Clustering of individuals with dementia based on MRI features

MSc Thesis Biomedical Engineering
BM51035 - Medical Physics

by

N.J. Koorn

to obtain the degree of Master of Science
at the Delft University of Technology
to be defended on Thursday May 30, 2024 at 10:00.

Chair and Supervisor:
F.M. Vos, Erasmus MC, TU Delft

Supervisors:
M.F. van Haaften, Erasmus MC
E.E. Bron, Erasmus MC

Independent Committee Member:
J. Neitzel, Erasmus MC

Clustering of individuals with dementia based on MRI features

N.J. Koorn[1,2], F.M. Vos[1,2], E.E. Bron[2] and M.F. van Haaften[2]

[1]Delft University of Technology
[2]Erasmus Medical Center, Rotterdam

Abstract

Dementia, characterized by a significant decline in cognitive abilities, encompasses various neurodegenerative disorders. This includes Alzheimer's disease (AD), frontotemporal dementia (FTD) and their clinical manifestation named primary progressive aphasia (PPA), mainly involving language difficulties. In general these diseases exhibit some disease specific clinical characteristics. However, considerable heterogeneity exists in both atrophy patterns and symptom expression within each disease. Additionally, the overlap in disease patterns between AD and FTD complicates diagnosis. In this study, an unsupervised clustering algorithm is used to identify distinct atrophy patterns within AD and FTD, aiming to explore heterogeneity as well as disease overlap. Through the clustering of grey matter volumes, we identified two primary clusters potentially reflective of disease stage. These clusters were externally validated. Additionally, the two primary clusters were subdivided into four distinct clusters, each representing an unique atrophy pattern. These four clusters exhibited differences in cognitive performance in the five cognitive domains: visuospatial functioning, language, memory/learning, processing speed, and attention/executive functioning. The identified clusters provide inside in the atrophy patterns in AD and FTD, as well as their relation to clinical diagnosis and cognitive performance.

Nomenclature

ACE	Alzheimer Center Erasmus MC
AD	Alzheimer's Disease
FTD	Frontotemporal Dementia
GMP	Grey Matter Percentile
NACC	National Alzheimer's Coordinating Center
PPA	Primary Progressive Aphasia

1 Introduction

Dementia is a syndrome characterised by a substantial deterioration in cognitive abilities. While atrophy, cognitive decline and vascular impairment are part of the normal ageing process, these factors are exacerbated in dementia and have a significant impact on daily activities [1].

The most common diseases causing dementia include Alzheimer's disease (AD), Frontotemporal dementia (FTD), vascular dementia (VaD), and dementia with Lewy Bodies (LBD) [2]. In all of them diseases, dementia is attributed to loss of brain neurons, caused by the accumulation of proteins in the brain, vascular damage, or both. Firstly, AD is the most prevalent disease causing dementia, and is characterized by amyloid beta and tau accumulation in parts of the brain. AD is in its typical form associated with hippocampal atrophy and memory loss [3], while in atypical forms of AD little to no hippocampal atrophy is presented. Examples of atypical forms of AD include posterior cortical atrophy (PCA) and logopenic progressive aphasia (LPA). PCA is characterized by impairment in visual and visuospatial skills, i.e. the ability to identify and analyse spaces [4]. The most prominent feature of LPA is language difficulties [1]. Secondly, FTD is primarily caused by the accumulation of tau or TDP-43 proteins, and presents with executive dysfunction and personality changes. Furthermore, FTD presents clinically with subtypes associated with changes in behavior and issues in language processing [1], of which the language variants include semantic dementia (SD) and progressive nonfluent aphasia (PNFA). Collectively, all language variants of FTD (SD and PNFA) and AD (LPA) are classified as primary progressive aphasia (PPA). In general, AD and FTD manifest with these disease specific clinical patterns. However, considerable heterogeneity is found within each disease, both in symptom expression and atrophy patterns [1].

MRI images can be used to identify these atrophy patterns and play an important role in detecting the cause of dementia in the memory clinic [1]. These atrophy patterns differ partly between diseases, but there is also overlap. In general, in AD, the hippocampus is frequently affected, while FTD shows atrophy in the frontal and/or temporal lobes. The language variants (PPA) of these diseases are characterised by temporal lobe atrophy.

Diagnosing and understanding of dementia is complex due to multiple factors. These include the ob-

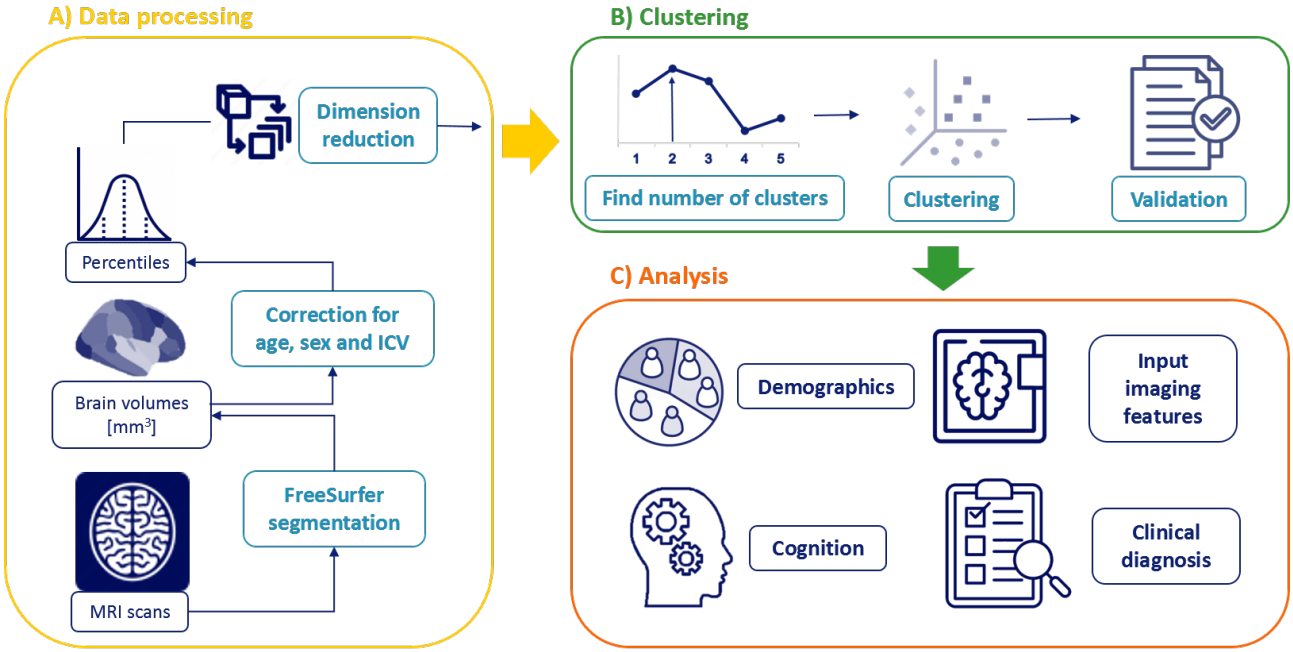


Figure 1: *The cluster analysis pipeline involves several steps. First, the MRI data is processed to derive grey matter percentiles, corrected for age, sex and intracranial volume (ICV) (A). Secondly, the grey matter percentiles are clustered for the optimal number of clusters and the resulting clusters are validated (B). Finally, the clusters are compared for demographic, imaging, cognitive and clinical features (C).*

served heterogeneity among patients with the same disease, the occurrence of mixed pathology, and the overlap in disease expressions between different types of dementia. Mixed pathology, which refers to the coexistence of multiple underlying pathologies, mainly involves a combination of AD and VaD [5]. This is most often seen in elderly subjects, since the vascular damage is increased with age [1][2]. The overlap in disease expression particularly, is causing high uncertainty in the etiological diagnosis.

To address the challenge of heterogeneity, current literature [3][6][7][8][9][10][11][12][13] suggests the application of unsupervised learning techniques. Unsupervised learning can reveal patterns in data and aid in understanding the heterogeneity within dementia without the use of a priori information. The problem with using labels lies in the uncertainty associated with clinical diagnoses of dementia patients. Previous research has mainly focused on the identification of clusters within specific diseases, especially AD [14] [15] [7] [16][8], FTD [9] [11] and LBD [13]. Comprehensive analyses that include multiple diseases that cause dementia are limited. The inclusion of more than one etiological diagnosis of dementia in a cluster analysis can provide insight into the overlap of atrophy patterns between these diseases, as well as insight to the heterogeneity within these diseases.

This study aims to gain more insight into the complexity within and between diseases causing dementia, with a focus on atrophy patterns within AD and FTD. This goal is pursued by the use of a clustering algorithm. This algorithm will make clusters in a dataset based on grey matter volumes. The resulting clusters will be analysed by exploring their relation with cognitive performance and clinical diagnoses.

2 Methods

2.1 Datasets

The study used data from the Alzheimer Center Erasmus MC (ACE) database. This database comprised patients who visited the memory clinic due to suspected neurodegenerative disorders. The data analysis encompassed a total of 297 subjects, with 51.9% of them being female. These subjects were diagnosed with AD, FTD or PPA. For PPA cases, subtypes were specified if available: PNFA (N = 15), SD (N = 27), LPA (N = 13), and PPA not otherwise specified (PPA-nos) (N = 20). Their mean age was 66.4 $\pm$ 9.6 (ranging from 32 to 95 years), and they had a median score of 5 on the scale of Verhage, which indicates an average level of completed secondary education [17].

To validate the results found in this study, an external dataset was used. This consisted of a large cohort from the National Alzheimer's Coordinating Centre (NACC) in the United States, included from memory clinics in the United States [18]. This data comprised 1375 subjects with an AD (N=1304), behavior variant of FTD (bvFTD, N = 38), or PPA (N = 34) diagno-

sis. The average age of the subjects was 75.1 $\pm$ 9.4 years (ranging from 35 to 100 years), and 54.5% of the subjects were female. The Clinical Dementia Rating (CDR) score was used to analyse disease progression in subjects of the NACC dataset. The CDR score is a measure for disease progression on a scale from 0 to 3: no impairment (CDR = 0), questionable impairment (CDR = 0.5), mild impairment (CDR = 1), moderate impairment (CDR = 2), and severe impairment (CDR = 3).

Additionally, a subset of the data from the Rotterdam Study was used for data normalisation. For this purpose only cognitively normal subjects were considered (N = 11728) [19]. Their mean age was 64.7 $\pm$ 9.8 (ranging from 45 to 100 years) and 54.5% was female.

Characteristics of the datasets used in this study can be found in Table 1 in Appendix A.

2.2 MRI Data

T1 weighted images were acquired from all subjects included in this study, in the period of 2011 to 2021. Neuropsychological test scores were available for a subset of the subjects. The MRI scans were acquired as part of routine medical examinations. White matter, grey matter, and cerebrospinal fluid were automatically segmented from the MRI images with FreeSurfer (version 6.0). Grey matter volumes (GMVs) from 80 cortical and subcortical regions and 4 ventricle volumes were included.

Subsequently, these volumes were normalized for intracranial volume (ICV), by calculating the volume fraction through

$$V_{ICVadjusted} = \frac{V_{region}}{ICV}. \quad (1)$$

Thereafter, these volume fractions were corrected for age and sex. To establish a reference for age and sex, a quantile regression model was generated using data from cognitively normal subjects in the Rotterdam Study. For each age group, the percentiles of their volume fractions were calculated and combined them to form a second order regression curve (Fig. 2). This calculation was performed separately for males and females. Subsequently, the ICV-adjusted volumes from study subjects were individually fitted to these quantile regression lines. This process yielded a percentile value for every brain region of each subject. Essentially, these percentile values represented the degree of deviation of the GMVs with respect to the cognitively normal population at that age, and will be referred to as grey matter percentile (GMP).

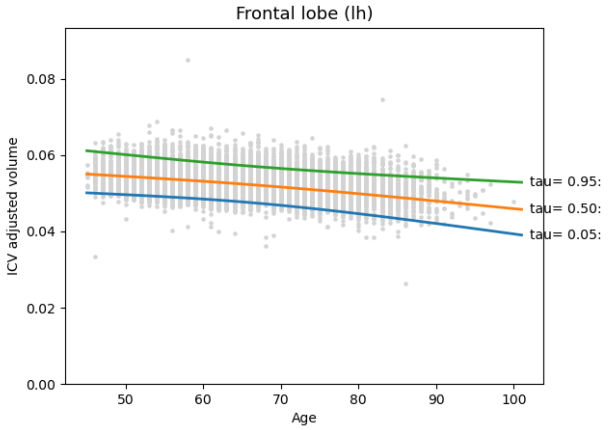


Figure 2: *Quantile regression plot of the frontal lobe in the left hemisphere in men, including tau-value (percentile) trajectories. Age is on the x-axis and the ICV-adjusted volume at the y-axis. Tau = 0.5 is the average of the cognitively normal subjects.*

2.3 Neuropsychological Test Scores

Neuropsychological tests were conducted inside the Erasmus MC as part of the standard clinical work-up. These tests were categorized into five cognitive domains: visuospatial functioning, language, memory/learning, processing speed and attention/executive functioning. Visuospatial functioning is the ability to make a visual representation of an object, or to identify an object and their location [4]. Language skills were assessed through tasks requiring subjects to name objects. Memory refers to the process of remembering and recalling information [20], while processing speed indicates the time it takes to process or respond to the information [21]. Finally, executive functioning is a broad term that contains multiple cognitive aspects under which concentration, planning and organisation, and cognitive flexibility [22]. These cognitive domains and their tests are outlined in Table 2.

The raw neuropsychological test scores were first adjusted for age, sex and level of education (Verhage scale for education [17]), using a general linear model based on the study cohort itself. Subsequently, the adjusted scores were transformed into z-scores and the resulting scores were combined (mean) per domain, according to the grouping in Table 2 in Appendix C. After the combination the scores, the resulting scores (no longer z-scores) were again z-transformed.

2.4 Cluster Analysis

The cluster analysis was done in Python using hierarchical clustering from Scikit-learn (version 1.4.1)[23]. Hierarchical clustering is a technique that clusters data points by splitting or merging the points based on similarity. This similarity was measured by the

Euclidean distance. The clustering process started with each datapoint as an individual cluster. At each step in the clustering, the two clusters with the smallest distance were merged into a singular cluster. This process continued until one large cluster was formed. For the computation of distances involving multiple points within a cluster, Wards method was employed. Wards method relies on minimizing the sum of squared distances to the center point of the newly formed cluster. Ward is the most advanced and most used, since it takes into account the variance within clusters.

The number of clusters (K) was computed through performance metrics across a range from 2 to 10 clusters. These performance metrics included the Silhouette index (SI) [24], Davies-Bouldin index (DB) [25], GAP-statistics [26], and Dunn index (DI) [27]. The optimal number of clusters was determined by either the minimal or maximal value of each metric. The SI, DI, and DB are all measures of intra- and inter-cluster distances. For each metric, these distances were measured differently, yet for all metrics the ratio of the measures of intra- and inter-cluster distances is a measure of fit of the obtained clustering. Finally, GAP reflects the cluster tendency of the data compared to random data (uniform distribution). For all metrics, the most prevalent optimal cluster number was selected for further analysis. Lastly, internal validation encompassed the inspection of the silhouette plot. The SI value for each subject was plotted individually onto the range of -1 to 1. This SI score for each subject provides a measure of fit to their cluster. A SI score of 1 indicates a perfect fit in the cluster, while a SI score below 0 suggests a better fit in another cluster.

2.5 Main and subanalyses

Our main analysis focused on the clustering of brain lobe and hippocampus GMPs. Therefore, regional GMVs were combined into brain lobe volumes for each hemisphere by summing the GMV of the constituting regions (see Appendix B). The hippocampus volume was also separately considered for both hemispheres. The GMPs of these brain regions were used as input for the clustering. This analysis is referred to as the clustering of global GMPs (GLOB). The quantile regression plots for all lobes and the hippocampus can be found in Appendix G.

Secondly, principal component analysis (PCA) was employed for dimension reduction of the data. PCA transformed the original feature space into a new, lower-dimensional space while preserving the directions of the largest variance in the data. These principal components (PCs) are mutually orthogonal and are ordered by the amount of variance. A subset of the PCs were selected, explaining the majority of the variance. In the first clustering approach this subset was clustered, referring to 'PCA1'. In a second clustering, the first and largest PC was left out of the input, named 'PCA2'. This approach was taken as we expected that the first PC might merely contain physiological atrophy.

Lastly, the NACC dataset was used for external validation. The data from NACC was clustered in two ways. First, the data from NACC was clustered according to the same pipeline as in GLOB, resulting in 'NACC' clusters. Secondly, the datapoints from NACC were assigned to the clusters from ACE (GLOB). This assignment was done by the k-nearest neighbor approach, which considered the 10 nearest neighbours of the GLOB clusters to cluster each subject from NACC. The degree of alignment was evaluated using the Jaccard index [28] and the Adjusted Rand Index [29]. For both these measures of alignment, the range was from 0 to 1, with 1 indicating a perfect alignment.

2.6 Statistical Analysis

Statistical analyses included the t-test or one-way ANOVA (parametrical), and Mann-Whitney U or Kruskal-Wallis tests (non-parametrical) for the cognitive domains. Only the Mann-Whitney U and Kruskal-Wallis test were used for the imaging features and age, because the assumptions of normal distribution and equal variances were not met. Post-hoc tests for numerical variables were done using the Turkeys HSD (parametrical) and Dunn's test (non-parametrical). The categorical variables (sex, CDR score and education) were tested with the Chi-squared test, including the post-hoc tests. All post-hoc tests included a correction for multiple comparisons.

3 Results

3.1 GMPs per clinical diagnosis

The GMP scores of all diagnoses and subdiagnoses present in the ACE dataset are depicted in Fig. 15 in Appendix D.

3.2 Clusters based on global GMPs (GLOB)

Clusters were generated based on the GMPs of the brain lobes and the hippocampi of both hemispheres. Two out of the four performance metrics indicated 2 as the optimal number of clusters (see Fig. 11 in Appendix D). Furthermore, to investigate the specifics

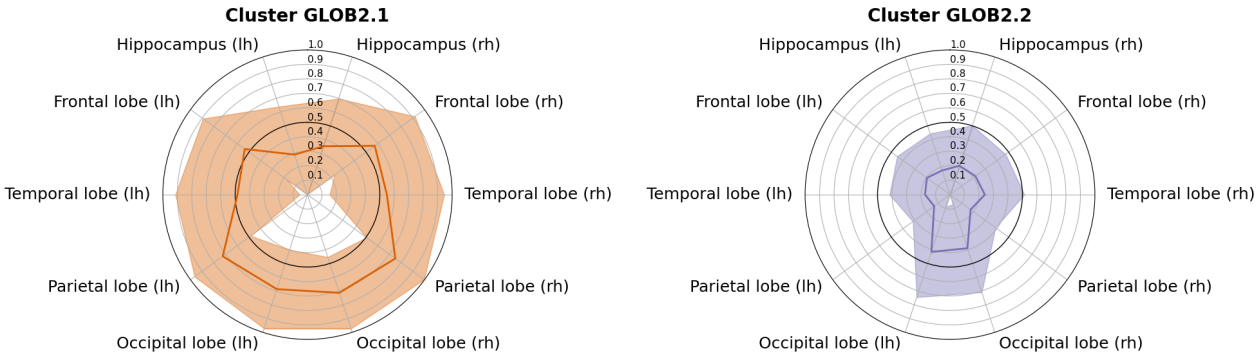
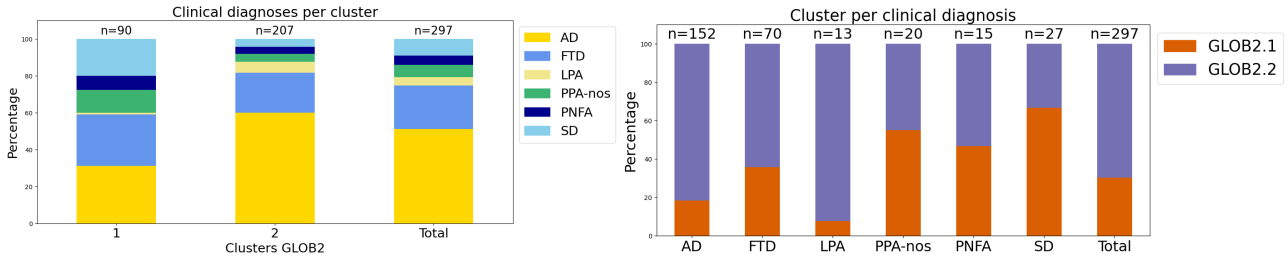


Figure 3: *Grey matter percentiles (GMPs) of the input features for the left hemisphere (lh) and right hemisphere (rh) for the two clusters of GLOB2. The mean GMP is for each region depicted by the colored line, with a concomitant shaded standard deviation. The axes portray the GMPs measured from cognitively normal participants in the Rotterdam Study, with the black 0.5 line representing the average. A GMP value below 0.5 indicates a smaller than average brain volume in a certain region, while a value above 0.5 indicates a larger than average brain volume.*



(a) *The percentage of the clinical diagnoses within each cluster.*

(b) *The percentage of clusters within each clinical diagnosis.*

Figure 4: *Distribution of the clinical diagnosis per cluster of GLOB2.*

of the dataset more deeply, a four-cluster solution was explored, emerging as the second-best option according to the Silhouette Index (SI = 0.24).

Two cluster solution (GLOB2) The radar plots in Fig. 3 compare the brain regions for the two clusters. Cluster GLOB2.1 (N = 90) showed GMPs in the range of 0.48 to 0.57 for the frontal and temporal lobe, and between 0.69 and 0.75 for the parietal and occipital lobe. The hippocampal GMPs were lower (0.29 and 0.35 for left and right respectively), indicating relatively high hippocampal atrophy compared to other regions. In contrast, cluster GLOB2.2 (N = 207) showed more global atrophy, with a relatively spared occipital lobe compared to other regions. The GMPs of all 10 brain regions are significantly different between the two clusters ($p < 0.01$). Regarding clinical diagnoses (Fig. 4a), the percentage of FTD and FTD-related language variants was higher in cluster GLOB2.1 (41.1%, n = 37) than in cluster GLOB 2.2 (19.3%, n = 40). The opposite trend was observed for AD, with 31.1% and 59.9% for GLOB2.1 and GLOB2.2 respectively.

The clusters were statistically compared for differences in demographics (age, sex, education), GMPs of the global brain regions, and cognitive performance, see Table 5 in Appendix D. For the cognitive domains, cluster GLOB2.2 demonstrated lower scores across all domains ($p < 0.001$) except for language ($p = 0.392$) (Fig. 12 in Appendix D). Notably, no significant differences were observed in age ($p = 0.446$), sex ($p = 0.119$) or education ($p = 0.854$) between the two groups.

Four cluster solution (GLOB4) Fig. 5 depicts the GMP scores from the global brain regions for the four cluster solution.

Cluster GLOB4.1 (N = 167) exhibited low GMPs for all brain regions, except for the relatively spared occipital lobe (0.37 and 0.36 for left and right respectively) compared to the other regions (0.07 - 0.20). Cluster GLOB4.2 (N = 46) displayed low GMPs in the temporal lobe (0.25 and 0.09 for left and right respectively) and hippocampus (0.07 and 0.17 for left and right respectively). The GMPs of cluster GLOB4.3 (N = 44) showed higher values on all brain regions compared to the average of cognitively normal population (GMP > 0.5). Cluster GLOB4.4 (N = 40) had a GMP under 0.5 for the parietal, temporal and frontal lobe, while the hippocampus and the left occipital lobe were above GMP of 0.5.

The GLOB4 clusters were statistically compared

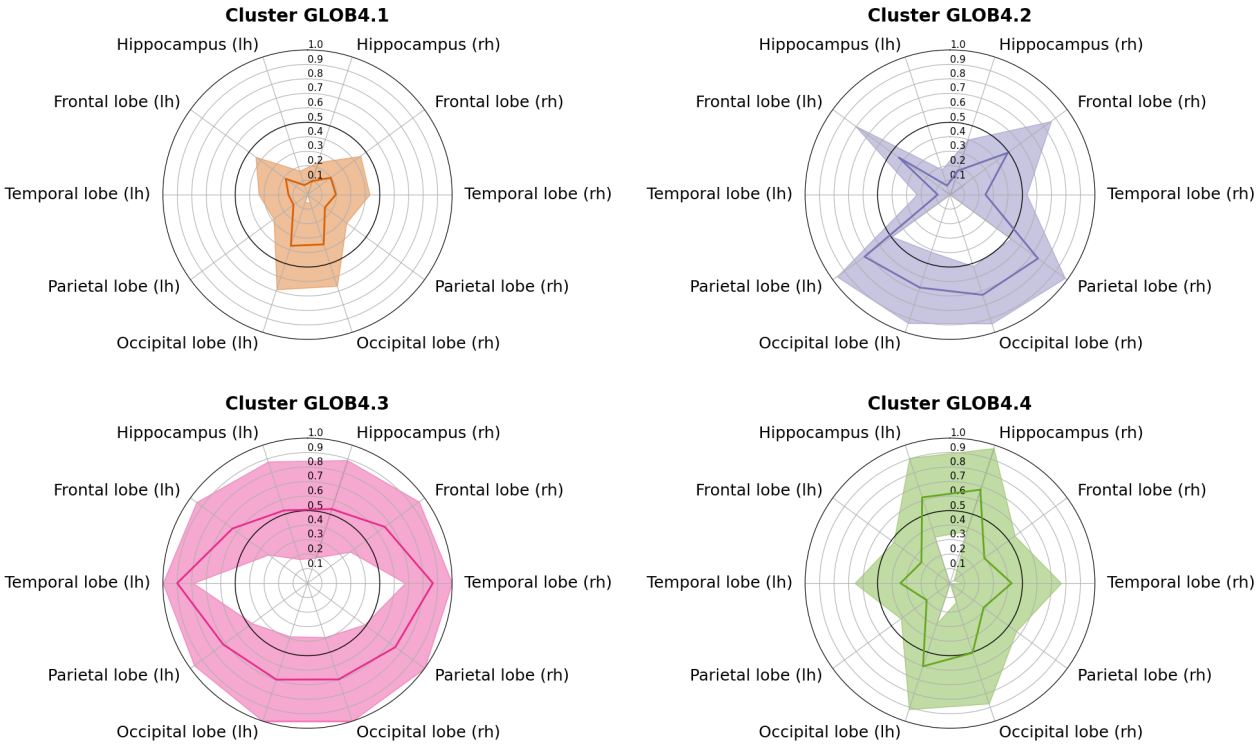


Figure 5: *GMPs of the input features for the left hemisphere (lh) and right hemisphere (rh) for the two clusters of GLOB4. The mean GMP is for each region depicted by the colored line, with a concomitant shaded standard deviation. The axes portray the GMPs measured from cognitively normal participants in the Rotterdam Study, with the black 0.5 line representing the average. A GMP value below 0.5 indicates a smaller than average brain volume in a certain region, while a value above 0.5 indicates a larger than average brain volume.*

for differences in demographics (age, sex, education), GMPs of the global brain regions, and cognitive performance, see Table 6 in Appendix D. Analysis of demographics indicated no statistically significant differences among the clusters in terms of age (p = 0.095), sex (p = 0.390), or education (p = 0.941). Results of the cognitive domain analysis are depicted in Fig. 6. Cluster GLOB4.2 had the highest performance for visuospatial functioning, processing speed and attention/executive functioning, especially compared to cluster GLOB4.1 and GLOB4.4. Additionally, cluster GLOB4.3 performed better on language compared to GLOB4.1 and GLOB4.2. In the memory/learning domain, cluster GLOB4.2 and GLOB4.3 scored higher than cluster GLOB4.1.

The missing z-scores for the cognitive domains for each cluster are reported in Table 3 (GLOB2), and Table 4 (GLOB4) in Appendix C.

3.3 Clusters based on principal components (PCA)

The GMPs of the brain regions (cortical, subcortical and ventricles) from the FreeSurfer segmentation were reduced to 18 principal components (PCs) using PCA. The 18 PCs explain 70% of the variance, capturing most of the variance while minimizing the dimensionality. This conclusion is based on a visual inspection of the scree plot (Fig. 16 in Appendix E). The clustering of these 18 PCs (PCA1) resulted into two clusters (Fig. 8). The performance metrics (Fig. 17a in Appendix E) were not unanimous. We decided to apply two clusters to make the analysis comparable to the one in the previous section. This was supported by visual inspection of the performance metrics in which we found that 2 clusters was the best or second best option according to 3 out of 4 performance metrics. Cluster PCA1.1 showed average GMP values around 0.5 (0.37 - 0.61). Cluster PCA1.2 had lower GMP values for all regions (0.09 - 0.17) except for the occipital lobe (0.40 for left and right).

In a second analysis (PCA2), the first principal component was left out (Fig. 9). The GMP values of cluster PCA2.1 were all between 0.2 and 0.4 which was similar for PCA2.2, except for the occipital lobe (0.67 and 0.68 for left and right respectively).

The 10 brain regions with the highest weighting in the first and second principal component are listed in Appendix E. Additionally, the diagnosis distribution for PCA1 and PCA2 are shown in Fig. 18.

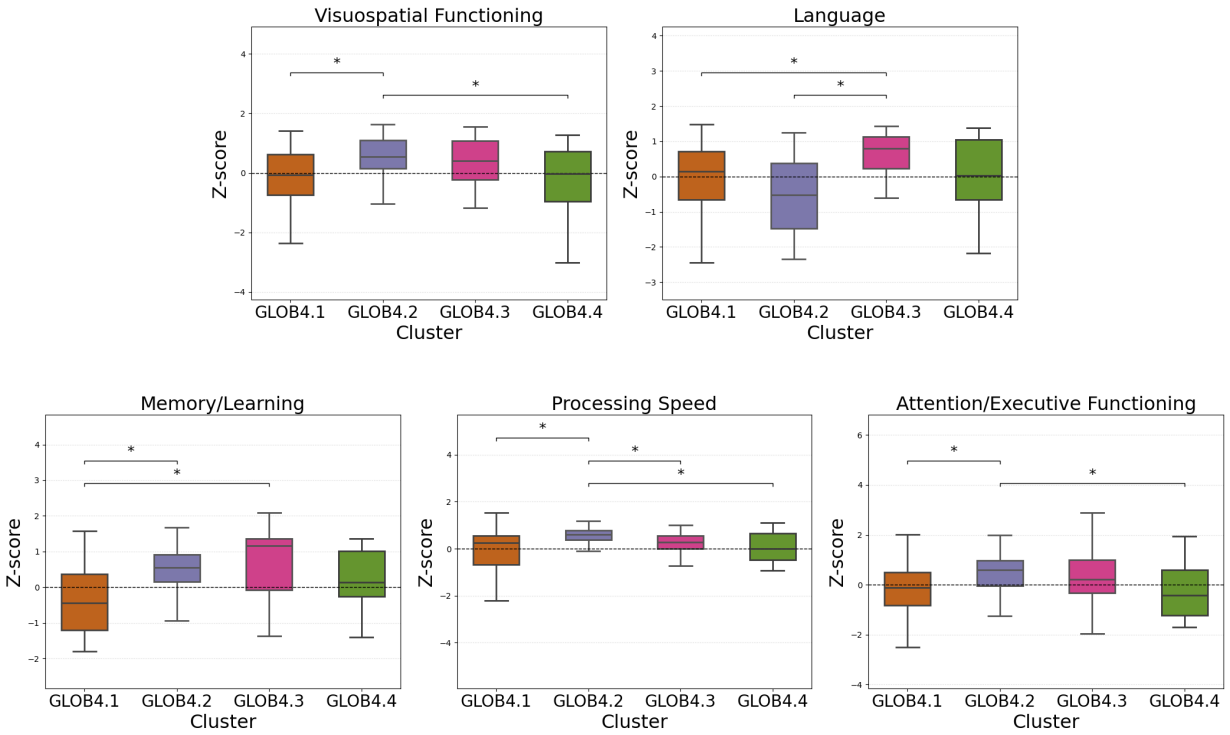
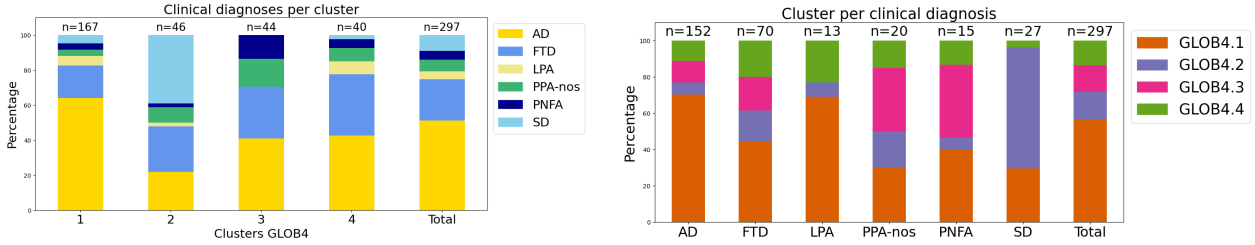


Figure 6: *Comparison of Z-scores for each cognitive domain for the four clusters in GLOB4. A higher score indicates a better performance for all cognitive domains. It is important to note that the z-score indicates performance relative to the dataset and should not be directly compared to subjects with normal cognitive function. An equal z-score suggests similar performance between the two groups without specifying the degree of impairment. Significant differences between clusters are denoted by *.*



(a) *The percentage of the clinical diagnoses within each cluster.*

(b) *The percentage of clusters within each clinical diagnosis.*

Figure 7: *The distribution of clinical diagnoses over the clusters of GLOB4.*

3.4 External validation using NACC data

First, the data from NACC was clustered according to the same analysis pipeline as done in ACE.

Two cluster solution (NACC2) In the analysis conducted on the NACC data, two distinct clusters emerged as optimal based on the performance metrics (Fig. 19 in Appendix F). Cluster NACC2.1 (Fig. 20 in Appendix F) had GMPs below the average of 0.5 in the hippocampus (0.31 and 0.35 for left and right respectively), around average in the parietal lobe and frontal lobe (0.45 - 0.50), and above average for the temporal and occipital lobe (0.61 - 0.66). In cluster NACC2.2, all GMPs were below average, although the occipital lobe seemed less affected (0.27 and 0.28 for left and right respectively) compared to other regions (0.08 - 0.11). Furthermore, cluster NACC2.2 comprised individuals who were younger (mean age: 74.4 $\pm$ 9.2 years) compared to cluster NACC2.1 (mean age: 76.4 $\pm$ 9.6 years, $p < 0.001$), while there was no significant difference in sex distribution between the clusters. Notably, cluster NACC2.2 demonstrated higher CDR scores (Table 10) compared to cluster 1 ($p < 0.001$), reflecting greater cognitive impairment.

Four clusters solution (NACC4) In the four cluster solution (Fig. 21 in Appendix F), cluster NACC4.1 exhibited higher GMPs in the temporal and occipital lobe (0.51 - 0.64), than in the hippocampus (0.18 and 0.22 for left and right respectively) and other lobes

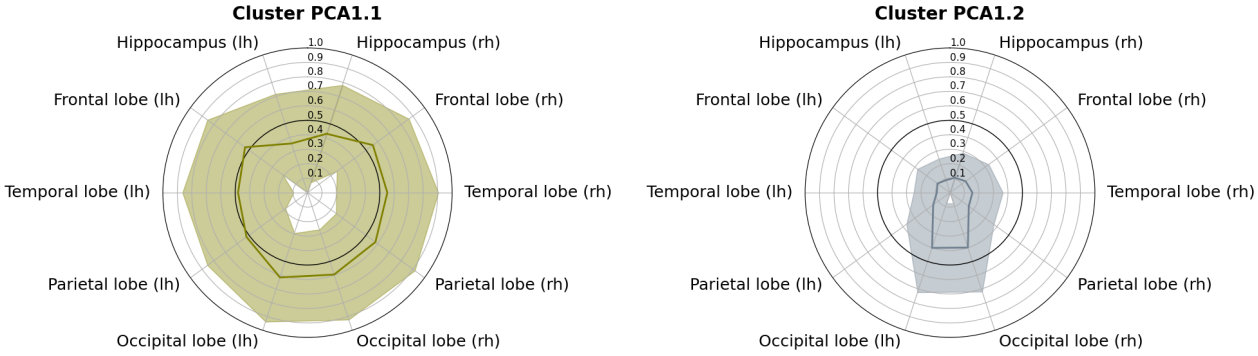


Figure 8: *GMPs of the global brain regions to compare GMP patterns the left (lh) and right hemisphere (rh) of PCA1 to the GLOB clusters. The mean GMP is for each region depicted by the colored line, with a concomitant shaded standard deviation. The axes portray the GMPs measured from cognitively normal participants in the Rotterdam Study (RS), with the black 0.5 line representing the average. A GMP value below 0.5 indicates a smaller than average brain volume in a certain region, while a value above 0.5 indicates a larger than average brain volume.*

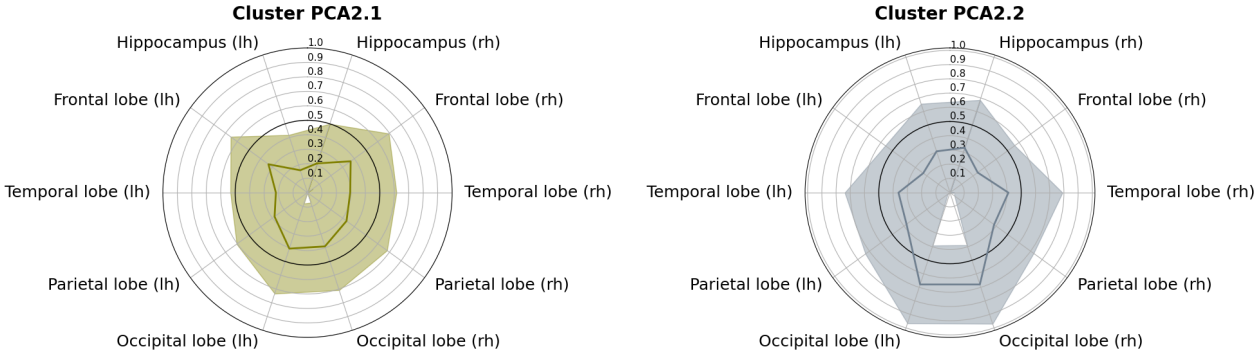


Figure 9: *GMPs of the global brain regions to compare GMP patterns the left (lh) and right hemisphere (rh) of PCA2 to the GLOB clusters. The mean GMP is for each region depicted by the colored line, with a concomitant shaded standard deviation. The axes portray the GMPs measured from cognitively normal participants in the Rotterdam Study (RS), with the black 0.5 line representing the average. A GMP value below 0.5 indicates a smaller than average brain volume in a certain region, while a value above 0.5 indicates a larger than average brain volume.*

(0.38 - 0.43). In cluster NACC4.2, the GMP value of the occipital lobe was average (0.5), but the other regions exhibited low GMP values (0.11 - 0.18). Cluster NACC4.3, had only GMPs of above 0.64, with the highest GMPs in the temporal lobe (0.88 and 0.89 for left and right respectively). Cluster NACC4.4 showed low GMPs in the range of 0.04 up to 0.14.

Regarding demographics, there were no significant differences in sex distribution among the clusters. However, differences were observed in age and CDR scores. Clusters NACC4.2 and NACC4.4 were significantly younger (mean age of 74.2 $\pm$ 9.3 and 74.5 $\pm$ 9.2 respectively) compared to subjects in cluster NACC4.3 (mean age of 77.2 $\pm$ 10.2, $p < 0.01$). The subjects in cluster NACC4.2 had a higher average CDR score compared to cluster NACC4.1 ($p < 0.01$), while cluster NACC4.4 had a higher CDR score than both cluster NACC4.1 and cluster NACC4.3 ($p < 0.001$ for both comparisons).

NACC subjects in ACE clusters Finally, the subjects from NACC were assigned to the clusters in ACE (GLOB2 and GLOB4 clusters). For the two-cluster scenario, most of the NACC subject were classified into cluster GLOB2.2 (N = 1114), while the rest was classified into cluster GLOB2.1 (N = 261). The comparison of the cluster assignment based on ACE and the NACC clustering yielded an ARI score of 0.44 and a Jaccard index of 0.80. When considering four clusters, most NACC subjects were clustered into GLOB4.1 (N = 1079), while the other clusters were less represented in NACC: GLOB4.2 (N = 57), GLOB4.3 (N = 174), and GLOB4.4 (N = 65). The comparison of this cluster assignment to the NACC4 clusters, resulted in an ARI score of 0.20 and a Jaccard index of 0.10.

4 Discussion

In this study, unsupervised cluster analysis was used to explore the heterogeneity within dementia. Our original aim was to study potential hidden patterns in the data that could provide insights into disease subtypes and the overlap in atrophy patterns across different forms of dementia. While previous research has predominantly focused on identifying clusters within either AD or FTD, this study took a broader approach by incorporating both etiological diagnoses. Our study aimed to examine the relation between the atrophy patterns of the identified clusters, cognitive performance, and clinical diagnoses.

Unlike conventional approaches such as z-scores or general linear models [30][31], this study employed a quantile regression model based on cognitively normal individuals to correct the MRI data for age and sex. This approach resulted in an informative measure and visualisation of the atrophy patterns within dementia. Additionally, the clusters were externally validated using an external dataset, strengthening the generalisability of the study.

4.1 Atrophy patterns

The atrophy patterns in the two cluster solution appeared to represent two stages of disease progression rather than specific disease manifestations. Cluster GLOB2.1 displays minimal atrophy at the given age, whereas cluster GLOB2.2 exhibits significant atrophy across various regions, with the exception of the occipital lobe.

Cluster GLOB2.1 exhibited greater hippocampal atrophy compared to the average at that age, which could signify an early stage of both AD [7] and FTD [32]. Furthermore, the atrophy pattern in cluster GLOB2.2 shows occipital lobe sparing, which has been associated with a further disease progression of AD [15] and FTD [32]. The association with two disease stages is confirmed by the cognitive domain scores (Table 5 in Appendix D). Cluster GLOB2.2, characterised by more pronounced atrophy, came with consistently lower scores than cluster GLOB2.1 in all cognitive domains except for language. This exception could be explained by the higher percentage of language variants in cluster GLOB2.1 (n = 37, 41.1%) compared to cluster GLOB2.2 (n = 38, 18.3%). Despite potentially being in an earlier disease stage, subjects in cluster GLOB2.1 already exhibit language difficulties at the time of diagnosis.

Interestingly, AD is more prevalent in cluster GLOB2.2 (n = 124, 60.0%) than in cluster GLOB2.1 (n = 28, 31.1%), which may be indicative of a more advanced disease stage in AD. AD generally has a shorter referral time compared to FTD [33][34]. Our analysis did not reveal a higher prevalence of FTD in the second cluster however, as might have been expected based on referral times. This may be attributed to ACE being a tertiary center, i.e., the ACE dataset may not necessarily contain subjects with AD who have short referral times.

Similar to the clusters in ACE, this potential clustering of disease progression stage was also observed in the cluster analysis of the NACC dataset. This further corroborates the findings in GLOB2. The higher CDR scores (Table 10 in Appendix F) in the atrophy cluster (NACC2.2) support the observation of atrophy patterns that reflect disease progression stages. Furthermore, the substantial overlap (Jaccard = 0.8) of NACC data with the GLOB2 clusters reinforces this observation.

The PCA also supports the hypothesis demonstrating that the most significant variance in the data distinguishes between the degree of atrophy. After omitting of the first principal component, the clusters no longer exhibited clear distinctions based merely on the level of atrophy (Fig. 9). Furthermore, the clinical diagnosis were more unequally distributed over the clusters of PCA2. As such, the other principal components could be informative of disease manifestations. The features with the highest weighting in the first principal components indicate that disease progression differences may be more pronounced in cortical brain regions, while discriminative features related to the disease could be located more subcortically. The regions are shown in Table 8 in Appendix E.

Exploring a four-cluster solution (GLOB4) provided additional insight into the atrophy patterns, potentially offering a more nuanced understanding beyond disease stage alone. In Fig. 5, it is unlikely that clusters GLOB4.2 and GLOB4.4 are two stages of an atrophy pattern, as the atrophy seen in GLOB4.2 in the temporal lobe and the hippocampus is not expected to progress over time to atrophy in the parietal and frontal lobe of cluster GLOB4.4. At the same time, though, GLOB4.3 could present an early stage of the disease with GMPs slightly above the average of 0.5. Furthermore, GLOB4.1 could reflect a later stage presenting more pronounced atrophy in the GMP pattern.

Cluster GLOB4.2 exhibited higher scores in processing speed, visuospatial function, and attention/executive function compared to Cluster GLOB4.4 (Fig. 6). This discrepancy may be attributed to differences in volume across brain regions, with cluster GLOB4.2 showing lower GMPs in the temporal lobe and hippocampus, while cluster GLOB4.4 exhibited lower GMPs in the parietal lobe and right occipital lobe. The performance difference on vi-

suospatial functioning could be associated with the difference in occipital GMPs [22].

Even though cluster GLOB4.2 displayed more pronounced atrophy in the temporal lobe and hippocampus, the performance across all cognitive domains, except for language, was notably strong compared to other clusters. The observed language impairment aligns with the identified temporal lobe atrophy and the high prevalence of language variants within this cluster (Fig. 7). Specifically, SD is prominently presented. This clustering pattern of SD is consistent with findings from another clustering study involving individuals with PPA [35]. A radar plot illustrating the atrophy pattern of SD is presented in Fig. 15 in Appendix D. In this figure a distinct atrophy pattern of SD is observable compared to other etiological diagnoses. In particular, the pattern is marked by hippocampal and temporal lobe atrophy. This is in line with existing literature, suggesting that hippocampal atrophy in SD may surpass that observed in AD [36][37]. Previous studies [36][38][39] have consistently reported bilateral asymmetrical temporal lobe atrophy as a characteristic feature of SD. Although the variation between the right (GMP = 0.26) and left (GMP = 0.22) temporal lobe is minimal in SD, it is interesting to note that in cluster GLOB4.2 the difference between right (GMP = 0.25) and left (GMP = 0.09) is more pronounced.

Another notable discovery was that cluster GLOB4.2 exhibited better performance in memory/learning compared to cluster GLOB4.1, despite similar levels of hippocampal atrophy. The hippocampus is known for its role in the brain's memory network [40]. Part of the rationale behind this observation could be the cognitive profile associated with SD. Individuals with SD often show a specific pattern of memory deficits, wherein recent event memory remains relatively preserved [36]. However, we should be cautious with drawing conclusions from these observations since there is a large number of missing values in the cognitive performance scores, see Table 4.

The clusters identified in this study align with clusters reported in literature including exclusively individuals with AD or FTD. Cluster GLOB4.2 exhibits an atrophy pattern resembling typical AD and limbic predominant clusters, characterized by hippocampal atrophy with progression to the temporal lobe and atrophy of the entorhinal cortex progressing in the temporal lobe respectively [7][16]. Simultaneously, the atrophy pattern of GLOB4.2 could resemble the temporal dominant cluster identified in bvFTD [11][39][41]. Cluster GLOB4.3 exhibit minimal atrophy, suggesting it may represent an earlier stage of the disease progression. Another interpretation of minimal atrophy is that cluster GLOB4.3 may partly comprise the subcortical atrophy cluster, observed in AD [11] and FTD [39][41], or the minimal atrophy cluster, observed in AD [7]. Cluster GLOB4.4 demonstrates a similar atrophy pattern as the hippocampal sparing AD cluster, presenting cortical atrophy while the hippocampus is relatively spared [42]. This pattern of cortical atrophy is akin to the fronto-tempo-parietal cluster observed in previous clustering studies of individuals with bvFTD [41][9]. Most subjects (N = 167) however, were clustered into cluster GLOB4.1, which appears to identify a later disease stage. This observation suggests that in advanced disease stages, disease specific atrophy patterns may become less distinct due to generalized atrophy.

The considerable variation in standard deviations of the GMPs (see Fig. 15) confirms the known heterogeneity in each etiological diagnosis as well as the disease overlap. This overlap of disease presentations is also evident in the silhouette scores, which measures the clustering fit. As illustrated in Fig. 13 in Appendix D, some individuals are assigned silhouette scores below zero, suggesting that they might be better assigned to another cluster. This discrepancy arises from the fact that the SI value is based on the distance from each datapoint to the cluster centers, while hierarchical clustering operates based on the variance within clusters. Consequently, individuals with silhouette scores around zero might actually be located between clusters.

4.2 Limitations

One of the limitations in this study is the difficulty of accurately identifying the etiological diagnosis or diagnoses.

Secondly, the presence of vascular pathology in dementia was not addressed in this study. Previous research has seen frequent co-occurrence of pathologies, especially AD and vascular pathologies [5][42][43][44]. This co-existence of pathologies, could be at the basis of the observed heterogeneity. The inclusion of vascular pathologies in the identification of subgroups, could provide a more accurate and and more detailed outcome of the clustering.

For this study, multiple test scores were combined to construct cognitive domains. However, it is important to acknowledge that some nuances may have been lost in the averaging of z-scores. The cognitive domains were derived from the available scores from the neuropsychological examination. It is worth noting that not all tests were applied in every subject, leading to missing scores, as indicated in Table 3 and Table 4 in Appendix C. The memory/learning domain were particularly underrepresented, partly due

to the tendency to omit the Boston Naming Test in cases of language variants, since language difficulties can interfere with speaking memory tasks. This could have influenced the results in particular in relation to cluster GLOB4.2 and GLOB4.1 on memory/learning.

Finally, a notable limitation stems from the predominantly AD-diagnosed subjects in the available NACC data used for validation. Diversifying the external validation set might give a more accurate view of the generalisability of the clustering.

4.3 Future research

Future research could focus on alternative feature selection. Using only small regional volumes could introduce potential noise into the dataset, while the global volumes of the brain lobe could result in discarding detailed information. In a subanalysis of this study, PCA was used to select most informative features based on variance. The second analysis, where the first PC was left out, suggests that there is subcortical involvement in disease-specific atrophy patterns. PCA or other feature selection methods could identify the most informative and discriminative features. The clustering of these features may yield valuable insights for future studies.

Additionally, certain co-pathologies exist in dementia, as described in previous literature [1][5][42][44]. For future research, it would be valuable to further explore co-pathologies, especially the overlap of dementia with vascular pathology and its effect on disease presentation and progression.

5 Conclusion

In this study, we explored atrophy patterns within AD and FTD by employing cluster analysis, and we analyzed their relation to cognition and clinical diagnosis. The analysis appears to reveal distinct atrophy patterns corresponding to different stages of disease progression rather than specific disease manifestations. Validation of these findings through external datasets further strengthened the robustness of the study, confirming the consistency of disease stage clusters. Additionally, specific disease patterns were found in SD, showing a high contribution to one cluster, presenting with mainly hippocampal and temporal atrophy. The clusters and their atrophy patterns provide not only an overview of heterogeneity within AD and FTD, but also information about potential disease overlap.

References

[1] J. Hodler, R. A. Kubik-Huch, G. K. V. Schulthess, and G. K. V. S. Editors. *Diseases of the Brain, Head and Neck, Spine 2020–2023*. 2020, pp. 135–146. URL: http://www.springer.com/series/15856.

[2] S. Baratono and D. Press. *What Are the Key Diagnostic Cognitive Impairment and Dementia Subtypes and How to Integrate all of the Diagnostic Data to Establish a Diagnosis?* Feb. 2023. DOI: 10.1016/j.cger.2022.08.002.

[3] K. Poulakis, J. B. Pereira, P. Mecocci, B. Vellas, M. Tsolaki, I. Kłoszewska, H. Soininen, S. Lovestone, A. Simmons, L.-O. Wahlund, and E. Westman. "Heterogeneous patterns of brain atrophy in Alzheimer's disease". In: *Neurobiology of Aging* 65 (May 2018), pp. 98–108. ISSN: 01974580. DOI: 10.1016/j.neurobiolaging.2018.01.009.

[4] N. B. M. Quental, S. M. D. Brucki, and O. F. A. Bueno. "Visuospatial function in early Alzheimer's disease: Preliminary study". In: *Dementia & Neuropsychologia* 3 (3 Sept. 2009), pp. 234–240. ISSN: 1980-5764. DOI: 10.1590/S1980-57642009DN30300010.

[5] K. A. Jellinger and J. Attems. "Neuropathological evaluation of mixed dementia". In: *Journal of the Neurological Sciences* 257 (1-2 June 2007), pp. 80–87. ISSN: 0022510X. DOI: 10.1016/j.jns.2007.01.045.

[6] K. Poulakis, D. Ferreira, J. B. Pereira, Ö. Smedby, P. Vemuri, and E. Westman. "Fully bayesian longitudinal unsupervised learning for the assessment and visualization of AD heterogeneity and progression". In: *Aging* 12 (13 July 2020), pp. 12622–12647. ISSN: 1945-4589. DOI: 10.18632/aging.103623. URL: https://www.aging-us.com/lookup/doi/10.18632/aging.103623.

[7] D. Ferreira, A. Nordberg, and E. Westman. "Biological subtypes of Alzheimer disease". In: *Neurology* 94 (10 Mar. 2020), pp. 436–448. ISSN: 0028-3878. DOI: 10.1212/WNL.0000000000009058.

[8] M. ten Kate, E. Dicks, P. J. Visser, W. M. van der Flier, C. E. Teunissen, F. Barkhof, P. Scheltens, and B. M. Tijms. "Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline". In: *Brain* 141 (12 Dec. 2018), pp. 3443–3456. ISSN: 0006-8950. DOI: 10.1093/brain/awy264. URL: https://academic.oup.com/brain/article/141/12/3443/5142624.

[9] J. L. Whitwell, S. A. Przybelski, S. D. Weigand, R. J. Ivnik, P. Vemuri, J. L. Gunter, M. L. Senjem, M. M. Shiung, B. F. Boeve, D. S. Knopman, J. E. Parisi, D. W. Dickson, R. C. Petersen, C. R. Jack, and K. A. Josephs. "Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: A cluster analysis study". In: *Brain* 132 (11 2009), pp. 2932–2946. ISSN: 14602156. DOI: 10.1093/brain/awp232.

[10] M. Bocchetta et al. "From brain volumes to subgroup classification in genetic mutation carriers for frontotemporal dementia: A cluster analysis in the GENFI study". In: *Alzheimer's and Dementia* 17 (S5 Dec. 2021). ISSN: 1552-5260. DOI: 10.1002/alz.052189. URL: https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.052189.

[11] A. L. Young, R. V. Marinescu, N. P. Oxtoby, M. Bocchetta, K. Yong, N. C. Firth, D. M. Cash, D. L. Thomas, K. M. Dick, J. Cardoso, et al. "Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference". In: *Nature communications* 9.1 (2018), p. 4273.

[12] A. Inguanzo et al. "Dementia with Lewy bodies subtypes identified by cluster analysis on structural MRI". In: *Alzheimer's and Dementia* 17 (S4 Dec. 2021). ISSN: 1552-5260. DOI: 10.1002/alz.053573.

[13] A. Inguanzo et al. "MRI data-driven clustering reveals different subtypes of Dementia with Lewy bodies". In: *npj Parkinson's Disease* 9 (1 Dec. 2023). ISSN: 23738057. DOI: 10.1038/s41531-023-00448-6.

[14] C. Zheng, W. Zhao, Z. Yang, D. Tang, M. Feng, S. Guo, A. D. N. Initiative, et al. "Resolving heterogeneity in Alzheimer's disease based on individualized structural covariance network". In: *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 129 (2024), p. 110873.

[15] B. Zhang, L. Lin, L. Liu, X. Shen, and S. Wu. "Concordance of Alzheimer's Disease Subtypes Produced from Different Representative Morphological Measures: A Comparative Study". In: *Brain Sciences* 12 (2 Feb. 2022). ISSN: 20763425. DOI: 10.3390/brainsci12020187.

[16] K. Poulakis, J. B. Pereira, J. S. Muehlboeck, L. O. Wahlund, Ö. Smedby, G. Volpe, C. L. Masters, D. Ames, Y. Niimi, T. Iwatsubo, D. Ferreira, and E. Westman. "Multi-cohort and longitudinal Bayesian clustering study of stage and subtype in Alzheimer's disease". In: *Nature Communications* 13 (1 Dec. 2022). ISSN: 20411723. DOI: 10.1038/s41467-022-32202-6.

[17] F. Verhage. "Intelligentie en leeftijd: Onderzoek bij Nederlanders van twaalf tot zevenenzeventig jaar". In: *(No Title)* (1964).

[18] National Alzheimer's Coordinating Center. *Uniform Data Set 3*. Data file. 2019.

[19] M. A. Ikram, G. G. O. Brusselle, S. D. Murad, C. M. van Duijn, O. H. Franco, A. Goedegebure, C. C. W. Klaver, T. E. C. Nijsten, R. P. Peeters, B. H. C. Stricker, H. Tiemeier, A. G. Uitterlinden, and M. W. Vernooij. "The Rotterdam Study: 2018 update on objectives, design and main results". In: *European Journal of Epidemiology* 33.9 (2018), pp. 921–953. ISSN: 1573-7284. DOI: 10.1007/s10654-018-0422-2. URL: https://doi.org/10.1007/s10654-018-0422-2.

[20] M. D. Lezak, D. B. Howieson, E. D. Bigler, and D. Tranel. *Neuropsychological Assessment*. 5th. 2012. ISBN: 978–0–19–539552–5.

[21] V. S. Ramachandran. *Encyclopedia of human behavior*. Academic Press, 2012.

[22] A. Diamond. *Executive functions*. 2013. DOI: 10.1146/annurev-psych-113011-143750.

[23] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[24] P. J. Rousseeuw. *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*. 1987.

[25] M. Halkidi. *On Clustering Validation Techniques*. 2001.

[26] R. Tibshirani, G. Walther, and T. Hastie. "Estimating the number of clusters in a data set via the gap statistic". In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 63 (2 2001), pp. 411–423. ISSN: 13697412. DOI: 10.1111/1467-9868.00293.

[27] J. C. Dunn. "Well-separated clusters and optimal fuzzy partitions". In: *Journal of Cybernetics* 4 (1 Jan. 1974), pp. 95–104. ISSN: 00220280. DOI: 10.1080/01969727408546059.

[28] P. Jaccard. "Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines". In: *Bull Soc Vaudoise Sci Nat* 37 (1901), pp. 241–272.

[29] L. Hubert and P. Arabie. "Comparing partitions". In: *Journal of classification* 2 (1985), pp. 193–218.

[30] J. L. Whitwell, J. Xu, J. Mandrekar, B. F. Boeve, D. S. Knopman, J. E. Parisi, M. L. Senjem, D. W. Dickson, R. C. Petersen, R. Rademakers, et al. "Frontal asymmetry in behavioral variant frontotemporal dementia: clinicoimaging and pathogenetic correlates". In: *Neurobiology of aging* 34.2 (2013), pp. 636–639.

[31] Y. Noh, S. Jeon, J. M. Lee, S. W. Seo, G. H. Kim, H. Cho, B. S. Ye, C. W. Yoon, H. J. Kim, J. Chin, K. H. Park, K. M. Heilman, and D. L. Na. "Anatomical heterogeneity of Alzheimer disease". In: *Neurology* 83 (21 Nov. 2014), pp. 1936–1944. ISSN: 0028-3878. DOI: 10.1212/WNL.0000000000001003. URL: https://www.neurology.org/doi/10.1212/WNL.0000000000001003.

[32] A. Bejanin et al. "Longitudinal structural and metabolic changes in frontotemporal dementia". In: *Neurology* 95 (2 July 2020), E140–E154. ISSN: 1526632X. DOI: 10.1212/WNL.0000000000009760.

[33] M. Leroy et al. "Characteristics and progression of patients with frontotemporal dementia in a regional memory clinic network". In: *Alzheimer's Research and Therapy* 13 (1 Dec. 2021). ISSN: 17589193. DOI: 10.1186/s13195-020-00753-9.

[34] B. Draper, M. Cations, F. White, J. Trollor, C. Loy, H. Brodaty, P. Sachdev, P. Gonski, A. Demirkol, R. G. Cumming, and A. Withall. "Time to diagnosis in young-onset dementia and its determinants: the INSPIRED study". In: *International Journal of Geriatric Psychiatry* 31 (11 Nov. 2016), pp. 1217–1224. ISSN: 10991166. DOI: 10.1002/gps.4430.

[35] B. Taylor, M. Bocchetta, C. Shand, E. G. Todd, A. Chokesuwattanaskul, S. J. Crutch, J. D. Rohrer, J. D. Warren, C. J. Hardy, and N. P. Oxtoby. "Data-driven neuroanatomical subtype trajectories of primary progressive aphasia". In: *Alzheimer's & Dementia* 19 (S17 Dec. 2023). ISSN: 1552-5260. DOI: 10.1002/alz.076663.

[36] J. R. Hodges and K. Patterson. "Semantic dementia: a unique clinicopathological syndrome". In: *The Lancet Neurology* 6 (11 Nov. 2007), pp. 1004–1014. ISSN: 14744422. DOI: 10.1016/S1474-4422(07)70266-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S1474442207702661.

[37] M. Lehmann, A. Douiri, L. G. Kim, M. Modat, D. Chan, S. Ourselin, J. Barnes, and N. C. Fox. "Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements". In: *NeuroImage* 49 (3 Feb. 2010), pp. 2264–2274. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2009.10.056.

[38] C. E. Leyton, K. J. Ballard, O. Piguet, and J. R. Hodges. "Phonologic errors as a clinical marker of the logopenic variant of PPA". In: *Neurology* 82 (18 May 2014), pp. 1620–1627. ISSN: 0028-3878. DOI: 10.1212/WNL.0000000000000387. URL: https://www.neurology.org/doi/10.1212/WNL.0000000000000387.

[39] "Detecting frontotemporal dementia syndromes using MRI biomarkers". In: *NeuroImage: Clinical* 22 (Jan. 2019). ISSN: 22131582. DOI: 10.1016/j.nicl.2019.101711.

[40] C. M. Bird and N. Burgess. "The hippocampus and memory: insights from spatial processing". In: *Nature reviews neuroscience* 9.3 (2008), pp. 182–194.

[41] V. Vuksanovic, R. T. Staff, S. Morson, T. Ahearn, L. Bracoud, A. D. Murray, P. Bentham, C. M. Kipps, C. R. Harrington, and C. M. Wischik. "Degeneration of basal and limbic networks is a core feature of behavioural variant frontotemporal dementia". In: *Brain Communications* 3 (4 2021). ISSN: 26321297. DOI: 10.1093/braincomms/fcab241.

[42] D. Ferreira, S. Shams, L. Cavallin, M. Viitanen, J. Martola, T. Granberg, M. Shams, P. Aspelin, M. Kristoffersen-Wiberg, A. Nordberg, et al. "The contribution of small vessel disease to subtypes of Alzheimer's disease: a study on cerebrospinal fluid and imaging biomarkers". In: *Neurobiology of aging* 70 (2018), pp. 18–29.

[43] L. L. Barnes, S. Leurgans, N. T. Aggarwal, R. C. Shah, Z. Arvanitakis, B. D. James, A. S. Buchman, D. A. Bennett, and J. A. Schneider. "Mixed pathology is more likely in black than white decedents with Alzheimer dementia". In: *Neurology* 85 (6 Aug. 2015), pp. 528–534. ISSN: 0028-3878. DOI: 10.1212/WNL.0000000000001834.

[44] J. B. Toledo, H. Liu, M. J. Grothe, T. Rashid, L. Launer, L. M. Shaw, H. Snoussi, S. Heckbert, M. Weiner, J. Q. Trojanwoski, S. Seshadri, and M. Habes. "Disentangling tau and brain atrophy cluster heterogeneity across the Alzheimer's disease continuum". In: *Alzheimer's and Dementia: Translational Research and Clinical Interventions* 8 (1 2022). ISSN: 23528737. DOI: 10.1002/trc2.12305.

[45] A. Rey. "L'examen clinique en psychologie." In: (1958).

[46] J. Lindeboom, B. Schmand, L. Tulner, G. Walstra, and C. Jonker. “Visual association test to detect early dementia of the Alzheimer type”. In: *Journal of neurology, neurosurgery, and psychiatry* 73.2 (2002), p. 126.

[47] A. I. T. Battery. *Manual of directions and scoring*. 1944.

[48] L. L. Thurstone. “Primary mental abilities.” In: *Psychometric monographs* (1938).

[49] D. Wechsler and J. Uterwijk. *WAIS-III: technische handleiding*. Harcourt Test Publ, 2005.

[50] J. R. Stroop. “Studies of interference in serial verbal reactions.” In: *Journal of experimental psychology* 18.6 (1935), p. 643.

[51] E. Kaplan, H. Goodglass, and S. Weintraub. “Boston naming test”. In: *The Clinical Neuropsychologist* (2001).

[52] D. R. Royall, J. A. Cordes, and M. Polk. “CLOX: an executive clock drawing task”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 64.5 (1998), pp. 588–594.

Appendix A: Dataset specifics

Inclusion criteria

The study included participants diagnosed with AD, FTD, or a language variant of either AD or FTD. These language variants, collectively named PPA, presents with different subtypes: SD, PNFA, and LPA. Individuals diagnosed with PPA but for whom a subtype was not specified are referred to as; PPA not otherwise specified (PPA-nos).

The diagnosis recorded in the datasets represents the last known diagnosis of the subject. The scan closest to the date of diagnosis was considered. For individuals who transitioned from Mild Cognitive Impairment (MCI) to dementia, a scan within 90 days of their dementia diagnosis was included. In this way people diagnosed with MCI at time of scan were not considered. Finally, most scans were conducted at the Erasmus Medical Center, with a few exceptions outside of the Erasmus.

Dataset statistics

	ACE	NACC	RS
N (m/f)	297 (143/154)	1375 (625/750)	11728 (5332/6396)
Age mean	66.4 $\pm$ 9.8	75.1 $\pm$ 9.4	64.7 $\pm$ 9.8
Age range	32-95	35-100	45-100
Age males	66.4 $\pm$ 10.6	75.1 $\pm$ 8.8	64.8 $\pm$ 9.7
Age females	64.5 $\pm$ 8.7	75.1 $\pm$ 9.8	64.7 $\pm$ 9.9
Scan date	2011 - 2021	2002 - 2019	2005 - 2015
Diagnosis			
AD	152	1304	
FTD	70		
bvFTD		38	
PPA-nos	20	14	
PNFA	15	8	
SD	27	8	
LPA	13	4	

Table 1: *Dataset specifics for ACE, NACC and the Rotterdam Study (RS).*

Appendix B: Brain volumes

The regional brain volumes of the Desikan-Killian atlas, combined according to FreeSurfer.
Parietal Lobe:

- inferiorparietal
- superiorparietal
- supramarginal
- postcentral
- precuneus

Temporal Lobe

- bankssts
- fusiform
- inferiortemporal
- middletemporal
- superiortemporal
- parahippocampal
- transversetemporal
- entorhinal
- temporalpole

Occipital Lobe

- cuneus
- lateraloccipital
- lingual
- pericalcarine

Frontal Lobe

- caudalmiddlefrontal
- frontalpole
- lateralorbitofrontal
- medialorbitofrontal
- paracentral
- parsopercularis
- parsorbitalis
- rostralmiddlefrontal
- superiorfrontal
- parstriangularis
- precentral

Appendix C: Cognitive domains

The cognitive domains are constructed using a combination of test scores, as outlined in Table 2. The mean z-scores are calculated, excluding any missing values.

Cognitive domain	Neuropsychological tests
Memory/learning	Dutch Rey Auditory Verbal Learning Test (RAVLT) [45]
	Visual Association Test (VAT) [46]
Attention/ Executive function	Trail making Test (TMT) [47]
	Letter fluency [48]
	WAIS-III Digit Span [49]
	Stroop Color-Word Test [50]
Language	The Boston Naming Test (BNT) [51]
Processing speed	TMT
	Stroop Color-Word Test
Visuospatial function	Clock drawing [52]

Table 2: *Included test scores per cognitive domain.*

TMT and Stroop tests

In the Trail Making Test (TMT), an '888' value may signify either a failed test or a time exceeding the limit. Specifically, if TMT A returns '888', both TMT A and TMT B are excluded from the cognitive domain scores. For TMT B, any '888' values are substituted with the maximum allowable time for the test, set at 300 seconds. If the time taken exceeds this limit, it is truncated to 300 seconds.

Similarly, in the Stroop test, an '888' value indicates an error and is replaced by the highest Stroop test score observed across participants:

- Stroop test 1 max score: 174.0.
- Stroop test 2 max score: 274.0.
- Stroop test 3 max score: 656.0.

It is important to note that in both the TMT and Stroop tests, lower scores reflect better performance in contrast to the other neuropsychological test scores. To compare and combine all tests, the TMT test and Stroop test scores are inverted, such that all scores are aligned in the same direction.

Missing z-scores per cognitive domain

Cluster	Etiological diagnosis	Total (N)	Visuospatial Functioning	Language	Memory / Learning	Processing Speed	Attention / Executive Functioning
GLOB2.1	AD	28	7 (25.0%)	6 (21.4%)	17 (60.7%)	9 (32.1%)	6 (21.4%)
	FTD	25	9 (36.0%)	7 (28.0%)	19 (76.0%)	9 (36.0%)	7 (28.0%)
	LPA	1	1 (100.0%)	0 (0.0%)	1 (100.0%)	1 (100.0%)	0 (0.0%)
	PPA	11	1 (9.1%)	0 (0.0%)	6 (54.5%)	0 (0.0%)	0 (0.0%)
	PNFA	7	0 (0.0%)	0 (0.0%)	3 (42.9%)	0 (0.0%)	0 (0.0%)
	SD	18	5 (27.8%)	7 (38.9%)	9 (50.0%)	4 (22.2%)	3 (16.7%)
GLOB2.2	AD	124	39 (31.5%)	38 (30.6%)	77 (62.1%)	45 (36.3%)	30 (24.2%)
	FTD	45	25 (55.6%)	20 (44.4%)	34 (75.6%)	20 (44.4%)	18 (40.0%)
	LPA	12	4 (33.3%)	1 (8.3%)	8 (66.7%)	4 (33.3%)	3 (25.0%)
	PPA	9	1 (11.1%)	0 (0.0%)	5 (55.6%)	1 (11.1%)	0 (0.0%)
	PNFA	8	2 (25.0%)	2 (25.0%)	5 (62.5%)	2 (25.0%)	1 (12.5%)
	SD	9	5 (55.6%)	7 (77.8%)	7 (77.8%)	6 (66.7%)	5 (55.6%)
Total		297	99 (33.3%)	88 (29.6%)	191 (64.3%)	101 (34.0%)	73 (24.6%)

Table 3: *The missing values (n, percentage) in the z-scores for every cognitive domain and clinical diagnosis in the GLOB2 clusters.*

Cluster	Etiological diagnosis	Total (N)	Visuospatial Functioning		Language		Memory / Learning		Processing Speed		Attention / Executive Functioning	
			n		n		n		n		n	
GLOB4.1	AD	107	35	32.7%	35	32.7%	66	61.7%	40	37.4%	28	26.2%
	FTD	31	15	48.4%	13	41.9%	21	67.7%	13	41.9%	11	35.5%
	LPA	9	4	44.4%	1	11.1%	5	55.6%	4	44.4%	3	33.3%
	PPA	6	1	16.7%	0	0.0%	4	66.7%	1	16.7%	0	0.0%
	PNFA	6	2	33.3%	2	33.3%	4	66.7%	2	33.3%	1	16.7%
	SD	8	5	62.5%	7	87.5%	6	75.0%	6	75.0%	5	62.5%
GLOB4.2	AD	10	1	10.0%	1	10.0%	4	40.0%	2	20.0%	1	10.0%
	FTD	12	3	25.0%	2	16.7%	9	75.0%	3	25.0%	2	16.7%
	LPA	1	1	100.0%	0	0.0%	1	100.0%	1	100.0%	0	0.0%
	PPA	4	0	0.0%	0	0.0%	1	25.0%	0	0.0%	0	0.0%
	PNFA	1	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	SD	18	5	27.8%	7	38.9%	9	50.0%	4	22.2%	3	16.7%
GLOB4.3	AD	18	6	33.3%	5	27.8%	13	72.2%	7	38.9%	5	27.8%
	FTD	13	6	46.2%	5	38.5%	10	76.9%	6	46.2%	5	38.5%
	PPA	7	1	14.3%	0	0.0%	5	71.4%	0	0.0%	0	0.0%
	PNFA	6	0	0.0%	0	0.0%	3	50.0%	0	0.0%	0	0.0%
GLOB4.4	AD	17	4	23.5%	3	17.6%	11	64.7%	5	29.4%	2	11.8%
	FTD	14	10	71.4%	7	50.0%	13	92.9%	7	50.0%	7	50.0%
	LPA	3	0	0.0%	0	0.0%	3	100.0%	0	0.0%	0	0.0%
	PPA	3	0	0.0%	0	0.0%	1	33.3%	0	0.0%	0	0.0%
	PNFA	2	0	0.0%	0	0.0%	1	50.0%	0	0.0%	0	0.0%
	SD	1	0	0.0%	0	0.0%	1	100.0%	0	0.0%	0	0.0%
Total		297	99	33.3%	88	29.6%	191	64.3%	101	34.0%	73	24.6%

Table 4: *The missing values (n, percentage) in the z-scores for every cognitive domain and clinical diagnosis in the GLOB4 clusters.*

Appendix D: ACE data results - GLOB

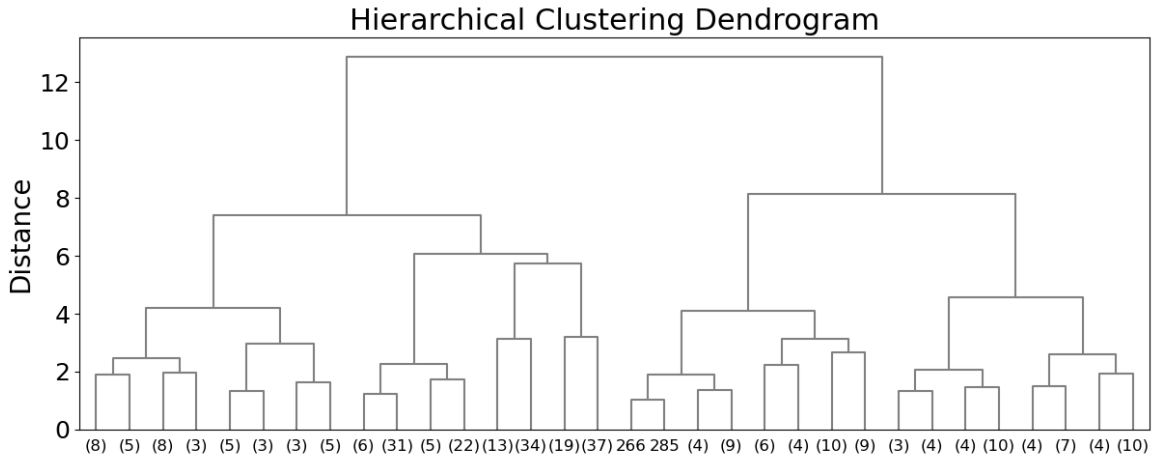


Figure 10: *Dendrogram of the hierarchical clustering. Between brackets is the number of subjects in that leaf. In case there is only one subject, only the index number is depicted.*

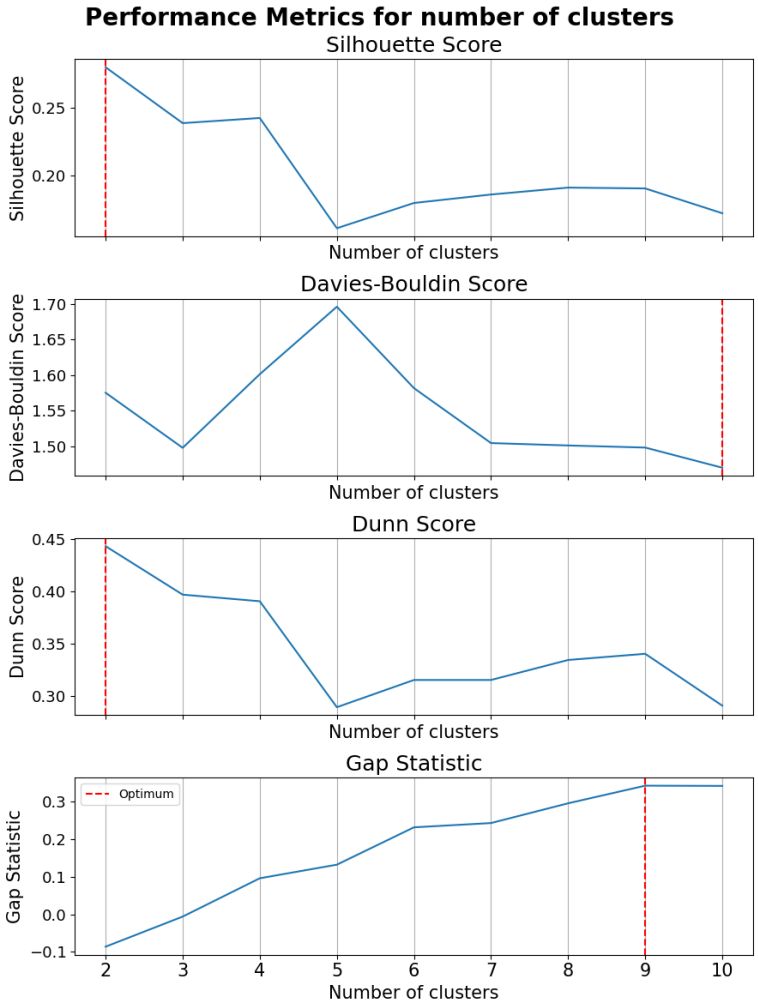


Figure 11: *Performance metric scores in clustering the ACE dataset. The red dotted line indicates the optimal number of clusters according to each performance metric. Silhouette and Dunn indicate 2 as the optimum, while Davies-Bouildin and GAP indicate 10 and 9 clusters respectively as the optimal number.*

	GLOB2.1	GLOB2.2	Statistic (p-value)
N	90	207	
Age	66.8 ± 9.8	66.3 ± 9.6	9834.0 (p=0.446)
Sex (m/f)	50/40	93/114	2.428 (p=0.119)
Education (median)	5	5	3.316 (p=0.854)
Parietal lobe (rh)	0.75 ± 0.24	0.18 ± 0.21	17462.0 (p<0.001)
Parietal lobe (lh)	0.72 ± 0.24	0.14 ± 0.18	17762.0 (p<0.001)
Temporal lobe (rh)	0.55 ± 0.40	0.24 ± 0.27	13320.5 (p<0.001)
Temporal lobe (lh)	0.48 ± 0.43	0.17 ± 0.24	12717.5 (p<0.001)
Occipital (rh)	0.71 ± 0.26	0.39 ± 0.32	14482.5 (p<0.001)
Occipital (lh)	0.69 ± 0.29	0.42 ± 0.33	13697.0 (p<0.001)
Frontal (rh)	0.57 ± 0.34	0.22 ± 0.26	14868.5 (p<0.001)
Frontal (lh)	0.54 ± 0.36	0.20 ± 0.25	14586.5 (p<0.001)
Hippocampus (rh)	0.35 ± 0.35	0.21 ± 0.29	11622.5 (p<0.001)
Hippocampus (lh)	0.29 ± 0.35	0.18 ± 0.27	11026.5 (p<0.01)
Memory/learning	0.5 ± 0.9	-0.3 ± 0.9	1801.0 (p<0.001)
Attention/Executive function	0.4 ± 0.9	-0.2 ± 1.0	3.421 (p<0.001)
Language	0.1 ± 1.1	0.0 ± 1.0	5218.5 (p=0.392)
Processing speed	0.3 ± 0.6	-0.2 ± 1.1	5615.0 (p<0.001)
Visuospatial function	0.4 ± 0.9	-0.2 ± 1.0	5962.0 (p<0.001)

Table 5: *Demographic, GMPs, and cognition characteristics of the two cluster in GLOB2. Statistical tests are done with the Mann-Whitney U test (age and GMPs), the t-test (attention/executive functioning), and the Chi-squared test (sex, education and other cognitive domains).*

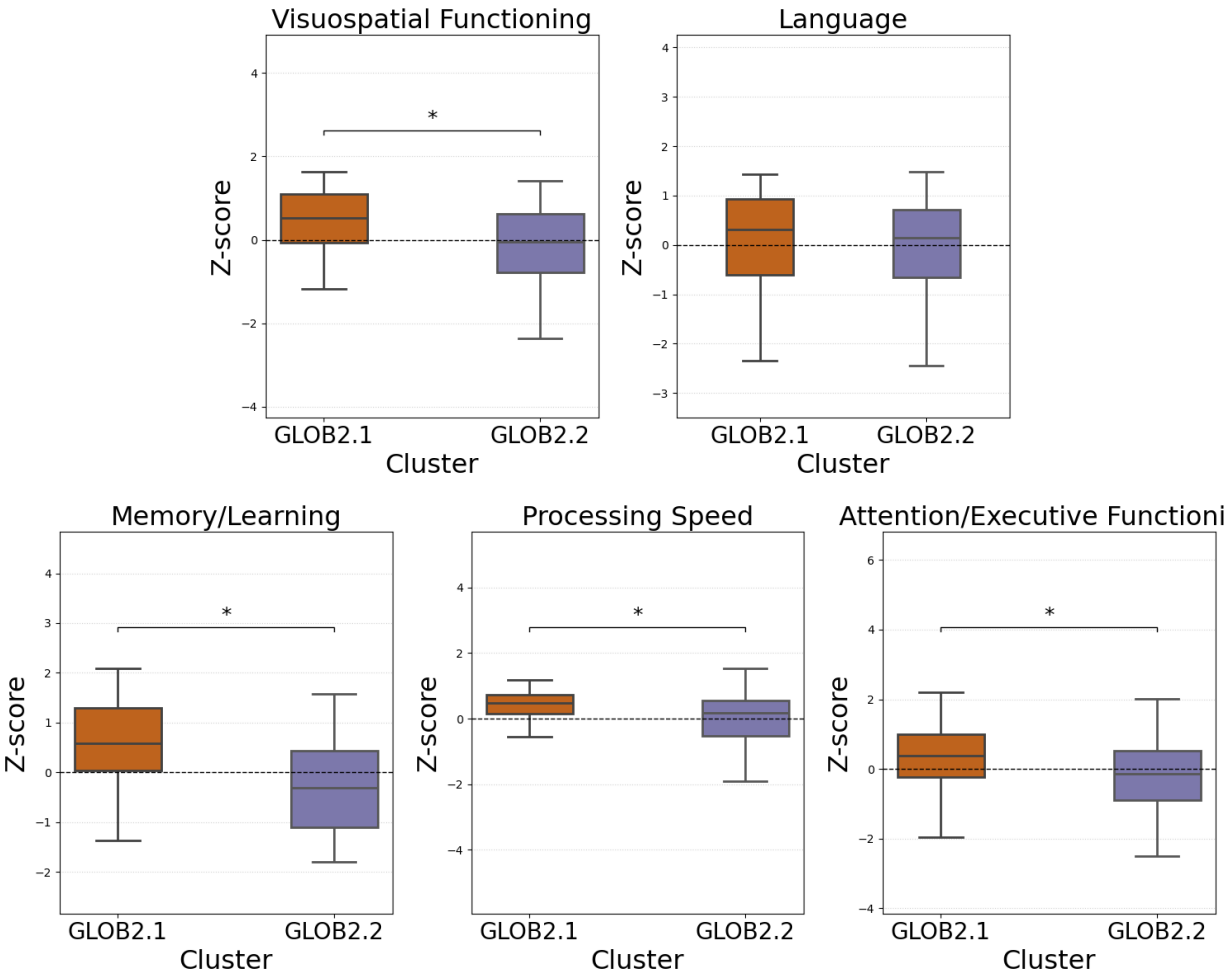


Figure 12: *Comparison of Z-scores for each cognitive domain for the two clusters in GLOB2. A higher score indicates a better performance for all cognitive domains. It is important to note that the z-score indicates performance relative to the dataset and should not be directly compared to subjects with normal cognitive function. An equal z-score suggests similar performance between the two groups without specifying the degree of impairment. Significant differences between clusters are denoted by *.*

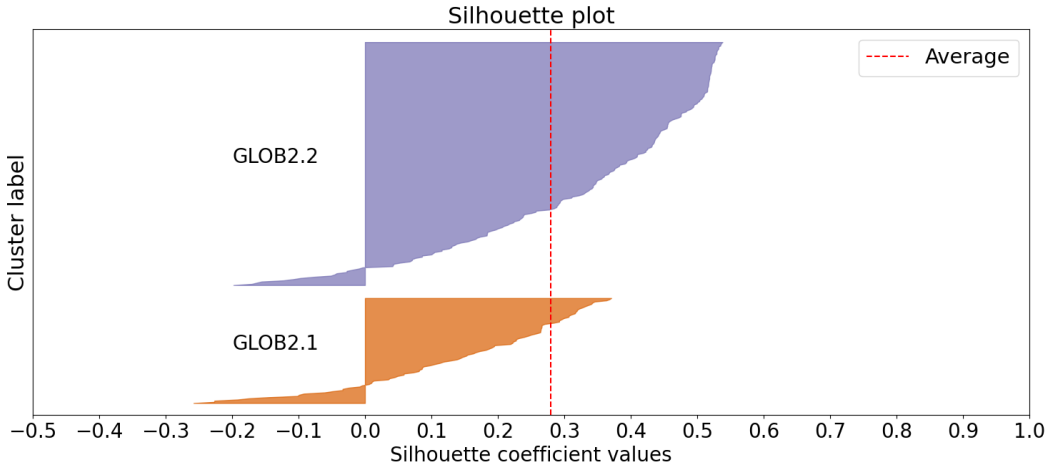


Figure 13: *Silhouette value for each subject for the two clusters in GLOB2. The average over all subjects is illustrated with the red line.*

	C1 (GLOB4.1)	C2 (GLOB4.2)	C3 (GLOB4.3)	C4 (GLOB4.4)	Statistic (p-value)	post-hoc (p-value)
N	167	46	44	40		
Age	66.6 ± 9.5	64.9 ± 9.0	68.8 ± 10.4	64.9 ± 10.0	6.357 (p=0.095)	
Sex (m/f)	76/91	25/21	25/19	17/23	3.010 (p=0.390)	
Education (median)	5	5	5	5	11.949 (p=0.941)	
Parietal lobe (rh)	0.15 ± 0.18	0.75 ± 0.23	0.75 ± 0.26	0.29 ± 0.28	151.310 (p<0.001)	C1<C2 (p<0.001) C1<C3 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)
Parietal lobe (lh)	0.12 ± 0.17	0.73 ± 0.23	0.72 ± 0.25	0.20 ± 0.22	159.791 (p<0.001)	C1<C2 (p<0.001) C1<C3 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)
Temporal lobe (rh)	0.20 ± 0.23	0.25 ± 0.29	0.87 ± 0.19	0.43 ± 0.34	103.997 (p<0.001)	C1<C3 (p<0.001) C1<C4 (p<0.01) C2<C3 (p<0.001) C2<C4 (p<0.001) C3>C4 (p<0.05)
Temporal lobe (lh)	0.13 ± 0.21	0.09 ± 0.14	0.90 ± 0.11	0.34 ± 0.32	129.095 (p<0.001)	C1<C3 (p<0.001) C1<C4 (p<0.01) C2<C3 (p<0.001) C2<C4 (p<0.001) C3>C4 (p<0.001)
Occipital (rh)	0.36 ± 0.30	0.73 ± 0.21	0.70 ± 0.31	0.50 ± 0.37	63.896 (p<0.001)	C1<C2 (p<0.001) C1<C3 (p<0.001) C2>C4 (p<0.05) C3>C4 (p<0.05)
Occipital (lh)	0.37 ± 0.32	0.68 ± 0.26	0.70 ± 0.31	0.60 ± 0.31	56.737 (p<0.001)	C1<C2 (p<0.001) C1<C3 (p<0.001) C1<C4 (p<0.001)
Frontal (rh)	0.20 ± 0.25	0.49 ± 0.37	0.66 ± 0.29	0.29 ± 0.26	75.904 (p<0.001)	C1<C2 (p<0.001) C1<C3 (p<0.001) C3>C4 (p<0.001)
Frontal (lh)	0.19 ± 0.26	0.44 ± 0.37	0.64 ± 0.31	0.24 ± 0.24	74.208 (p<0.001)	C1<C2 (p<0.001) C1<C3 (p<0.001) C3>C4 (p<0.001)
Hippocampus (rh)	0.10 ± 0.14	0.17 ± 0.23	0.54 ± 0.35	0.68 ± 0.30	113.793 (p<0.001)	C1<C3 (p<0.001) C1<C4 (p<0.001) C2<C3 (p<0.001) C2<C4 (p<0.001)
Hippocampus (lh)	0.07 0.10	0.07 ± 0.12	0.53 ± 0.35	0.62 ± 0.29	133.478 (p<0.001)	C1<C3 (p<0.001) C1<C4 (p<0.001) C2<C3 (p<0.001) C2<C4 (p<0.001)

Table 6: *Demographics, and GMPs of the four cluster of GLOB4. Significant differences in sex and education are tested using the Chi-squared test, while differences in age and GMPs are statistically tested by the Kruskal-Wallis test and the Dunn's test (post-hoc).*

	C1 (GLOB4.1)	C2 (GLOB4.2)	C3 (GLOB4.3)	C4 (GLOB4.4)	Statistic (p-value)	post-hoc (p-value)
N	167	46	44	40		
Memory/ learning	-0.34 ± 0.94	0.43 ± 0.83	0.68 ± 1.07	0.25 ± 0.88	17.259 (p<0.001)	C1<C2 (p<0.01) C2<C3 (p<0.01)
Attention/ Executive function	-0.17 ± 0.97	0.52 ± 0.73	0.21 ± 1.09	-0.24 ± 1.07	2.737 (p<0.001)	n.s.
Language	-0.03 ± 0.96	-0.46 ± 1.08	0.61 ± 0.70	-0.02 ± 1.06	20.323 (p<0.001)	C1<C3 (p<0.01) C2<C3 (p<0.001)
Processing speed	-0.17 ± 1.11	0.54 ± 0.43	0.11 ± 0.76	-0.20 ± 1.14	18.849 (p<0.001)	C1<C2 (p<0.001) C2>C3 (p<0.05) C2>C4 (p<0.01)
Visuospatial function	-0.21 ± 1.00	0.47 ± 0.89	0.35 ± 0.83	-0.23 ± 1.06	17.459 (p<0.001)	C1<C2 (p<0.01) C2>C4 (p<0.05)

Table 7: *Cognitive characteristics of the four clusters of GLOB4. Differences between clusters in attention/executive functioning was statistically tested by the one-way ANOVA and Turkeys HSD (post-hoc), while the differences in other cognitive domains were tested by the Kruskal-Wallis and the Dunn's test (post-hoc).*

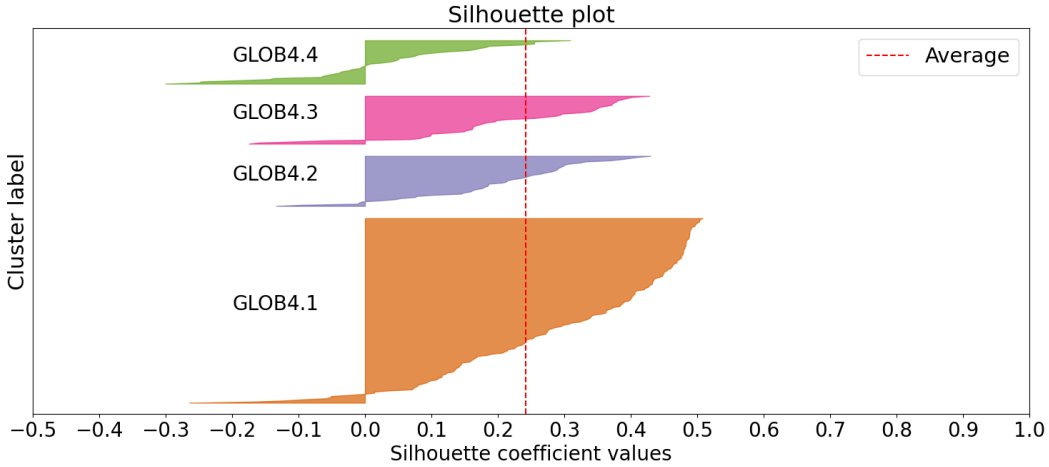


Figure 14: *Silhouette value for each subject for the two clusters in GLOB2. The average over all subjects is illustrated with the red line.*

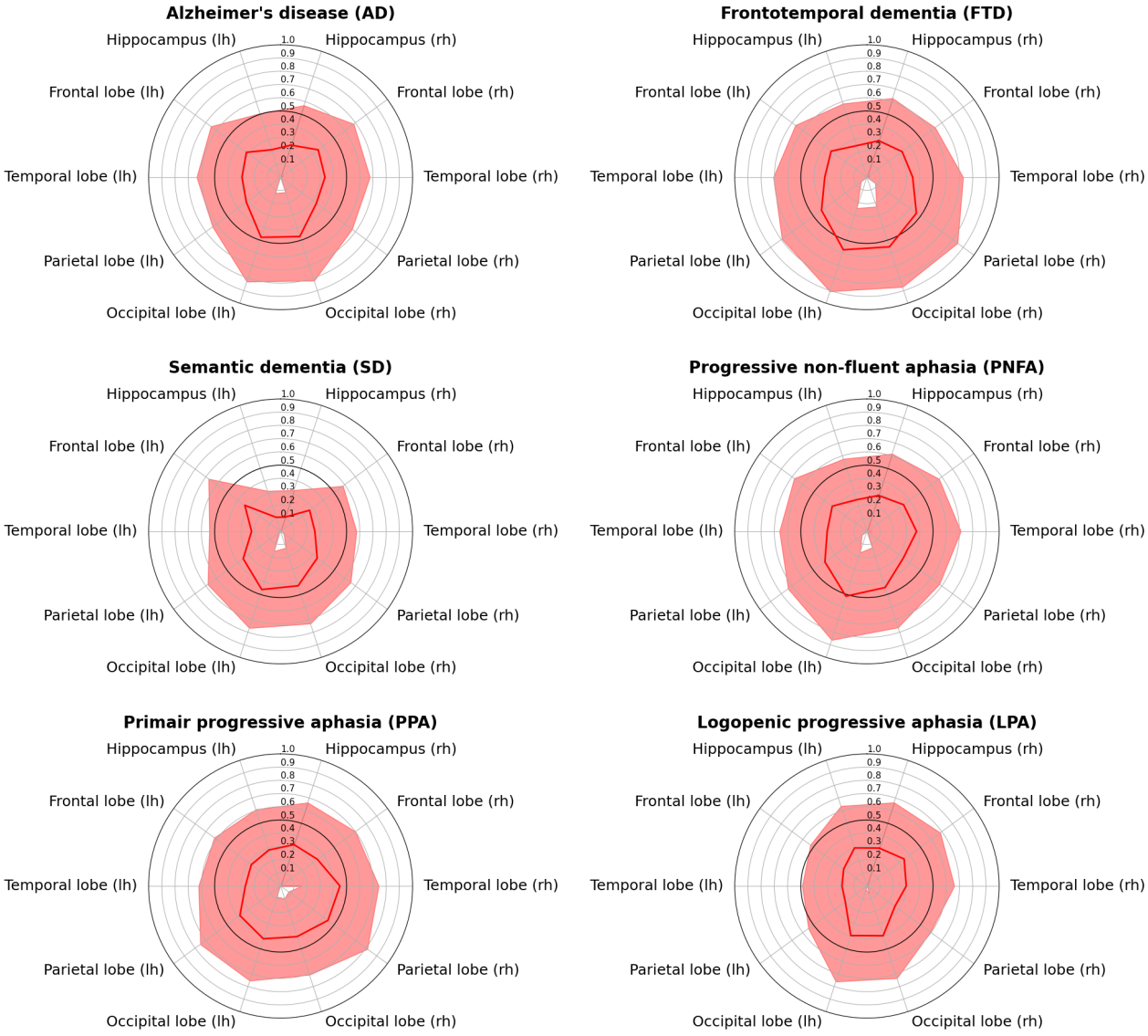


Figure 15: *Atrophy patterns for each etiological diagnosis in the ACE dataset.*

Appendix E: ACE data results - PCA

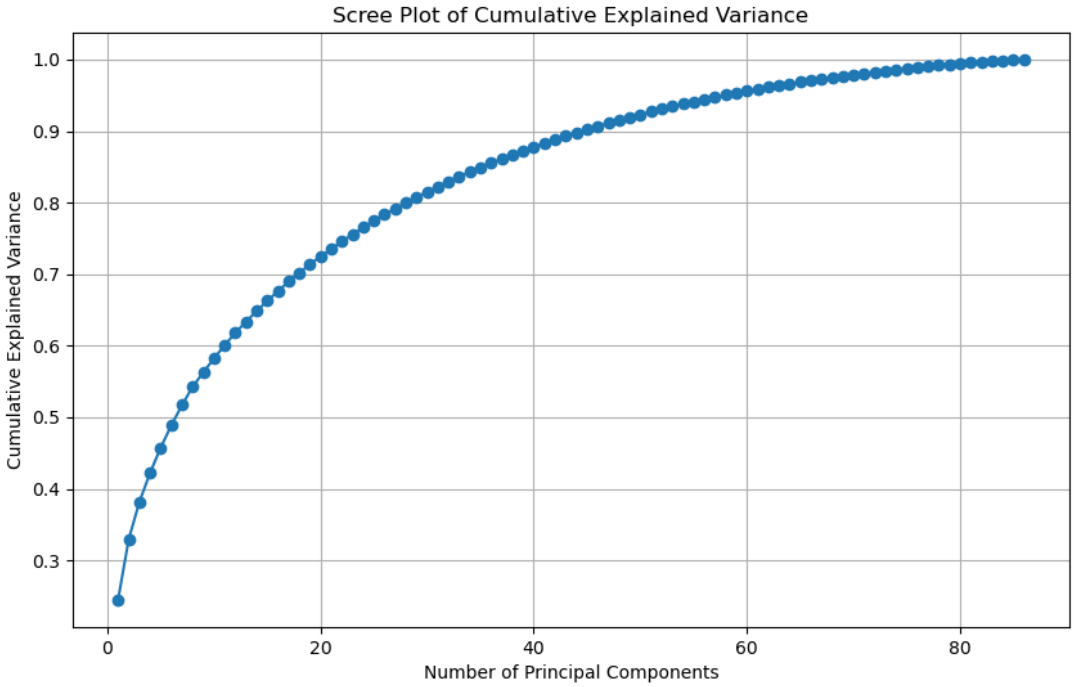
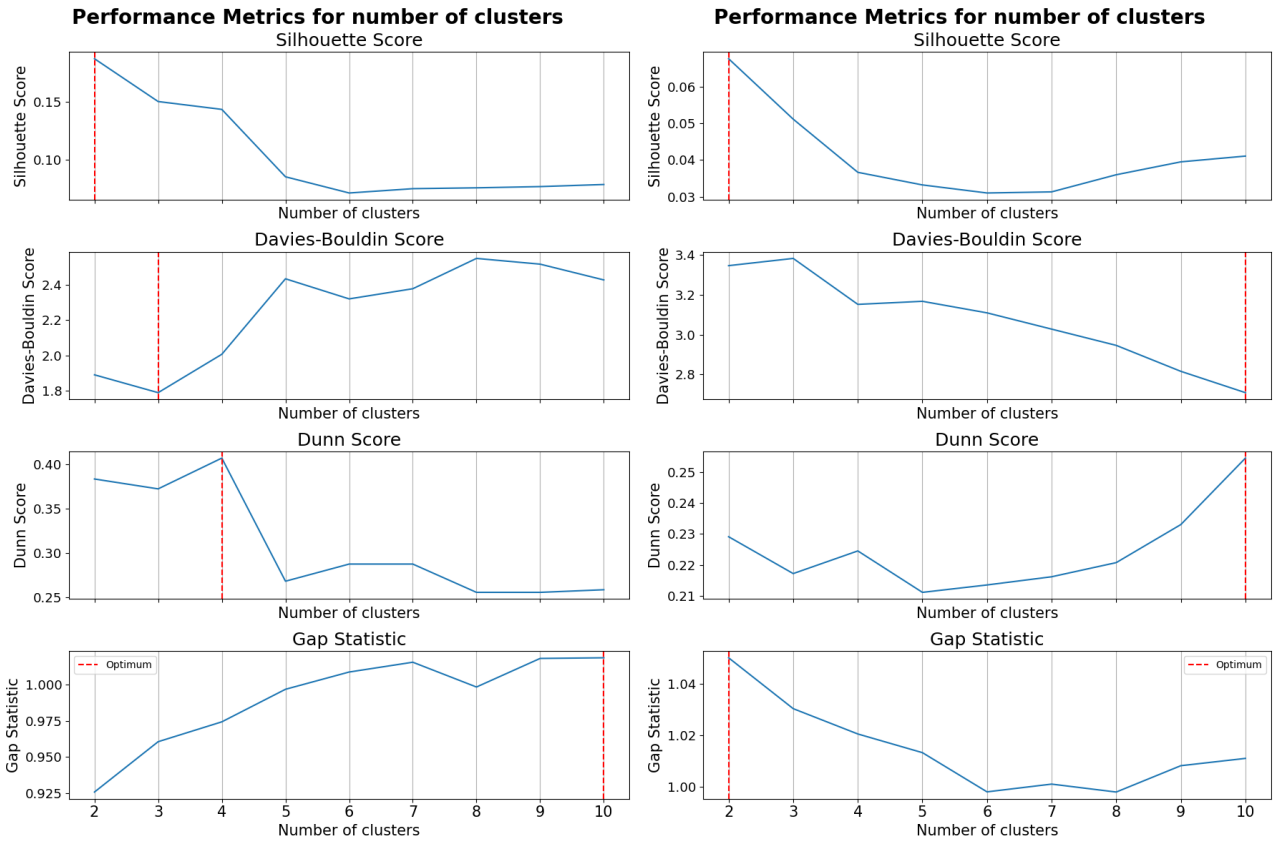


Figure 16: *Scree plot of PCA*

PC1	PC2
Lateralorbitofrontal (rh)	Entorhinal (lh)
Lateralorbitofrontal (lh)	Temporalpole (lh)
Superiorfrontal (lh)	Entorhinal (rh)
Insula (rh)	Temporalpole (rh)
Insula (lh)	Cuneus (rh)
Superiortemporal (rh)	Parahippocampal (lh)
Middletemporal (rh)	Superiorparietal (lh)
Fusiform (rh)	Superiorparietal (rh)
Parsorbitalis (rh)	Cuneus (lh)
Superiortemporal (lh)	Amygdala (lh)

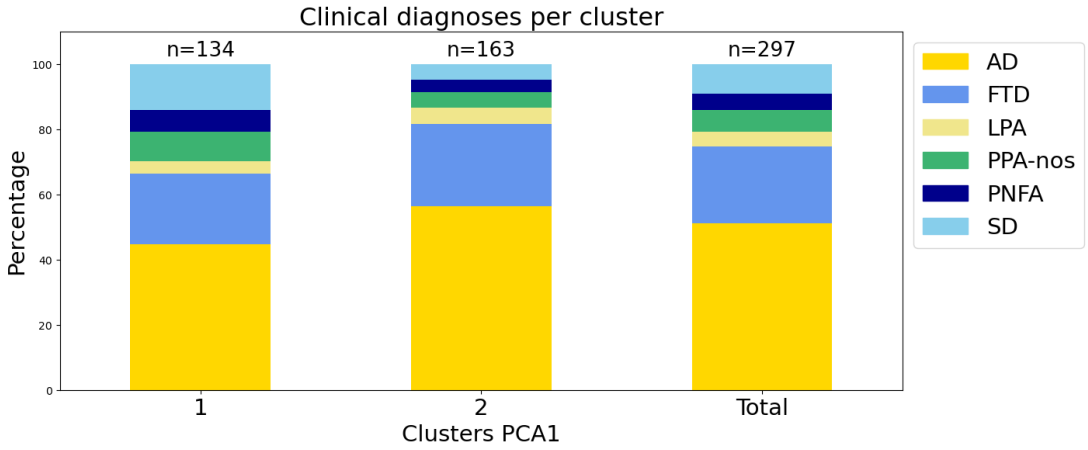
Table 8: *The 10 most contributing features, i.e. having the highest weighting, of PC1 and PC2.*



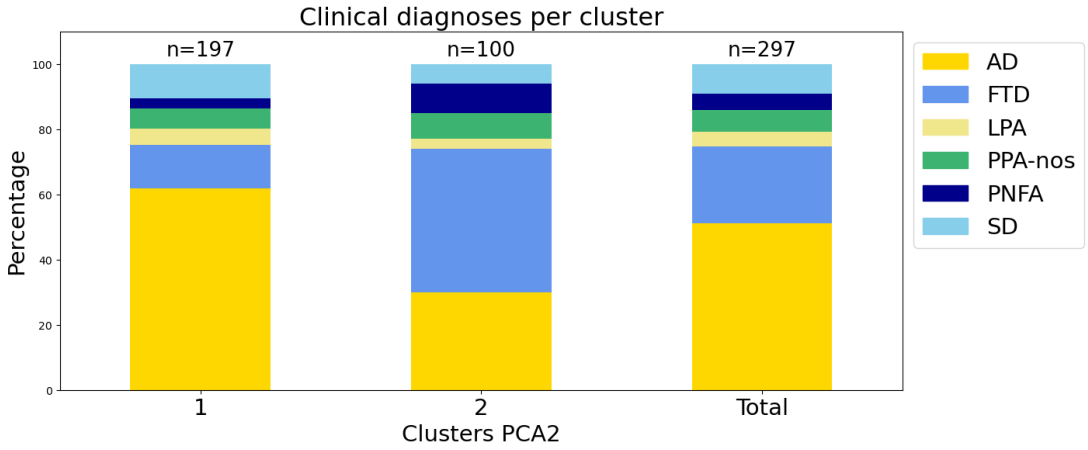
(a) *PCA1. Silhouette indicates 2, while Davies-Bouildin and Dunn indicate 3 and 4 as the optimal number of clusters respectively. GAP indicates 10 as the optimum.*

(b) *PCA2. Silhouette and GAP indicate 2 as the optimum, while Davies-Bouildin and Dunn indicate 10 clusters as the optimal number of clusters.*

Figure 17: *Performance metric scores in clustering the ACE dataset based on the selected PCs. The red dotted line indicates the optimal number of clusters according to each performance metric.*



(a) *PCA1*



(b) *PCA2*

Figure 18: *Diagnosis distribution per cluster of PCA1 and PCA2.*

Appendix F: NACC data results

Performance metrics

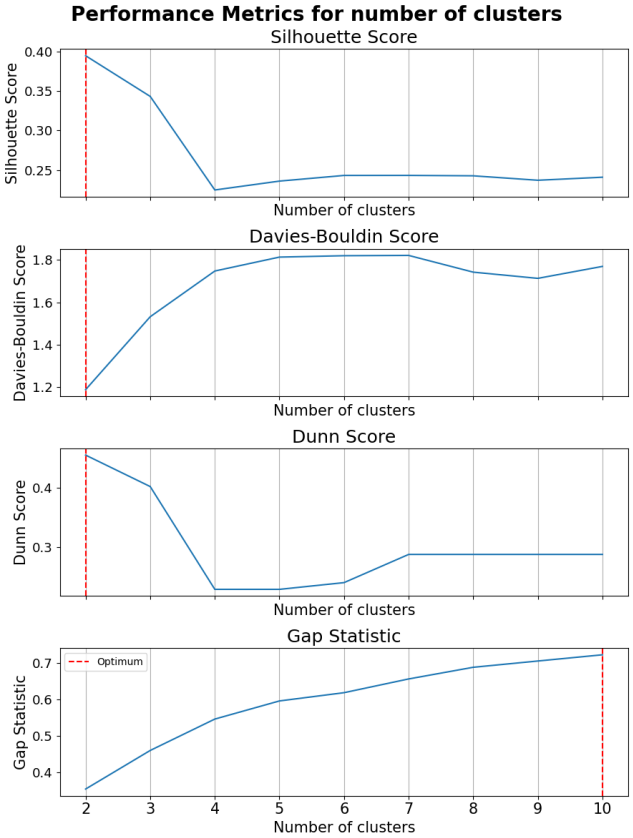


Figure 19: *Performance metric scores in clustering the NACC dataset. The red dotted line indicates the optimal number of clusters according to each performance metric. Silhouette, Dunn and GAP indicate 2 as the optimum, while Davies-Bouildin indicates 10 as the optimal number of clusters.*

NACC2

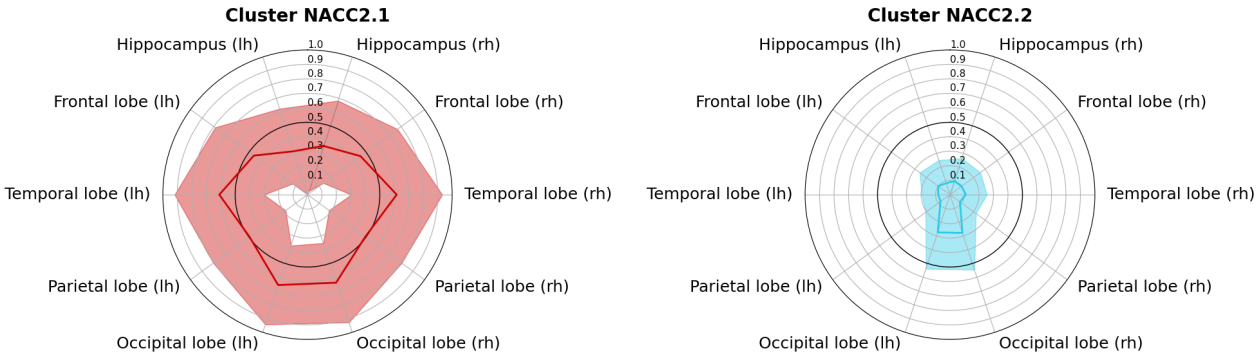


Figure 20: *GMPs of the input features for the left hemisphere (lh) and right hemisphere (rh) for the two clusters of NACC2. The mean GMP is for each region depicted by the colored line, with a concomitant shaded standard deviation. The axes portray the GMPs measured from cognitively normal participants in the Rotterdam Study, with the black 0.5 line representing the average. A GMP value below 0.5 indicates a smaller than average brain volume in a certain region, while a value above 0.5 indicates a larger than average brain volume.*

	NACC2.1	NACC2.2	Statistic (p-value)
N	476	899	
Age	76.4 ± 9.6	74.4 ± 9.2	242047.5 (p<0.001)
Sex (m/f)	208/268	417/482	0.801 (p = 0.371)
CDR (median)	0.5	1	33.79 (p < 0.001)
Parietal lobe (rh)	0.49 ± 0.31	0.08 ± 0.14	385454.0 (p<0.001)
Parietal lobe (lh)	0.50 ± 0.31	0.08 ± 0.13	386451.0 (p<0.001)
Temporal lobe (rh)	0.62 ± 0.31	0.11 ± 0.15	388526.0 (p<0.001)
Temporal lobe (lh)	0.61 ± 0.31	0.08 ± 0.11	396589.0 (p<0.001)
Occipital (rh)	0.64 ± 0.29	0.28 ± 0.27	349774.0 (p<0.001)
Occipital (lh)	0.66 ± 0.29	0.27 ± 0.27	354565.0 (p<0.001)
Frontal (rh)	0.45 ± 0.32	0.10 ± 0.16	365580.5 (p<0.001)
Frontal (lh)	0.46 ± 0.33	0.10 ± 0.16	360552.0 (p<0.001)
Hippocampus (rh)	0.35 ± 0.33	0.10 ± 0.17	330895.5 (p<0.001)
Hippocampus (lh)	0.31 ± 0.31	0.08 ± 0.17	335882.0 (p<0.001)

Table 9: *Demographic, GMPs, and cognition characteristics of the two cluster of NACC2. Differences in age and GMPs are statistically tested by Mann-Whitney U test, while differences in sex and CDR are tested by the Chi-squared test.*

Cluster	CDR score	count (n)	percentage of cluster (%)
NACC2.1 (N = 476)	0	1	0.2
	0.5	237	49.8
	1	188	39.5
	2	38	8.0
	3	12	2.5
NACC2.2 (N = 899)	0	1	0.1
	0.5	320	35.6
	1	400	44.5
	2	139	15.5
	3	39	4.3

Table 10: *CDR scores of cluster NACC2.1 and NACC2.2. The CDR score ranges from 0 to 3: no impairment (CDR = 0), questionable impairment (CDR = 0.5), mild impairment (CDR = 1), moderate impairment (CDR = 2), and severe impairment (CDR = 3).*

NACC4

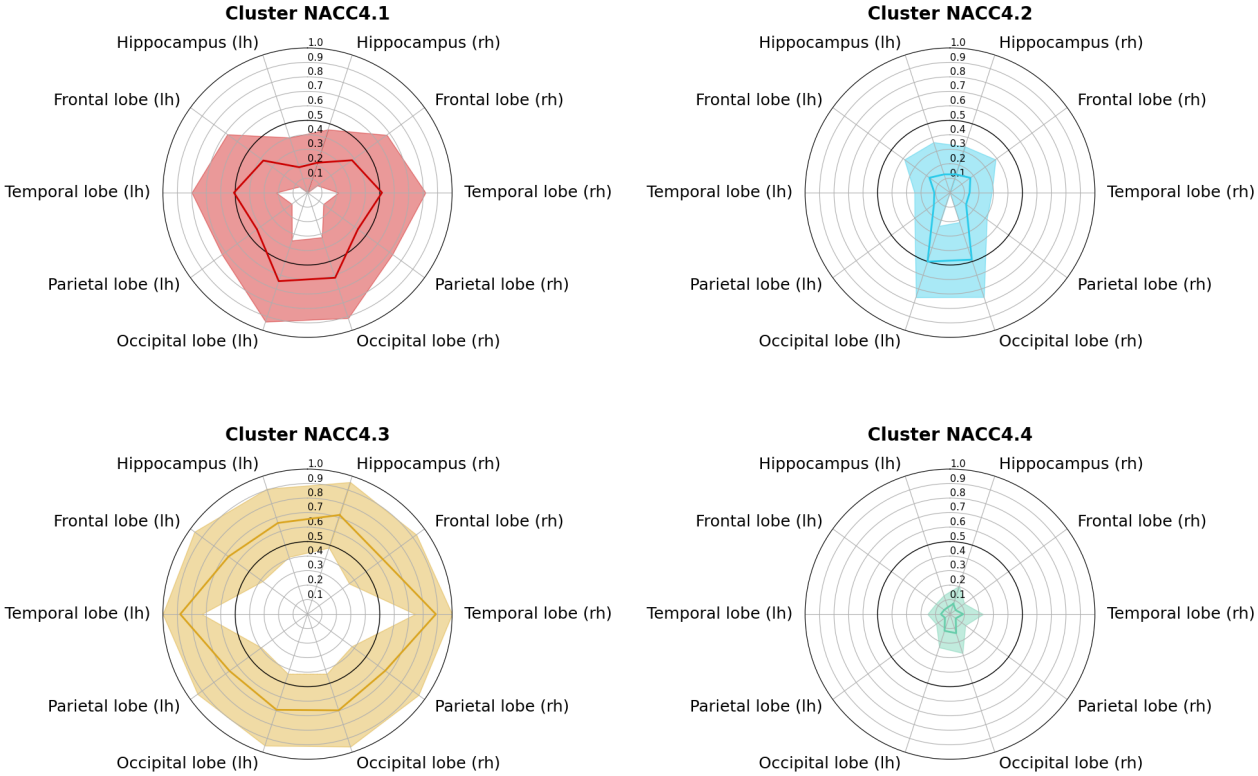


Figure 21: *GMPs of the input features for the left hemisphere (lh) and right hemisphere (rh) for the two clusters of NACC4. The mean GMP is for each region depicted by the colored line, with a concomitant shaded standard deviation. The axes portray the GMPs measured from cognitively normal participants in the Rotterdam Study, with the black 0.5 line representing the average. A GMP value below 0.5 indicates a smaller than average brain volume in a certain region, while a value above 0.5 indicates a larger than average brain volume.*

	C1 (NACC4.1)	C2 (NACC4.2)	C3 (NACC4.3)	C4 (NACC4.4)	Statistic (p-value)	post-hoc (p-value)
N	346	362	130	537		
Age	76.1 ± 9.3	74.2 ± 9.3	77.2 ± 10.2	74.5 ± 9.2	19.081 ($p < 0.001$)	C2<C3 (p<0.01) C3<C4 (p<0.01)
Sex (m/f)	157/189	172/190	51/79	245/292	2.657 ($p = 0.448$)	
CDR (median)	0.5	1	0.5	1	42.957 ($p < 0.001$)	C1<C2 (p<0.05) C4<C3 (p<0.001) C4<C1 (p<0.001)
Parietal lobe (rh)	0.43 ± 0.30	0.14 ± 0.18	0.66 ± 0.29	0.05 ± 0.08	709.407 (p<0.001)	C1>C2 (p<0.001) C1<C3 (p<0.001) C1>C4 (p<0.001) C2<C3 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)
Parietal lobe (lh)	0.43 ± 0.30	0.14 ± 0.16	0.67 ± 0.27	0.04 ± 0.07	740.674 (p<0.001)	C1>C2 (p<0.001) C1<C3 (p<0.001) C1>C4 (p<0.001) C2<C3 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)
Temporal lobe (rh)	0.51 ± 0.30	0.14 ± 0.15	0.89 ± 0.14	0.09 ± 0.14	702.777 (p<0.001)	C1>C2 (p<0.001) C1<C3 (p<0.001) C1>C4 (p<0.001) C2<C3 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)
Temporal lobe (lh)	0.51 ± 0.29	0.11 ± 0.14	0.88 ± 0.15	0.06 ± 0.09	758.697 (p<0.001)	C1>C2 (p<0.001) C1<C3 (p<0.001) C1>C4 (p<0.001) C2<C3 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)
Occipital (rh)	0.62 ± 0.29	0.49 ± 0.27	0.70 ± 0.26	0.14 ± 0.14	661.901 (p<0.001)	C1>C2 (p<0.001) C1<C3 (p<0.001) C1>C4 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)
Occipital (lh)	0.64 ± 0.30	0.50 ± 0.26	0.69 ± 0.26	0.12 ± 0.12	733.272 (p<0.001)	C1>C2 (p<0.001) C1<C3 (p<0.001) C1>C4 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)
Frontal (rh)	0.38 ± 0.30	0.17 ± 0.22	0.64 ± 0.29	0.05 ± 0.07	606.996 (p<0.001)	C1>C2 (p<0.001) C1<C3 (p<0.001) C1>C4 (p<0.001) C2<C3 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)

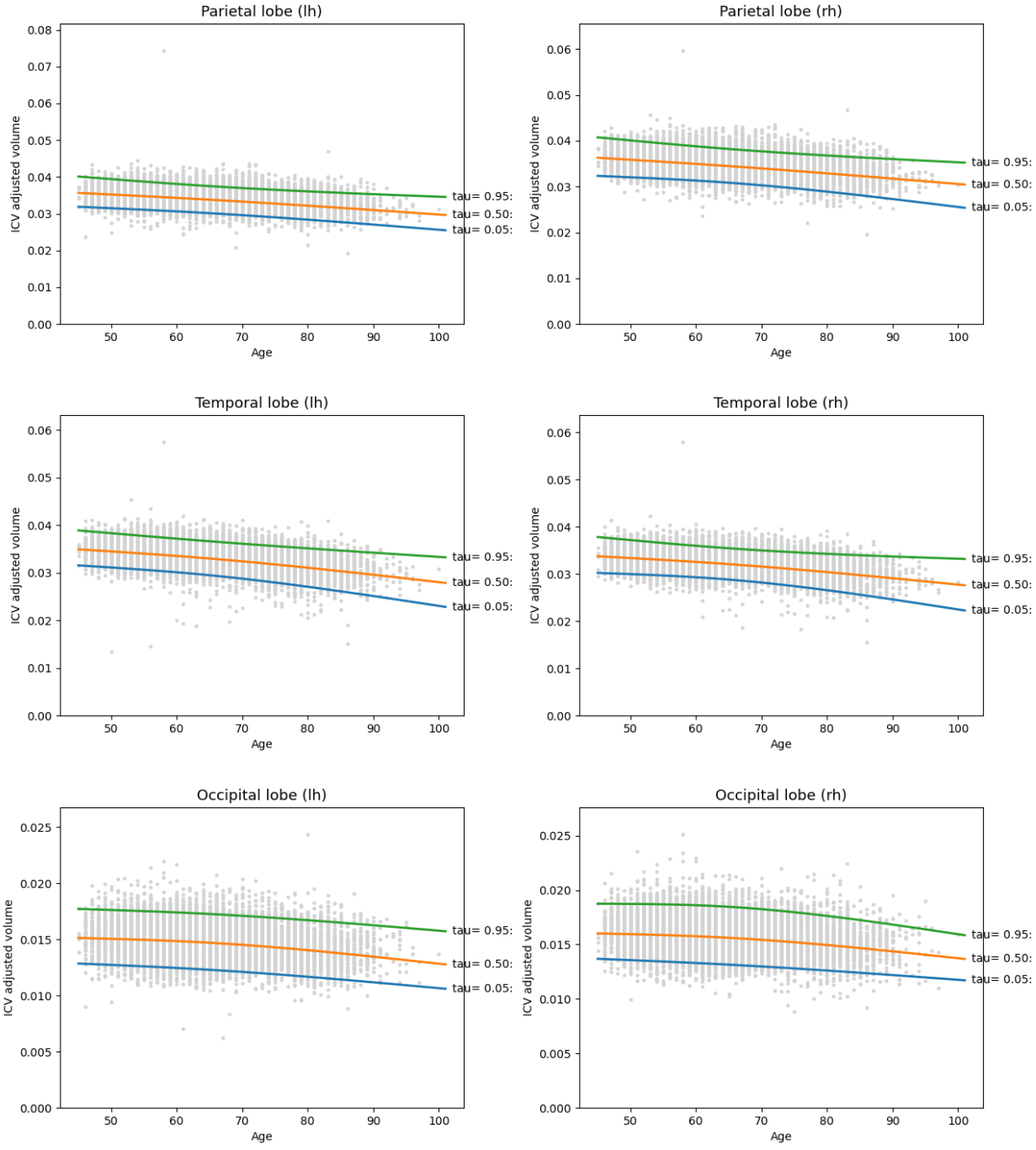
	C1 (NACC4.1)	C2 (NACC4.2)	C3 (NACC4.3)	C4 (NACC4.4)	Statistic (p-value)	post-hoc (p-value)
Frontal (lh)	0.38 ± 0.31	0.18 ± 0.21	0.68 ± 0.29	0.05 ± 0.07	585.528 (p<0.001)	C1>C2 (p<0.001) C1<C3 (p<0.001) C1>C4 (p<0.001) C2<C3 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)
Hippocampus (rh)	0.22 ± 0.24	0.13 ± 0.20	0.72 ± 0.24	0.07 ± 0.13	436.884 (p<0.001)	C1>C2 (p<0.001) C1<C3 (p<0.001) C1>C4 (p<0.001) C2<C3 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)
Hippocampus (lh)	0.18 ± 0.22	0.13 ± 0.23	0.66 ± 0.25	0.05 ± 0.09	461.238 (p<0.001)	C1>C2 (p<0.001) C1<C3 (p<0.001) C1>C4 (p<0.001) C2<C3 (p<0.001) C2>C4 (p<0.001) C3>C4 (p<0.001)

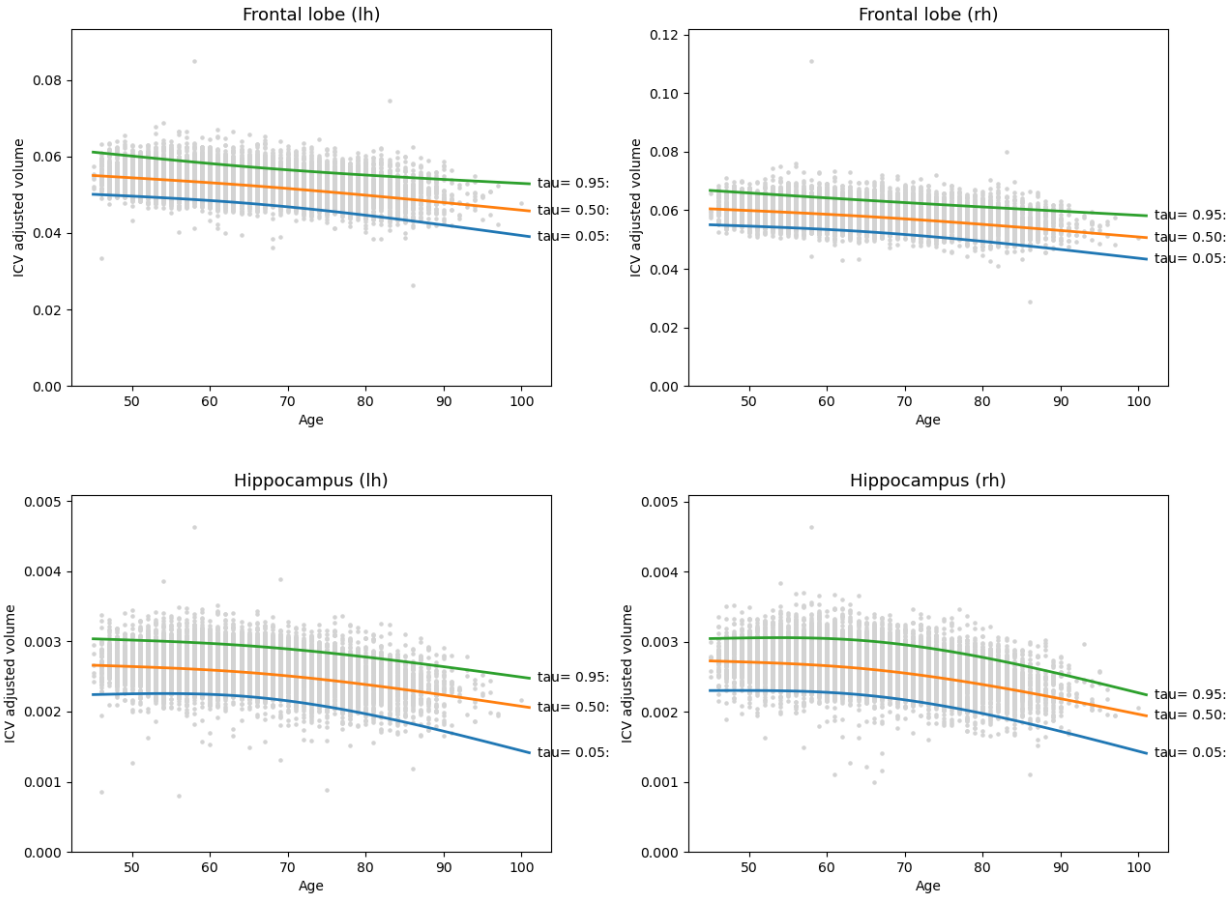
Table 11: *Demographic, GMPs, and cognitive characteristics of the four cluster of NACC4. Differences in age and GMPs are statistically tested by the Kruskal-Wallis test and the Dunn's test (post-hoc), while differences in sex and CDR are tested by the Chi-squared test.*

Cluster	CDR score	count (n)	percentage of cluster (%)
NACC4.1 (N = 346)	0	1	0.3
	0.5	170	49.1
	1	140	40.5
	2	29	8.4
	3	6	1.7
NACC4.2 (N = 362)	0	1	0.3
	0.5	138	38.1
	1	163	45.0
	2	47	13.0
	3	13	3.6
NACC4.3 (N = 130)	0.5	67	51.5
	1	48	36.9
	2	9	6.9
	3	6	4.6
NACC4.4 (N = 537)	0.5	182	33.9
	1	237	44.1
	2	92	17.1
	3	26	4.8

Table 12: *CDR scores of the four clusters of NACC4. The CDR score ranges from 0 to 3: no impairment (CDR = 0), questionable impairment (CDR = 0.5), mild impairment (CDR = 1), moderate impairment (CDR = 2), and severe impairment (CDR = 3).*

Appendix G: Quantile regression




Frontal lobe (lh)
ICV adjusted volume
Age
tau= 0.95:
tau= 0.50:
tau= 0.05:
Frontal lobe (rh)
ICV adjusted volume
Age
tau= 0.95:
tau= 0.50:
tau= 0.05:
Hippocampus (lh)
ICV adjusted volume
Age
tau= 0.95:
tau= 0.50:
tau= 0.05:
Hippocampus (rh)
ICV adjusted volume
Age
tau= 0.95:
tau= 0.50:
tau= 0.05: