

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

Neurological Disease Subtyping and Prediction

Session 12

Authors: Ruiming Wu, Bing He, Bojian Hou, Andrew J Saykin, Jingwen Yan, Li Shen

Presenter: Bojian Hou

University of Pennsylvania

wuru330@gmail.com

Shen Lab



I and my partner have no relevant relationships with commercial interests to disclose.

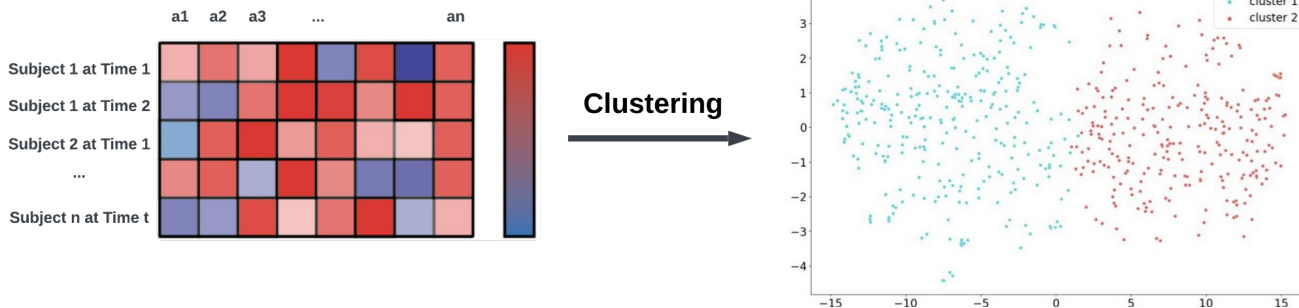
Learning Objectives

After participating in this session the learner should be better able to:

- Apply dimensionality reduction and unsupervised learning on high-dimensional biomarker measurements of Alzheimer's Disease and perform data-driven methods to study and analyze these data.
- Implement a pipeline of identifying novel subtypes and verify the effectiveness of new results through several quantitative and qualitative metrics using longitudinal and multimodal data.
- Use functional annotation to identify relevant biological processes based on significant SNPs from a genome-wide association study and explain the difference of subtype populations in a genetic and a biological pathway basis.

Background: Traditional clinical evaluation

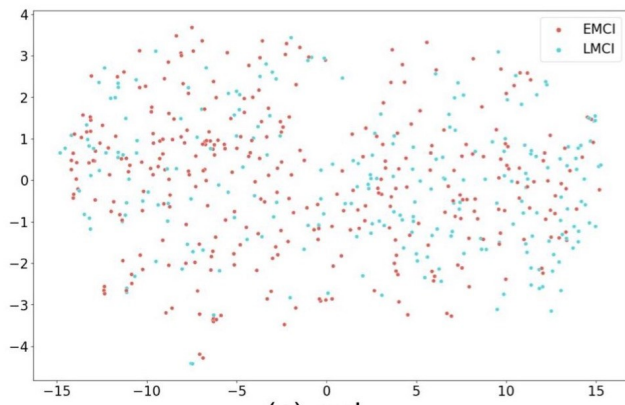
Quantifying the progression of MCI individuals is difficult because the brain alterations in MCI are often subtle. Traditional method uses cognitive score supported by a subjective clinical test to assess the progression of MCI. Multimodal imaging and biomarker datasets are used to differentiate early and late MCI subjects.



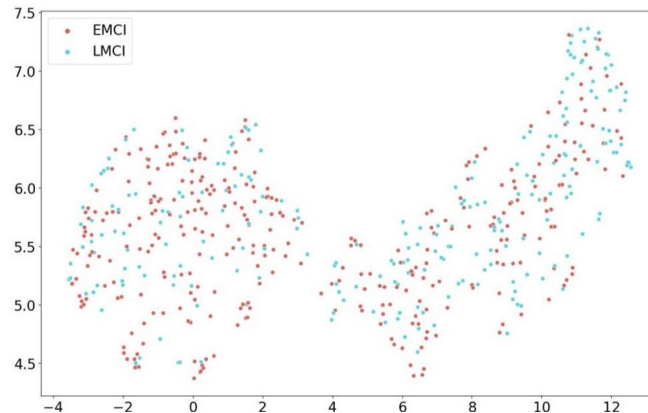
Challenge: Current diagnosis fails to identify subtypes in the amyloid domain.

Proposed new subtype: amyloid-specific: Our study partitions the MCI population into amyloid positive and negative subtypes and identifies greater statistical significance in longitudinal risks, SNPs, genetic pathways, and biomarkers compared to diagnosis benchmark.

Challenge :



a) t-SNE



b) UMAP

Dimensionality reduction shows that diagnosis from traditional clinic evaluation cannot identify distinct clusters in Euclidean manifolds (as shown in t-SNE results) or Riemann manifolds (as shown in UMAP results).

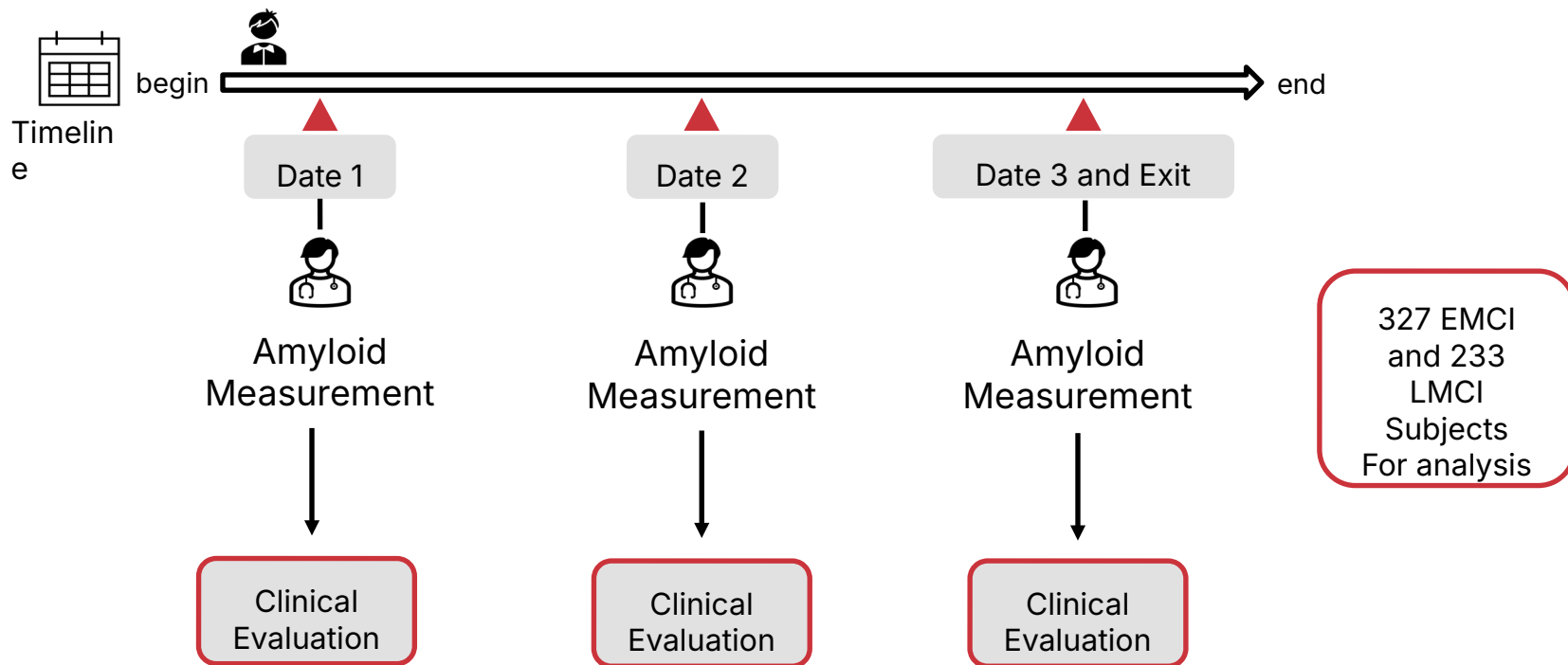
As a result, a novel subtype is proposed in the amyloid domain to differentiate the MCI population from the amyloid perspective, which could provide important insights.

We propose the following pipeline to identify novel subtype and interpret biological insights:

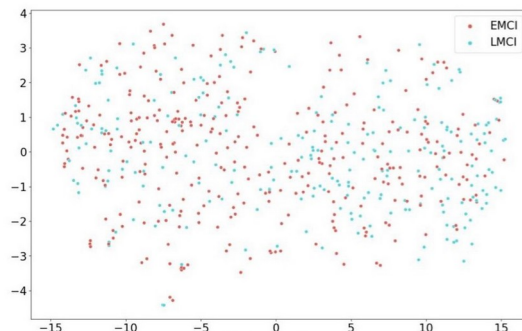
- Use a variety of clustering algorithms to identify subtypes in the population within the amyloid domain
 - K-means
 - Spectral
 - Agglomerative
- Validate the clustering results qualitatively with dimensionality reduction
 - t-SNE
 - UMAP
- Visualize risk development through Kaplan-Meier plot to evaluate the final clustering performance across time
- Identify significant SNPs, biological pathways, and biomarkers with computed p-values

The results demonstrate that using various clustering methods on samples within the amyloid space identify a robust and novel subtype compared to the conventional subtype MCI labels.

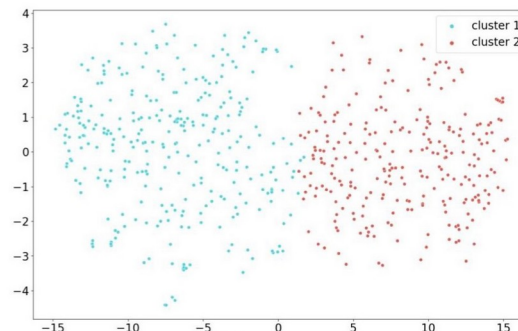
Longitudinal Study of Patients



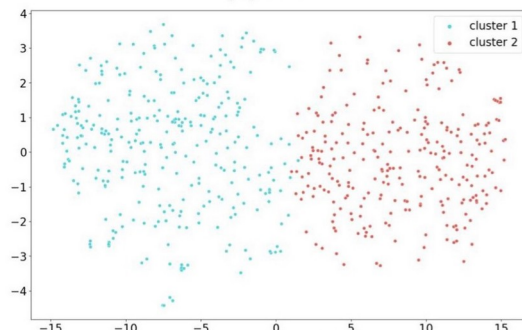
Result: t-SNE



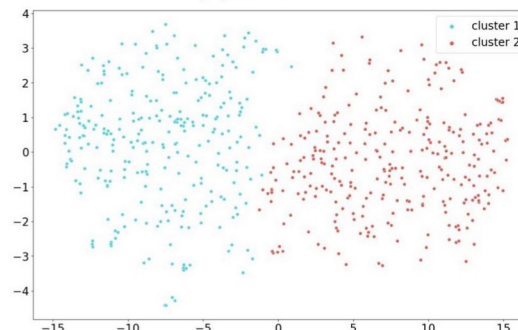
(a) mci



(b) k-means

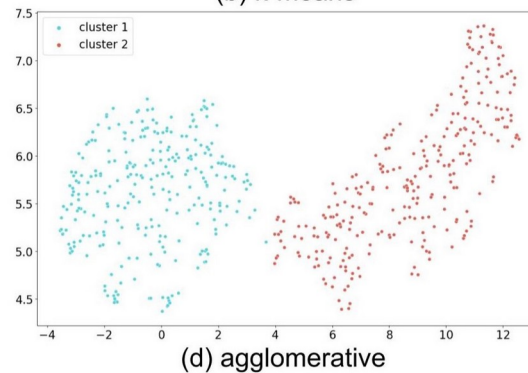
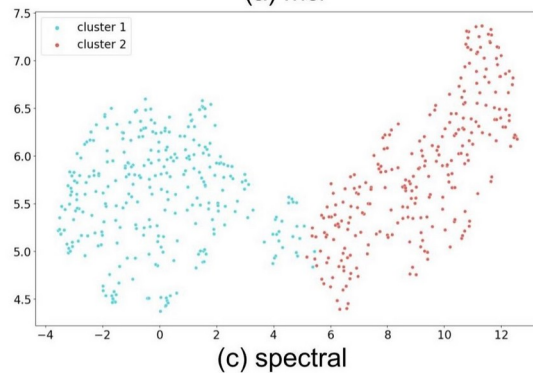
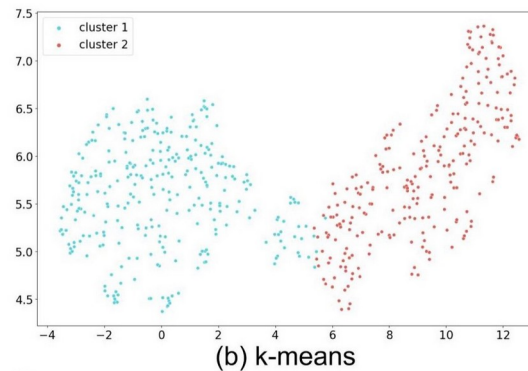
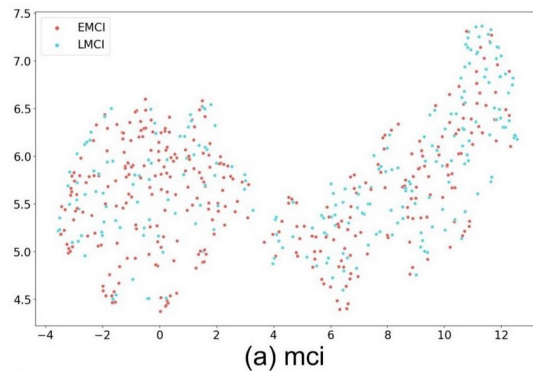


(c) spectral

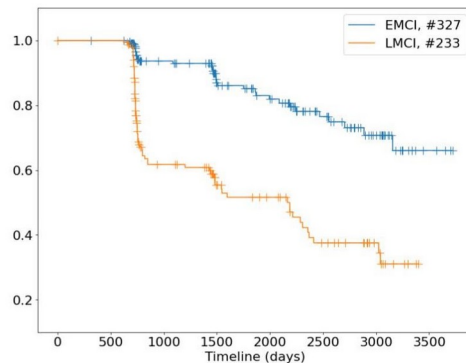


(d) agglomerative

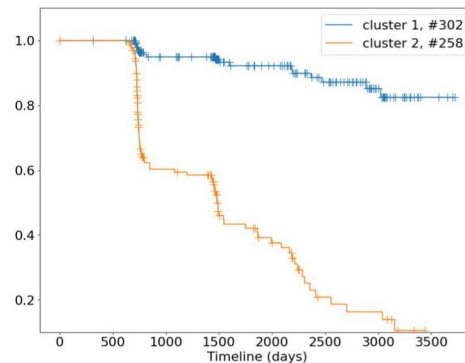
Result: UMAP



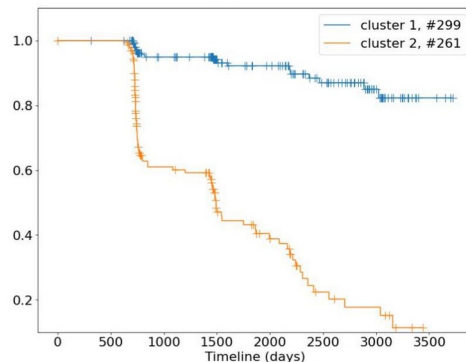
Result: Kaplan-Meier Plot



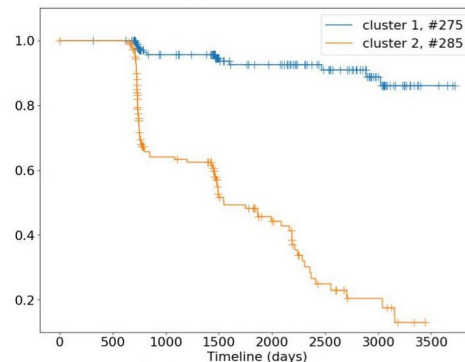
(a) mci



(b) k-means



(c) spectral



(d) agglomerative

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						

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	

P-values indicates the probability that a SNP is differentially expressed within two populations. The smaller the p value, the better the two subtypes are differentiated in the genetic domain. Functional annotation is performed based on the list of SNPs and outputs the most correlated biological pathways. Because most of those pathways are related to Alzheimer's disease, our results demonstrate that our methods can better stratify the MCI patients for better subtype discovery.

Conclusion

- This paper highlights the potential of amyloid-specific subtypes to reveal important insights of Alzheimer's pathogenesis.
- By performing clustering on the MCI population, we generate a similar partition of amyloid positive and negative subtypes across three common algorithms. Validated qualitatively by dimensionality reduction and visualization, the novel subtypes identified significant longitudinal risk development, SNPs and biological pathways between one another .
- By providing valuable insights from the amyloid perspective, this study reveals the power of domain-specific subtypes in studying the longitudinal development of disease.

Limitations and Future Directions

- The current study uses Kaplan-Meier plot to visualize the chance of survival across time. A traditional or deep learning based Cox model can be fitted to the dataset to quantify the significance of each amyloid feature.
- The current study focuses on the amyloid domain only. Future studies can use a similar pipeline to reveal pathogenesis insights across multimodal domains and combine those information.

Thank you!

