Cluster Analysis of Cortical Amyloid Burden for Identifying Imaging-driven Subtypes in Mild Cognitive Impairment

Ruiming Wu, MS[1], Bing He, MS[2], Bojian Hou, PhD[1], Andrew J Saykin, PsyD[2], Jingwen Yan, PhD[2], Li Shen, PhD[1]
[1]University of Pennsylvania, Philadelphia, PA; [2]Indiana University, Indianapolis, IN

Abstract

Over the past decade, Alzheimer's disease (AD) has become increasingly severe and gained greater attention. Mild Cognitive Impairment (MCI) serves as an important prodromal stage of AD, highlighting the urgency of early diagnosis for timely treatment and control of the condition. Identifying the subtypes of MCI patients exhibits importance for dissecting the heterogeneity of this complex disorder and facilitating more effective target discovery and therapeutic development. Conventional method uses clinical measurements such as cognitive score and neurophysical assessment to stratify MCI patients into two groups with early MCI (EMCI) and late MCI (LMCI), which shows their progressive stages. However, such clinical method is not designed to de-convolute the heterogeneity of the disorder. This study uses a data-driven approach to divide MCI patients into a novel grouping of two subtypes based on an amyloid dataset of 68 cortical features from positron emission tomography (PET), where each subtype has a homogeneous cortical amyloid burden pattern. Experimental evaluation including visual two-dimensional cluster distribution, Kaplan-Meier plot, genetic association studies, and biomarker distribution analysis demonstrates that the identified subtypes performs better across all metrics than the conventional EMCI and LMCI grouping.

Introduction

Alzheimer's disease (AD), a degenerative neurological condition, is currently the fifth most common cause of death among individuals aged 65 and above in the United States [1]. It results in irreversible cognitive deterioration, marked by a slow decline in cognitive and behavioral abilities [2]. According to the World Health Organization (WHO)[1], dementia has impacted 55 million individuals globally in 2023. The figure may rise to 139 million by 2050 due to an aging population. Over two-thirds of dementia cases are attributed to Alzheimer's disease (AD), making it the leading cause. Nonetheless, AD is a complicated brain disorder with a mechanism that is still not well understood. Identifying the subtypes of AD and understanding the genetic and biomarker components could aid in the creation of new medications and provide direction for early-stage treatments for this serious disease.

Mild Cognitive Impairment (MCI) serves as an important prodromal stage of Alzheimer's Disease (AD), highlighting the urgency of early diagnosis for timely treatment and control of the condition [3]. However, quantifying the progression of MCI individuals is difficult because the brain alterations in MCI are often subtle. Traditional method uses cognitive score supported by a subjective clinical test to assess the progression of MCI [4]. In addition, biomarker measurements can also be used to trace MCI development, which is an additional information source for MCI subtype discovery [5]. Normalized amyloid measurement from PET is a common quantitative trait that is investigated throughout AD diagnosis.

Due to the high-dimensional nature of biomarker measurements, only a few important features are used during clinical evaluation. To account for this limit, several studies combine multi-modal data source and use a data-driven approach in order to identify more effective subtype groupings. Feng et al [6] uses clustering to identify subtypes via multimodal data, performs survival analysis, learns a nonlinear embedding, and analyzes canonical correlation . In order to study the data in the multimodal domain, the intersection of all modalities restricts the number of valid subjects in the study. This paper uses a much larger cohort in a single modality and proves that the pattern we observe generalizes well in domains of all other modalities.

In this paper, we leverage a set of data analysis tools specialized for high-dimensional data to identify novel subtypes in the MCI population. K-means clustering, spectral clustering, and agglomerative clustering are used as an attempt to

[1] https://www.who.int/news-room/fact-sheets/detail/dementia

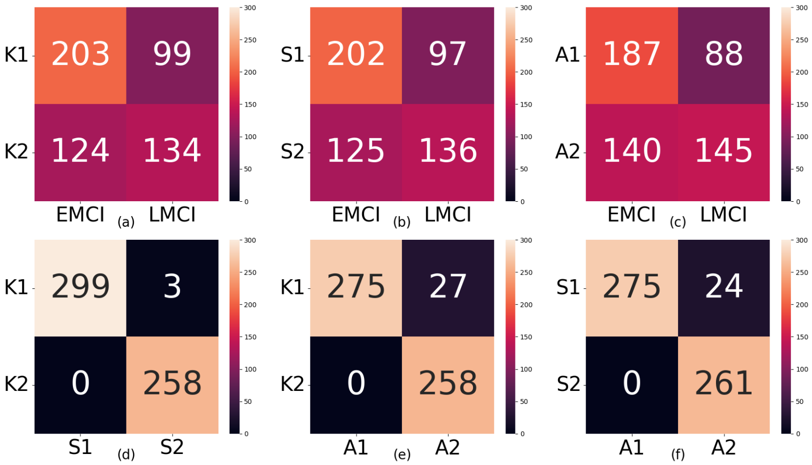


Figure 1: Confusion matrices between MCI benchmark and the three clustering results. Category K1 and K2 represent k-means cluster 1 and 2. Category S1 and S2 represent spectral cluster 1 and 2. Category A1 and A2 represent agglomerative cluster 1 and 2. Subfigures a), b), and c) compares k-means, spectral, and agglomerative clustering results with MCI benchmark respectively. Subfigures d), e), and f) performs pairwise comparisons between the three clustering results.

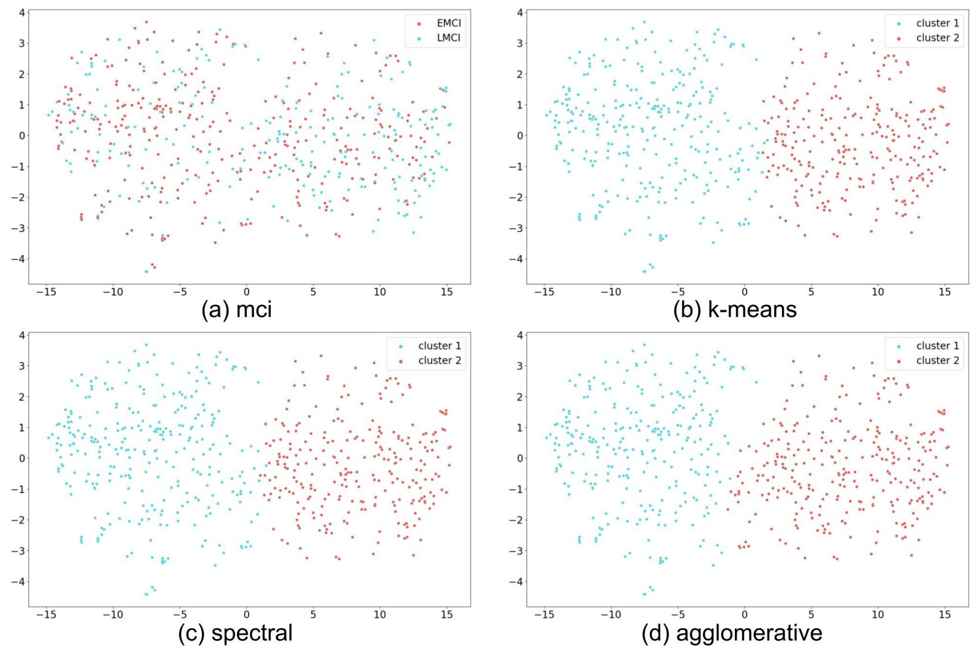
generate new subtypes. t-SNE and UMAP are applied to perform dimension reduction such that local neighborhood information across a high-dimensional manifold can be visualized. Clustering quality can be assessed qualitatively by observing the patterns in the projected 2-dimensional space. A traditional survival analysis technique called Kaplan-Meier [7] is applied as a quantitative metrics to check whether the two groups are stratified with significant risk discrepancy. After an effective subtyping result is discovered, a genome-wide association study (GWAS) is performed to find significant single nucleotide polymorphisms (SNPs) that distinguish the two subtypes. Aside from the genetic domain, important biomarker measurements are compared between the two subtypes to study the results in the imaging domain. Throughout this study, the EMCI and LMCI diagnosis labels are used as the benchmark grouping strategy.

Methods

Data: Data used in this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://adni.loni.usc.edu/) database [8]. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. All participants provided written informed consent and study protocols were approved by each participating site's Institutional Review Board (IRB). For up-to-date information, see www.adni-info.org. In this paper, we use the longitudinal amyloid imaging data and we used baseline data which totally have 560 subjects including 327 early mild cognitive impairment (EMCI), 233 late MCI (LMCI). The detailed demographic information of gender, age and education years are shown in Table 1. The race of this population is non-Hispanic white.

Table 1: Demographic information of Amyloid imaging data.

subjects	**EMCI**	**LMCI**
Number	327	233
Gender(M/F)	141/186	96/137
Age(mean±std)	72.03±7.30	74.34±8.23
Education(mean±std)	16.08±2.64	16.23±2.77



(a) mci

(b) k-means

(c) spectral

(d) agglomerative

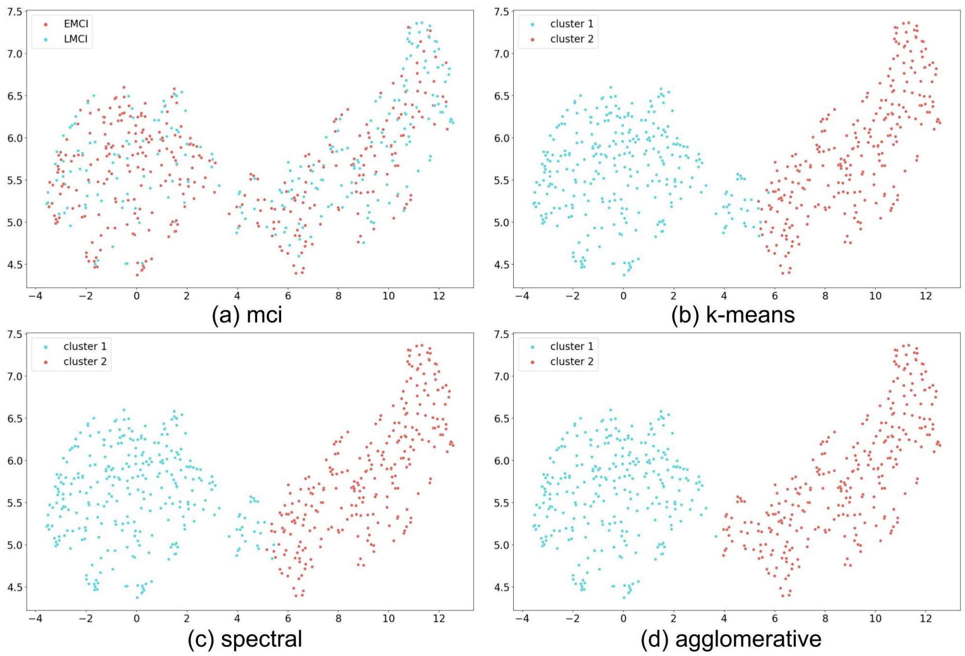
Figure 2: t-SNE visualization of 68 amyloid features of 327 EMCI subjects and 233 LMCI subjects. For each clustering algorithm, cluster 1 refers to the amyloid negative subtype and cluster 2 refers to the amyloid positive subtype.

Amyloid Imaging Data Preprocessing: Amyloid imaging data have been downloaded from the ADNI website as preprocessed. Briefly, Amyloid PET used florbetapir (18F) as a tracer to measure amyloid-β (Aβ) plaques [9]. For each subject, brain regions of interest (ROIs) were defined from structural MRI through segmentation and parcellation using Freesurfer [10]. Then, each florbetapir scan was coregistered to the corresponding MRI and calculated the mean florbetapir uptake within the predefined ROIs. All the regional Amyloid deposition was re-normalized using whole cerebellum as reference region. Finally, we have amyloid measurement in 68 cortical ROIs. More detailed image processing information can be found in [11, 12]. To remove potential bias, we then did pre-adjusted using baseline age, gender with the weight derived from healthy controls. Finally, they were normalized to zero mean and unit variance for subsequent analysis.

Subjects of Interest: 68 amyloid features are used for t-Distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction visualization. t-SNE performs dimensionality reduction such that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points with high probability. UMAP performs dimensionality reduction such that a Riemannian manifold is modeled with a fuzzy topological structure. No distinct pattern is observed between groups of EMCI and LMCI patients identified by current diagnosis labels. As a result, this study aims for identifying effective patient subtypes for the MCI cohort. Patients that are diagnosed as EMCI or LMCI at baseline, including 327 EMCI subjects and 233 LMCI subjects, are thus selected as the subjects of interest for this study.

Clustering Algorithms: Kmeans clustering, spectral clustering, and agglomerative clustering are selected as the clustering algorithms of this study. Instead of performing an empirical experiment to determine the optimal number of clusters, we fix the number of clusters to be 2 for this study for the two reasons. First, we would like to use the EMCI and LMCI diagnosis label as a benchmark grouping strategy for this study. Second, 2 distinct clusters can be visually observed in the 2D space from multiple trials of t-SNE and UMAP dimensionality reduction. All clustering algorithms are used on 560 MCI subjects at baseline, each subject with 68 amyloid features.

Visualization of Identified Clusters: t-SNE and UMAP dimensionality reduction algorithms are used to transform identified clusters into the 2D space for qualitative assessment and visualization. The visualization that reveals a more



(a) mci

(b) k-means

(c) spectral

(d) agglomerative

Figure 3: UMAP visualization of 68 amyloid features of 327 EMCI subjects and 233 LMCI subjects. Subfigures b), c), and d) shows k-means, spectral, and agglomerative clustering results respectively. For each clustering algorithm, cluster 1 refers to the amyloid negative subtype and cluster 2 refers to the amyloid positive subtype.

obvious pattern of separating two groups should have a higher clustering quality.

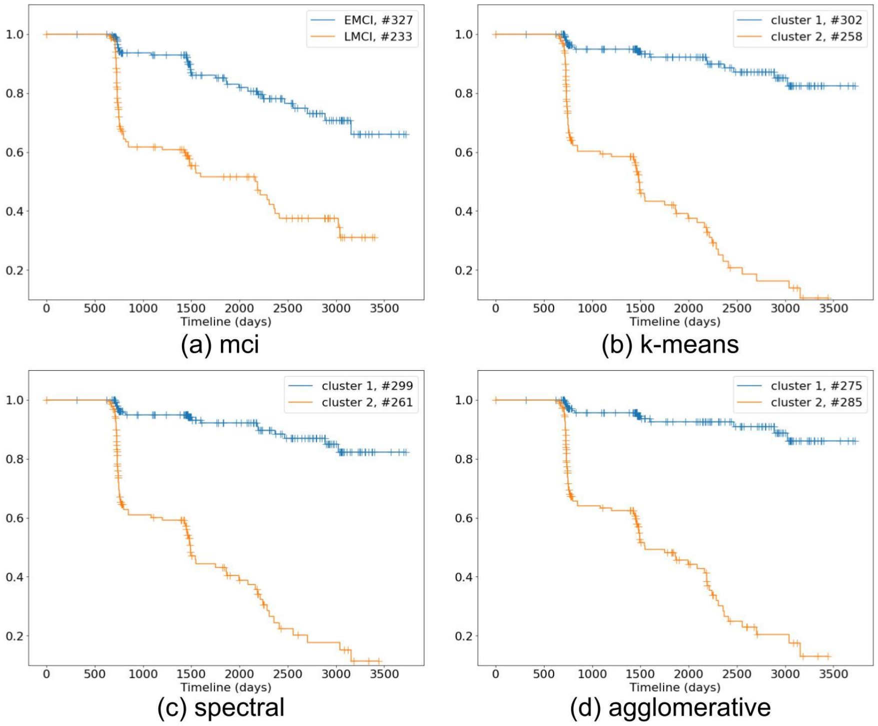
Kaplan Meier Plot of Survival Analysis of Identified Clusters: Kaplan-Meier Plot of survival analysis is performed on the two clusters to quantify the effectiveness of clustering. The event is defined as the transition from any MCI state to the AD state. Censorship refers to the transition of the subject from the MCI state to the AD state before the last observation. If the conversion occurs, the subject is considered uncensored, and if the conversion does not occur, the subject is considered censored. The Kaplan-Meier plot is created for the result of each clustering algorithm as a qualitative visualization of survival probability between the two clusters at each time stage. P-value is used as a quantitative metric to evaluate the quality of clusters.

Genetic Analysis of Identified Clusters and Functional Annotation: Genome-wide association study (GWAS) [13, 14, 15] is performed to identify SNPs with the most significant variance between the two clusters. Plink is used as the execution tool of GWAS for this study. A total of 565373 SNPs were used in this GWAS study. A logistic regression is performed with cluster 1 patients as phenotype value 0 and cluster 2 patients as phenotype value 1. Gender, age, and education are selected as the covariates of this study. P-values of the regression are used to identify the most significant SNPs. Manhattan plots are produced to qualitatively assess the quality of clustering. Functional annotation is performed through functional mapping and annotation of genetic associations with FUMA (FUMA GWAS) [16] [2]. The significant SNPs are projected to nearest genes defined by base pair distances. Biological pathways associated with the two clusters are identified such that those claimed to be relevant with AD pathogenesis in prior literature serve as a indicator as the effectiveness of our results.

Biomarker Analysis of Identified Clusters: Relevant AD biomarkers examined in the QT-Pad challenge are used to evaluate the promise of our proposed clusters [3]. If the identified cluster produces a difference between groups that resembles or surpasses the observed pattern in the MCI benchmark, a more effective cluster result is obtained. 12 out

[2] https://fuma.ctglab.nl

[3] http://www.pi4cs.org/qt-pad-challenge



(a) mci

(b) k-means

(c) spectral

(d) agglomerative

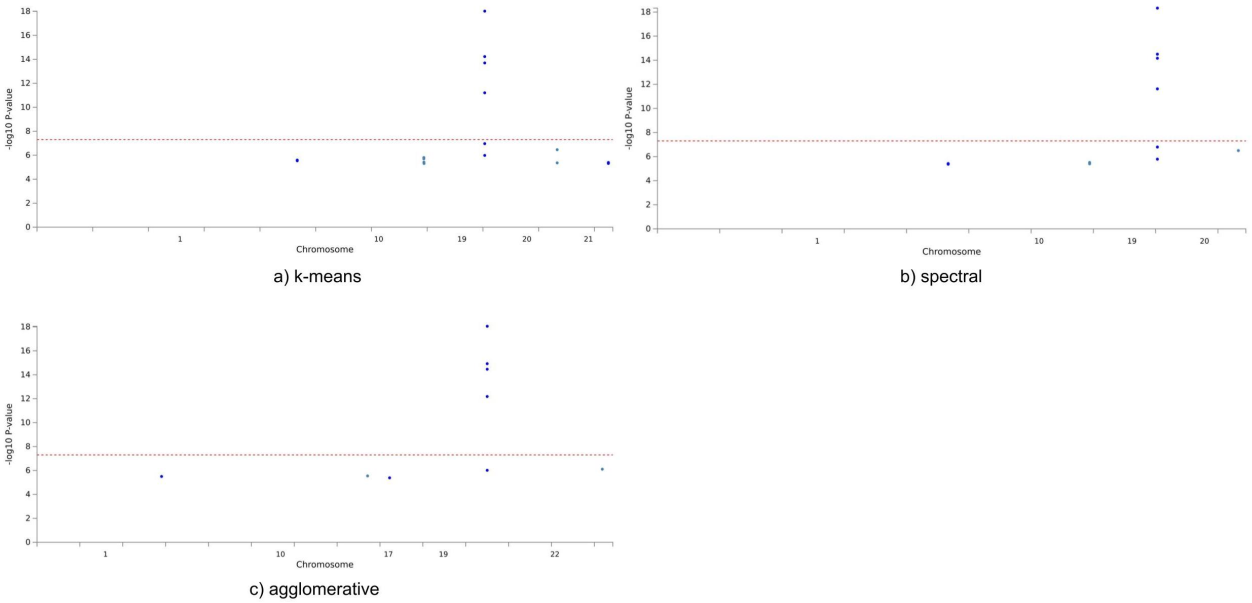
Figure 4: Kaplan-Meier survival curve of 68 amyloid features of 327 EMCI subjects and 233 LMCI subjects. For each clustering algorithm, cluster 1 refers to the amyloid negative subtype and cluster 2 refers to the amyloid positive subtype

of 15 biomarkers other than AV45, which is an amyloid-based biomarker, are used in this study.

Results

Confusion Matrix of Different Clustering Algorithms: The cluster labels generated from the algorithms are aligned with MCI benchmark such that cluster 1 contains a higher portion of EMCI subjects and cluster 2 contains a higher portion of LMCI subjects. Figure 1(a, b, c) compares the results of clustering algorithms with the MCI benchmark. All confusion matrices have chi-square test p-values less than 1e-4, which indicate that the identified clusters are different from the benchmark at a statistically significant level. Figure 1(d, e, f) shows a pairwise comparison between the clustering results. Figure 1(d) indicates that the clusters identified by k-means and spectral clustering are close to identical with 3 subjects as exceptions, which is a 0.536% difference. Figure 1(e, f) suggest that agglomerative clustering identifies a slightly different subtype than that of k-means and spectral clustering. While over 90% of the clusters still align, there is a close to 10% variance where agglomerative clustering classifies about 20 subjects as the subtype with higher risk in MCI development.

t-SNE and UMAP Visualization: Figure 2 shows the t-SNE dimensionality reduction to 2-dimensional space while Figure 3 shows the UMAP dimensionality reduction to 2-dimensional space. While randomized initialization leads to different projects every time the algorithm is executed, a consistent pattern of 2 distinct clusters can be observed in the projected population, which explains why we decide the number of clusters to be two in this study. Both visualizations qualitatively evaluate the clustering qualities of each result: the MCI benchmark exhibits no significant pattern in both t-SNE and UMAP while all clustering algorithms yield significant pattern of separation between the two subtypes in the projected space. The comparison provides strong evidence that our study identifies a more effective subtype in the high-dimensional data manifold. As discussed in the confusion matrix session, k-means and spectral clustering produce an extremely similar result with some minor disagreeement in the border area while agglomerative tend to



a) k-means

b) spectral

c) agglomerative

Figure 5: Manhattan Plot of GWAS summary statistics. Only SNPs with p-values less than 5e-6 are selected as candidate significant SNPs. The final cutoff uses a conventional threshold of 5e-8 to filter important SNPs.

classify more subjects in "zone of confusion" as cluster 2 where subjects are considered to carry more risk of AD pathogenesis.

Table 2: Kaplan-Meier plot p-value

	MCI	k-means	spectral	agglomerative
p-value	3.136e-12	**3.511e-29**	7.427e-28	7.393e-26

Kaplan-Meier Plot as Visualization and Validation: While the previous analysis uses only the measurements at baseline, we could also leverage the longitudinal data of the followup visits to analyze the quality of clustering. Survival analysis is a common approach that assesses and predicts the risk of populations over time. Figure 4 shows the Kaplan-Meier plots of the MCI benchmark and the proposed subtypes. The event is defined as transition to AD while censorship is defined as whether the subject transitions to AD before it leaves inspection. Each survival analysis calculates a p-value that quantifies how much the two subtypes differs in survival risks. A lower p-value indicates a more significant difference of survival risk between the two subtypes, which refers to a more effective subtype. The MCI benchmark has a p-value of 3.136e-12, which is a sanity check that the EMCI and LMCI population differs significantly in the probability of transitioning to AD. K-means, spectral, and agglomerative clustering have p-values of 3.511e-29, 7.427e-28 and 7.393e-26. Quantitatively, the clustering algorithms identify subtypes that are more different in the risk of transitioning to AD. Qualitatively, we can visually observe that the subtypes start to diverge more significantly than the benchmark from day 1500 through the end of the duration. The Kaplan-Meier plot provides both quantitative and qualitative that the new subtypes are more effectively not only at the baseline level, but also throughout the progression of MCI in a duration of 10 years.

Overview of Evaluation: While all qualitative and quantitative metrics suggest that the newly proposed subtypes are more effective than that of the MCI benchmark, it is important to interpret how the novel subtypes divide the population into groups with different genetic basis, biomarker measurements, and biological pathways. The introduction of additional datasets not only provides alternative perspectives of explaining our subtypes but also validates the clus-

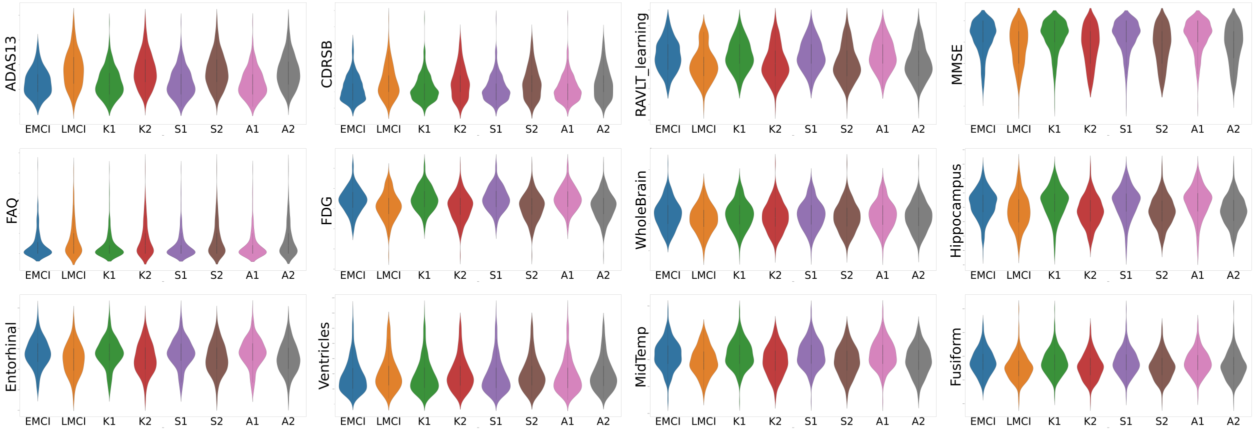


Figure 6: Violin plot of 12 biomarkers examined in the QT-PAD challenge. Category K1 and K2 represent k-means cluster 1 and 2. Category S1 and S2 represent spectral cluster 1 and 2. Category A1 and A2 represent agglomerative cluster 1 and 2.

Table 3: Significant SNPs identified by Plink GWAS analysis

k-means			**spectral**			**agglomerative**		
chr	SNP (Posi)	p-value	chr	SNP (Posi)	p-value	chr	SNP (Posi)	p-value
1	6700744	2.905e-06	1	6700744	4.307e-06	1	11588804	3.167e-06
1	4649215	2.544e-06	1	4649215	3.802e-06	10	1325904	2.857e-06
10	10509549	2.001e-06	10	10509549	3.971e-06	17	4790408	4.113e-06
10	10509550	1.571e-06	10	10509550	3.077e-06	19	2075650	**6.578e-13**
10	1325904	3.799e-06	19	2075650	**2.389e-12**	19	157582	**1.189e-15**
10	10749593	4.904e-06	19	157582	**3.141e-15**	19	1160985	9.661e-07
19	2075650	**6.233e-12**	19	8106922	1.633e-06	19	769449	**3.442e-15**
19	157582	**5.962e-15**	19	1160985	1.611e-07	19	4420638	**8.843e-19**
19	8106922	1.037e-06	19	769449	**6.846e-15**	22	5771761	7.804e-07
19	1160985	1.085e-07	19	4420638	**4.714e-19**			
19	769449	**2.017e-14**	20	1557183	3.174e-07			
19	4420638	**9.626e-19**						
20	6089530	4.359e-06						
20	1557183	3.473e-07						
21	2836445	4.861e-06						
21	2836469	4.048e-06						

tering quality through an additional source. In the genetic domain, Plink is used to perform GWAS, identify SNPs that differ significantly between the two subtypes. More effective subtype should identify more significant SNPs as the subtypes should differ more in the genetic domain. These SNPs can be mapped to nearest genes with relevant biological pathways to interpret how the two subtypes differ genetically and functionally. Alternatively, we could also use alternative biomarker measurements that are significant and compare the new subtype with MCI benchmark to validate our results.

Genetic Domain: In the genetic domain, GWAS was performed to analyze the genetic basis of the subtypes proposed by the clustering algorithms. As shown in Figure 5, only SNPs with p-values less than 5e-6 are selected as candidate significant SNPs, where a complete list is present in Table 3. The MCI benchmark is not included in Figure 5 because its summary statistics from GWAS does not output any SNPs with p-values less than 5e-6. A final p-value threshold

Table 4: Biological processes identified by significant SNPs from FUMA-GWAS.

k-means	**spectral**	**agglomerative**
Very Low Density Lipoprotein Particle Clearance	Very Low Density Lipoprotein Particle Clearance	Very Low Density Lipoprotein Particle Clearance
Triglyceride Rich Lipoprotein Particle Clearance	Triglyceride Rich Lipoprotein Particle Clearance	Triglyceride Rich Lipoprotein Particle Clearance
Triglyceride Metabolic Process	Triglyceride Metabolic Process	Phospholipid Efflux
Lipid Catabolic Process	Lipid Catabolic Process	High Density Lipoprotein Particle Remodeling
Phospholipid Efflux	Phospholipid Efflux	
High Density Lipoprotein Particle Remodeling	High Density Lipoprotein Particle Remodeling	
Neutral Lipid Metabolic Process	Neutral Lipid Metabolic Process	

of 5e-8 is used to obtain the significant SNPs, where all three subtypes identify rs2075650, rs769449, rs157582, and rs4420638 at chromosome 19.

Single Nucleotide Polymorphysim: Chromosome 19 and the Apolipoprotein E4 (APOE4) gene has always been considered as one of the most important genetic basis indicator of AD [17, 18]. TOMM40 is an adjacent gene of APOE4, and its allele has been shown to encode high likelihood of Alzheimer's onset in former studies [19]. rs2075650 is a polymorphism located at TOMM40 highly correlated with AD in various populations and potentially causes neuroinflammation in AD [20, 21]. rs769449 is a SNP located at APOE and a proxy of rs429358 which encodes the APOE4 allele [22]. rs157582 is another SNP located at TOMM40 that's significantly correlated with dementia and contributes to a high polygenetic risk score for AD [23, 24]. rs4420638 is also a proxy SNP to rs429358 with a minor allele G that leads to lower mini-mental state examination (MMSE) score, higher AD assessment scale-cognitive subscale 11 (ADAS-cog 11) score, and smaller entorhinal volume [25]. All of the SNPs identified in the GWAS analysis are supported by literature reviews and earlier studies that they are highly correlated with AD and dementia, which shows that the new subtypes preserve genetic basis pattern of AD pathogenesis.

Biological Processes: Functional Annotation is performed through FUMA GWAS. A list of biological processes identified by the significant SNPs are shown in Table 4. Amyloid-beta accumulation in the brain is known to play a central role in the development of AD. High-density lipoproteins (HDL) are crucial for maintaining cholesterol balance. APOE is also a key HDL-associated protein involved in lipid transport in both the periphery and central nervous systems. Former research revealed that Changes in high density lipoprotein (HDL) particles are specific to the APOE genotype in Alzheimer's disease, and HDL function and size are highly correlated with cognitive ability [26]. Very low density lipoproteins particle clearance, triglyceraide rich lipoprotein particle clearance, and high density lipoprotein particle remodeling are all lipoprotein-related biological processes that could affect AD in a complicated process. On the other hand, brain is highly enriched by lipids as it is the important component of forming cell membranes. Disruption of lipid homeostasis is associated with neurologic disorders and neurodegenerative diseases such as AD [27]. Aside from the crucial role of phospholipid and lipid in membrane formation, lipid also plays an important role of cell signaling and other physiological functions. Prior research has shown that abnormal lipid synthesis, degradation, traffic, and modification can result in AD pathogenesis through mechanisms including: 1) amyloid-beta production, aggregation, and clearance, 2) APOE isoform neuroinflammation, tau phosphorylation, and 3) synaptic function, learning, and memory [28]. Lipid catabolic process, phospholipid efflux, and neutral lipid metabolic process can all be indicators of lipid disruptions. All biological processes are involved in identified mechanisms of AD pathogenesis, which shows how the new subtype differentiates the MCI population based on biological processes.

Table 5: Biomarker distribution negative base 10 logarithm of p-value

	MCI	**k-means**	**spectral**	**agglomerative**
Average negative base10 logarithm of p-value	5.679	13.835	14.073	**14.122**

Biomarker Domain: In the biomarker domain, 12 out of 16 biomarkers identified in the QT-PAD challenge are selected to analyze the new subtypes compared to the MCI benchmark. The biomarker amyloid PET is the biomarker used in this study while biomarkers CSG ABETA, CSF TAU, CSF PTAU include missing data and are excluded from this study to avoid bias. As shown in Figure 6, the directional patterns of the MCI benchmark subtypes are preserved across all 12 biomarkers, with a substantial improvement of difference in ADAS13, CDRSB and FS Entorhinal. While the violin plot serves as a qualitative evidence of the effectivesness of our results, t-test is also performed for each of the 12 biomarkers. The average of the 12 negative base10 logarithms of the p-values from the t-test are used as a quantitative metric. While the MCI benchmark has an average negative base10 logarithm of 5.679, k-means, spectral, and agglomerative subtypes have average negative base10 logarithms of 13.835, 14.073 and 14.122. This metric shows that the subtypes from clustering algorithms differ more on a population basis in all 12 biomarkers identified by QT-PAD, which validates the quality of the new subtypes.

Discussion

Limit of This Study and Future Directions: The Kaplan-Meier Plot in this study is used as a visualization and qualitative assessment of risk over time. In order to assess the development of AD-transition risk on a longitudinal basis, a Cox model survival analysis should be performed. As the conventional Cox survival analysis evaluates features on a linear basis, we will work on introducing non-linearity by implementing Cox survival analysis via deep neural networks. By projecting amyloid features into a nonlinear latent space, we could study the pathogenesis of AD along a amyloid defined manifold that provides novel insights.

Conclusion

Alzheimer's disease is a complex neurodegenerative disease that serves as one of the leading causes of death in the United States. Identify the stage of mild cognitive impairment progression is an important process for early diagnosis and timely treatment. Traditional diagnosis combines a cognitive score and biomarker measurements to divide MCI patients into early MCI and late MCI subjects. This study leverages a data-driven approach to identify a novel subtype in amyloid PET measurement that generalizes well in domains of other modalities including genetic basis, FDG, VBM, freesurfer, and a few other AD biomarkers of interest. A pattern of two distinct clusters are observed in the projected 2-dimensional nonlinear data manifold, which serves as a motivation of this study. Several common clusters lead to a similar subtype that follows the trend we observed in dimensionality reduction plots. Subsequent analysis show promising results across longitudinal risk of transitioning to AD, genetic pathway associated with biological pathways, and biomarker distribution studies compared to the MCI benchmark. The subtype proposed by the clustering algorithms outperforms the MCI benchmark across all qualitative and quantitative metrics and preserves or exemplifies directional patterns of the benchmark in all AD biomarkers of interest. This work provides important insights regarding future diagnosis of MCI subjects from a data-driven perspective that could assist clinical trials by providing a quantitative reference.

Acknowledgments

This work was supported in part by National Institutes of Health grants U01 AG068057, U01 AG066833, R01 AG071470 and RF1 AG068191; and National Science Foundation grant IIS 1837964. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH- 12-2-0012).

References

1. Association A, et al. 2012 Alzheimer's disease facts and figures. Alzheimer's & Dementia. 2012;8(2):131-68.
2. Uwishema O, Mahmoud A, Sun J, Correia IFS, Bejjani N, Alwan M, et al. Is Alzheimer's disease an infectious neurological disease? A review of the literature. Brain and Behavior. 2022;12(8):e2728.
3. Rasmussen J, Langerman H. Alzheimer's disease–why we need early diagnosis. Degenerative neurological and neuromuscular disease. 2019:123-30.
4. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive im-

pairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011 May;7(3):270-9.

5. Gauthier S, Patterson C, Gordon M, Soucy JP, Schubert F, Leuzy A. Commentary on "Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease." A Canadian perspective. Alzheimers Dement. 2011 May;7(3):330-2.
6. Feng Y, Kim M, Yao X, Liu K, Long Q, Shen L. Deep multiview learning to identify imaging-driven subtypes in mild cognitive impairment. BMC Bioinformatics. 2022 Sep;23(Suppl 3):402.
7. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). Bmj. 1998;317(7172):1572-80.
8. Weiner MW, Veitch DP, Aisen PS, et al. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement. 2013;9(5):e111-94.
9. Okamura N, Yanai K. Florbetapir (18F), a PET imaging agent that binds to amyloid plaques for the potential detection of Alzheimer's disease. IDrugs. 2010;13(12):890-9.
10. Fischl B. FreeSurfer. Neuroimage. 2012;62(2):774-81.
11. Landau SM, Breault C, Joshi AD, Pontecorvo M, Mathis CA, Jagust WJ, et al. Amyloid-β imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. Journal of Nuclear Medicine. 2013;54(1):70-7.
12. Landau S, Jagust W. Flortaucipir (AV-1451) processing methods. Alzheimer's Disease Neuroimaging Initiative. 2016.
13. Shen L, Thompson PM. Brain Imaging Genomics: Integrated Analysis and Machine Learning. Proceedings of the IEEE. 2020;108(1):125-62.
14. Saykin AJ, Shen L, Yao X, et al. Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. Alzheimers Dement. 2015;11(7):792-814.
15. Shen L, Thompson PM, Potkin SG, et al. Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. Brain Imaging Behav. 2014;8(2):183-207.
16. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017 Nov;8(1):1826.
17. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet. 1994 Jun;7(2):180-4.
18. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993 Aug;261(5123):921-3.
19. Soyal SM, Kwik M, Kalev O, Lenz S, Zara G, Strasser P, et al. A TOMM40/APOE allele encoding APOE-E3 predicts high likelihood of late-onset Alzheimer's disease in autopsy cases. Mol Genet Genomic Med. 2020 Aug;8(8):e1317.
20. Huang H, Zhao J, Xu B, Ma X, Dai Q, Li T, et al. The TOMM40 gene rs2075650 polymorphism contributes to Alzheimer's disease in Caucasian, and Asian populations. Neurosci Lett. 2016 Aug;628:142-6.
21. Chen YC, Chang SC, Lee YS, Ho WM, Huang YH, Wu YY, et al. Genetic Variants Cause Neuroinflammation in Alzheimer's Disease. Int J Mol Sci. 2023 Feb;24(4).
22. Del-Aguila JL, ndez MV, Schindler S, Ibanez L, Deming Y, Ma S, et al. Assessment of the Genetic Architecture of Alzheimer's Disease Risk in Rate of Memory Decline. J Alzheimers Dis. 2018;62(2):745-56.
23. Driscoll I, Snively BM, Espeland MA, Shumaker SA, Rapp SR, Goveas JS, et al. A candidate gene study of risk for dementia in older, postmenopausal women: Results from the Women's Health Initiative Memory Study. Int J Geriatr Psychiatry. 2019 May;34(5):692-9.
24. Wang T, Han Z, Yang Y, Tian R, Zhou W, Ren P, et al. Polygenic Risk Score for Alzheimer's Disease Is Associated With Ch4 Volume in Normal Subjects. Front Genet. 2019;10:519.
25. Guo Y, Xu W, Li JQ, Ou YN, Shen XN, Huang YY, et al. Genome-wide association study of hippocampal atrophy rate in non-demented elders. Aging (Albany NY). 2019 Nov;11(22):10468-84.
26. Hong BV, Zheng J, Agus JK, Tang X, Lebrilla CB, Jin LW, et al. Genotype-Specific. Biomedicines. 2022 Jun;10(7).
27. Kao YC, Ho PC, Tu YK, Jou IM, Tsai KJ. Lipids and Alzheimer's Disease. Int J Mol Sci. 2020 Feb;21(4).
28. Liu Q, Zhang J. Lipid metabolism in Alzheimer's disease. Neurosci Bull. 2014 Apr;30(2):331-45.