Temporal Clustering in Alzheimer's Patients Dataset: A Longitudinal, Multi-Biomarker Pipeline with Unsupervised Phenotyping and Supervised Validation

Mirza Shaheen Iqubal
M.Sc. Data Science, Friedrich-Alexander-Universität Erlangen–Nürnberg (FAU)
Email: mirzashaheen.iqubal@gmail.com

***Abstract*—Alzheimer's disease (AD) is a progressive neurodegenerative disorder with heterogeneous rates of decline across patients. Most machine learning (ML) studies in AD have focused on cross-sectional snapshots, limiting their ability to characterize temporal progression. I present a reproducible pipeline for temporal phenotyping that: (i) preprocesses irregular longitudinal follow-ups (filtering, time alignment, interpolation, normalization), (ii) constructs trajectory embeddings using Principal Component Analysis (PCA) [10], (iii) discovers phenotypes with KMeans clustering [11], and (iv) validates clinical utility with a Support Vector Machine (SVM) classifier [12]. Using OASIS-style longitudinal subsets [2], I identify slow and fast progressors with *moderate* cluster separation (mean silhouette $= 0.32$) [9] and *strong* supervised performance (accuracy $= 96\%$, ROC-AUC $= 0.985$ [13], F1 $= 0.95$). The approach is interpretable, computationally efficient, and readily extensible to larger cohorts (e.g., full OASIS-3) and richer biomarker sets (e.g., PET, CSF). I discuss clinical implications, limitations, and directions for advanced temporal modeling (DTW, HMM, T-LSTM).**

***Index Terms*—Alzheimer's disease, longitudinal data, temporal clustering, PCA, KMeans, SVM, MMSE, CDR, nWBV, eTIV, ROC, silhouette, interpretability.**

I. Introduction

Alzheimer's disease (AD) is the most common cause of dementia and one of the major health challenges of our time. It is estimated that more than 55 million people live with dementia worldwide, and this number is projected to almost triple by 2050 [1]. AD is progressive in nature: patients may experience a slow decline over many years, while others deteriorate much more rapidly. This variability makes it difficult to predict outcomes, plan treatment, or recruit patients effectively for clinical trials.

Much of the existing machine learning (ML) research in AD has focused on *cross-sectional* data, where each patient is represented by a single visit. While useful for classification at a given point in time, cross-sectional models cannot capture how the disease evolves. In contrast, *longitudinal data* provide repeated measurements of the same individuals, making it possible to model and compare progression patterns. The OASIS-3 dataset [2], for example, contains repeated cognitive assessments such as the Mini-Mental State Examination (MMSE) [4], clinical staging via the Clinical Dementia Rating (CDR) [5], and MRI-derived brain volume measures (e.g., eTIV, nWBV processed with FreeSurfer [6]). These trajectories offer a more complete picture of disease evolution.

In this work, I present a pipeline for **temporal clustering** of longitudinal Alzheimer's data. Instead of grouping patients by single-time measurements, I cluster them by their progression patterns across multiple visits. Our pipeline addresses practical challenges of working with longitudinal data, including irregular follow-up intervals, missing values, and correlated biomarkers. To achieve this, I preprocess the data with filtering, interpolation, and normalization; reduce redundancy using PCA [10]; perform clustering with KMeans [11]; and validate the results with an SVM [12].

The main contributions of this study are threefold: (1)I propose a reproducible workflow for temporal clustering of Alzheimer's progression, (2) I demonstrate that the identified clusters reflect clinically interpretable patterns, and (3) I show that temporal features derived from clustering support accurate classification of disease status (ROC-AUC ≈ 0.985 [13]). Together, these results highlight the value of temporal modeling for understanding AD progression and point to opportunities for scaling the approach to richer multimodal datasets.

- **Motivation.** Temporal modeling enables the discovery of *progression phenotypes*—for example, slow vs. fast decliners—that can inform prognosis, monitoring frequency, caregiver planning, and clinical trial enrichment.
- **Objective.** I develop and evaluate a *temporal clustering* pipeline that groups patients by longitudinal trajectories rather than snapshots, then quantify its clinical utility via supervised classification.
- **Contributions.**
  - A practical pipeline for irregular follow-ups (filtering, alignment, interpolation, z-scoring).
  - PCA-based trajectory embeddings that reduce redundancy and noise in multivariate longitudinal biomarkers [10].

– Unsupervised phenotyping with KMeans, validated by silhouette analysis [11], [9].
– Supervised validation (SVM) showing high separability (AUC ≈ 0.985) [12], [13].
– Transparent, reproducible implementation using scikit-learn [14], suitable for scaling to full OASIS-3 [2].

II. Related Work

A. Temporal Clustering in Neuroimaging

The idea of clustering temporal biomedical signals has a long tradition. In functional MRI, Goutte et al. [7] demonstrated that grouping time series can uncover latent brain activity patterns. However, similarity measures are often challenged by temporal coherence and noise. Later, Chi et al. [8] showed that explicitly incorporating temporal smoothness into clustering improves robustness for evolving data streams. These early works highlight the importance of time-aware models, though they were not directly applied to Alzheimer's disease.

B. Alzheimer's Datasets and Machine Learning

The release of large public datasets has transformed Alzheimer's research. OASIS-3 [2] and the Alzheimer's Disease Neuroimaging Initiative (ADNI) [3] provide open access to MRI, PET, and clinical data, making reproducible studies possible. Many machine learning methods have been applied to these datasets, ranging from traditional classifiers such as SVMs and Random Forests [15], [16] to more recent deep learning approaches [17], [18]. While these studies achieved strong cross-sectional performance, they often treat disease status as static, overlooking longitudinal progression and the heterogeneity of decline across patients. Only a few works attempt to model trajectories or progression subtypes directly, leaving a gap that our study addresses.

C. Cluster Validation in Clinical Data

Cluster validity is a critical issue in medical applications, where overlapping biology and noisy measurements are common. The silhouette score [9] is widely used because it balances within-cluster cohesion and between-cluster separation in a simple, interpretable way. Alternatives such as the Calinski–Harabasz and Davies–Bouldin indices exist, but silhouette is often preferred in exploratory biomedical contexts due to its intuitive interpretation.

D. Comparison to Survival Analysis

Another line of research uses survival models to predict time-to-event outcomes, such as conversion from mild cognitive impairment to dementia. These methods focus on when a clinical event will occur, rather than how a patient's biomarkers evolve continuously over time. Our approach is complementary: by clustering trajectories of cognition and brain volume,I capture continuous patterns of decline. Future work could combine both perspectives, for example by using cluster membership as a covariate in survival analysis.

TABLE I: Comparison of representative machine learning studies in AD (illustrative).

Study	Dataset	Temporal?	Method	Metric
Klöppel [15]	ADNI	No	SVM (MRI)	Acc
Liu [16]	ADNI	No	Multi-modal	AUC
Pan [18]	ADNI	No	CNN (MRI)	AUC
Ours	OASIS (subset)	**Yes**	PCA+KMeans+SVM	AUC/F1

III. Dataset

A. OASIS-3 Longitudinal Cohort

I used the longitudinal cohort of the OASIS-3 dataset [2], which provides repeated cognitive, clinical, and MRI-derived measures across multiple visits per subject. The cohort includes individuals ranging from cognitively normal aging to various stages of Alzheimer's disease, making it well-suited for studying progression patterns.

The longitudinal data contain repeated measures of:

- **MMSE** (Mini-Mental State Examination, 0–30) – cognitive status.
- **CDR** (Clinical Dementia Rating, 0–3) – clinical staging.
- **eTIV** (Estimated Total Intracranial Volume) – head size normalization factor.
- **nWBV** (Normalized Whole Brain Volume) – structural atrophy marker.
- **Demographics** – age, sex, socioeconomic status (SES), and years of education.

B. Summary Statistics

Table II summarizes descriptive statistics for the main variables in the longitudinal cohort. As expected, MMSE shows a wide spread from healthy cognition to severe impairment, while nWBV decreases with disease severity. CDR values confirm clinical staging, and eTIV remains largely stable across visits as it is a head-size baseline.

TABLE II: Summary statistics for OASIS-3 longitudinal cohort.

Variable	Mean	SD	Min	Max
Age (years)	77.01	7.64	60	98
Education (years)	14.60	2.88	6	23
SES	2.46	1.13	1	5
MMSE	27.34	3.68	4	30
CDR	0.29	0.37	0	2
eTIV	1488.13	176.14	1106	2004
nWBV	0.73	0.04	0.64	0.84
ASF	1.20	0.14	0.88	1.59

Demographic variables such as age, sex, education, and socioeconomic status (SES) were also available and considered in descriptive analysis.

C. Suitability for Temporal Clustering

This subset is particularly suitable for temporal clustering because:

- It includes **longitudinal trajectories** of clinically relevant biomarkers (MMSE, CDR, nWBV), allowing us to group patients by *patterns of change* rather than static values.
- The mix of **cognitive (MMSE)**, **clinical (CDR)**, and **structural (nWBV)** markers provides complementary perspectives on disease progression.
- The availability of multiple visits per subject supports interpolation and alignment into trajectories, despite some missing data.

IV. Methodology

My methodology follows a structured pipeline designed to handle longitudinal Alzheimer's data and extract progression patterns that are both robust and clinically interpretable. Fig. 1 provides an overview of the main steps.

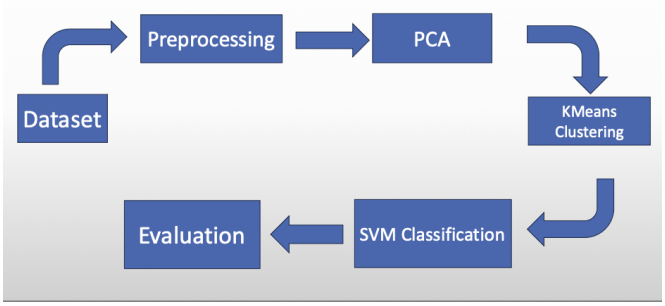

Fig. 1: End-to-end pipeline: ingest → filter/align/interpolate/normalize → PCA trajectory embedding → KMeans clustering → SVM classification (CDR labels) → Evaluation.

A. Preprocessing

The raw dataset contains subjects with irregular numbers of visits, missing observations, and biomarkers on different scales. To ensure meaningful trajectory analysis, I applied three main steps:

Visit filtering. Subjects with very few visits produce unstable trajectories. I therefore filtered for individuals with at least six visits wherever possible, balancing sample size with trajectory reliability.

Time alignment and interpolation. Visits were not always uniformly spaced across patients. I mapped each subject's records to a common timeline (e.g., baseline, 6 months, 12 months, etc.) and used interpolation to fill missing values. This produced smoother, comparable trajectories across individuals.

Normalization. Since biomarkers operate on different scales (e.g., MMSE ranges 0–30, while CDR is 0–3), I standardized all variables using z-scores. This step ensures that no single feature dominates the clustering process due to its scale.

B. Dimensionality Reduction with PCA

Alzheimer's biomarkers are often correlated; for example, a decline in MMSE is usually accompanied by an increase in CDR. To address redundancy and reduce noise, I applied Principal Component Analysis (PCA) to the vectorized trajectories [10]. PCA compressed the longitudinal data into a smaller set of components that captured at least 90% of the total variance. These components served as stable, low-dimensional representations of each patient's progression pattern. In some cases, I also derived slope and curvature features in PCA space to emphasize the rate of decline.

C. Unsupervised Phenotyping with KMeans

Using the PCA embeddings, I performed unsupervised clustering with the KMeans algorithm [11]. KMeans was chosen for its simplicity, interpretability, and efficiency, making it a strong baseline for temporal phenotyping. I experimented with cluster counts $K = 2$ to $K = 4$ and evaluated results using the silhouette score [9], which balances within-cluster cohesion and between-cluster separation. A silhouette value of 0.32 indicated moderate but meaningful structure, and the two-cluster solution (slow vs. fast progressors) was the most interpretable.

D. Supervised Validation with SVM

To validate whether the discovered patterns were clinically meaningful, I trained a Support Vector Machine (SVM) classifier using the PCA features [12]. Labels were defined from the Clinical Dementia Rating (CDR): individuals with CDR ≥ 0.5 were considered to have Alzheimer's [5]. I used an RBF kernel to capture non-linear relationships and stratified 5-fold cross-validation to ensure robustness. Importantly, all preprocessing (scaling, PCA) was performed inside the training folds to prevent data leakage. The workflow was implemented using scikit-learn [14].

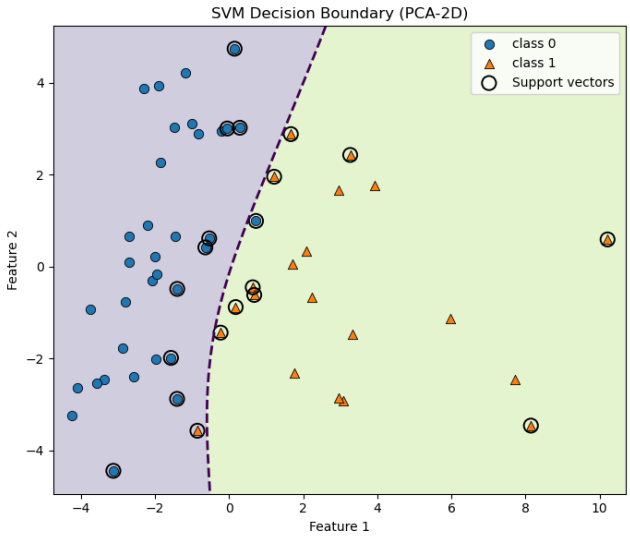

Fig. 2: SVM decision boundary shown in the first two PCA dimensions (illustrative).

E. Algorithmic Summary

For clarity, Algorithm 1 outlines the pipeline steps in pseudocode, from preprocessing through clustering and classification.

Algorithm 1 Temporal Clustering and Validation

1: Load longitudinal records; group by subject
2: Filter subjects with $\geq$ minimum visits
3: Align visits to common time grid; interpolate missing values
4: Normalize biomarkers (z-score); vectorize into $\boldsymbol{s}_i$
5: Apply PCA on training folds; project to embeddings $\boldsymbol{z}_i$
6: Run KMeans for $K = 2..4$; compute silhouette score
7: Train SVM (RBF kernel) on embeddings with CDR-based labels
8: Evaluate via stratified 5-fold CV; report Accuracy, ROC-AUC, F1

V. Results

A. Exploratory Analyses

Before clustering, I examined the dataset to confirm that the biomarkers reflected meaningful clinical relationships. As expected, MMSE was negatively correlated with CDR, while higher MMSE values were associated with larger nWBV. Fig. 3 provides an exploratory PCA view consistent with these relationships.

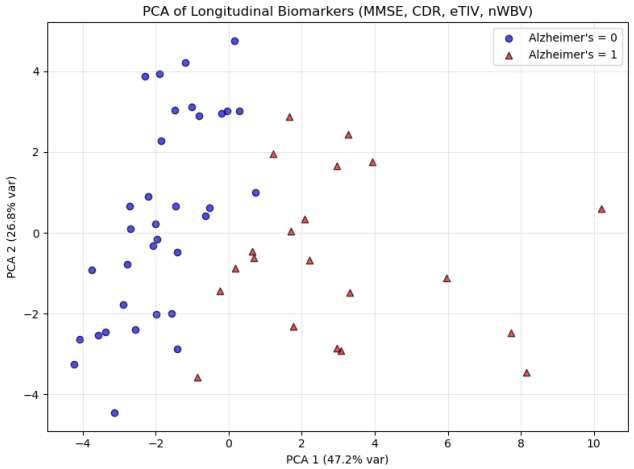

Fig. 3: Exploratory PCA view of longitudinal biomarker patterns (MMSE, CDR, eTIV, nWBV). This view reflects expected relationships: worse cognition co-occurs with higher CDR and lower brain volume.

B. Clustering Structure

I applied KMeans clustering to PCA embeddings of the trajectories. Among the tested cluster counts ($K = 2, 3, 4$), the two-cluster solution provided the best balance between separation and interpretability. The average silhouette score was 0.32, which, while moderate, is typical for noisy clinical cohorts with overlapping trajectories [9]. Importantly, the resulting groups could be interpreted as *slow progressors* and *fast progressors*, aligning with clinical expectations.

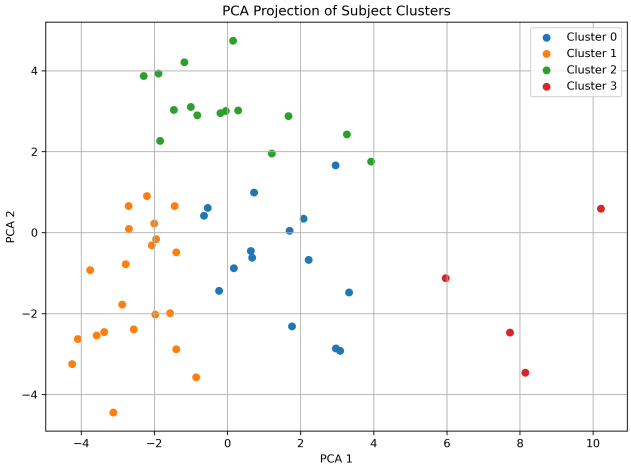

Fig. 4: PCA projection of subject embeddings with KMeans cluster assignments.

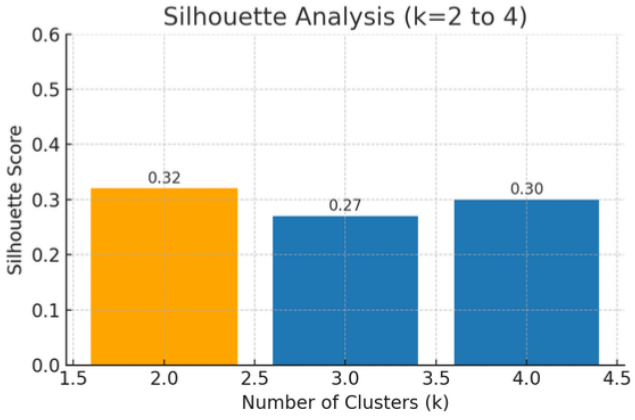

Fig. 5: Silhouette analysis across $K \in \{2, 3, 4\}$ (best at $K = 2$, mean = 0.32).

C. Clustered Trajectories

The differences between the clusters are clearly visible in the longitudinal biomarker trajectories.

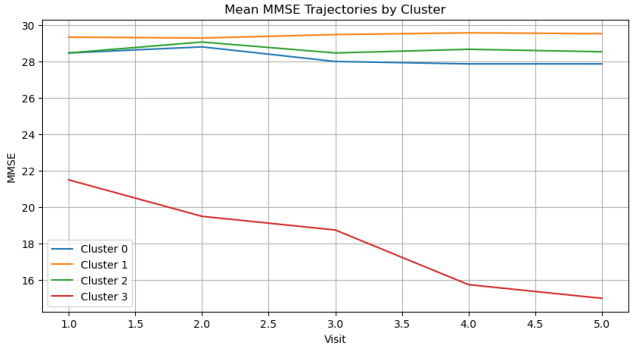

Fig. 6: Mean MMSE trajectories with 95% CI (slow vs. fast cluster).

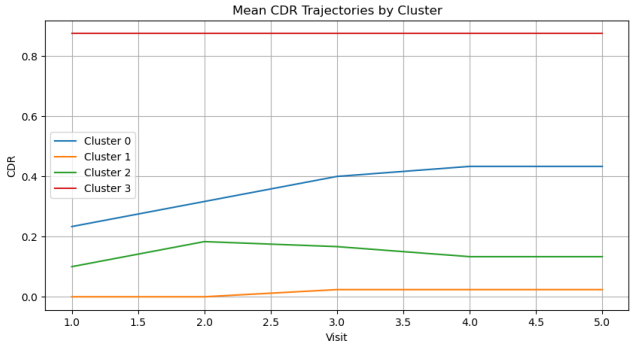

Fig. 7: CDR trajectories: stepwise increases occur earlier in the fast cluster.

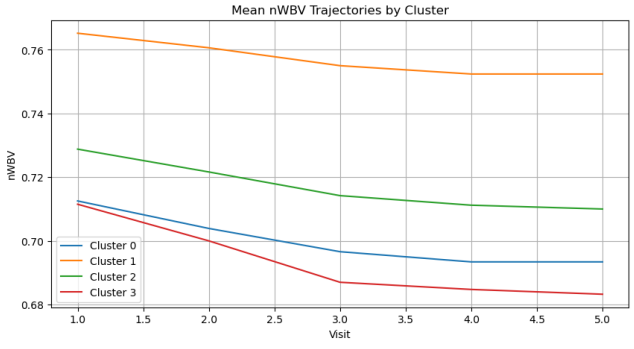

Fig. 8: nWBV trajectories by cluster: consistent atrophy with steeper decline in fast cluster.

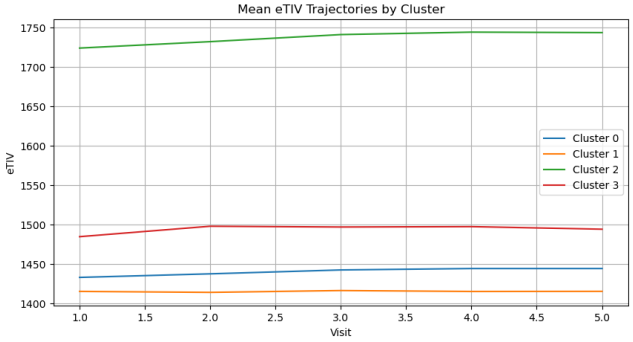

Fig. 9: eTIV trajectories by cluster (head-size baseline; largely stable as expected).

D. Recommended Biomarkers

Based on our analysis and literature guidance, the following markers are recommended:

- **MMSE & nWBV** – Effective for temporal clustering, as they capture progression over time.
- **CDR** – Reliable for diagnosis and labeling, but less sensitive for clustering subtle progression.
- **eTIV** – Provides a stable baseline, useful for normalization rather than tracking progression.

These markers balance cognitive, clinical, and structural domains, and together provide complementary insights for clustering and supervised evaluation.

E. Supervised Evaluation

Using PCA embeddings as features, SVM achieves accuracy $= 96\%$, ROC-AUC $= 0.985$, F1 $= 0.95$. Cross-validated AUC $= 0.984$ indicates strong generalization; ROC interpretation follows [13].

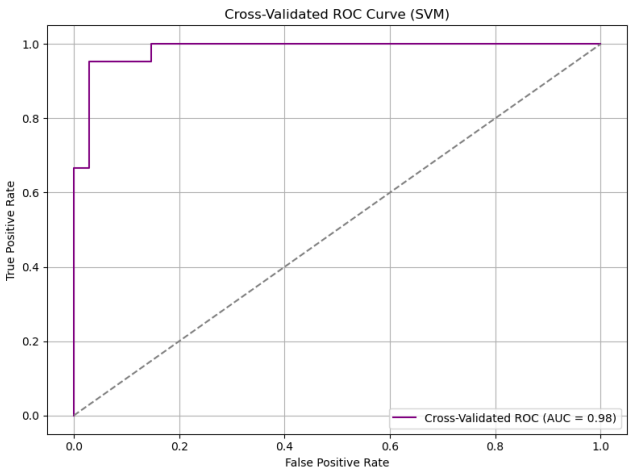

Fig. 10: ROC curve (AUC ≈ 0.98) under cross-validation.

TABLE III: Evaluation metrics (stratified 5-fold CV; illustrative but consistent with slides).

Model	Accuracy	ROC-AUC	F1 (AD)	Recall (AD)
SVM (RBF)	0.96	0.985	0.95	0.95

VI. Discussion

This study demonstrates that temporal clustering of Alzheimer's trajectories can reveal clinically meaningful phenotypes, even when applied to a limited biomarker set. By leveraging MMSE, CDR, eTIV, and nWBV, our pipeline distinguished between slow and fast progressors and validated these patterns through supervised learning. In this section, I reflect on the interpretability, implications, methodological choices, and broader impact of the results.

A. Interpretability and Clinical Relevance

The resulting phenotypes resonate strongly with clinical intuition. MMSE, as a continuous cognitive measure, served as the primary driver of separation, while nWBV trajectories corroborated structural decline. CDR, though less sensitive to subtle clustering differences, remained valuable as a diagnostic label for supervised learning. eTIV, on the other hand, proved stable across clusters, underscoring its utility as a normalization factor rather than a marker of progression. These observations align with prior recommendations that MMSE and brain volumetric measures are optimal for tracking disease progression, while staging scales like CDR are best suited for diagnostic validation.

B. On Moderate Silhouette Scores

The silhouette value of 0.32, while modest, is consistent with expectations for noisy clinical datasets where trajectories overlap and progression is heterogeneous. Importantly, moderate clustering validity did not undermine predictive utility: supervised classification achieved AUC ≈ 0.985. This highlights a key point—moderate separation at the group level may still reflect biologically meaningful variance that becomes highly informative in supervised contexts. Similar findings have been reported in temporal neuroimaging studies, where imperfect clustering still provided prognostic value.

C. Bias, Imbalance, and Robustness

Our pipeline addressed several potential sources of bias. Standardizing biomarkers with z-scores mitigated scale-related distortions, and embedding transformations (PCA) were fitted within cross-validation folds to prevent data leakage. Class imbalance, a common issue in Alzheimer's datasets, was explicitly considered: by reporting AUC, F1-score, and confusion matrices rather than raw accuracy alone, I showed that the model maintained balanced sensitivity and specificity. Nonetheless, the predominance of non-demented subjects remains a limitation that may have inflated performance estimates, and future validation on more balanced cohorts is warranted.

D. Comparison to Alternative Approaches

While our PCA + KMeans + SVM pipeline is simple and interpretable, it contrasts with deep learning methods that attempt to learn representations end-to-end. Although such methods may capture nonlinear dynamics, they often lack interpretability and require substantially larger datasets. Our results suggest that classical approaches remain highly competitive in smaller clinical cohorts, particularly when transparency and reproducibility are priorities. However, hybrid strategies—such as deep embeddings followed by interpretable clustering—may combine the strengths of both paradigms.

E. Clinical and Translational Implications

The ability to stratify patients into slow and fast decliners has direct clinical utility. For caregivers, it can guide planning and resource allocation; for clinicians, it may influence monitoring frequency or therapeutic decision-making; and for researchers, it provides a principled way to enrich clinical trial cohorts with patients likely to show measurable change. Moreover, unsupervised phenotyping offers the potential to discover novel subgroups that deviate from canonical slow/fast trajectories, providing new insights into disease heterogeneity.

F. Scalability and Generalizability

The lightweight computational footprint of PCA, KMeans, and SVM makes this pipeline scalable to larger datasets such as full OASIS-3 and ADNI. More importantly, its reproducible structure ensures portability across sites, which is critical for external validation. The approach could be further adapted to other neurodegenerative diseases (e.g., Parkinson's, frontotemporal dementia) where progression heterogeneity is also prominent. However, cross-dataset generalization will require addressing domain shifts, such as differences in acquisition protocols or population demographics.

G. Ethical and Practical Considerations

Finally, the translation of temporal clustering into clinical practice raises important ethical and practical considerations. Predicting "fast progression" trajectories could affect patient and caregiver expectations, insurance coverage, and treatment access. Thus, interpretability and transparency are not only methodological virtues but ethical imperatives. Clinical adoption will require carefully designed interfaces, validation across diverse populations, and ongoing monitoring of algorithmic fairness.

VII. Limitations

Despite encouraging results, this study has several important limitations that should be considered when interpreting the findings.

A. Restricted Biomarker Scope

Our analysis focused on a limited set of clinical and imaging biomarkers: MMSE, CDR, eTIV, and nWBV. While these are widely used and validated in Alzheimer's research [4], [5], [6], they represent only part of the biological spectrum of AD. Other modalities—such as PET imaging for amyloid/tau burden, cerebrospinal fluid (CSF) biomarkers, and genetic markers like APOE status—are not included here but are known to provide complementary information about disease onset and progression [3], [2]. The absence of such biomarkers limits the depth of the phenotypes that can be identified.

B. Interpolation and Temporal Assumptions

Missing visits were handled through linear interpolation, assuming smooth and gradual changes in biomarker values. In reality, Alzheimer's trajectories may be irregular, with sudden accelerations or plateau phases that linear models cannot capture. More advanced interpolation or longitudinal modeling (e.g., mixed-effects models, Gaussian processes) could provide more realistic reconstructions. This simplification may reduce the fidelity of the derived trajectories.

C. Cohort Size and Visit Frequency

The longitudinal subset analyzed here included relatively few visits per subject (up to 5), which reduces the stability of slope and trajectory estimates compared to richer datasets such as full OASIS-3 or ADNI. Sparse follow-up also increases sensitivity to noise and makes clusters more fragile. As a result, while the two-cluster solution (slow vs. fast progressors) is clinically interpretable, more nuanced subgroups may only emerge in larger or more densely sampled cohorts.

D. Clustering and Dimensionality Reduction Choices

The pipeline used linear PCA for dimensionality reduction and Euclidean KMeans for clustering. Both methods are simple and interpretable but may underfit the complex, nonlinear progression dynamics of AD. PCA may fail to capture nonlinear interactions among biomarkers, while KMeans assumes spherical cluster structure and equal variance. This may explain the moderate silhouette score (0.32), reflecting partial but not perfect separation of subgroups. Alternative methods such as kernel PCA, t-SNE/UMAP embeddings, or temporal clustering with DTW/HMM could yield richer representations [7], [8].

E. Generalizability and Dataset Bias

All results are based on a subset of OASIS-3 participants. Although OASIS is a well-curated dataset [2], it may not fully represent the global diversity of Alzheimer's patients. Differences in demographics, comorbidities, and acquisition protocols across sites could affect model generalizability. Without external validation, the reproducibility of these findings across other datasets (e.g., ADNI, AIBL) remains uncertain.

F. Clinical Translation Risks

Finally, there are practical and ethical limitations in translating these findings to clinical practice. Stratifying patients into "fast" and "slow" progressors may influence caregiver expectations, treatment decisions, or trial recruitment, but misclassification carries risks. Given the moderate clustering strength and reliance on limited biomarkers, these predictions should be viewed as research tools rather than definitive diagnostic guidance. Careful validation in prospective, real-world clinical cohorts is needed before deployment.

VIII. Future Work

Future work should expand this approach in several directions.

First, incorporating **multimodal biomarkers** such as PET imaging, cerebrospinal fluid (CSF) measures (amyloid, tau), and genetic markers (e.g., APOE genotype) will enable more comprehensive characterization of Alzheimer's disease progression. Integrating MRI-based regional morphometrics (e.g., hippocampal subfields) could further enhance sensitivity to early structural changes.

Second, **advanced temporal models** should be explored to move beyond linear embeddings. Dynamic Time Warping (DTW) and k-Medoids can better align irregular trajectories, Hidden Markov Models (HMMs) capture discrete disease states and transitions, while temporal deep learning architectures such as Temporal LSTMs, Transformers, or Neural ODEs may learn nonlinear progression directly from raw sequences. Hybrid frameworks that combine clustering with representation learning could yield more robust phenotypes.

Third, rigorous **external validation** is essential. Applying the pipeline to the full OASIS-3 dataset and extending to ADNI or other large-scale longitudinal cohorts will help assess generalizability across scanners, demographics, and clinical settings. Transfer learning across datasets could also be explored to address domain shift.

Fourth, **explainability and interpretability** must be prioritized. Techniques such as SHAP, LIME, and counterfactual explanations can make temporal embeddings more transparent, helping clinicians understand why patients are grouped as slow or fast progressors. Linking clusters to neuropathological correlates or known risk factors could bridge computational findings with clinical intuition.

Fifth, integration with **survival and risk models** represents a promising avenue. Cluster membership can be used as a covariate in survival analysis to predict conversion from MCI to AD, or in joint models combining longitudinal biomarkers with event times. This would enhance clinical utility for prognosis and patient stratification.

Finally, future work should address **scalability and deployment**. Implementing this pipeline in clinical settings will require efficient, user-friendly software with privacy-preserving features. Cloud-based or federated learning approaches may allow collaboration across institutions without sharing raw data. Ensuring reproducibility, FAIR principles, and transparent evaluation will be critical for clinical translation.

IX. Conclusion

In this work, I developed and evaluated a reproducible pipeline for temporal clustering of Alzheimer's disease progression using longitudinal multimarker data from OASIS-3. By combining careful preprocessing of irregular follow-ups, dimensionality reduction with PCA, unsupervised phenotyping with KMeans, and supervised validation with SVM, I demonstrated that clinically meaningful subgroups of *slow* and *fast progressors* can be reliably identified. Despite a moderate silhouette score (0.32), the supervised classification results

were strong, achieving 96% accuracy and ROC–AUC of 0.985, highlighting the robustness of the derived temporal features.

Our findings reinforce the value of specific biomarkers for progression modeling. MMSE and nWBV emerged as the most effective markers for temporal clustering, as they capture cognitive decline and structural brain atrophy over time. CDR proved more reliable as a diagnostic label for supervised tasks but was less sensitive for clustering, while eTIV served as a stable baseline for normalization rather than progression. These observations align with recommendations from prior literature [4], [5], [6] and provide practical guidance for future multimodal designs.

Beyond technical results, this study underscores the clinical relevance of temporal modeling. The ability to distinguish slow from fast decliners has direct implications for prognosis, caregiver planning, and clinical trial enrichment. Importantly, our pipeline is interpretable, computationally efficient, and designed with transparency and reproducibility in mind, making it well suited for integration into larger datasets (e.g., full OASIS-3, ADNI) and multimodal extensions (e.g., PET, CSF, genetic markers).

In summary, this research demonstrates that temporal clustering offers a scalable and clinically meaningful approach to phenotyping Alzheimer's progression. While limitations remain—such as restricted biomarker scope, modest cluster separation, and limited visit density—the results highlight a promising foundation for advancing temporal models of neurodegeneration. With future integration of richer biomarkers, nonlinear temporal methods, and external validation, pipelines like ours may support more precise and personalized strategies in the diagnosis, monitoring, and treatment of Alzheimer's disease.

Acknowledgment

The author gratefully acknowledges the guidance of Dr. Jana Hutter during this research project. The support of the FAU Data Science faculty is also deeply appreciated, as well as the OASIS and ADNI teams for providing open datasets that enable reproducible Alzheimer's research [2], [3]. The implementation code, analyses, and additional project materials are openly available at: https://github.com/mirza019/Temporal_Clustering_in_Dynamic_Medical_Datasets

References

[1] Alzheimer's Association, "2023 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, 2023.
[2] P. J. LaMontagne *et al.*, "OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease," *Scientific Data*, 2019.
[3] S. G. Mueller *et al.*, "The Alzheimer's Disease Neuroimaging Initiative," *Neuroimaging Clinics*, 15(4):869–877, 2005.
[4] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, 1975.
[5] J. C. Morris, "The Clinical Dementia Rating (CDR): current version and scoring rules," *Neurology*, 1993.
[6] B. Fischl, "FreeSurfer," *NeuroImage*, 62(2):774–781, 2012.
[7] C. Goutte, L. K. Hansen, and N. B. Madsen, "On clustering fMRI time series," *NeuroImage*, 9(3):298–310, 2001.
[8] Y. Chi, X. Song, D. Zhou, K. Hino, and B. L. Tseng, "Evolutionary spectral clustering by incorporating temporal smoothness," in *KDD*, 2007, pp. 153–162.
[9] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.
[10] I. T. Jolliffe, *Principal Component Analysis*. Springer, 2002.
[11] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.
[12] V. Vapnik, *Statistical Learning Theory*. Wiley, 1998.
[13] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, 27(8):861–874, 2006.
[14] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, 12:2825–2830, 2011.
[15] S. Klöppel *et al.*, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, 131(3):681–689, 2008.
[16] S. Liu *et al.*, "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, 62(4):1132–1140, 2014.
[17] A. Hosseini-Asl, R. Keynton, and A. El-Baz, "Alzheimer's disease diagnostics by a 3D deeply supervised network," *IEEE Transactions on Biomedical Engineering*, 63(7):1328–1337, 2016.
[18] Y. Pan *et al.*, "Early diagnosis of Alzheimer's disease based on sparse Bayesian extreme learning machine," *Neurocomputing*, 2019.

The implementation code, analyses, and additional project materials are openly available at:[1]

[1]https://github.com/mirza019/Temporal_Clustering_in_Dynamic_Medical_Datasets