

Uncovering feature significance in Alzheimer’s Disease via statistics and machine learning

Jenifer Roque-Martinez, Branden Crenshaw, UNC Charlotte
Dr. Zheming Gao, William States Lee College of Engineering, Department of Industrial and Systems Engineering



Introduction

Problem

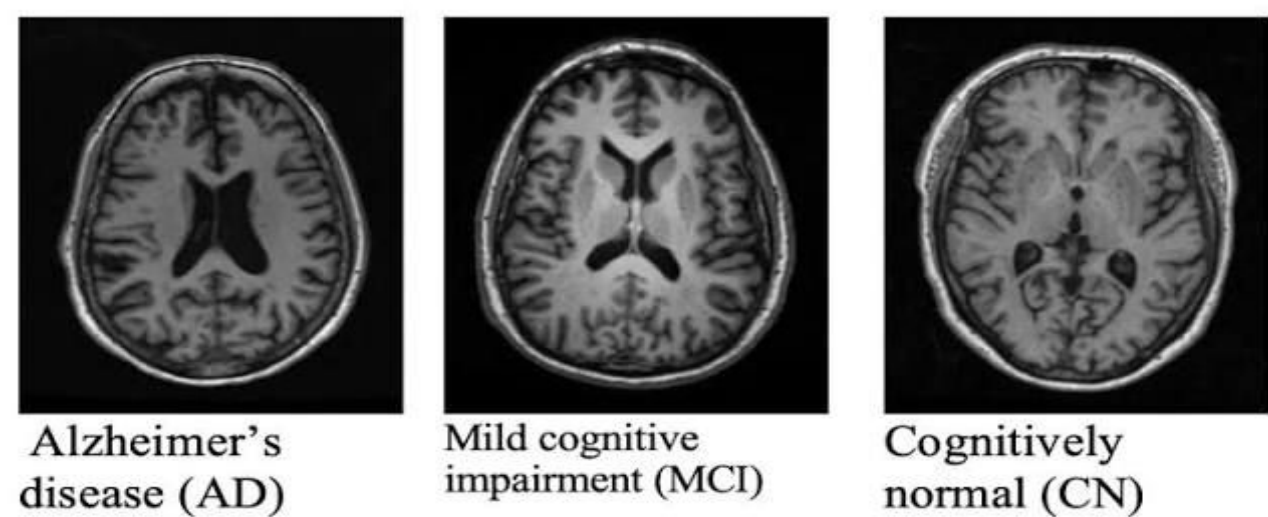
- **Alzheimer’s Disease (AD)** is a progressive, irreversible neurodegenerative disorder.
- Early detection at the Mild Cognitive Impairment (MCI) stage is critical before irreversible structural damage occurs.
- High-dimensional sMRI data contains hundreds of localized features, making it difficult to determine which ones are important for AD diagnosis.

Objective

Advance data-driven tools to find and analyze ROI features related to the detection of early Alzheimer’s Disease.

Data

MRI data is sourced from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database with 3 cohorts:



327 distinct structural brain features were grouped into 10 macro regions:

Macro Region	Feature Count
Temporal Lobe	50
Frontal Lobe	48
Cingulate Cortex/Limbic	37
Parietal Lobe	32
Occipital Lobe	24
Subcortical Deep Gray Matter	12
Ventricular/CSF Systems	8
Insular Cortex	8
Cerebellum/Brainstem	5
Other/Cortical WM Parcellations	103
Total Features:	327

Methodology

- Group normalized data by cohorts
- Run ANOVA on individual structural measurements
- Map significant features ($p < 0.05$) to their macro brain regions
- Compare results by dividing number of significant features retained in Best PCA model by same regions.

Feature Reduction Strategy (PCA)

$$PC_k = w_1(\text{Feature}_1) + w_2(\text{Feature}_2) + w_3(\text{Feature}_3) + w_4(\text{Feature}_4)$$

Feature_1 = ST32TA

Feature_2 = ST32TS

Feature_3 = ST32SA

Feature_4 = ST32CV

$$PC_k = \text{ST32}$$

w = coefficient
 k = top-k component

Feature Weighting Strategy

Individual Features:

ANOVA(Feature) = F-Statistic, P-Value

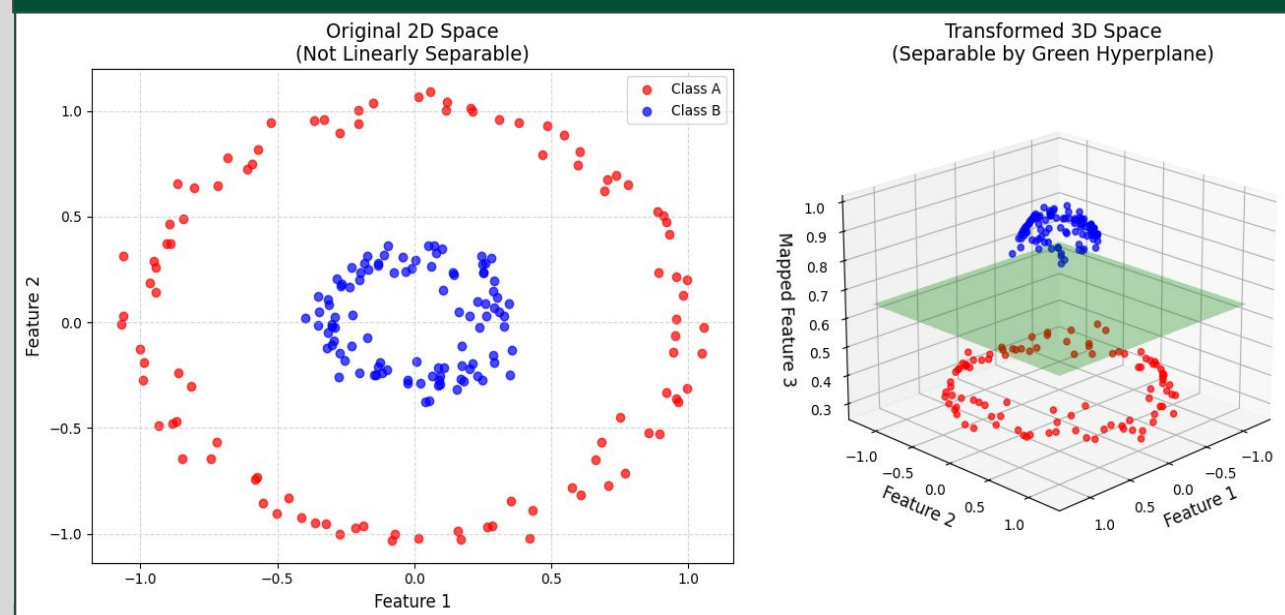
Score = Softmax(All F-Statistics) → Scores Add Up to 1

Regional/PCA:

ANOVA(Region) = Average F-Statistic of Region’s Features

Score = Softmax(All Averages) → Scores Add Up to 1

SVM Model (Example)



Weight Application During Training

- Calculated weights based on training data only
- Applied same weights to validation and testing data

SVM Input = (features) * (Individual/PCA weight) * (Regional weight)

Determined **best weights** by training many different SVM models

References

[1] Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., ... & Jack Jr, C. R. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*, 74(3), 201-209.

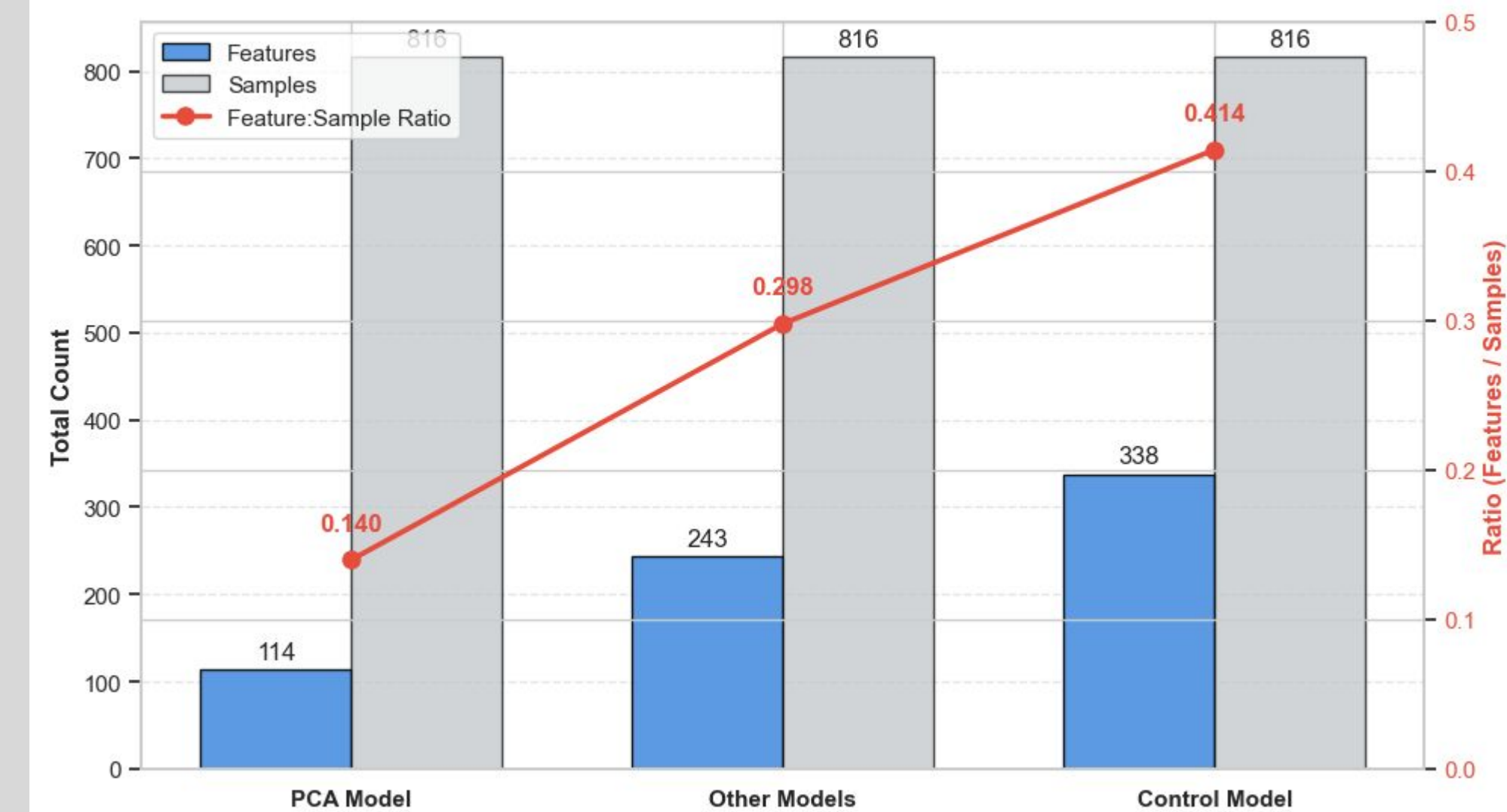
[2] Nichols, B. N., Mejino, J. L., Detwiler, L. T., Nilsen, T. T., Martone, M. E., Turner, J. A., Rubin, D. L., & Brinkley, J. F. (2014). Neuroanatomical domain of the foundational model of anatomy ontology. *Journal of biomedical semantics*, 5(1), 1. <https://doi.org/10.1186/2041-1480-5-1>

[3] Structural MRI scans highlighting dataset features from page 8 are sourced from Alwuthaynani et al. (2023). The figure illustrates key characteristics used for detecting cognitive decline. Access the full paper at arxiv.org/pdf/2301.13504

Results

Model	Accuracy	Precision	Recall	F1-Score	Characteristic
Control Model	73.00%	74.00%	73.00%	73.00%	Weakest Overall
PCA Model	82.00%	82.00%	82.00%	81.00%	Most Efficient
Regional Model	82.00%	82.00%	82.00%	82.00%	Most Balanced + Best AD Recall
Individual Model	84.00%	85.00%	84.00%	84%	Best Overall

Feature-to-Sample Counts and Ratios Across Models



Observations::

- PCA performed noticeably weaker on AD specific recall
- Recall is the ability of a model to correctly identify all true positive cases (missing an AD diagnosis is dire)
- Individual weighting had a bigger impact on regional weighting than PCA did
- Weighting features beyond individual weighting appeared to have no significant change in performance

Conclusions

ANOVA provides a largely stable method of ranking brain regions. Regional ranking is best used as an interpretive lens when using individual weights rather than using it as a weight itself. PCA provides a solid foundation for compressing complex features but falls short in performance.

Future Directions

- Using regional ranking as a means to further explore important features
- Finding similar compression algorithms such as PLS to reduce dimensionality