

# Cognitive Decline Prediction in Alzheimer's Disease Cohorts Using Baseline MRIs

Author: Nathaniel Putera  
Advisor: Mostafa Mehdiipour Ghazi

October 2024



# Contents

<b>1 Abstract</b>	<b>3</b>
<b>2 Introduction</b>	<b>3</b>
<b>3 Methods</b>	<b>4</b>
3.1 Dataset . . . . .	4
3.1.1 Data Statistics . . . . .	4
3.2 Clustering . . . . .	7
3.2.1 K-Means . . . . .	7
3.2.2 Dynamic Time Warping . . . . .	7
3.3 Classification . . . . .	7
3.3.1 Extreme Gradient Boost . . . . .	7
3.3.2 Autoencoder . . . . .	8
<b>4 Experimental Results</b>	<b>8</b>
4.1 Data Preprocessing . . . . .	8
4.2 Clustering . . . . .	9
4.2.1 K-Means . . . . .	9
4.2.2 Determining K . . . . .	10
4.2.3 Dynamic Time Warping . . . . .	11
4.3 Classification . . . . .	12
4.3.1 Extreme Gradient Boost . . . . .	12
4.3.2 Autoencoder . . . . .	14
<b>5 Discussion</b>	<b>16</b>
<b>6 Conclusion</b>	<b>17</b>
<b>A K-Means</b>	<b>19</b>
A.1 Fluctuating CDRSB patients . . . . .	19
A.2 K-Means Experiments . . . . .	19
<b>B XGBoost</b>	<b>19</b>
B.1 Modality sets . . . . .	21
B.2 Significance of latent space . . . . .	22
<b>C Linear Regression</b>	<b>22</b>
C.1 Regression Results . . . . .	22

# 1 Abstract

Alzheimer’s disease (AD) is the leading cause of dementia worldwide, yet early diagnosis remains a challenge. Identifying structural brain changes through non-invasive imaging techniques such as MRI offers a promising approach to detecting AD progression. However, leveraging these imaging biomarkers effectively requires advanced data-driven techniques, such as clustering and machine learning models, to extract meaningful patterns from complex dataset.

Among the analyzed datasets—cognitive, cerebrospinal fluid, positron emission tomography, MRI, and risk factors—MRI-based features emerged as strong predictors of cognitive decline, reinforcing their relevance as non-invasive biomarkers. Additionally, APOE4 status was identified as a significant genetic risk factor, consistent with prior research linking it to accelerated dementia progression.

Our findings demonstrate that integrating MRI features with machine learning models enhances predictive accuracy for AD progression, providing a scalable and non-invasive approach for early detection. This research establishes a solid foundation for future work incorporating 3D MRI imaging and deep learning techniques to further refine predictive precision and improve clinical applicability.

# 2 Introduction

Alzheimer’s disease (AD) is the most common form of dementia, affecting millions globally. It is a progressive neurodegenerative disease that impacts cognitive function and eventually leads to death. Researchers and clinicians have used various methods to better understand the progression of AD.

The early detection of Alzheimer’s disease is critical for developing interventions before irreversible cognitive decline occurs. Recent studies emphasize the need for more research into preclinical AD diagnosis, particularly in refining biomarker-based approaches and machine learning models[1, 2]. MRI-based imaging and cognitive assessments have been widely explored as potential diagnostic tools, yet further advancements in computational approaches are required to improve accuracy and accessibility.

Recent advancements in machine learning (ML) and deep learning (DL) have shown promise in identifying early-stage AD, particularly through MRI-based analysis. Convolutional neural networks (CNNs), such as ResNet18 and AlexNet, have been successfully applied to MRI data for detecting neurodegenerative patterns[3, 4]. These deep learning models leverage hierarchical feature extraction, enabling the identification of subtle structural changes in brain imaging that may correlate with cognitive decline. Moreover, the integration of multimodal data, including MRI scans, cognitive scores, and genetic risk fac-

tors, has the potential to enhance the predictive power of these models[2]. Understanding how different modalities contribute to AD progression remains an essential area of ongoing research.

In this project, we aim to build upon these advancements by first clustering Alzheimer’s progression using tabular patient data, including cognitive scores and biomarker information. Following this, we will apply machine learning models to classify AD progression based on MRI data, leveraging deep learning methods to detect patterns that may be missed by traditional diagnostic approaches. By combining both tabular and MRI data, this study seeks to improve our understanding of AD progression and contribute to the development of more reliable diagnostic tools, particularly for earlier stages of the disease.

## 3 Methods

### 3.1 Dataset

For this project, we are using a dataset from the Alzheimer’s Disease Neuroimaging Initiative (ADNI), which includes tabular data detailing patients’ conditions at each visit. While the dataset contains extensive information, our focus is on key variables such as patient age, sex, education level, APOE4 gene status (a gene associated with Alzheimer’s risk), Clinical Dementia Rating Sum of Boxes (CDRSB), a measure of cognitive function, and biomarkers, including the size of specific brain regions like the ventricles and hippocampus, as well as overall brain volume. For this research we will split the dataset to be train set and validation set.

#### 3.1.1 Data Statistics

Before going to the experiment, we did some analysis of how this data look like, how is the distribution, what each columns means, etc, so we will add some interesting information that we got from the data that we used. First, we examined the gender distribution within the ADNI dataset. As shown in Figure 1, the number of male and female participants is relatively balanced.

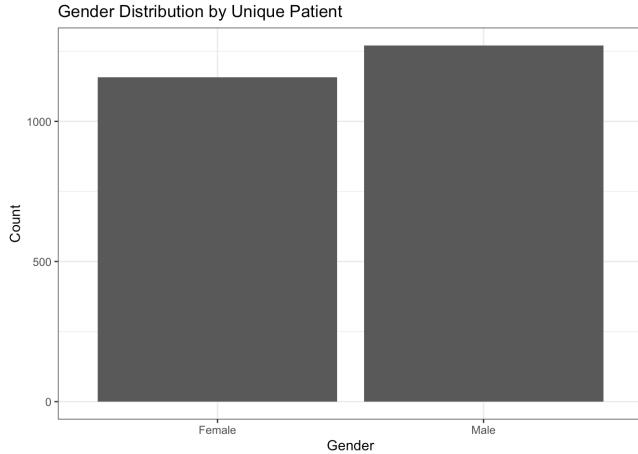


Figure 1: Gender distribution of participants in the ADNI dataset

The following analysis examines the distribution of patients' ages and their diagnoses, as shown in Figure 2. Most patients have not yet been diagnosed with Alzheimer's disease. However, as age increases, there is a noticeable trend of cognitive decline, with a rising proportion of patients being diagnosed with Mild Cognitive Impairment (MCI) across older age groups. This pattern suggests a correlation between aging and the progression toward cognitive impairment.

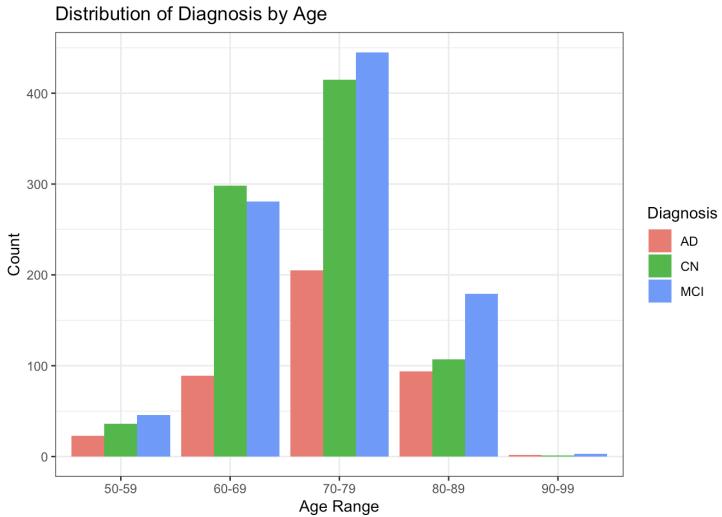


Figure 2: Diagnosis distribution of participants in the ADNI dataset

To group patients based on the progression of cognitive decline, we need a metric that effectively represents their cognitive status. Ideally, this measure should be non-invasive, which is why we selected the CDRSB score. A higher CDRSB score indicates greater cognitive impairment, with worsening cognitive conditions as the score increases. The relationship between CDRSB scores and diagnoses is shown in Figure 3, where we can

see how CDRSB scores tend to rise with increasing severity of cognitive impairment and dementia. This distribution emphasizes the role of CDRSB as a marker for assessing the progression of cognitive decline.

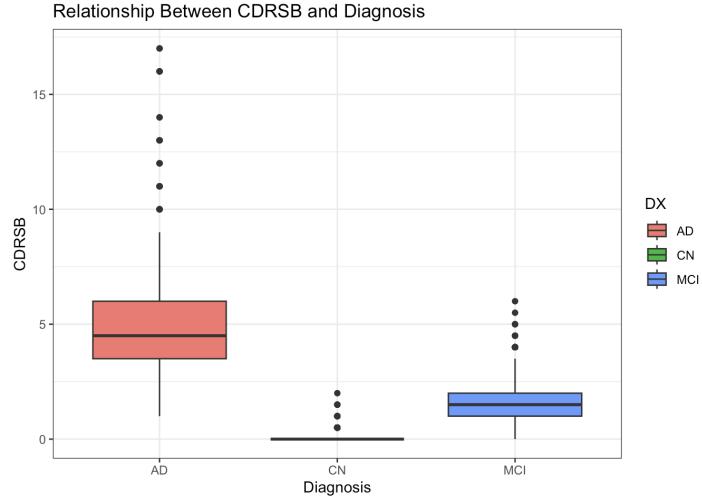


Figure 3: CDRSB and Diagnosis distribution of participants in the ADNI dataset

This data also show how the volume of the whole brain changes overtime, so by analyzing each visit of the patient, as we can see on Figure 4 that whole brain size of the patients tend to shrink over time.

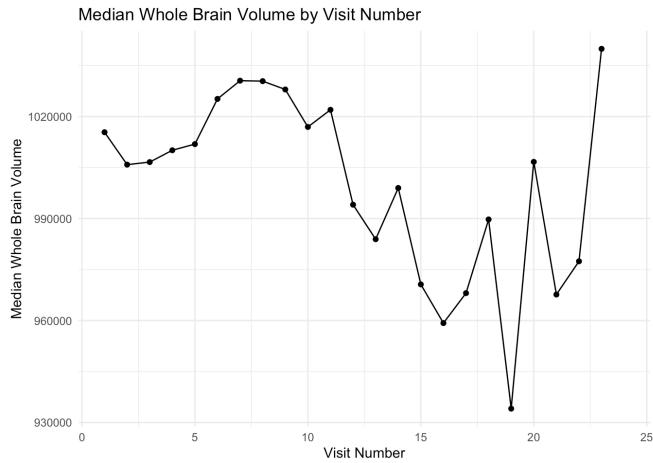


Figure 4: Median of WholeBrain size of participants in the ADNI dataset

Additionally, we can observe the distribution of CDRSB values in relation to APOE4, a genetic variant associated with Alzheimer's disease. Patients carrying one or two alleles of APOE4 tend to exhibit higher CDRSB values, indicating a stronger association with disease progression.

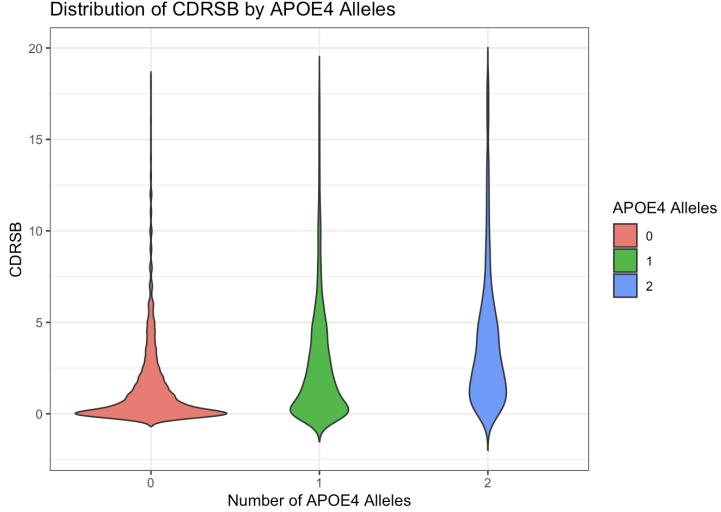


Figure 5: CDRSB and APOE4 distribution of participants in the ADNI dataset

## 3.2 Clustering

### 3.2.1 K-Means

Since we aim to create our own classification, we opted for unsupervised learning. We chose K-Means because it is a simple yet effective method, particularly well-suited for tabular data. To identify distinct cognitive decline trajectories, we apply K-means clustering using CDRSB scores. The optimal number of clusters is determined via the elbow method, resulting in four groups: Stable, Mild Progression, Moderate Progression, and Severe Progression.

### 3.2.2 Dynamic Time Warping

To incorporate the time aspect of cognitive decline, we employed Dynamic Time Warping (DTW). DTW is used to align patient trajectories over time, improving cluster consistency by handling variations in visit frequency and duration. This technique ensures that patients with similar patterns of cognitive decline are grouped together, making clustering results more robust and meaningful.

## 3.3 Classification

### 3.3.1 Extreme Gradient Boost

We used Extreme Gradient Boosting (XGBoost) to classify patients into cognitive decline categories based on MRI features. This method was selected for its high predictive accuracy and ability to handle complex feature interactions. Additionally, XGBoost provides feature importance scores, allowing us to identify which MRI biomarkers contribute most

to cognitive decline predictions.

### 3.3.2 Autoencoder

An autoencoder was trained to extract latent space representations from MRI data, reducing dimensionality while preserving essential information for classification tasks. The extracted latent representations were subsequently used in downstream classification tasks, ensuring an optimal balance between feature compression and information retention.

## 4 Experimental Results

### 4.1 Data Preprocessing

To ensure consistency across clustering and classification experiments, we applied several data preprocessing steps to prepare the dataset:

- **Handling Missing Values:** Any rows with missing data were removed to avoid potential biases in model training.
- **Normalization:** Continuous features (e.g., whole brain volume, age) were scaled using Min-Max normalization to bring values between 0 and 1, preventing models from being biased toward larger numerical ranges.
- **Modality Splitting:** We make 5 subsets from dataset representing feature modalities that we will use later on in XGBoost part:
  - **Cognitive:** (e.g., CDRSB scores)
  - **MRI:** (e.g., hippocampus, ventricles, fusiform volume)
  - **CSF:** (e.g., tau and amyloid-beta levels)
  - **PET:** (e.g., FDG, FBB)
  - **Risk Factors:** (e.g., APOE4, age, gender, education)
- **Train-Validation Splitting:** The dataset was split into 70% training and 30% validation.

These preprocessing steps ensured that input features were clean, normalized, and properly structured for clustering and classification tasks.

## 4.2 Clustering

### 4.2.1 K-Means

To assess patients' cognitive decline, the easiest and most non-invasive approach is to evaluate their cognitive test scores. The CDRSB demonstrates high consistency in repeated assessments, ensuring reliable tracking of cognitive performance over time[5]. Additionally, CDRSB effectively captures the progression of cognitive decline, making it a valuable tool for monitoring disease progression in clinical settings. Therefore, we chose to use CDRSB as the basis for our clustering approach

Before proceeding with K-means clustering, we first determined the optimal number of clusters (K). Initially, we hypothesized that using four clusters would best capture the different stages of Alzheimer's progression: Stable, Mild Progression, Moderate Progression, and Severe Progression. Our objective was to identify which features would be the most meaningful patient clusters. We experimented with both original and normalized data, ultimately testing four clustering approaches: (1) using the first and last CDRSB values, (2) using the mean and standard deviation of CDRSB, (3) combining all four values, and (4) incorporating a time element to account for the number of visits and intervals between them (Experiment results of other approach can be seen on Appendix).

To evaluate the clustering quality, we calculated the Within-Cluster Sum of Squares (WCSS) values for each method, which are essential metrics in K-means clustering to measure the compactness of the clusters. The WCSS values obtained were 13.82328, 5.543241, 19.92309, and 1169.3 for the respective methods. A lower WCSS indicates more compact and well-defined clusters.

To ensure the validity of these results, we conducted a sanity check through manual inspection, assessing whether the clusters aligned with expected cognitive progression patterns (see Table 1). Sanity check data refers to cases where the CDRSB scores fluctuate over time. In these cases, the initial and final scores remain approximately the same, but the standard deviation exceeds 1, indicating variability in cognitive assessment results. Based on this validation, we ultimately chose the third method, which combined all four values (mean, standard deviation, first, and last CDRSB scores). This method struck a balance between clustering quality and interpretability, providing insights into the progression of cognitive decline.

PTID	cdrsb_first	cdrsb_last	cdrsb_mean	cdrsb_sd	Cluster
011_S_0010	5.0	5.0	4.125000	1.181454	Moderate Progression
057_S_0839	1.0	1.5	1.250000	1.172604	Stable
098_S_4506	0.5	0.0	1.166667	1.607275	Stable
099_S_4202	1.0	1.0	1.666667	1.154701	Stable
109_S_0950	0.5	0.5	1.500000	1.172604	Stable
109_S_1343	1.5	2.0	1.166667	1.004833	Stable
109_S_4499	0.0	0.0	0.875000	1.181454	Stable
114_S_0374	4.5	4.5	4.375000	1.436141	Moderate Progression
123_S_4806	2.5	2.0	2.833333	1.004833	Mild Progression
127_S_2234	2.5	2.5	1.000000	1.048809	Mild Progression

Table 1: CDRSB Scores and Clustering Results

From the previous analysis, we observed that patients' whole brain size tends to shrink over time, as shown in Figure 4. When examining this trend across different clusters, illustrated in Figure 6, it becomes evident that the most pronounced decline occurs in patients classified under the moderate or severe progression clusters. In contrast, patients in the stable and mild progression clusters generally shows either minimal decline or maintain relatively consistent brain sizes. This distinction support the relationship between the severity of cognitive decline and the extent of brain atrophy, also can be said that this result support the sanity test that we did.

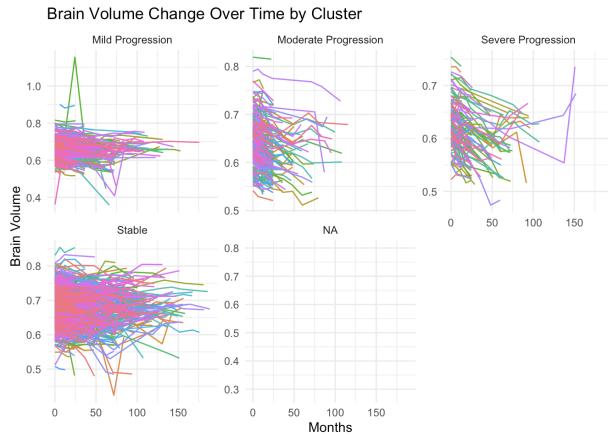


Figure 6: Whole brain size of participants in the ADNI dataset by cluster

#### 4.2.2 Determining K

To support decision on using  $K=4$ , we tried to search best k using an elbow plot (refer to image 7) on one of our K-means clustering that will be discuss more on next section. The plot shows that the WCSS decreases significantly up to  $K = 4$  or  $5$ , after which the

changes become marginal. This suggests that choosing  $K = 4$  or  $5$  would be appropriate as adding more clusters does not significantly improve the clustering quality.

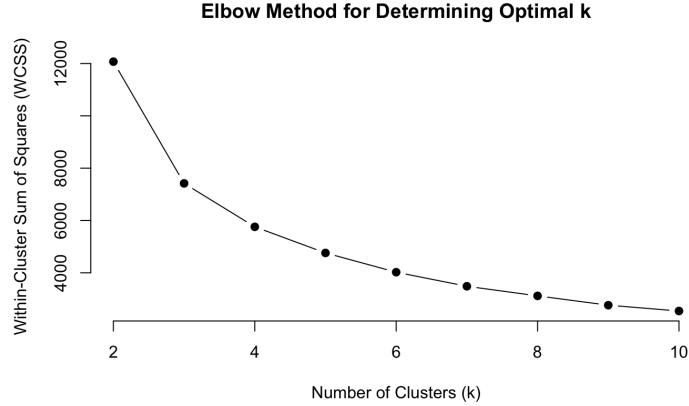


Figure 7: Elbow Method Plot

#### 4.2.3 Dynamic Time Warping

After determining the optimal  $K$ , we decided to revisit the time aspect of the data using DTW. In this phase, after normalizing the data, we performed clustering again using DTW. The resulting clusters were then used as new labels for further analysis.

DTW is advantageous when working with datasets where individuals do not share the same number of visits, as it enables alignment of sequences of varying lengths while preserving their temporal structure[6]. By leveraging DTW, we aim to achieve more accurate and meaningful clustering results that account for the time aspect of our data.

We can observe the comparison between the DTW-based clustering and the simple K-means approach (as described in Section K-Means) in Figure 8. The simple K-means method is referred to as such because it directly applies the K-means algorithm using the selected tabular data without considering temporal alignment. In contrast, the DTW-based method incorporates a dynamic time warping algorithm to account for temporal variations in the data, offering a different perspective on clustering patients based on their cognitive progression patterns. This highlights the methodological differences and potential impacts on the clustering results.

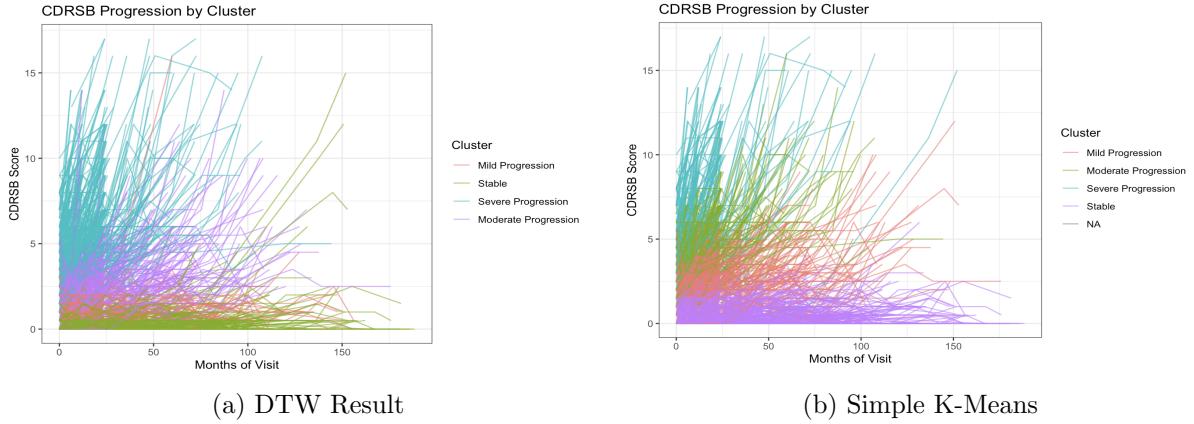


Figure 8: Comparison simple clustering vs DTW

### 4.3 Classification

In this section we are trying to find which variables are actually really significant to determine which patient belong to which class. By using the cluster we got from the Clustering section, we are trying to process the data by only taking the baseline data of every patients, as our goal is to determine how the cognitive decline progress based on the baseline data only.

#### 4.3.1 Extreme Gradient Boost

For this XGBoost model to reduce the bias, we are excluding the CDRSB values as it is the values that we used for clustering. To better analyze the contribution of different data modalities, we split the dataset into five distinct sets, each representing a different aspect of Alzheimer's disease progression. Here we are using the 5 sets that we got from the modality splitting.



Figure 9: XGBoost results on Cognitive set

An interesting observation from Figure 9 is that FAQ emerges as the most significant variable in the XGBoost analysis of the cognitive set. However, when we try to remove FAQ from the feature set to assess its impact, the significance of other variables shifts noticeably. LDELTOTAL becomes more influential, while DIGITSCORE loses some of its importance. To evaluate the predictive performance of each feature set, we compared their Area Under the Curve (AUC) values using XGBoost. Figure 10 illustrates the AUC scores for different datasets, highlighting which modality contributes most to classification accuracy. Among the datasets, the cognitive test set achieves the highest accuracy, which is expected given that cognitive assessments directly reflect the severity of Alzheimer's progression. Surprising result from the CSF set is not as good as we expect, considering that CSF testing is often a standard procedure for patients showing Alzheimer's symptoms.

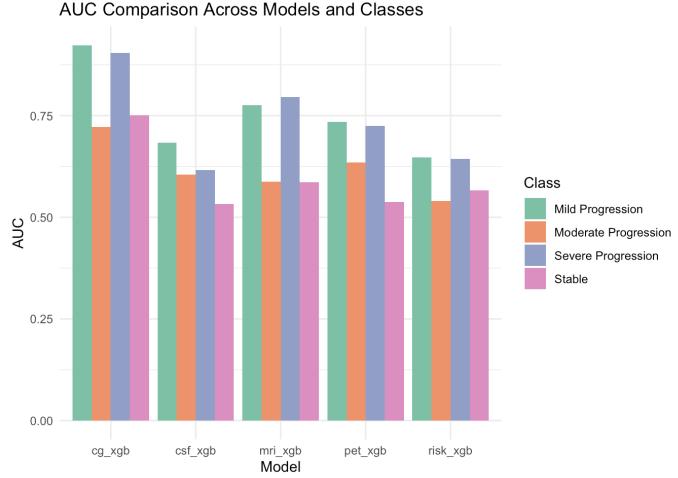


Figure 10: AUC scores from XGBoost across different feature sets (Cognitive, CSF, PET, Risk Factors, and MRI). The x-axis represents different feature sets, while the y-axis indicates the AUC score, measuring model performance

After analyzing all five individual sets, we became curious about the potential outcome of combining the MRI and risk factor sets, as these are the most commonly available data

from patients. We chose AUC (Area Under the Curve) as our scoring method because it effectively handles class imbalance and provides a clinically relevant performance measure. Unlike standard accuracy, which only considers whether predictions are correct or incorrect, AUC accounts for both false positives and false negatives, offering a more comprehensive evaluation. Additionally, we prefer AUC over ROC (Receiver Operating Characteristic) since ROC can be misleading when dealing with highly skewed datasets. The combined analysis yielded promising results, with AUC scores of 0.62, 0.81, 0.60, and 0.80 for each respective class. This indicates that integrating MRI and risk factor data provides a more comprehensive view of patient progression, as it is enhancing the model’s accuracy too if we compare it to use MRI and risk factor separately.

One key observation from all the results is that the AI performs well in distinguishing between mild and severe progression; however, the accuracy is notably lower when differentiating between stable and moderate progression. This could be due to these categories representing a transition area, where stable patients may be on the verge of progressing to mild, or moderate patients could be bordering on either mild or severe progression. This overlap likely introduces ambiguity, making it challenging for the model to draw clear distinctions between these intermediate stages.

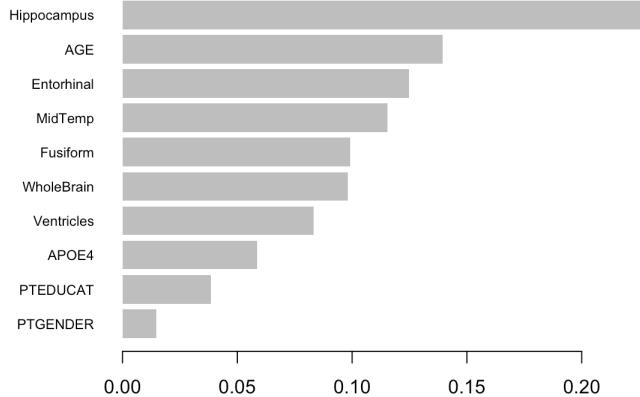


Figure 11: Feature importance scores for the Risk and MRI set, as determined by the XGBoost model. The x-axis represents relative feature importance, while the y-axis lists the features used in the model. Hippocampus volume emerges as the most significant predictor of cognitive decline, followed by age and entorhinal cortex volume.

#### 4.3.2 Autoencoder

We used the Autoencoder for dimensionality reduction and fed the resulting latent space into XGBoost to evaluate the significance of each latent space dimension to the clustering.

We experimented with latent spaces of 4, 8, and 16 dimensions, ultimately selecting 4 dimensions as it provided the best accuracy.

The Autoencoder yielded some interesting results. In the cognitive set, we observed how each latent space dimension contributed meaningfully. However, in the MRI set, only latent 3 showed significant values, while the other latent dimensions remained at zero. Even though latent 4 having nonzero values, it was fully covered by latent 3 due to their near-perfect correlation (0.9997), rendering it redundant. This highlights a stark difference in how information is represented across these datasets.

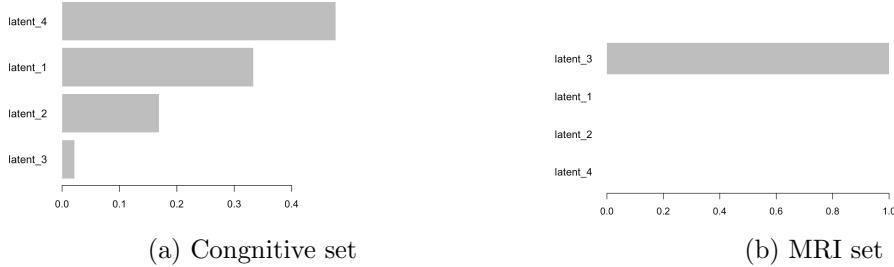


Figure 12: Latent space representations from the autoencoder for the cognitive and MRI datasets. The x-axis represents different latent space dimensions, while the y-axis indicates the contribution of each dimension to clustering

Building upon our XGBoost results, we further evaluated model performance using different feature sets. Figure 13 summarizes the AUC scores for models trained on cognitive, CSF, PET, risk factor, and MRI datasets.

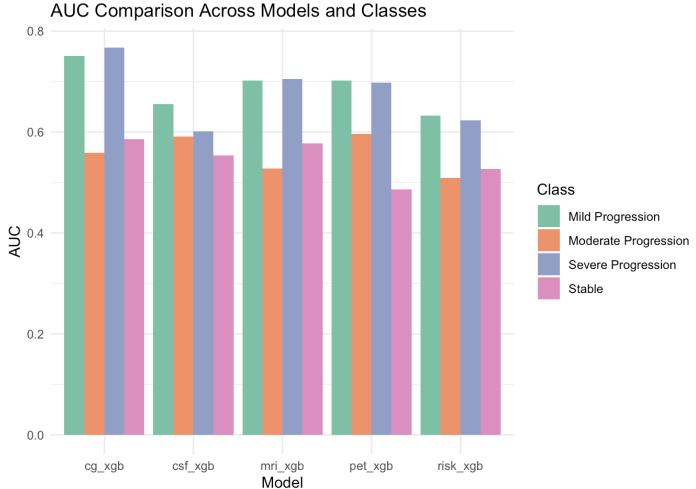


Figure 13: AUC scores for different predictive models trained on individual feature sets. The x-axis represents the feature sets, while the y-axis shows the corresponding AUC score, reflecting model performance.

For this final result, we can also see it clearer in table 2.

Feature Set	Stable	Mild	Moderate	Severe
Cognitive	0.5859	0.7504	0.5589	0.7675
CSF	0.5541	0.6554	0.5906	0.6015
PET	0.4865	0.7023	0.5966	0.6974
Risk	0.5267	0.6329	0.5093	0.6229
MRI	0.5773	0.7018	0.5282	0.7049
MRI + Risk	0.5506	0.6755	0.5388	0.6441

Table 2: AUC Scores for Different Feature Sets Across Classes

## 5 Discussion

Rather than drawing definitive conclusions from the experimental results, our focus is on interpreting the findings in the context of cognitive decline prediction. This study aimed to assess how different data modalities—MRI, cognitive scores, CSF, and PET scans—contribute to predicting Alzheimer’s disease progression. Among these, MRI emerged as a promising non-invasive modality, offering a viable alternative to more invasive and costly procedures like CSF and PET imaging.

Our findings demonstrate that MRI-derived features effectively capture neurodegenerative patterns associated with cognitive decline. Specifically, our classification models identified ventricular enlargement and whole-brain atrophy as key indicators of Alzheimer’s

progression. These results align with prior studies showing that MRI-based biomarkers can distinguish between different stages of AD, reinforcing their role in non-invasive diagnostics[7]. Compared to CSF and PET, MRI-based classification maintained competitive AUC scores, further validating its clinical applicability.

To further validate our findings, we conducted a linear regression analysis on multiple datasets. While APOE4 remained a significant genetic risk factor, MRI-based features exhibited stronger associations with cognitive decline, highlighting their importance as structural biomarkers. This reinforces the potential of MRI as a primary predictive tool, especially when integrated with patient demographics. (Full regression details are available in Appendix)

Additionally, our results suggest that combining MRI with key risk factors (age, APOE4, gender, and education) enhances predictive accuracy. Unlike PET and CSF tests, which require invasive procedures, MRI provides a practical and scalable alternative for clinical applications. Future research should focus on refining MRI-based feature extraction, leveraging 3D imaging, and integrating deep learning methods for more precise cognitive decline prediction.

## 6 Conclusion

This study highlights the significance of MRI as a key modality for predicting cognitive decline in Alzheimer’s disease. While other biomarkers such as CSF and PET scans offer valuable insights, they are often limited by their invasive nature and high costs. Our findings demonstrate that MRI-based features correlate well with cognitive decline, making it a viable, non-invasive tool for early diagnosis and monitoring of Alzheimer’s progression.

Furthermore, by incorporating machine learning models, we effectively classified patients into different cognitive decline trajectories, leveraging MRI data and risk factors. The combined analysis of MRI and risk sets showed strong predictive capability, reinforcing the idea that a multi-modal approach enhances diagnostic accuracy.

Our findings highlight MRI’s potential for early AD detection, particularly in distinguishing subtle structural changes in brain regions. Future work should explore multi-modal approaches, integrating MRI with genetic risk factors and cognitive assessments to enhance predictive power. Additionally, the use of deep learning models on high-resolution 3D MRI scans could further improve classification accuracy.

## References

- [1] Reisa A. Sperling et al. “Toward defining the preclinical stages of Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease”. In: *Alzheimer’s Dementia* 7.3 (2011), pp. 280–292. doi: 10.1016/j.jalz.2011.03.003.
- [2] Mostafa Mehdipour Ghazi and et al. “Cognitive aging and reserve factors in the Metropolit 1953 Danish male cohort”. In: *Frontiers in Aging Neuroscience* 16 (2024), p. 1345417. doi: 10.3389/fnagi.2024.1345417.
- [3] Modupe Odusami et al. “Analysis of features of Alzheimer’s disease: Detection of early stage from functional brain changes in magnetic resonance images using a fine-tuned ResNet18 network”. In: *Diagnostics* 11.6 (2021), p. 1071. doi: 10.3390/diagnostics11061071.
- [4] L. S. Kumar et al. “AlexNet approach for early stage Alzheimer’s disease detection from MRI brain images”. In: *Materials Today: Proceedings* 51 (2022), pp. 58–65. doi: 10.1016/j.matpr.2021.08.321.
- [5] F. McDougall et al. “Psychometric Properties of the Clinical Dementia Rating - Sum of Boxes and Other Cognitive and Functional Outcomes in a Prodromal Alzheimer’s Disease Population”. In: *Journal of Prevention of Alzheimer’s Disease* 8.2 (2021), pp. 151–160. doi: 10.14283/jpad.2020.73.
- [6] Danilo Avola et al. “Signal enhancement and efficient DTW-based comparison for wearable gait recognition”. In: *Computers & Security* 137 (2024), p. 103643. doi: 10.1016/j.cose.2023.103643. URL: <https://doi.org/10.1016/j.cose.2023.103643>.
- [7] Clifford R. Jack et al. “NIA-AA Research Framework: Toward a biological definition of Alzheimer’s disease”. In: *Alzheimer’s Dementia* 15.4 (2019), pp. 535–562. doi: 10.1016/j.jalz.2018.02.018.

## A K-Means

### A.1 Fluctuating CDRSB patients

We are searching for patients that has fluctuating CDRSB value for us to verify sanity test earlier in the project. In here we are searching for patients that has difference between first and last cdrsb below 1, but standard deviation above 1.

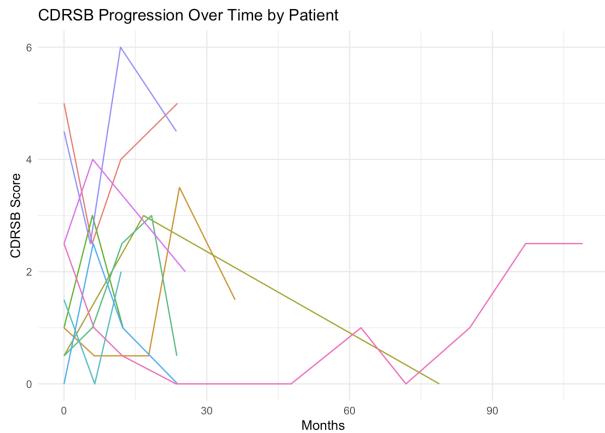


Figure 14: Identification of fluctuating CDRSB values for sanity testing. Patients were selected based on having a difference of less than 1 between their first and last CDRSB scores but a standard deviation above 1, indicating fluctuations in cognitive assessment over time.

### A.2 K-Means Experiments

As mentioned above, we did some experiments with the K-Means before decide to use DTW.

## B XGBoost

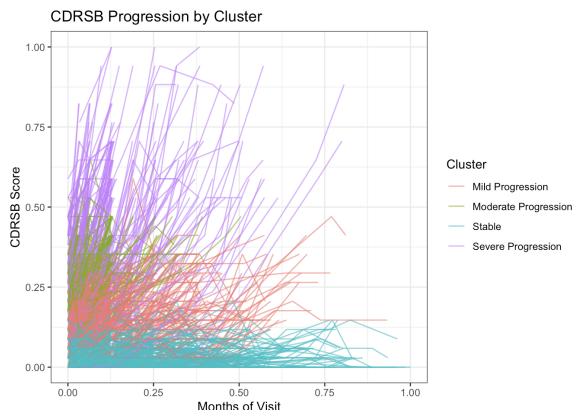


Figure 15: K-Means clustering based on first and last CDRSB values

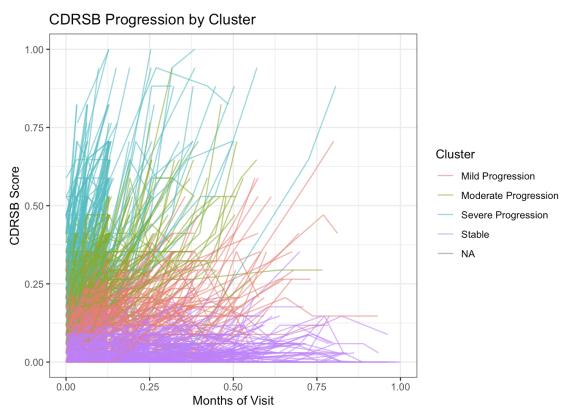


Figure 16: K-Means clustering using mean and standard deviation of CDRSB

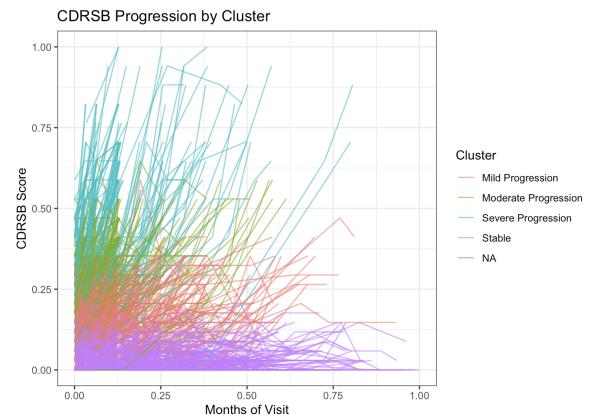


Figure 17: K-Means clustering using all four values (first, last, mean, standard deviation of CDRSB)

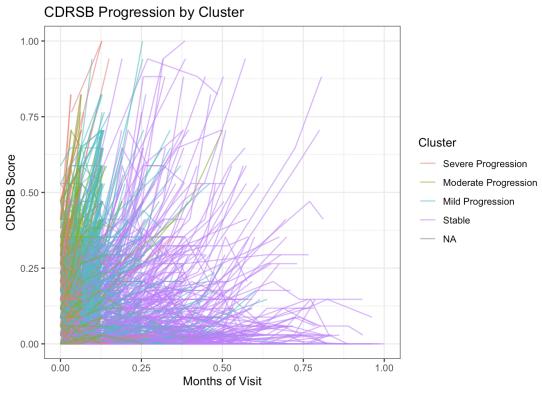


Figure 18: K-Means clustering incorporating a time-based aspect, accounting for the number of visits and time intervals between them

## B.1 Modality sets

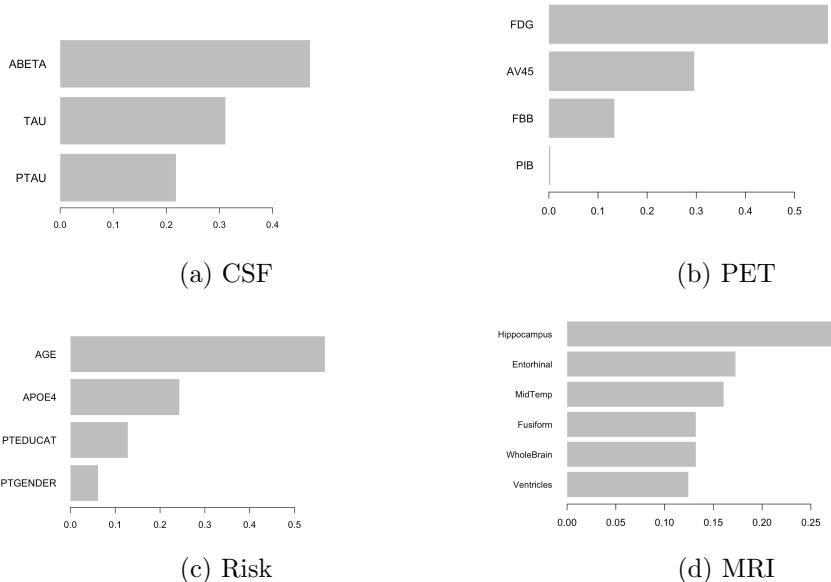


Figure 19: Feature importance analysis using XGBoost across different data modalities: (a) CSF, (b) PET, (c) Risk Factors (age, APOE4, education, gender), and (d) MRI features (hippocampus, whole brain, ventricles, etc.). The model evaluates which features contribute most to predicting cognitive decline

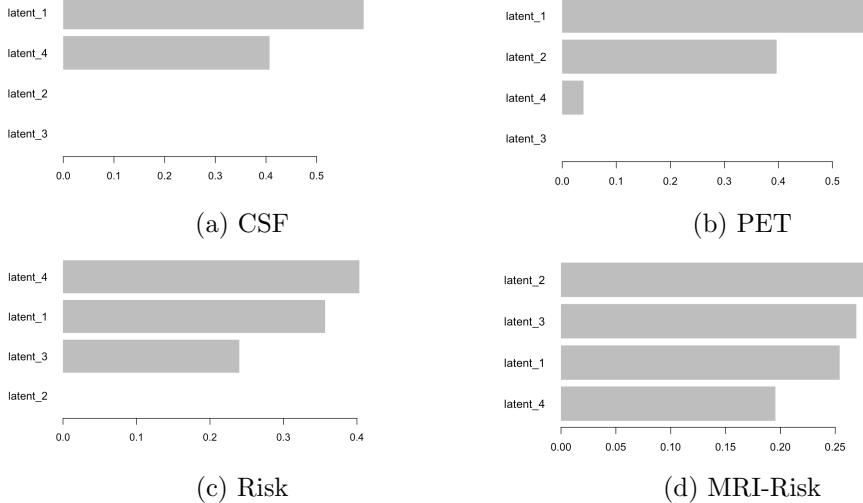


Figure 20: Visualization of latent space representations across different modalities: (a) CSF, (b) PET, (c) Risk Factors, and (d) Combined MRI-Risk set. These representations highlight the underlying feature distribution and structure for each dataset

## B.2 Significance of latent space

## C Linear Regression

To further explore the relationship between key biomarkers and cognitive decline, we conducted a linear regression analysis on multiple datasets. The analysis focused on assessing the influence of APOE4, age, gender, and MRI-based features on cognitive status.

### C.1 Regression Results

These results highlight that while genetic factors play a role, MRI-based features provide essential structural markers of disease progression\*\*, supporting their inclusion in predictive models.

```

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.592989  0.033722 17.584 <2e-16 ***
## latent_1    -0.008174  0.019853 -0.412   0.681
## latent_4     0.002469  0.114663  0.022   0.983
## cluster_set  0.006001  0.004018 1.494   0.135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1787 on 1704 degrees of freedom
## Multiple R-squared:  0.01075, Adjusted R-squared:  0.00904
## F-statistic:  6.17 on 3 and 1704 DF, p-value: 0.0003613

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.46441   0.09307 15.735 <2e-16 ***
## latent_1    -0.12849   0.05479 -2.345  0.0191 *
## latent_4     0.62590   0.31645  1.978  0.0481 *
## cluster_set  0.02142   0.01109  1.932  0.0535 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4933 on 1704 degrees of freedom
## Multiple R-squared:  0.01454, Adjusted R-squared:  0.01281
## F-statistic:  8.382 on 3 and 1704 DF, p-value: 1.568e-05

(a) Age                                (b) Gender

```

---

```

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.105993  0.063549  1.668 0.095524 .
## latent_1    -0.155137  0.037765 -4.108 4.19e-05 ***
## latent_4     0.784671  0.217833  3.602 0.000325 ***
## cluster_set  0.051229  0.007634  6.710 2.66e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3312 on 1649 degrees of freedom
## (55 observations deleted due to missingness)
## Multiple R-squared:  0.06367, Adjusted R-squared:  0.06197
## F-statistic: 37.38 on 3 and 1649 DF, p-value: < 2.2e-16

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.527100  0.014081 37.434 <2e-16 ***
## latent_1    0.031785  0.011311  2.810 0.00501 **
## latent_2    -0.031745  0.008600 -3.691 0.00023 ***
## cluster_set  0.007333  0.003888  1.886 0.05944 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1786 on 1704 degrees of freedom
## (3 observations deleted due to missingness)
## Multiple R-squared:  0.0126, Adjusted R-squared:  0.01086
## F-statistic:  7.25 on 3 and 1704 DF, p-value: 7.833e-05

(c) APOE4

```

Figure 21: Linear regression results analyzing cognitive set variables: (a) Age, (b) Gender, and (c) APOE4 status. While genetic factors play a role, MRI-based features demonstrate stronger associations with cognitive decline

```

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.231721  0.018221 12.717 <2e-16 ***
## latent_1    0.011087  0.001403  7.902 4.96e-15 ***
## latent_4    -0.016896  0.002116 -7.985 2.60e-15 ***
## cluster_set  0.045099  0.007265  6.208 6.78e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.323 on 1652 degrees of freedom
## (55 observations deleted due to missingness)
## Multiple R-squared:  0.1092, Adjusted R-squared:  0.1076
## F-statistic: 67.51 on 3 and 1652 DF, p-value: < 2.2e-16

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.527100  0.014081 37.434 <2e-16 ***
## latent_1    0.031785  0.011311  2.810 0.00501 **
## latent_2    -0.031745  0.008600 -3.691 0.00023 ***
## cluster_set  0.007333  0.003888  1.886 0.05944 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1786 on 1704 degrees of freedom
## (3 observations deleted due to missingness)
## Multiple R-squared:  0.0126, Adjusted R-squared:  0.01086
## F-statistic:  7.25 on 3 and 1704 DF, p-value: 7.833e-05

(a) CSF                                (b) PET

```

---

```

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.266370  0.056949  4.677 3.14e-06 ***
## latent_3    -0.102662  0.061538 -1.668 0.0955 ***
## cluster_set  0.060514  0.007521  8.046 1.62e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3343 on 1653 degrees of freedom
## (55 observations deleted due to missingness)
## Multiple R-squared:  0.04493, Adjusted R-squared:  0.04377
## F-statistic: 38.88 on 2 and 1653 DF, p-value: < 2.2e-16

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.527100  0.014081 37.434 <2e-16 ***
## latent_1    0.031785  0.011311  2.810 0.00501 **
## latent_2    -0.031745  0.008600 -3.691 0.00023 ***
## cluster_set  0.007333  0.003888  1.886 0.05944 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1786 on 1704 degrees of freedom
## (3 observations deleted due to missingness)
## Multiple R-squared:  0.0126, Adjusted R-squared:  0.01086
## F-statistic:  7.25 on 3 and 1704 DF, p-value: 7.833e-05

(c) MRI

```

Figure 22: Linear regression results comparing different data modalities: (a) CSF, (b) PET, (c) MRI. The results highlight the varying degrees of association between cognitive decline and different biological markers