Towards Multimodal Longitudinal Analysis for Predicting Cognitive Decline Using Neuroimaging Biomarkers, Cognitive Assessments, and Demographic Data

Rayan M. Ashish[1]* and Jessica A. Turner[2]

[1]*The Head-Royce School, Oakland CA*

[2]*Department of Psychiatry and Behavioral Health, Wexner Medical Center, The Ohio State University, Columbus, OH*

ABSTRACT

Understanding and predicting cognitive decline in Alzheimer's disease (AD) is crucial for timely intervention and management. While neuroimaging biomarkers and clinical assessments are valuable individually, their combined predictive power and interaction with demographic and cognitive variables remain underexplored. This study lays the groundwork for comprehensive longitudinal analyses by integrating neuroimaging markers and clinical data to predict cognitive changes over time. Using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), feature-driven supervised machine learning techniques were applied to assess cognitive decline predictability. Results showed that while imaging biomarkers alone offered moderate predictive capabilities, including cognitive and demographic factors significantly improved model performance, with the Random-Forest model achieving near-perfect accuracy. The analysis highlighted the leading role of neuroimaging biomarkers as predictors, along with the importance of incorporating cognitive scores and demographic data to enhance predictability. Explainability analyses further revealed that clinical and demographic data could estimate brain imaging metrics, emphasizing the bidirectional nature of these relationships. This study underscores the need for integrating multi-dimensional data in future longitudinal research to capture time-dependent patterns in cognitive decline and guide the development of targeted intervention strategies. We also introduce and provide NeuroLAMA, an open and extensible data engineering and machine-learning system, to support the continued investigation by the community

---

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

INTRODUCTION

Alzheimer's disease (AD) is one of the most prevalent neurodegenerative disorders, characterized by progressive cognitive decline and structural brain changes (Scheltens et al., 2021). Early detection and accurate monitoring of cognitive impairment are vital for managing disease progression and evaluating the efficacy of therapeutic interventions. Biomarkers derived from neuroimaging, such as the Spatial-Pattern-of-Abnormality-for-Recognition-of-Early-Alzheimer's-Disease (**SPARE_AD**), have been established as reliable indicators of brain atrophy associated with AD (Contador et al., 2021). These imaging-based biomarkers provide quantitative measures that reflect underlying neurodegenerative changes, making them invaluable tools for research and clinical practice.

Cognitive and functional performance is often assessed using clinical tools like the Clinical-Dementia-Rating-Sum-of-Boxes (**CDRSB**), which evaluates cognitive decline through a comprehensive set of cognitive and functional domains. The *change* in CDRSB scores over time (**CDRSBDIFF**) serves as a measure of disease progression, offering insights into how patients' cognitive and functional abilities evolve (Yang et al., 2021). Despite the known value of neuroimaging biomarkers and clinical assessments individually, understanding their predictive relationship and how they interact with other factors such as demographic and cognitive variables remains an area ripe for exploration. Recent studies underscore the importance of integrating neuroimaging biomarkers with clinical assessments and demographic factors to enhance the prediction of cognitive decline. For instance, a 2024 study demonstrated that combining plasma biomarkers with traditional imaging techniques improves the prediction of cognitive decline in non-demented individuals, highlighting the value of a multifaceted approach (Mendes et al., 2024). Similarly, research published in 2023 found that multi-modal neuroimaging signatures, when combined with clinical data, effectively predict cognitive decline in individuals with multiple sclerosis, emphasizing the need to consider various factors in predictive modeling (Al-iedani et al., 2024). These findings suggest that while neuroimaging and clinical assessments are valuable individually, their predictive power is significantly enhanced when analyzed in conjunction with demographic and cognitive variables.

This study investigates the predictive power of SPARE_AD for changes in cognitive performance (CDRSBDIFF) and extends the analysis to include a combination of demographic and cognitive variables, such as age, education length, and Mini-Mental-State-Examination (MMSE) scores. Additionally, the study explores the ability of clinical assessments and demographic variables to

predict imaging-derived SPARE_AD scores and evaluates broader cognitive measures such as physical activity and cognitive or psychological scores, to understand their association with neuroimaging and clinical performance.

To address these research questions, multiple feature-driven models, including Linear-Regression and Random-Forest Regression, were applied (Uddin et al., 2024). The integration of machine learning approaches provided an opportunity to capture non-linear relationships and assess the predictive power of multi-dimensional data. This comprehensive analysis aims to elucidate how well imaging biomarkers and cognitive scores, in conjunction with demographic factors, can predict cognitive decline, brain atrophy, and overall cognitive health.

By leveraging *longitudinal* data from ADNI, this work contributes to a deeper understanding of the complex relationships among neuroimaging, cognitive performance, and demographic variables. The findings highlight the potential for combining imaging and clinical data to enhance early detection, monitoring, and prediction of Alzheimer's disease progression, ultimately informing more effective intervention strategies.

METHODOLOGY

ADNI Overview

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal, multi-center study launched in 2004 with the aim of developing clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease (AD) (Veitch et al., 2024). Funded by various public and private partners, ADNI has become a cornerstone in the field of AD research, providing an extensive dataset that supports the exploration of disease progression from preclinical stages through mild cognitive impairment (MCI) to advanced AD. The study collects data from a wide range of participants, including cognitively normal older adults, individuals with MCI, and patients with AD, enabling researchers to investigate patterns of cognitive decline and identify potential early indicators of the disease. The comprehensive data provided by ADNI includes neuroimaging scans (such as MRI and PET), cognitive assessments, genetic information, and clinical evaluations, making it an invaluable resource for developing predictive models and testing hypotheses related to the diagnosis, monitoring, and treatment of Alzheimer's disease.

The Longitudinal Nature of the Data

The dataset utilized in this study was derived from the Alzheimer's Disease Neuroimaging Initiative (ADNI), which conducts longitudinal assessments of participants. The Clinical Dementia Rating Sum of Boxes (CDRSB) scores, used to calculate the derived variable 'CDRSBDIFF' which is the change in CDSRB (at that assessment) over the *first assessment* in the study data (our baseline). For this analysis we selected individuals that had **at least 6 and at most 10** assessments over the years with recorded CDRSB scores, resulting in a cohort of 661 subjects. The table below provides the basic age and gender distribution of these subjects, as well as the statistical distribution of CDSRB and CDRSBDIFF in the data analyzed.

AGE (years)			GENDER		CDRSB				CDRSBDIFF			
MIN	**MAX**	**AVG**	**M**	**F**	**MIN**	**MAX**	**AVG**	**STD**	**MIN**	**MAX**	**AVG**	**STD**
53	109	85	359	302	0.0	18.0	1.88	2.75	0	17.5	3.29	3.93

Selected Features and Tables from ADNI for This Study

In this study, specific features from the ADNI dataset were chosen to investigate the relationships between neuroimaging biomarkers, cognitive performance, and demographic characteristics. The primary neuroimaging variable of interest is **SPARE_AD**, an index derived from MRI scans that reflects Alzheimer's disease-related brain atrophy. SPARE_AD has been established as an effective biomarker for detecting early structural brain changes and tracking disease progression. The cognitive outcome of interest is **CDRSBDIFF**, which measures the change in the Clinical Dementia Rating Sum of Boxes (CDRSB) over time, providing a metric for assessing the degree of cognitive decline. Demographic variables included **age** and **education level (PTEDUCAT)**, which are critical for understanding individual differences in cognitive reserve and susceptibility to cognitive decline. The **MMSCORE (Mini-Mental State Examination score)** was selected as a cognitive performance measure that offers a snapshot of overall cognitive function. To gain insight into the potential interaction between cognitive and physical health, **PEDAVG** (physical activity average) and **AMASAVG** (a cognitive or psychological average score) were included as additional features. These variables help to contextualize the relationship between brain structure, cognitive performance, and functional health.

Investigation, and Analysis System

Our long term goal is that of uncovering specific predictors, including combinations thereof, towards cognitive decline and by leveraging heterogenous potential factors. Figure 1 provides a

schematic architecture of the first version of **NeuroLAMA** (for Neuro Longitudinal and Multimodal Analysis), which is a machine learning system for such analyses.

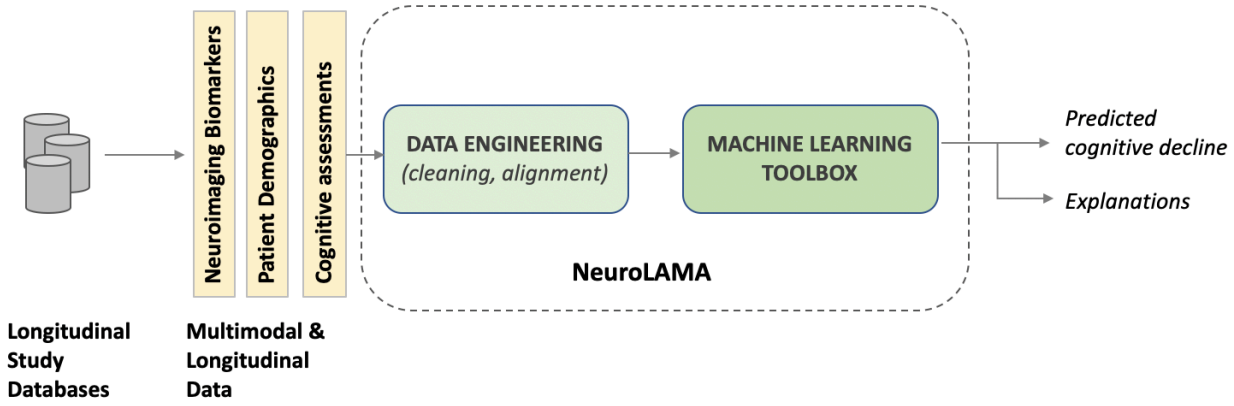


Figure 1: The NeuroLAMA System for Multimodal Longitudinal Predictive Analysis

NeuroLAMA is an open and extensible machine-learning resource for analysis over prominent datasets of longitudinal Neuroscience datasets (such as ADNI), that we have developed as part of this investigation. It currently provides the feature-driven machine-learning algorithms that were employed in the analysis reported in this paper. The system is implemented in Python and is built upon the Pandas library for data processing, and the SciKit Learn library for machine learning. It is available at ww.neurolama.org[2].

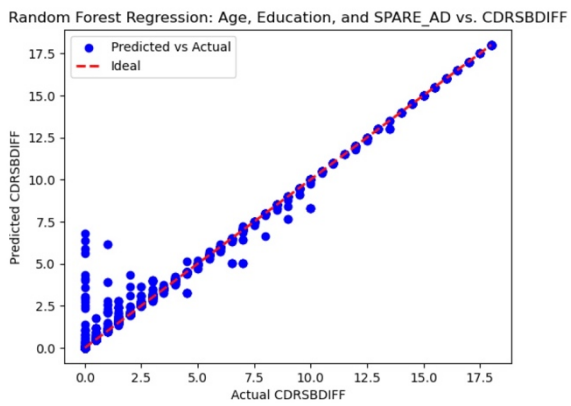
RESULTS

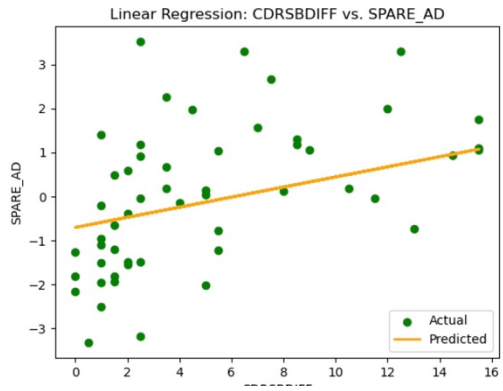
1. **Prediction of CDRSBDIFF using SPARE_AD (only)**

This analysis focused on predicting changes in cognitive performance (CDRSBDIFF) using SPARE_AD as the primary predictor. SPARE_AD, a biomarker derived from MRI data indicating brain atrophy associated with Alzheimer's disease, was analyzed for its effectiveness in predicting clinical outcomes. Linear Regression and Random Forest Regression models were employed, revealing the extent to which structural brain changes could act as standalone predictors of cognitive decline (Table 1).

[2] www.neurolama.org

Table 1: Association of SPARE_AD with CDRSBDIFF





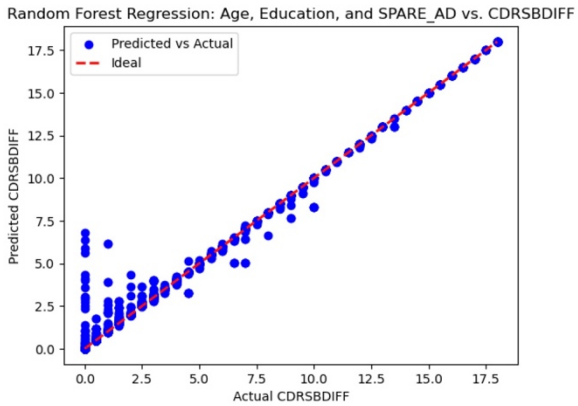
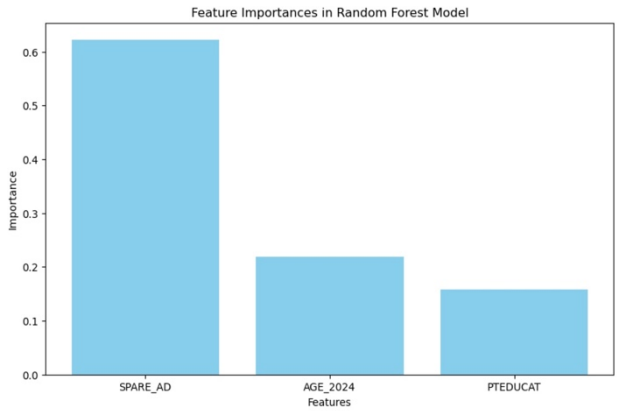
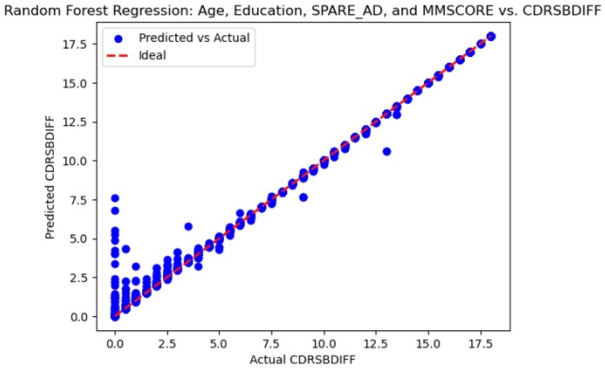
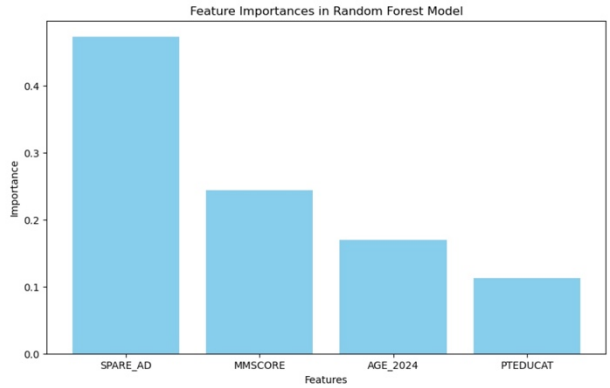
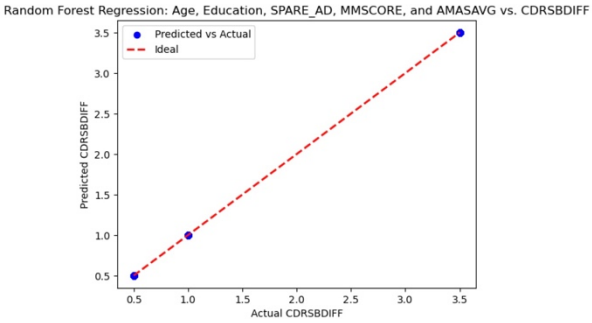
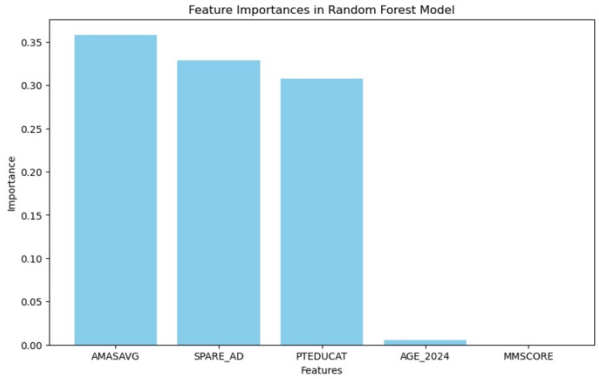
2. **Investigating different combinations of factors**

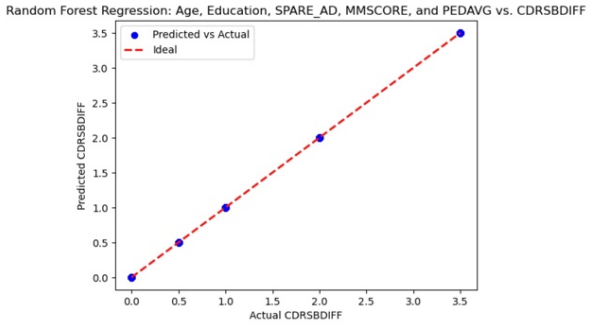
We then investigated employing different combinations of variables for predicting CDRSBDIFF. Table 2(a) provides the five different sets of variables we evaluated, followed by the results in Table 2(b) (the results include the predictive accuracy as well as feature importance of the variables employed).

Table 2(a): Variables sets for predicting CDRSBDIFF

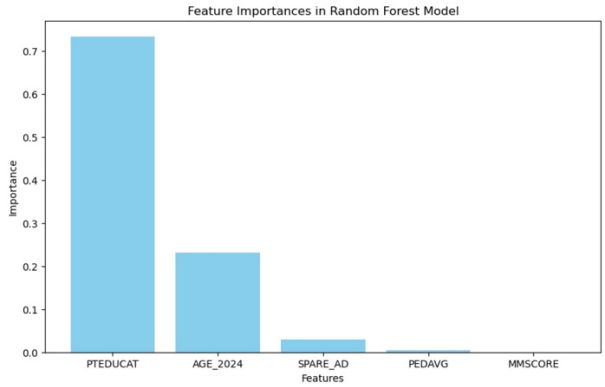
Variables Sets
{SPARE_AD, AGE_2024, PTEDUCAT}
{SPARE_AD, MMSCORE, AGE_2024, PTEDUCAT}
{AMASAVG, SPARE_AD, MMSCORE, AGE_2024, PTEDUCAT}
{PEDAVG, SPARE_AD, MMSCORE, AGE_2024, PTEDUCAT}
{PEDAVG, MMSCORE, AGE_2024, PTEDUCAT}

Table 2(b): Investigating different variable combinations for predicting CDRSBDIFF

Predictive accuracy	**Feature Importance**
$y\ axis$: Predicted CDRSBDIFF $x\ axis$: Actual CDRSBDIFF	$y\ axis$: Feature Importance $x\ axis$: Features
 **MSE:** 0.21 **R^2**: 0.99	 **SPARE_AD:** 0.42 **AGE_2024**: 0.22 **PTEDUCAT:** 0.16
 **MSE:** 0.19 **R^2**: 0.99	 **SPARE_AD:**0.47 **MMSCORE:**0.24 **AGE_2024**: 0.17 **PTEDUCAT:** 0.11
 **MSE:** 0.0 **R^2**: 1.0	 **AMASAVG:**0.36 **SPARE_AD:**0.33 **PTEDUCAT:**0.31 **AGE_2024**: 0.01 **MMSCORE:**0.00



MSE: 0.0 **R^2**: 1.0



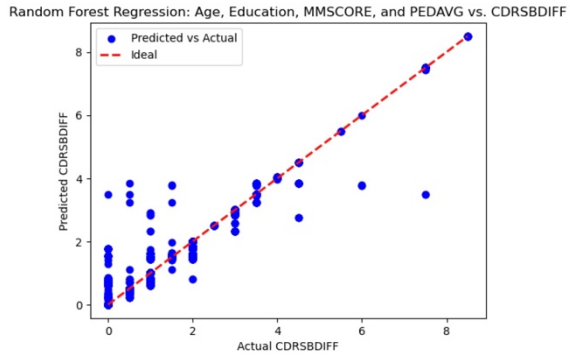
PTEDUCAT:0.73 **AGE_2024**:0.23 **SPARE_AD:**0.03 **PEDAVG:**0.00 **MMSCORE:**0.00

The results from Table 2(b) demonstrate that the Random-Forest Regression model proved particularly effective, capturing non-linear interactions and providing high predictive accuracy, as evidenced by an R-squared value approaching 0.99 in some models. The feature importance analysis underscored SPARE_AD's role as the most influential predictor, supporting its clinical utility in monitoring AD progression. The incorporation of cognitive scores like MMSCORE and demographic variables such as age and PTEDUCAT into predictive models offers a more nuanced view of how cognitive reserve and structural changes interact to influence cognitive outcomes over time.

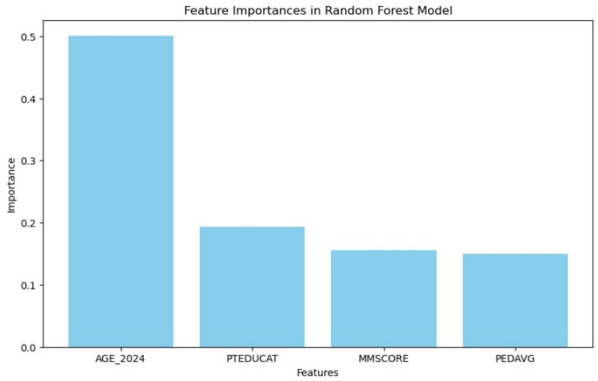
3. Prediction using more accessible (non-imaging variables)

In the final analysis, a model was developed to predict the CDR (Clinical Dementia Rating) score using only *non-imaging* variables: **AGE**, **PTEDUC (education level)**, **MMSCORE (Mini-Mental State Examination score)**, and **PEDAVG (physical activity average)**. The model's performance was evaluated across different diagnostic categories—Alzheimer's disease (AD), mild cognitive impairment (MCI), and cognitively normal (CN). We explored both the Random-Forest as well as XGBoost classifiers with the results provided in Table 3 below.

Table 3: Predicting the CDR score with non-imaging variables



MSE: 0.0 9 **R^2**: 0.94



AGE_2024:0.50 **PTEDUCAT:**0.19 **MMSCORE:**0.16 **PEDAVG:**0.15

Category	Precision	Recall	F1-Score	Support
AD	0.90	0.92	0.91	38
CN	0.98	0.92	0.95	194
MCI	0.92	0.97	0.95	204

Confusion Matrix:

```
[[35   0   3]
 [1   179  14]
 [3   3   198]]
```

The results demonstrate impressive predictive accuracy despite not including neuroimaging data such as SPARE_AD. Achieving high accuracy using only these demographic and cognitive assessment variables highlights the potential for more accessible and cost-effective prediction models. This approach underscores the importance of well-chosen non-imaging variables that can contribute significantly to early identification and classification in AD research and clinical practice.

DISCUSSION

The findings from this study highlight the importance of integrating neuroimaging biomarkers, cognitive scores, and demographic variables to predict changes in cognitive function, as measured by CDRSBDIFF. The analyses demonstrated that SPARE_AD, a robust imaging biomarker reflecting Alzheimer's-related brain atrophy, serves as a significant predictor of cognitive decline. However, the predictive power was enhanced when additional variables, such

as age, education level, and MMSCORE, were included. This multi-dimensional approach aligns with the understanding that cognitive decline is influenced by an interplay of structural brain changes, cognitive reserve, and demographic factors.

These result with non-imaging variable illustrate that even without neuroimaging biomarkers, accurate prediction of cognitive status across AD, MCI, and CN groups is possible using age, education, cognitive scores, and physical activity data. This finding holds significance for developing scalable, resource-efficient screening tools in clinical and community settings.

Despite the promising results, some limitations should be noted. While the models performed well on the provided data, external validation with other cohorts is necessary to confirm the generalizability of the findings. Additionally, potential confounding factors not included in the current analysis, such as *lifestyle attributes* (Bhatti et al., 2020) as well as *genetic predispositions* (such as APOE-ε4 status) (Qian et al., 2021), may further elucidate the complex relationships affecting cognitive decline.

Future Work

Given the longitudinal nature of the data used in this study, future work should emphasize building predictive models that leverage the temporal dimension to enhance understanding of disease progression over time. Advanced machine learning approaches, such as *Recurrent-Neural-Networks* (RNNs) and *Transformer Models* (Hu et al., 2023), could be employed to model the sequential aspects of cognitive decline and brain changes. These methods would be particularly effective in capturing time-dependent patterns that static models may overlook, offering a more dynamic prediction of future cognitive states.

Further research can also investigate the integration of additional longitudinal variables that contribute to cognitive resilience and decline, such as physical activity patterns, social engagement, and comprehensive lifestyle measures. Including genetic information (e.g., APOE-ε4 genotype) and other biomarkers (e.g., tau and amyloid levels) could provide a more detailed picture of the interplay between genetic risk factors and observed brain and cognitive changes over time. Moreover, longitudinal data can be used to analyze how changes in variables like SPARE_AD interact with shifts in *cognitive reserve* (Alvares Pereira et al., 2022) and other protective factors over multiple time points. Understanding these interactions will be essential for

developing targeted interventions aimed at delaying the onset of cognitive decline and mitigating the impact of Alzheimer's disease.

Our continued work will focus on developing both the machine-learning as well as the clinical analysis thrusts of the work. In machine-learning we are working on developing and providing more advanced analysis algorithms, specifically RNNs and Transformers, in the NeuroLAMA system. In the clinical analysis domain our focus will be on applying these longitudinal models to diverse populations, to ensure findings are applicable across different racial, ethnic, and socio-economic groups. This would enhance the equity and generalizability of the predictive models and inform culturally tailored interventions that address the unique needs of various populations.

CONCLUSION

This study sets a foundation for comprehensive longitudinal analysis by integrating neuroimaging biomarkers, cognitive assessments, and demographic data to predict cognitive decline. The NeuroLAMA system introduced as part of this research will serve as an open and extensible machine learning and data engineering framework for the community.

REFERENCES

Al-iedani, O., Lea, S., Alshehri, A., Maltby, V. E., Saugbjerg, B., Ramadan, S., Lea, R., & Lechner-Scott, J. (2024). Multi-modal neuroimaging signatures predict cognitive decline in multiple sclerosis: A 5-year longitudinal study. *Multiple Sclerosis and Related Disorders*, *81*. https://doi.org/10.1016/j.msard.2023.105379

Alvares Pereira, G., Silva Nunes, M. V., Alzola, P., & Contador, I. (2022). Cognitive reserve and brain maintenance in aging and dementia: An integrative review. *Applied Neuropsychology: Adult*, *29*(6), 1615–1625. https://doi.org/10.1080/23279095.2021.1872079

Bhatti, G. K., Reddy, A. P., Reddy, P. H., & Bhatti, J. S. (2020). Lifestyle Modifications and Nutritional Interventions in Aging-Associated Cognitive Decline and Alzheimer's Disease. *Frontiers in Aging Neuroscience*, *11*, 369. https://doi.org/10.3389/fnagi.2019.00369

Contador, J., Pérez-Millán, A., Tort-Merino, A., Balasa, M., Falgàs, N., Olives, J., Castellví, M., Borrego-Écija, S., Bosch, B., Fernández-Villullas, G., Ramos-Campoy, O., Antonell, A., Bargalló, N., Sanchez-Valle, R., Sala-Llonch, R., & Lladó, A. (2021). Longitudinal brain atrophy and CSF biomarkers in early-onset Alzheimer's disease. *NeuroImage: Clinical*, *32*, 102804. https://doi.org/10.1016/j.nicl.2021.102804

Hu, Z., Wang, Z., Jin, Y., & Hou, W. (2023). VGG-TSwinformer: Transformer-based deep learning model for early Alzheimer's disease prediction. *Computer Methods and Programs in Biomedicine*, *229*, 107291. https://doi.org/10.1016/j.cmpb.2022.107291

Mendes, A. J., Ribaldi, F., Lathuiliere, A., Ashton, N. J., Zetterberg, H., Abramowicz, M., Scheffler, M., Assal, F., Garibotto, V., Blennow, K., & Frisoni, G. B. (2024). Comparison of plasma and neuroimaging biomarkers to predict cognitive decline in non-demented memory clinic patients. *Alzheimer's Research & Therapy*, *16*(1), 110. https://doi.org/10.1186/s13195-024-01478-9

Qian, J., Betensky, R. A., Hyman, B. T., & Serrano-Pozo, A. (2021). Association of *APOE* Genotype With Heterogeneity of Cognitive Decline Rate in Alzheimer Disease. *Neurology*, *96*(19). https://doi.org/10.1212/WNL.0000000000011883

Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., Cummings, J., & van der Flier, W. M. (2021). Alzheimer's disease. *The Lancet*, *397*(10284), 1577–1590. https://doi.org/10.1016/S0140-6736(20)32205-4

Uddin, K. M. M., Al Mamun, A., Chakrabarti, A., Mostafiz, R., & Dey, S. K. (2024). An ensemble machine learning-based approach to predict cervical cancer using hybrid feature selection. *Neuroscience Informatics*, *4*(3), 100169. https://doi.org/10.1016/j.neuri.2024.100169

Veitch, D. P., Weiner, M. W., Miller, M., Aisen, P. S., Ashford, M. A., Beckett, L. A., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Landau, S. M., Morris, J. C., Nho, K. T., Nosheny, R., Okonkwo, O., Perrin, R. J., Petersen, R. C., Rivera Mindt, M., Saykin, A., … for the Alzheimer's Disease Neuroimaging Initiative. (2024). The Alzheimer's Disease Neuroimaging Initiative in the era of Alzheimer's disease treatment: A review of ADNI studies from 2021 to 2022. *Alzheimer's & Dementia*, *20*(1), 652–694. https://doi.org/10.1002/alz.13449

Yang, Y.-W., Hsu, K.-C., Wei, C.-Y., Tzeng, R.-C., & Chiu, P.-Y. (2021). Operational Determination of Subjective Cognitive Decline, Mild Cognitive Impairment, and Dementia Using Sum of Boxes of the Clinical Dementia Rating Scale. *Frontiers in Aging Neuroscience*, *13*, 705782. https://doi.org/10.3389/fnagi.2021.705782