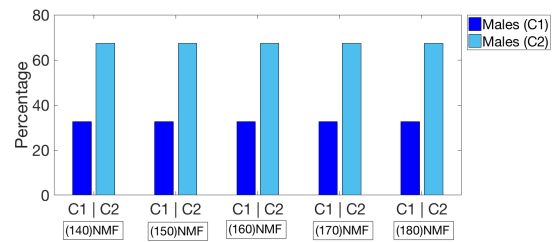
(a) Female distribution for *CHIMERA* clusters based on PCA components(b) Male distribution for *CHIMERA* clusters based on PCA components(c) Female distribution for *K-means* clusters based on PCA components(d) Male distribution for *K-means* clusters based on PCA components(e) Female distribution for *CHIMERA* clusters based on NMF components(f) Male distribution for *CHIMERA* clusters based on NMF components(g) Female distribution for *K-means* clusters for NMF components(h) Male distribution for *K-means* clusters based on NMF components

Figure 6.8: Illustration of the percentage of males and females in clusters C1 and C2 obtained from both *CHIMERA* and *K-means* based on multiple NMF and PCA components. Correction for possible confounds was patients and healthy controls. The number of components is denoted in (·).

6.6.2 Scanner

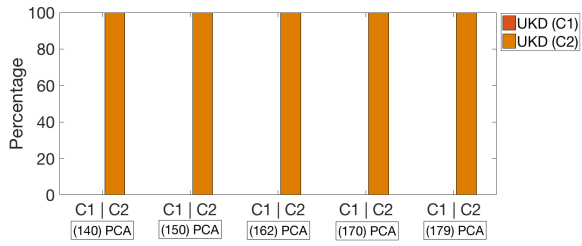
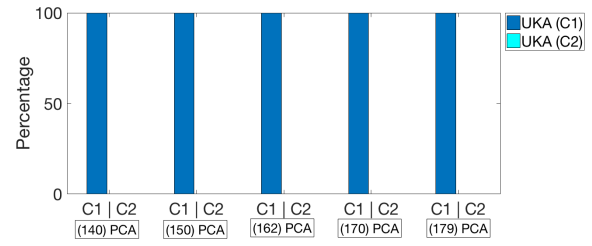
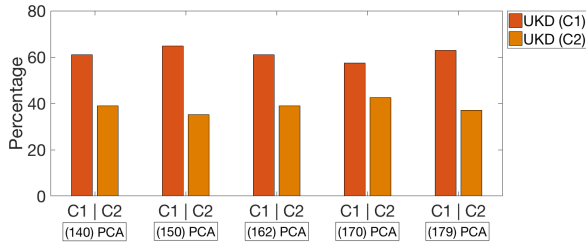
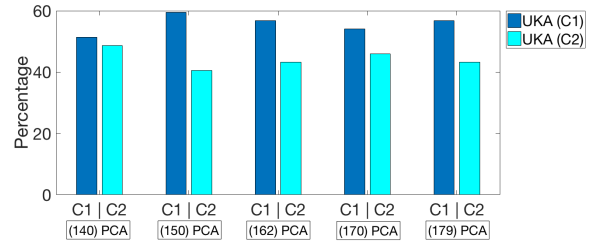
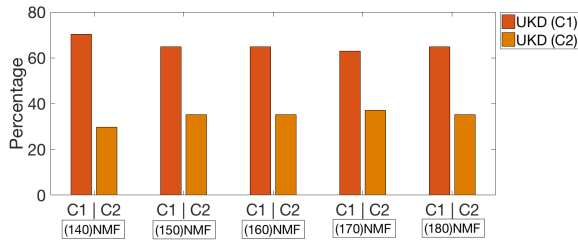
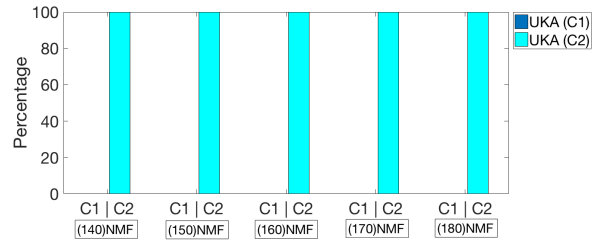
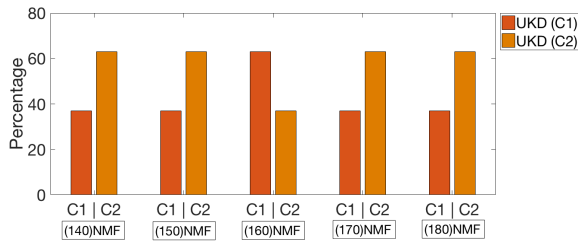
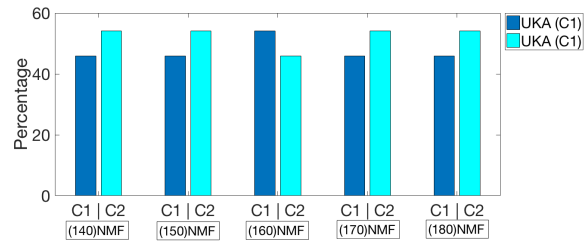
(a) Site distribution UKD *CHIMERA*, PCA(b) Site distribution UKA *CHIMERA*, PCA(c) Site distribution UKD *K-means*, PCA(d) Site distribution UKA *K-means*, PCA(e) Site distribution UKD *CHIMERA*, NMF(f) Site distribution UKA *CHIMERA*, NMF(g) Site distribution UKD *K-means*, NMF(h) Site distribution UKA *K-means*, NMF

Figure 6.9: Illustration of the percentage of patients scanned at *Universitätsklinikum Aachen* (UKA) or *Universitätsklinikum Düsseldorf* (UKD) in clusters C1 and C2 obtained from both *CHIMERA* and *K-means* for multiple NMF and PCA components. Correction for possible confounds were based on patients and healthy controls. The number of components is denoted in (·).

List of Figures

2.1	Illustration of T_1 weighted physiological brain images. Images were obtained from the clinical data set provided for this work	12
2.2	Illustration of T_2 weighted pathological brain images [26].	12
2.3	(a) the problem setting is displayed where \mathbf{X} denotes the control distribution and \mathbf{Y} the patient population respectively. (b) the model assumption: \mathbf{X} is transformed into a distribution \mathbf{X}' covering the distribution \mathbf{Y} by a set of K different transformations [17].	15
3.1	illustrates the step-wise procedure of this work. The different process steps are colored as follows: green indicating preprocessing, blue corresponding to the actual clustering and red reflecting the postprocessing.	19
3.2	sagittal (left), coronal (middle) and axial (right) illustrations of a single subjects brain. Images are T1-weighted brain MR-images where the mask used in this work was overlayed.	20
3.3	the magenta colored line denotes the lower dimensional <i>principal subspace</i> that <i>principle component analysis</i> seeks to find. The subspace is chosen in such a way that the orthogonal projection of the red colored data points onto the new subspace maximizes the variance of the projected green points. The alternative formulation of <i>PCA</i> that is based on minimizing the sum-of-squares of the projection errors is illustrated by the blue lines [59].	24
3.4	Visualization of K-means: (a) two-dimensional data in the Euclidean space (green points). Not yet clustered. The initial centroids or prototypes μ_1 and μ_2 are denoted by the red and blue crosses respectively. (b) the initial E-step assigns each data point to the red or the blue cluster according to the closest centroid. The perpendicular bisector of the two cluster centers, shown by the magenta denotes the decision boundary. (c) each cluster center is recomputed (M-step) to be the mean of the points assigned to the corresponding cluster (d) - (i) alternately E and M steps until final convergence of <i>K-means</i> is reached [69].	32
3.5	Illustration of elements involved in the computation of the silhouette $s(o)$. A, B and C are clusters, o is the observation that is considered.	34
4.1	Scanning site comparison of subjects	39

4.2	Boxplots: Age and disease duration. The p-value corresponds to a two-sample t-test of the respective distributions.	39
4.3	Gray matter mean and variance of every subject per voxel	40
4.4	180 samples in the 2D feature space. Each sample in the original space was represented by 20792 (number of voxels) grey matter values. Here, every sample is characterized by two values that were obtained by <i>multidimensional scaling</i> . Points close together in the original space are also close in the 2D space.	41
4.5	Variance of the feature matrix retained for each principal component	43
4.6	Contribution of each voxel to the components in different NMF granularities	44
4.7	2D visualization of <i>CHIMERA</i> and <i>K-means</i> clusterings based on different data representations. For <i>K-means</i> possible confounds were corrected according to the patient population. The number of components for each reduction method is denoted in (\cdot).	46
4.8	2D visualization of <i>CHIMERA</i> and <i>K-means</i> clusterings based different data representations. For <i>K-means</i> possible confounds were corrected based on the total population. The number of components for each reduction method is denoted in (\cdot).	47
4.9	Adjusted Rand index between <i>CHIMERA</i> and different schemes of <i>K-means</i> . The number of components is denoted in (\cdot).	48
4.10	Silhouette for <i>CHIMERA</i> and <i>K-means</i> for two different NMF granularities. For <i>K-means</i> possible confounds were corrected based on the patients distribution, s denotes the <i>silhouette coefficient</i>	50
4.11	Silhouette for <i>CHIMERA</i> and <i>K-means</i> for two different NMF granularities. For <i>K-means</i> possible confounds were corrected based on the patients and healthy distribution, s denotes the <i>silhouette coefficient</i>	50
4.12	Boxplots of the age distribution in clusterings (C1,C2) obtained from both, <i>CHIMERA</i> and <i>K-means</i> for multiple NMF and PCA components. P-value was obtained by a two sample t-test. Correction for possible confounds was based on the patient samples. The number of components is denoted in (\cdot).	51
4.13	Illustration of the disease duration of patients in clusterings (C1,C2) obtained from both <i>CHIMERA</i> and <i>K-means</i> for multiple NMF and PCA components. The correction for possible confound was based on the patient population. P-values were calculated by performing a two sampled t-test for the specific clustering. The number of components is denoted in (\cdot).	52
4.14	Illustration of the percentage of males and females in clusters C1 and C2 obtained from both <i>CHIMERA</i> and <i>K-means</i> based on multiple NMF and PCA components. Correction for possible confounds was based on patient samples. The number of components is denoted in (\cdot).	53

4.15	Illustration of the percentage of patients scanned at <i>Universitätsklinikum Aachen</i> (UKA) and <i>Universitätsklinikum Düsseldorf</i> (UKD) in clusters C1 and C2 obtained from both <i>CHIMERA</i> and <i>K-means</i> for multiple NMF and PCA components. Correction for possible confounds were based on patient samples. The number of components is denoted in (\cdot).	54
6.1	Scree Plot	62
6.2	2D visualization of <i>CHIMERA</i> vs. <i>K-means</i> clusterings with different granularities. For <i>K-means</i> without any correction for confounds. The number of components is denoted in (\cdot).	63
6.3	Rand index between <i>CHIMERA</i> and different confound correction schemes of <i>K-means</i> . The number of components is denoted in (\cdot).	64
6.4	Silhouette for <i>CHIMERA</i> and <i>K-means</i> for two different NMF granularities. For <i>K-means</i> confounds were removed based on the total population.	65
6.5	Silhouette for <i>CHIMERA</i> and <i>K-means</i> for two different PCA granularities. For <i>K-means</i> confounds were removed based on the total population.	65
6.6	Boxplots regarding age in the clusters for multiple NMF and PCA granularities. Confound correction based on the total population. P value refers to a two-sample t-test between both clusterings. The number of components is denoted in (\cdot).	66
6.7	Boxplots regarding patients disease duration for both clustering methods and multiple NMF and PCA granularities. For <i>K-means</i> confounds correction was based on the total population. P value refers to a two-sample t-test between both clusterings. The number of components is denoted in (\cdot).	67
6.8	Illustration of the percentage of males and females in clusters C1 and C2 obtained from both <i>CHIMERA</i> and <i>K-means</i> based on multiple NMF and PCA components. Correction for possible confounds was patients and healthy controls. The number of components is denoted in (\cdot).	68
6.9	Illustration of the percentage of patients scanned at <i>Universitätsklinikum Aachen</i> (UKA) or <i>Universitätsklinikum Düsseldorf</i> (UKD) in clusters C1 and C2 obtained from both <i>CHIMERA</i> and <i>K-means</i> for multiple NMF and PCA components. Correction for possible confounds were based on patients and healthy controls. The number of components is denoted in (\cdot).	69

List of Tables

3.1 contingency table for clusterings X and Y 35

4.1 Cluster consistency in *CHIMERA* clusters. The number of components for each reduction method is denoted in (\cdot) 49

4.2 Cluster consistency in *K-means* clusterings, possible confounds were corrected based on patients. The number of components for each reduction method is denoted in (\cdot) 49

4.3 Cluster consistency in *K-means* clusterings, possible confounds were corrected based on patients and healthy controls. The number of components for each reduction method is denoted in (\cdot) 49

Bibliography

- [1] Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol.* 2011 Sep;10(9):785–796.
- [2] Geschwind DH, Levitt P. Autism spectrum disorders: developmental disconnection syndromes. *Curr Opin Neurobiol.* 2007 Feb;17(1):103–111.
- [3] Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol.* 2014 Feb;10(2):74–81.
- [4] Lewis SJ, Foltynie T, Blackwell AD, Robbins TW, Owen AM, Barker RA. Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. *J Neurol Neurosurg Psychiatry.* 2005 Mar;76(3):343–348.
- [5] Buchanan RW, Carpenter WT. Domains of psychopathology: an approach to the reduction of heterogeneity in schizophrenia. *J Nerv Ment Dis.* 1994 Apr;182(4):193–204.
- [6] Zhang T, Koutsouleris N, Meisenzahl E, Davatzikos C. Heterogeneity of structural brain changes in subtypes of schizophrenia revealed using magnetic resonance imaging pattern analysis. *Schizophr Bull.* 2015 Jan;41(1):74–84.
- [7] Nenadic I, Sauer H, Gaser C. Distinct pattern of brain structural deficits in subsyndromes of schizophrenia delineated by psychopathology. *Neuroimage.* 2010 Jan;49(2):1153–1160.
- [8] Nikolaos Koutsouleris MJRBTFSHGJESTZBBJSCBMRHJMEMMristian Gaser. Structural correlates of psychopathological symptom dimensions in schizophrenia: A voxel-based morphometric study. *ELSEVIER.* 2007;.
- [9] Graham JM, Sagar HJ. A data-driven approach to the study of heterogeneity in idiopathic Parkinson's disease: identification of three distinct subtypes. *Mov Disord.* 1999 Jan;14(1):10–20.
- [10] Yamasue H, Kasai K, Iwanami A, Ohtani T, Yamada H, Abe O, et al. Voxel-based analysis of MRI reveals anterior cingulate gray-matter volume reduction in posttraumatic stress disorder due to terrorism. *Proc Natl Acad Sci USA.* 2003 Jul;100(15):9039–9043.
- [11] Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, et al. Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci.* 1999 Oct;2(10):861–863.
- [12] Thirion B, Pinel P, Meriaux S, Roche A, Dehaene S, Poline JB. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage.* 2007 Mar;35(1):105–120.

- [13] Ashburner J, Friston KJ. Voxel-based morphometry—the methods. *Neuroimage*. 2000 Jun;11(6 Pt 1):805–821.
- [14] Honea R, Crow TJ, Passingham D, Mackay CE. Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am J Psychiatry*. 2005 Dec;162(12):2233–2245.
- [15] Goodlett CB, Fletcher PT, Gilmore JH, Gerig G. Group analysis of DTI fiber tract statistics with application to neurodevelopment. *Neuroimage*. 2009 Mar;45(1 Suppl):S133–142.
- [16] Varol E, Sotiras A, Davatzikos C. HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *Neuroimage*. 2017 Jan;145(Pt B):346–364.
- [17] Dong A, Honnorat N, Gaonkar B, Davatzikos C. CHIMERA: Clustering of Heterogeneous Disease Effects via Distribution Matching of Imaging Patterns. *IEEE Trans Med Imaging*. 2016 Feb;35(2):612–621.
- [18] Pourhoseingholi MA, Baghestani AR, Vahedi M. How to control confounding effects by statistical analysis. *Gastroenterol Hepatol Bed Bench*. 2012;5(2):79–83.
- [19] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature*. 1999 10;401:788 EP –. Available from: <http://dx.doi.org/10.1038/44565>.
- [20] Bishop CM. Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc.; 2006. Chapter 12.1.
- [21] Krishnan KRR. Chapter 6 - Structural imaging in psychiatric disorders. In: Aminoff MJ, Boller F, Swaab DF, editors. *Neurobiology of Psychiatric Disorders*. vol. 106 of *Handbook of Clinical Neurology*. Elsevier; 2012. p. 89 – 95. Available from: <http://www.sciencedirect.com/science/article/pii/B9780444520029000061>.
- [22] Rinck P. Magnetic Resonance in Medicine. The Basic Textbook of the European Magnetic Resonance Forum. 11th ed. Electronic version; 1 June 2017. Chapter: 11-04 Functional Imaging.
- [23] Donald W McRobbie MJGMR Elizabeth A Moore. MRI from Picture to Proton. Cambridge University Press; 2007.
- [24] Joseph P Hornak PD. The Basics of NMR. Rochester Institute of Technology; 1997-2017. <https://www.cis.rit.edu/htbooks/nmr/inside.htm>.
- [25] Donald W McRobbie MJGMR Elizabeth A Moore. MRI from Picture to Proton. vol. pp. 33-35. Cambridge University Press; 2007.
- [26] Chavhan GB, Babyn PS, Thomas B, Shroff MM, Haacke EM. Principles, techniques, and applications of T2*-based MR imaging and its special applications. *Radiographics*. 2009;29(5):1433–1449.
- [27] Davatzikos C, Genc A, Xu D, Resnick SM. Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. *Neuroimage*. 2001 Dec;14(6):1361–1369.
- [28] Ashburner J, Hutton C, Frackowiak R, Johnsrude I, Price C, Friston K. Identifying global anatomical differences: deformation-based morphometry. *Hum Brain Mapp*. 1998;6(5-6):348–357.

- [29] Chung MK, Worsley KJ, Paus T, Cherif C, Collins DL, Giedd JN, et al. A unified statistical approach to deformation-based morphometry. *Neuroimage*. 2001 Sep;14(3):595–606.
- [30] Chung MK, Worsley KJ, Robbins S, Paus T, Taylor J, Giedd JN, et al. Deformation-based surface morphometry applied to gray matter deformation. *Neuroimage*. 2003 Feb;18(2):198–213.
- [31] Job DE, Whalley HC, McConnell S, Glabus M, Johnstone EC, Lawrie SM. Structural gray matter differences between first-episode schizophrenics and normal controls using voxel-based morphometry. *Neuroimage*. 2002 Oct;17(2):880–889.
- [32] Giuliani NR, Calhoun VD, Pearlson GD, Francis A, Buchanan RW. Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. *Schizophr Res*. 2005 May;74(2-3):135–147.
- [33] Ecker C, Marquand A, Mourao-Miranda J, Johnston P, Daly EM, Brammer MJ, et al. Describing the brain in autism in five dimensions—magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. *J Neurosci*. 2010 Aug;30(32):10612–10623.
- [34] Sabuncu MR, Balci SK, Shenton ME, Golland P. Image-driven population analysis through mixture modeling. *IEEE Trans Med Imaging*. 2009 Sep;28(9):1473–1487.
- [35] Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehericy S, Habert MO, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage*. 2011 May;56(2):766–781.
- [36] Davatzikos C, Fan Y, Wu X, Shen D, Resnick SM. Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol Aging*. 2008 Apr;29(4):514–523.
- [37] McEvoy LK, Fennema-Notestine C, Roddey JC, Hagler DJ, Holland D, Karow DS, et al. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. *Radiology*. 2009 Apr;251(1):195–205.
- [38] Noh Y, Jeon S, Lee JM, Seo SW, Kim GH, Cho H, et al. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. *Neurology*. 2014 Nov;83(21):1936–1944.
- [39] Whitwell JL, Petersen RC, Negash S, Weigand SD, Kantarci K, Ivnik RJ, et al. Patterns of atrophy differ among specific subtypes of mild cognitive impairment. *Arch Neurol*. 2007 Aug;64(8):1130–1138.
- [40] Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Mov Disord*. 2014 Nov;29(13):1583–1590.
- [41] Braak H, Del Tredici K, Rub U, de Vos RA, Jansen Steur EN, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging*. 2003;24(2):197–211.
- [42] Sveinbjornsdottir S. The clinical symptoms of Parkinson's disease. *Journal of Neurochemistry*. 2016;139:318–324. Available from: <http://dx.doi.org/10.1111/jnc.13691>.

- [43] Ashburner J, Friston KJ. Unified segmentation. *NeuroImage*. 2005;26(3):839 – 851. Available from: <http://www.sciencedirect.com/science/article/pii/S1053811905001102>.
- [44] Rosenberg-Katz K, Herman T, Jacob Y, Giladi N, Hendler T, Hausdorff JM. Gray matter atrophy distinguishes between Parkinson disease motor subtypes. *Neurology*. 2013 Apr;80(16):1476–1484.
- [45] Christian Gaser RD. Manual Computational Anatomy Toolbox - CAT12. Jena University Hospital, Departments of Psychiatry and Neurology; 2017. Available from: <http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf>.
- [46] Groenen IBPJF. Modern Multidimensional Scaling. 2nd ed. Springer-Verlag New York; 2005.
- [47] Hubert L, Arabie P. Comparing partitions. *Journal of Classification*. 1985 Dec;2(1):193–218. Available from: <https://doi.org/10.1007/BF01908075>.
- [48] Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987;20:53 – 65. Available from: <http://www.sciencedirect.com/science/article/pii/0377042787901257>.
- [49] Sotiras A, Resnick SM, Davatzikos C. Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization. *Neuroimage*. 2015 Mar;108:1–16.
- [50] Chris Ding HDS Xiaofeng He. On the equivalence of nonnegative matrix factorization and spectral clustering. *Proceedings of the 2005 SIAM International Conference on Data Mining*. Chris Ding HDS Xiaofeng He, editor. Society for Industrial and Applied Mathematics; 2005/4/21. Pp. 606-610.
- [51] S Tsuge SK M Shishibori. Dimensionality reduction using non-negative matrix factorization for information retrieval. In: *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*; 06 August 2002. .
- [52] Lee DD, Seung HS. Algorithms for Non-negative Matrix Factorization. In: *In NIPS*. MIT Press; 2001. p. 556–562.
- [53] Yang Z, Oja E. Linear and Nonlinear Projective Nonnegative Matrix Factorization. *Trans Neur Netw*. 2010 May;21(5):734–749. Available from: <http://dx.doi.org/10.1109/TNN.2010.2041361>.
- [54] Boutsidis C, Gallopoulos E. SVD Based Initialization: A Head Start for Nonnegative Matrix Factorization. *Pattern Recogn*. 2008 Apr;41(4):1350–1362. Available from: <http://dx.doi.org/10.1016/j.patcog.2007.09.010>.
- [55] Daubechies I, Roussos E, Takerkart S, Benharrosh M, Golden C, D'Ardenne K, et al. Independent component analysis for brain fMRI does not select for independence. *Proc Natl Acad Sci USA*. 2009 Jun;106(26):10415–10422.
- [56] Jolliffe I. Principal component analysis. New York: Springer Verlag; 2002.
- [57] Pearson K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*. 1901;2(6):559–572.
- [58] H H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*. 1933;24.

- [59] Bishop CM. Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc.; 2006. P. 561.
- [60] Ruigrok AN, Salimi-Khorshidi G, Lai MC, Baron-Cohen S, Lombardo MV, Tait RJ, et al. A meta-analysis of sex differences in human brain structure. *Neurosci Biobehav Rev*. 2014 Feb;39:34–50.
- [61] Terribilli D, Schauffelberger MS, Duran FL, Zanetti MV, Curiati PK, Menezes PR, et al. Age-related gray matter volume changes in the brain during non-elderly adulthood. *Neurobiol Aging*. 2011 Feb;32(2):354–368.
- [62] Maclaren J, Han Z, Vos SB, Fischbein N, Bammer R. Reliability of brain volume measurements: a test-retest dataset. *Sci Data*. 2014;1:140037.
- [63] David M Diez MCR Christopher D Barr. OpenIntro Statistics Third Edition. Third edition ed. OpenIntro; 2017. Chapter 5.
- [64] Myronenko A, Song X. Point set registration: coherent point drift. *IEEE Trans Pattern Anal Mach Intell*. 2010 Dec;32(12):2262–2275.
- [65] Bishop CM. Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc.; 2006. Chapter 9.
- [66] Moon TK. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*. 1996 Nov;13(6):47–60.
- [67] Neal RM, Hinton GE. In: Jordan MI, editor. A View of the Em Algorithm that Justifies Incremental, Sparse, and other Variants. Dordrecht: Springer Netherlands; 1998. p. 355–368. Pp. 355–368. Available from: https://doi.org/10.1007/978-94-011-5014-9_12.
- [68] Bishop CM. Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc.; 2006. Chapter 9.1.
- [69] Bishop CM. Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc.; 2006. P. 426.
- [70] Rand WM. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*. 1971;66(336):846–850. Available from: <http://www.jstor.org/stable/2284239>.
- [71] Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*. 1964 Mar;29(1):1–27. Available from: <https://doi.org/10.1007/BF02289565>.
- [72] Refaeilzadeh P, Tang L, Liu H. In: LIU L, ÖZSU MT, editors. Cross-Validation. Boston, MA: Springer US; 2009. p. 532–538. Available from: https://doi.org/10.1007/978-0-387-39940-9_565.
- [73] Lewis MM, Du G, Lee EY, Nasrallah Z, Sterling NW, Zhang L, et al. The pattern of gray matter atrophy in Parkinson's disease differs in cortical and subcortical regions. *J Neurol*. 2016 Jan;263(1):68–75.
- [74] Varikuti DP, Genon S, Sotiras A, Schwender H, Hoffstaedter F, Patil KR, et al. Evaluation of non-negative matrix factorization of grey matter in age prediction. *NeuroImage*. 2018; Available from: <http://www.sciencedirect.com/science/article/pii/S1053811918301927>.