Temporal Clustering in Alzheimer's Dataset

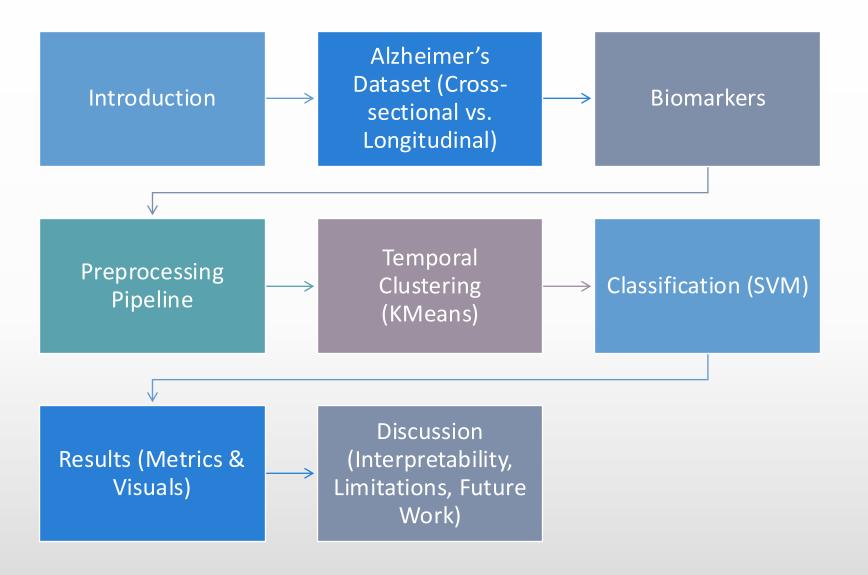
Mirza Shaheen Iqubal · M.Sc. Data Science, FAU

Matriculation No. 22998316





AGENDA



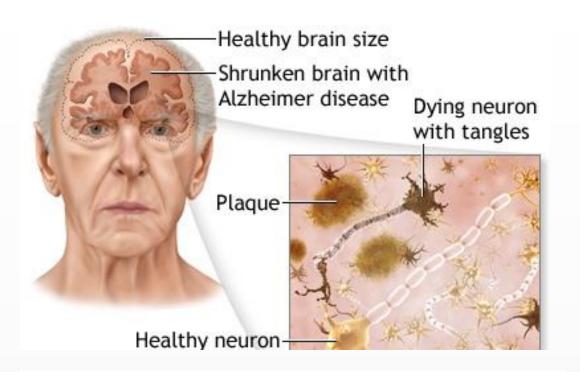
Introduction

Temporal clustering

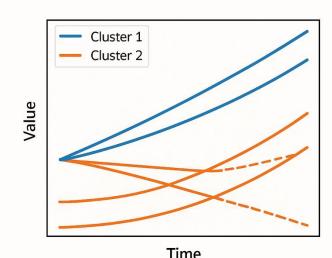
- Groups data on patterns change over time,
- Considers trajectories or sequences.

What is Alzheimer's?

- Neurodegenerative disorder
- Leads to progressive memory loss and cognitive decline
- Most common cause of dementia worldwide



Temporal Clustering



INTRODUCTION

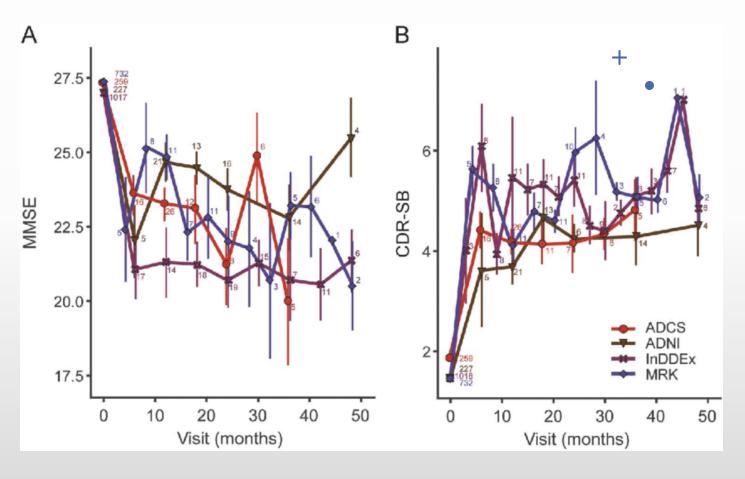
Alzheimer's → progressive neurodegenerative disease.

Most ML studies: single snapshot (cross-sectional).

Challenge: disease evolves -> requires longitudinal modeling.

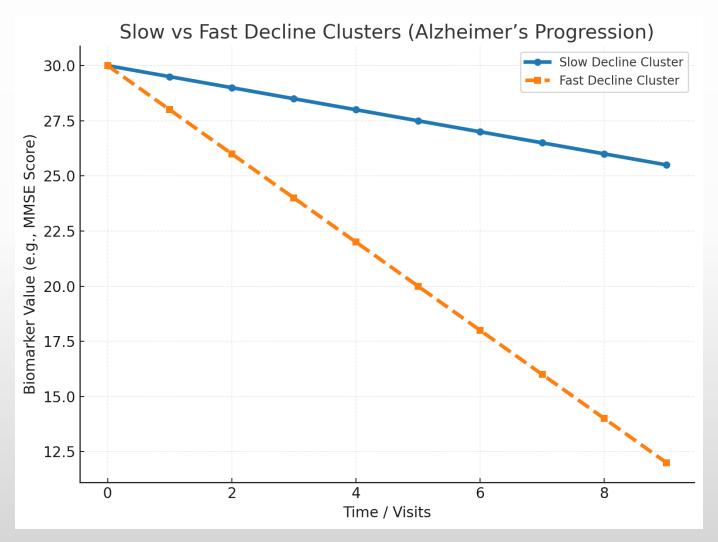
Introduction

- Patient biomarkers evolve across visits
 - MMSE (cognitive score)
 - CDR (dementia rating)
 - Brain volume



Introduction

- Temporal clustering groups patients into:
 - Slow decline cluster
 - Fast decline cluster



DATASET

Dataset: Alzheimer's dataset (OASIS-3, via Kaggle).

Cross-sectional: single visit per subject.

Longitudinal: repeated MRI + cognitive assessments across years.

Why longitudinal? → progression tracking.

Dataset Structure 1,000+ participants, ages 42-95.

Each subject: 2–15 visits.

Follow-up period: up to 10 years.

Data includes
MRI scans +
cognitive test
scores +
demographics.

BIOMARKERS

CDR: Clinical Dementia Rating (0 \rightarrow normal, 0.5+ \rightarrow dementia).

MMSE: cognitive ability (0–30 scale).

eTIV: estimated intracranial volume (brain size baseline).

nWBV: normalized whole brain volume.

PREPROCESSING PIPELINE

Filter: subjects with ≥6 visits.

Interpolation: fill missing visits→ regular timeline.

Normalization: Z-score across biomarkers.

PCA: reduce dimensionality of biomarker time series.

Why PCA?



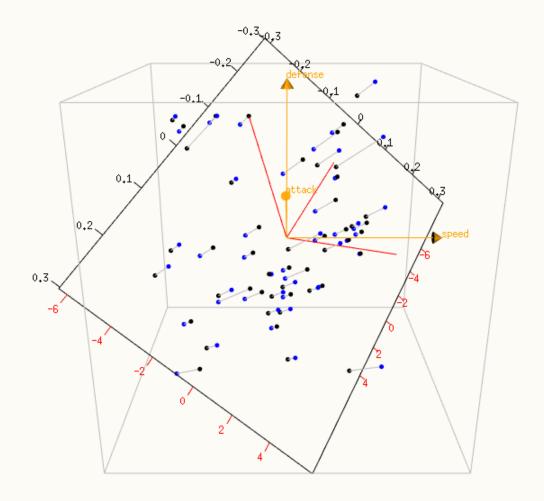
Alzheimer's biomarkers correlate.



PCA removes redundancy, reduces noise.

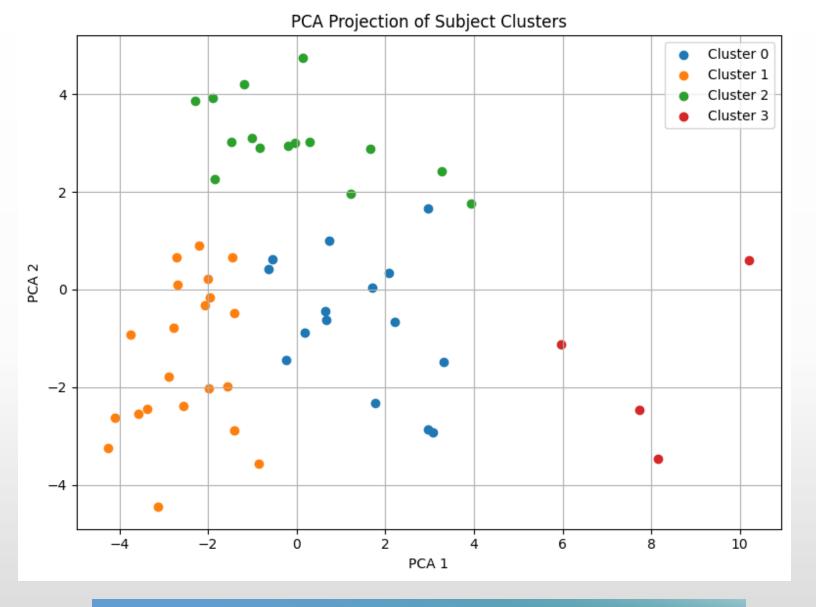


Trajectories become more stable + easier to cluster.



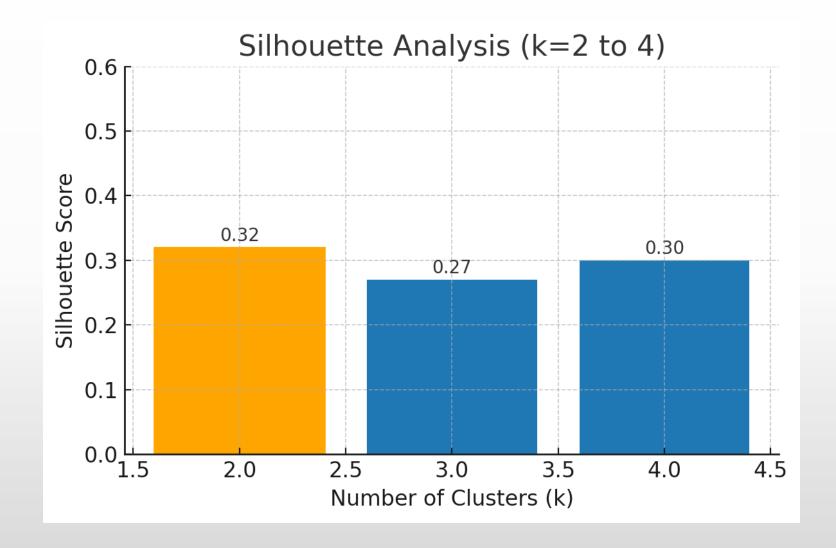
Temporal Clustering

- Algorithm: **KMeans** on PCA-based trajectories.
- Tried on k=2-4 clusters.
- Evaluated using Silhouette Score.

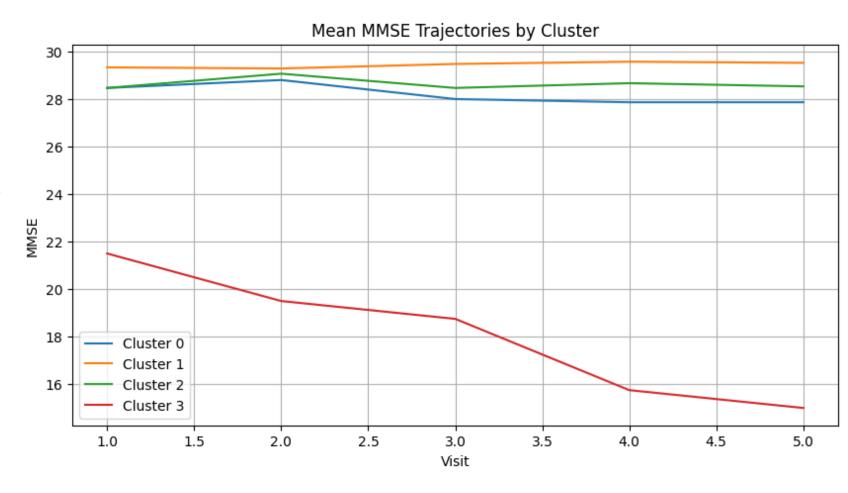


Silhouette Analysis

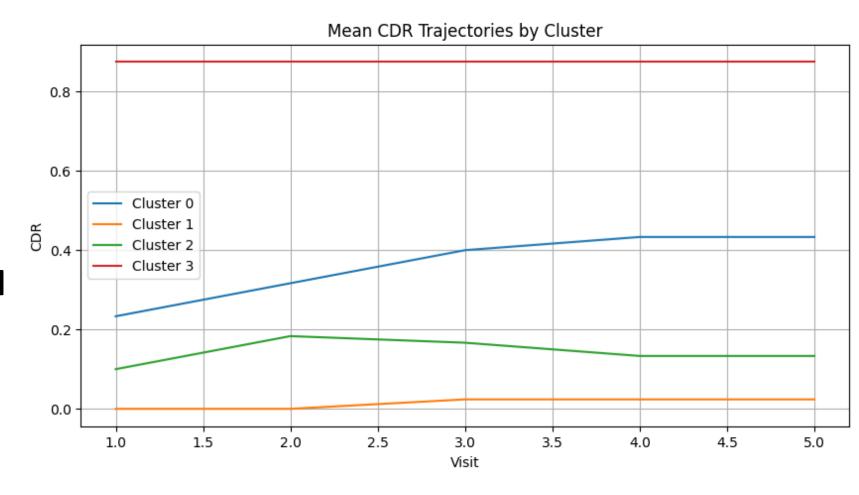
- Obtained value: 0.32
- Range: $-1 \rightarrow +1$
- ~0.0 → overlapping clusters (poor separation)
- 0.25–0.50 → reasonable structure (moderate separation)
- >0.50 → strong cluster separation



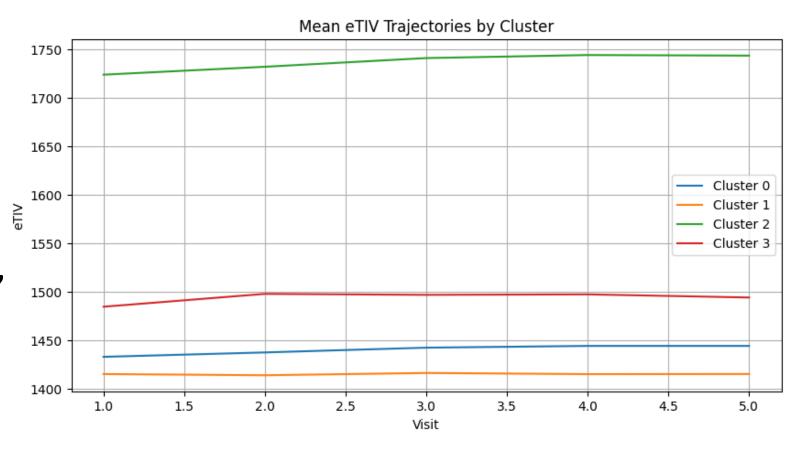
- Stable vs. steep decline
 → clear slow vs fast
 clusters
- Strong cognitive progression marker



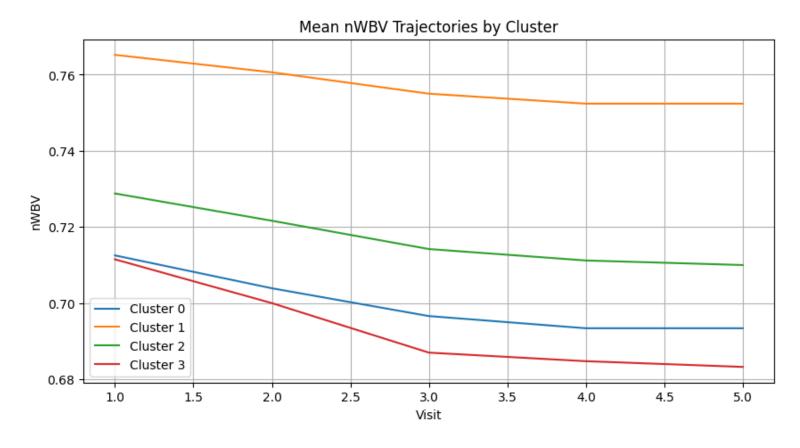
- Separates normal vs dementia stages
- Less sensitive to gradual progression



- Stable baseline across visits
- Useful for normalization, not progression



- Shows consistent decline over visits
- Strong structural progression marker



RECOMMENDED MARKERS

MMSE & nWBV → effective for temporal clustering (progression captured).

CDR → reliable for diagnosis/labels, less sensitive for clustering.

eTIV → stable baseline, useful for normalization, not progression.

Goal: detect Alzheimer's vs. non-Alzheimer's.

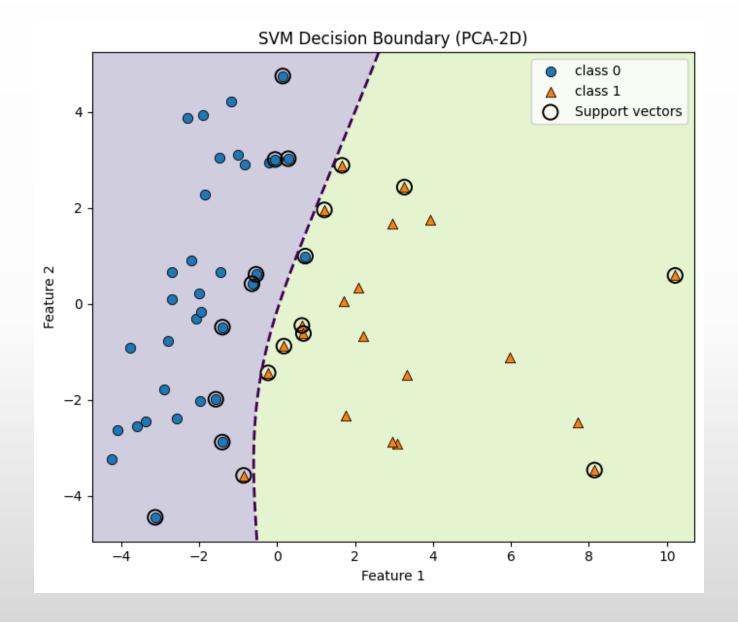
Label: Alzheimer's if CDR ≥0.5.

CLASSIFICATION SETUP

Model: Support Vector Machine (SVM).

Features: PCA components from biomarker sequences.

SUPPORT VECTOR MACHINE (SVM)



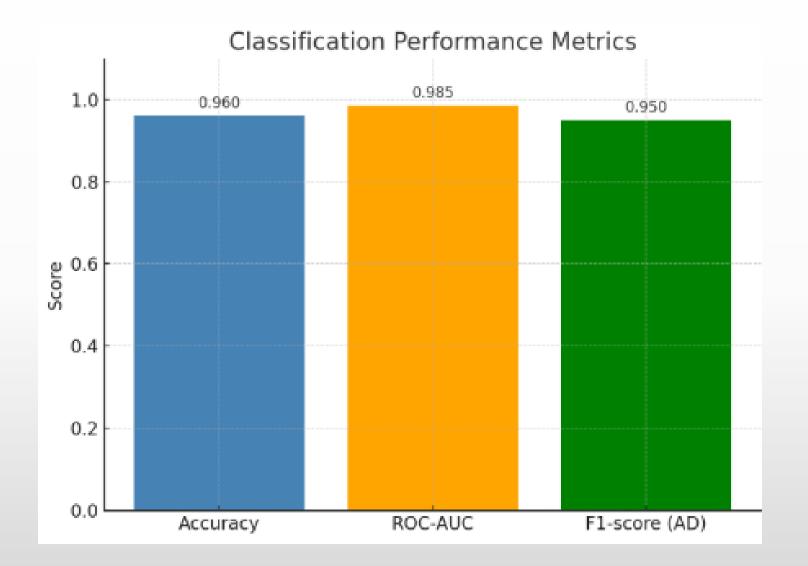
Evaluation Metrics

• Accuracy: 96%

• ROC-AUC: **0.985**

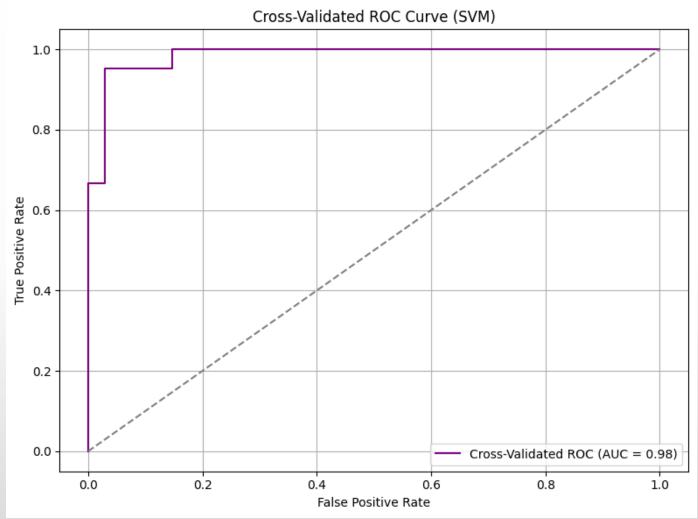
• F1-score (AD): **0.95**

• Strong predictive performance.



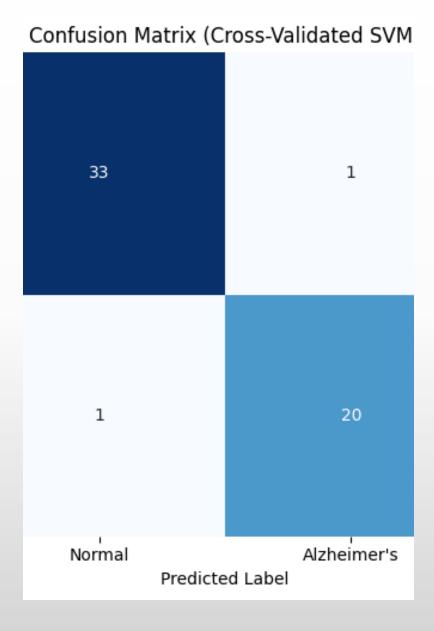
Cross Validation on ROC Curve

- **AUC = 0.98** → excellent classification performance.
- Curve lies close to the **top-left corner**, showing strong separation



Cross-validation

- 5-fold validation for robustness.
- Cross-validated AUC: 0.984.
- Confirms generalization of model.



INTERPRETABILITY

Two clusters = interpretable subgroups: fast vs. slow progressors.

Supports clinical decision-making.

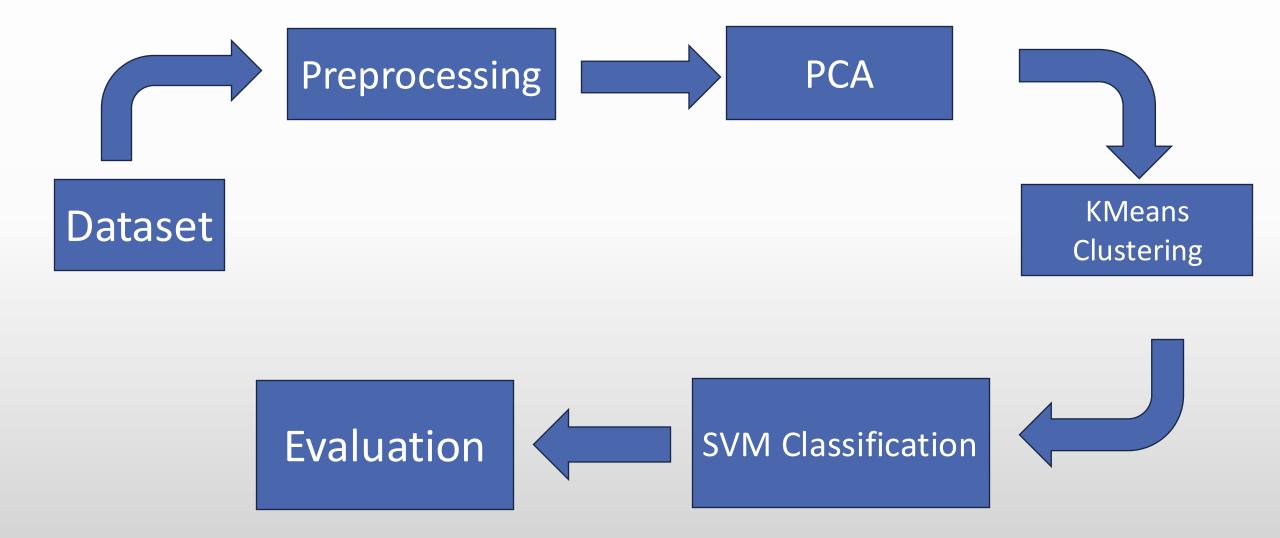
ML outputs aligned with medical knowledge.

TECHNOLOGY STACK Python 3.x, Jupyter Notebook scikit-learn (PCA, KMeans, SVM)

NumPy, Pandas Matplotlib, Seaborn

GitHub for reproducibility

Pipeline Flow



LIMITATIONS

Restricted biomarker set

Interpolation assumptions

Class imbalance

Model simplicity

Why This Work Stands Out

Most studies → cross-sectional (single visit, one biomarker)

My work → longitudinal + multi-biomarker analysis

Combines clustering + classification (SVM)

Validated with Silhouette, ROC, Confusion Matrix, CV

FUTURE WORK

Extend biomarkers

→ PET, advanced

MRI features.

Apply sequence models → HMM, LSTM, DTW clustering.

External dataset validation.

Explainability → SHAP, LIME.

SUMMERY

MMSE & nWBV best markers for progression.

Clustering distinguishes slow vs. fast decline.

SVM achieves high accuracy (96%, AUC=0.985).

Approach demonstrates temporal clustering is effective for Alzheimer's biomarker analysis.

REFERENCES

Goutte, C., et al. (2001). Clustering fMRI time series. Neurolmage.

Chi, Y., et al. (2007). Evolutionary spectral clustering by incorporating temporal smoothness. KDD.

Marcus, D. S., et al. (2007). Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle-aged, nondemented, and demented older adults. J. Cognitive Neuroscience.

McKhann, G. M., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the NIA-Alzheimer's Association workgroup. Alzheimer's & Dementia.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Computational and Applied Mathematics.

GitHub Repository (Code & Documentation):



THANK YOU

Any Questions!