

Multimodal deep learning algorithm as a tool for Dementia clinical trial patient disease screening

William Jin (1), Ying Liu (2), Polina Vyniavska (2)

(1) West Windsor - Plainboro High School North, N/A

(2) Princeton Pharmatech LLC, N/A

Keywords: multimodal deep learning algorithm, Dementia clinical trial, disease screening, sensitivity, specificity, PPV and NPV, real-world

Abstract: Dementia is a complex disease due to various etiologies. New multimodal deep learning algorithms were developed to improve the diagnosis of dementia into different categories of normal cognition (NC), mild cognitive impairment (MCI), AD, and non-AD dementias (nADD). One of the core difficulties in implementing Dementia clinical trials, especially the AD trials lie in the diagnostic ambiguity of Alzheimer's, where symptomatic overlap with other cognitive disorders often leads to misdiagnosis. Dementia clinical trials usually have high screen failure rates and burden for the sponsor for the manual inclusion screening verification. In our work, we explore the use of this multimodal deep learning algorithm as a tool for the clinical trial patient disease screening verification to reduce the cost of the clinical study while improving the quality. We will present the accuracy assessment of the deep learning algorithm compared to the neurologist assessment based on the sensitivity, specificity, PPV, and NPV in the real-world clinical trial setting. We will explore the optimal set of input variables used for the algorithm to balance the accuracy, cost, and time of the medical exams.

Table of Contents

1. INTRODUCTION.....	4
1.1 Literature Review.....	4
1.2 Gap.....	6
1.3 Research Question or Project Goal.....	7
2. METHOD.....	8
2.1 Methods.....	8
2.1.1 Data Preprocessing	8
2.1.2 Feature Comparison.....	8
2.1.3 Machine Learning.....	8
2.1.4 Evaluation.....	9
2.2 Justification for chosen method	9
2.3 Data Collection	10
2.4 Prediction Models	11
3. RESULTS	12
3.1 Findings	12
3.2 Quantitive or qualitive	13
3.3 Graph(s), Tables, and/or Figures, etc.	14
4. DISCUSSION	23
4.1 Significance.....	23
4.2 Interpretations	23
4.3 Connections.....	25

5. CONCLUSION	25
5.1 Limitations	25
5.2 Implications	26
5.3 Future Research	26
6. REFERENCES/BIBLIOGRAPHY	27

1. INTRODUCTION

1.1 Literature Review

Dementia, as defined by the International Statistical Classification of Diseases and Related Health Problems (ICD), is a syndrome due to brain disease, usually of a chronic or progressive nature, in which there is a disturbance of multiple higher cortical functions, including memory, thinking, orientation, comprehension, calculation, learning capacity, language, and judgment (World Health Organization, 2019). Representing a clinical syndrome characterized by a constellation of symptoms and functional impairments, dementia is not indicative of a singular etiology but can be caused by more than 55 diseases (Geldmacher & Whitehouse, 1996). Alzheimer's disease (AD), a primary cerebral disorder, stands as the most prevalent cause of dementia (Blennow et al., 2006).

Alzheimer's disease is a progressive neurodegenerative disorder characterized by the accumulation of amyloid- β plaques and aggregation of tau protein tangles within the brain (Scheltens et al., 2021). It primarily affects the elderly population over 65 years of age (Hendrie, 1998). AD is a significant public health concern, with a global prevalence of around 44 million people, a number that is expected to triple by 2050 worldwide (Dumurgier 2020; Scheltens et al., 2021). Furthermore, Alzheimer's disease is a significant economic burden, with direct medical costs, such as nursing home care, and indirect costs, such as lost productivity, contributing to its high societal cost (Meek, 1998). Despite numerous treatment approaches that have been investigated over the years, there is still no curative treatment (Passeri et al., 2022). In addition to that, the escalating prevalence and financial strain underscore the urgent necessity for further research to understand the nature of AD and discover effective treatment.

The availability of effective treatment relies on accurate clinical trials, particularly patient enrollment selection, which is based on medical diagnosis in the

first place. The current diagnosis of dementia due to Alzheimer's disease mainly relies on the National Institute on Aging and Alzheimer's Association NINCDS-ADRDA (National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorders Association) criteria. These criteria are provided for general healthcare providers without access to neuropsychological testing, advanced imaging, and cerebrospinal fluid (CSF) measures. Additionally, NINCDS-ADRDA criteria represent a framework of probable AD dementia (McKhann et al., 2011).

Other diagnostic tools are much more expensive, but also more accurate. These tools include neuropsychological assessments, such as Mini-Mental State Examination (MMSE), the Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog), and the Clinical Dementia Rating scale (CDR) (Balsis et al., 2015); the detection of amyloid-beta and tau proteins in CSF; brain imaging techniques like magnetic resonance imaging (MRI) and positron emission tomography (PET) (Walhovd et al., 2010).

Despite the array of available biomarkers, Alzheimer's disease is often misdiagnosed with other dementias. For example, the differential diagnosis between Alzheimer's disease and vascular dementia remains challenging due to symptom and risk factor overlap (Kalaria, 2002). Moreover, dementia can present as a mixed type, combining Alzheimer's disease, vascular dementia, and other dementia disorders, that requires distinct treatment strategies (Jellinger, 2007).

The diagnostic challenges have unveiled a neuropathological heterogeneity in AD, suggesting that the classification based solely on biomarkers like amyloid-beta, tau protein, and cognitive assessments may lack accuracy. Consequently, a new era of Alzheimer's biomarker identification has arrived which contributes a tremendous step in understanding AD progression and pathophysiology and improving the diagnostic accuracy for the disease (Hurtado et al., 2018; Bai et al., 2021; Chang et

al., 2021; Zetterberg & Blennow, 2021; Budelier & Bateman, 2019; Aerqin et al., 2022; Badhwar et al., 2019; Tan et al., 2021; Xie et al., 2021).

The shift to computational studies in dementia research has been transformative, enabling the application of multimodal deep-learning algorithms to improve patient disease screening accuracy (Wong-Lin et al., 2020). These algorithms can integrate diverse data types, such as neuroimaging, body fluid biomarkers, genetic or proteomic data, and even speech and facial appearance, to detect early-stage dementia (Senanarong, 2021; Karako et al., 2023). Supervised learning methodologies, a type of machine learning, have been particularly effective in biomarker discovery and neuroimaging studies (Skolariki & Exarchos, 2020). These advancements have the potential to revolutionize dementia care, making it more efficient, transparent, and personalized.

1.2 Gap

The current diagnostic landscape for Alzheimer's disease presents significant barriers to the advancement of effective preventative therapies. While early diagnosis is crucial for timely intervention, accurately distinguishing AD from other dementias remains a challenge. The redundancy of biomarkers available for AD diagnosis reflects the disease's multifactorial nature; however, the most accurate biomarkers are often the most expensive. This includes imaging techniques like MRI and PET scans, neuropsychological tests, and cerebrospinal fluid (CSF) analysis. Unfortunately, no single test can capture all aspects of the disease, and the high costs associated with these comprehensive assessments make it impractical to screen all potential participants in clinical trials. This inaccuracy in diagnostic tools leads to misdiagnosis, compromising the enrollment of appropriate participants in clinical trials. Such participants, though presenting similar symptoms, may have distinct underlying pathologies that can distort statistical outcomes and mask the true

effectiveness and safety of investigational treatments. Hence, there is a critical need for improved diagnostic methods that enable the precise identification of AD in its early stages. This will not only facilitate the development of preventative therapies but also ensure the success of future clinical trials by reducing the financial burden of extensive testing and improving the selection process for trial participants.

1.3 Research Question or Project Goal

The central question guiding this research is: How can a multimodal deep learning algorithm, designed for explainable artificial intelligence, enhance the accuracy of dementia screening in clinical trial settings, particularly for Alzheimer's disease?

The goal of this project is to develop and validate a multilayer multimodal detection and prediction model that utilizes explainable artificial intelligence to accurately categorize dementia into normal cognition (NC), mild cognitive impairment (MCI), Alzheimer's disease (AD), and non-AD dementias (nADD). Utilizing data from our clinical study, we aim to demonstrate the model's efficacy in diagnosing enrolled patients, thereby reducing the screen failure rates and associated costs in dementia clinical trials. We will evaluate the model's performance by comparing its accuracy to traditional neurologist assessments, focusing on sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) in a real-world clinical trial environment. Additionally, we will explore the optimal combination of input variables to achieve a balance between diagnostic accuracy, cost-efficiency, and time-effectiveness of medical examinations.

2. METHOD

2.1 Methods

The experiment conducted a comprehensive analysis of Alzheimer's Disease (AD) data using a Random Forest (RF) classifier. The data was sourced from two distinct datasets: the Alzheimer's Disease Neuroimaging Initiative (ADNI) and a second real-world dataset called Lab Data. The methodology of the experiment was meticulously designed and executed in several stages.

2.1.1 Data Preprocessing

The ADNI data was loaded from a CSV file and cleaned. This cleaning process involved handling missing values using a custom function based on data type and removing columns with a high missing value ratio. Labels were assigned based on diagnosis codes. The Lab Data was loaded from an Excel file, with features selected and renamed to match those from the ADNI data. Missing values were handled, and categorical features were converted using LabelEncoder.

2.1.2 Feature Comparison

Kernel Density Estimation (KDE) plots were generated to visually compare the distribution of common features between the ADNI and Lab Data datasets.

2.1.3 Machine Learning

The experiment was conducted in two scenarios. In the first scenario, six common features were used. The ADNI data was split into training and validation sets, and MinMaxScaler was used for normalization. A baseline RF model was trained and evaluated using 10-fold repeated stratified cross-validation with accuracy as the metric. GridSearchCV was performed to find the best

hyperparameters for the RF model, which was then trained with the optimized hyperparameters and evaluated. The second scenario involved repeating these steps using 10 features from both datasets.

2.1.4 Evaluation

The accuracy (mean and standard deviation) of the RF models for both scenarios was reported. Once the model is trained and evaluated, it was applied to the preprocessed Lab Data. The features from the Lab Data were fed into the model, and the model predicted diagnoses for the patients in the Lab Data.

2.2 Justification for chosen method

The methodological backbone of our research is the multilayer multimodal detection and prediction model described by (El-Sappagh et al., 2021). This model is predicated on the integration of explainable artificial intelligence (XAI) principles, ensuring that each diagnostic decision is accompanied by a comprehensible rationale, thereby fostering trust and transparency in clinical settings.

Unlike traditional approaches that rely on a single data source, typically neuroimaging or CSF concentration of biomarkers, this model encompasses a diverse array of biomarkers and cognitive assessments, reflecting the heterogeneous nature of AD. RF can effectively handle both numerical and categorical features without extensive feature engineering.

A core feature of the model is its use of the SHapley Additive exPlanations (SHAP) framework, which provides both global and instance-based explanations for the Random Forest classifier's decisions. While not as interpretable as simpler models like decision trees, RFs offer better interpretability compared to black-box models. Feature importances can be extracted from RFs to understand which features contribute most to the model's predictions. This aligns with the growing

demand for XAI in healthcare, where understanding the ‘why’ behind a diagnosis is as important as the diagnosis itself.

The presence of irrelevant features in the dataset can negatively impact some machine learning algorithms. RF's inherent ability to handle irrelevant features makes it a good choice for this scenario. The model demonstrates high accuracy and F1-scores in cross-validation settings, indicating its robustness and reliability. More importantly, the explanations generated are consistent with medical literature, enhancing its clinical applicability.

In our study, we extend the application of El-Sappagh et al.'s model to the realm of clinical trial screening. RFs efficiently handle datasets with a high number of features, making them suitable for this application. By adapting this model to our dataset, we aim to reduce the high screen failure rates endemic to dementia trials. We hypothesize that the model's multimodal, explainable, and accurate nature will translate into more efficient and cost-effective patient screening processes, ultimately accelerating the development of AD therapeutics.

2.3 Data Collection

The model has been validated using the Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset, which includes a wide spectrum of AD-related cases. We used the same ADNI database for building our model, in which the models are trained and validated with 992 subjects: 183 cognitively normal, 372 stable mild cognitive impairment (MCI), 240 progressive MCI, and 197 AD.

We also test our models with a real-world case study. This data (Lab Data) was collected as a part of the clinical trial of NE3107 in Alzheimer's disease (Reading et al., 2021). Lab data contains 439 subjects who enrolled the study based on AD diagnosis. Inclusion criteria included male and female subjects aged 60 to 85; NINCDS-ADRDA criteria (McKhann et al., 2011); CDR Standard

Global score of 1 to 2; MMSE score ≥ 14 and ≤ 24 ; amyloid-beta (A β) PET scan; evidence of progressive cognitive decline in last year, etc. During study, different modalities were collected such as medical history, FDG-PET, MRI, lab tests, neuropsychological assessments, genetics.

Our clinical study data (Lab Data) for testing was collected as a part of the clinical trial of NE3107 in Alzheimer's disease (Reading et al., 2021).

Among the most discriminant and informative 28 features for the first layer, we selected common features with our study dataset: ADAS-Cog 11, MMSE, CDR sum of boxes (CDRSB), gender, and ethnicity. Because of a lack of other significant features for AD, we selected other available common features which include APOE $\epsilon 4$ genotype, MRI volumes of whole brain, hippocampus, ventricles, and intracranial region.

In result, our model integrates selected 10 features and 992 subjects from ADNI dataset: 183 cognitively normal, 372 stable mild cognitive impairment (MCI), 240 progressive MCI, and 197 AD; and 439 subjects diagnosed as AD from Lab Data.

2.4 Prediction Models

Random forest is used as the diagnostic algorithm in this research, due to its robust and accurate characteristics among machine learning models. Four experiments are conducted to fully utilize the features and maximize the number of participants: i) using 6 features without model optimization, ii) using 6 features with model optimization, iii) using 10 features without model optimization, and iv) using 10 features with model optimization. The 6-feature experiments build the random forest model using APOE $\epsilon 4$ genotype (APOE4), CDR Sum of Boxes (CDRSB), 11-item version of ADAS-Cog (ADAS11), MMSE, gender (PTGENDER), ethnicity (PTETHCAT), while the 10-feature experiments add

four more features, which are Ventricle, Hippocampus, Whole Brain, intracranial volume (ICV).

The models are trained and validated using the ADNI datasets with five-fold cross-validation and then used to diagnose participants in the second dataset. The evaluation metric used in this research is accuracy, which is the percentage of correctly classified instances out of the total instances.

3. RESULTS

3.1 Findings

The distribution of the same features in the 2 datasets is significantly different except for APOE4 (Fig. 1).

Table 1 summarizes the accuracy metrics of the four experiments: i) random forest detection model using 6 features and without model optimization, ii) random forest detection model using 6 features and with grid search model optimization, iii) random forest detection model using 10 features and without model optimization, iv) random forest detection model using 10 features and with grid search model optimization. It is observed that the grid search optimization improves the detection accuracy by 0.1% (Tab. 1). Additionally, using extra MRI-derived features will improve the detection accuracy by 0.9%. The detailed detection results are shown in (Fig. 2 – Fig. 5).

To validate predictions, we used green, red, and black site classifications based on the trustworthiness of collected data during the study. Green site is the most reliable class, red site is less reliable class, and black sites are expected to have completely falsified data. Despite the insignificant improvement of optimization, the prediction of 6-feature model has extremely significant differences in all classes (Tab. 2 – Tab. 3). In 6-feature model without

optimization, $\approx 99\%$ of predicted misdiagnosis (MCI) among green and red sites decreased to $\approx 50\%$. However, in 10-feature model both with optimization and without, results are the same: all subjects are predicted to be misdiagnosed (Tab. 4 – Tab. 5).

In total, 6-feature model without optimization predicts that that 95.5% of patients were misdiagnosed with accuracy 82.5%. However, 6-feature model with optimization misdiagnosis prediction decreases to 53.9% with slightly increased accuracy to 82.6%. Both 10-feature model predict that all 47 patients who got MRI assessment are misdiagnosed with MCI despite optimization of model.

In the second part, we added 2 more features: amyloid- β 42/40 ratio in plasma and CDR Global score. However, due to lack of data of amyloid- β (A β) 42/40 ratio for significant number of subjects for these 2 features, we ran experiments from the first part one more time but with decreased number of subjects so that we can compare importance of added new features (Tab. 6).

We have also compared distributions for those 6 models (Fig. 6 – Fig. 11).

3.2 Quantitive or qualitive

Our study's methodology is rooted in quantitative research, as evidenced by the systematic computational analysis of numerical data from the ADNI dataset and our clinical study. The selection of discriminant features for data collection, such as ADAS-Cog 11, MMSE, and CDRSB scores, alongside genetic and neuroimaging data, exemplifies a quantitative strategy. These features are quantifiable, allowing for statistical analysis.

The construction of the random forest detection model, a machine learning algorithm, is based on the manipulation of numerical variables to predict

outcomes. This process is quantitative, involving the computation of probabilities and the generation of decision trees from the input data.

Evaluating the model's performance through accuracy metrics underscores the quantitative nature of the study. These metrics are numerical representations of the model's diagnostic capabilities.

The use of grid search optimization to enhance model performance is a quantitative method. It systematically searches through a specified parameter space to find the combination that yields the highest accuracy.

Lastly, the comparison of feature distributions between the ADNI dataset and our clinical study data involves statistical tests to determine significance. This hallmark of quantitative research ensures that the results provide objective, measurable insights into the effectiveness of the multimodal deep learning algorithm for dementia screening in a clinical trial context.

3.3 Graph(s), Tables, and/or Figures, etc.

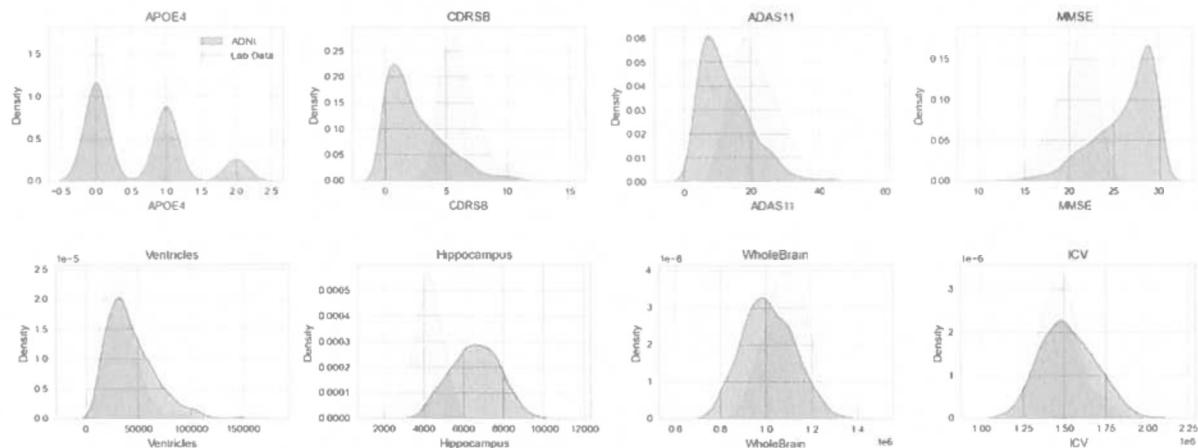


Figure 1. Distributions of the same feature in two datasets.

Number of features	Model optimization	Accuracy
6	No	82.5% (2.5%)
6	Yes	82.6% (2.8%)
10	No	83.3% (2.7%)

10	Yes	83.4% (2.9%)
----	-----	--------------

Table 1. Detection accuracy of different experiments.

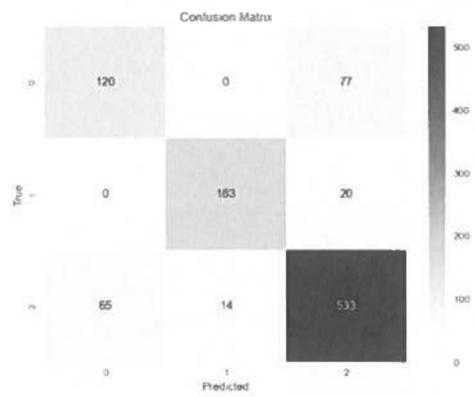


Figure 2 .Confusion matrix of the detection results by random forests with 6 features and no optimization.

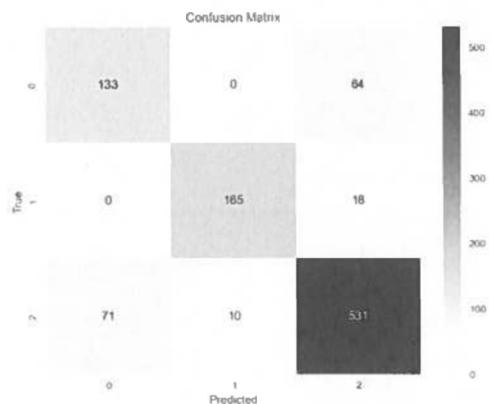


Figure 4. Confusion matrix of the detection results by random forests with 10 features and no optimization.

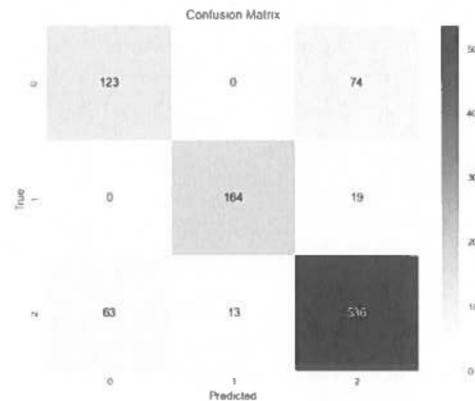


Figure 3. Confusion matrix of the detection results by random forests with 6 features and optimization.

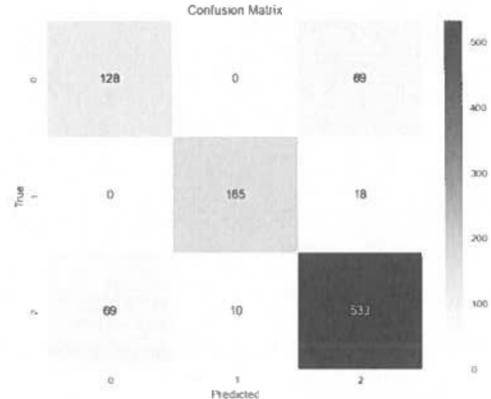


Figure 5. Confusion matrix of the detection results by random forests with 10 features and optimization.

Site group	AD (n predicted, % within a site group)	MCI (n predicted, % within a site group)
Green site	1 (0.8%)	126 (99.2%)
Red site	15 (16.7%)	75 (83.3%)
Black site	3 (1.44%)	205 (98.6%)
In total	19 (4.5%)	406 (95.5%)

Table 2. Prediction results by random forests with 6 features and no optimization.

Site group	AD (n predicted, % within a site group)	MCI (n predicted, % within a site group)
Green site	62 (48.8%)	65 (51.2%)
Red site	65 (31.1%)	62 (68.9%)
Black site	106 (51%)	102 (49%)
In total	196 (46.1%)	229 (53.9%)

Table 3. Prediction results by random forests with 6 features and optimization.

Site group	AD (n predicted, % within a site group)	MCI (n predicted, % within a site group)
Green site	0	25 (100%)
Red site	0	4 (100%)
Black site	0	18 (100%)
In total	0	47 (100%)

Table 4. Prediction results by random forests with 10 features and no optimization.

Site group	AD (n predicted, % within a site group)	MCI (n predicted, % within a site group)
Green site	0	25 (100%)
Red site	0	4 (100%)
Black site	0	18 (100%)
In total	0	47 (100%)

Table 5. Prediction results by random forests with 10 features and optimization.

n features	n participants in training dataset	n participants in testing dataset	CN	MCI	AD	Optimization	Accuracy	n predicted AD	n predicted MCI
6	1640	446	462	894	284	-	0.79	309	137
6	1640	446	462	894	284	+	0.8	0	446
10	1640	50	462	894	284	-	0.8	0	50
10	1640	50	462	894	284	+	0.8	0	50
7	1640	446	462	894	284	-	0.83	421	25
7	1640	446	462	894	284	+	0.83	44	402
8	283	356	88	187	8	-	0.88	0	356
8	283	356	88	187	8	+	0.89	0	356

11	1640	50	462	894	284	-	0.82	49	1
11	1640	50	462	894	284	+	0.82	0	50
12	283	26	88	187	8	-	0.89	0	26
12	283	26	88	187	8	+	0.89	0	26

Table 6. Summary of 12 experiments due to adding A β 42/40 ratio and CDR Global. 6 features: APOE ϵ 4 genotype (APOE4), CDR Sum of Boxes (CDRSB), 11-item version of ADAS-Cog (ADAS11), MMSE, gender (PTGENDER), ethnicity (PTETHCAT). 7 features: 6 features + CDR Global. 8 features: 7 features + A β 42/40 ratio. 10 features: 6 features + 4 MRI features. 11 features: 7 features + 4 MRI features. 12 features: 8 features + 4 MRI features.

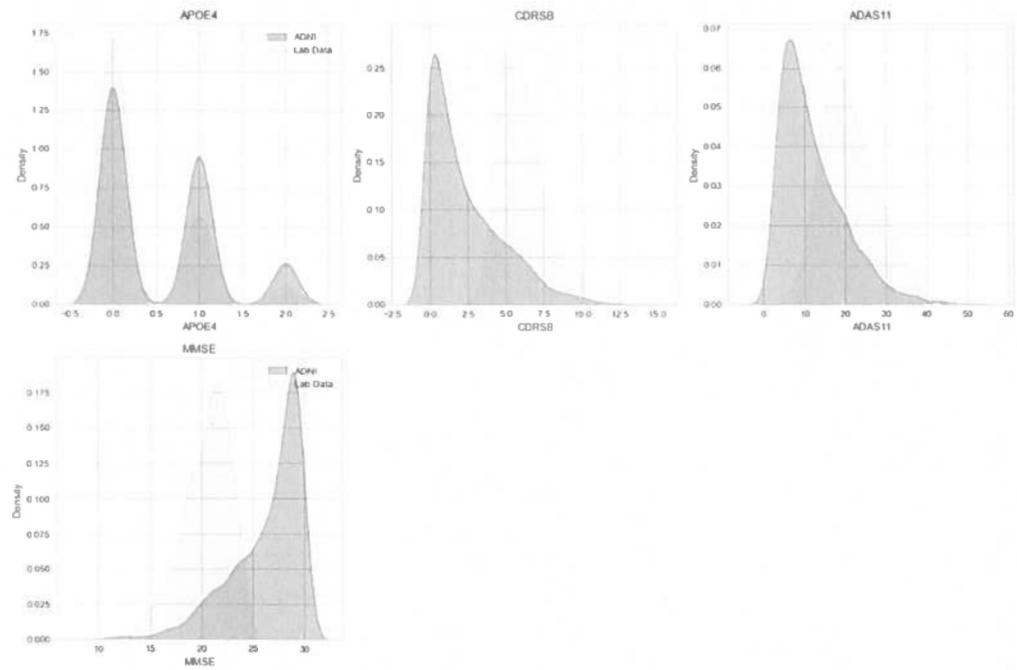


Fig. 6. Distributions of the same features in training (ADNI) and testing (Lab Data) for 6-feature model.

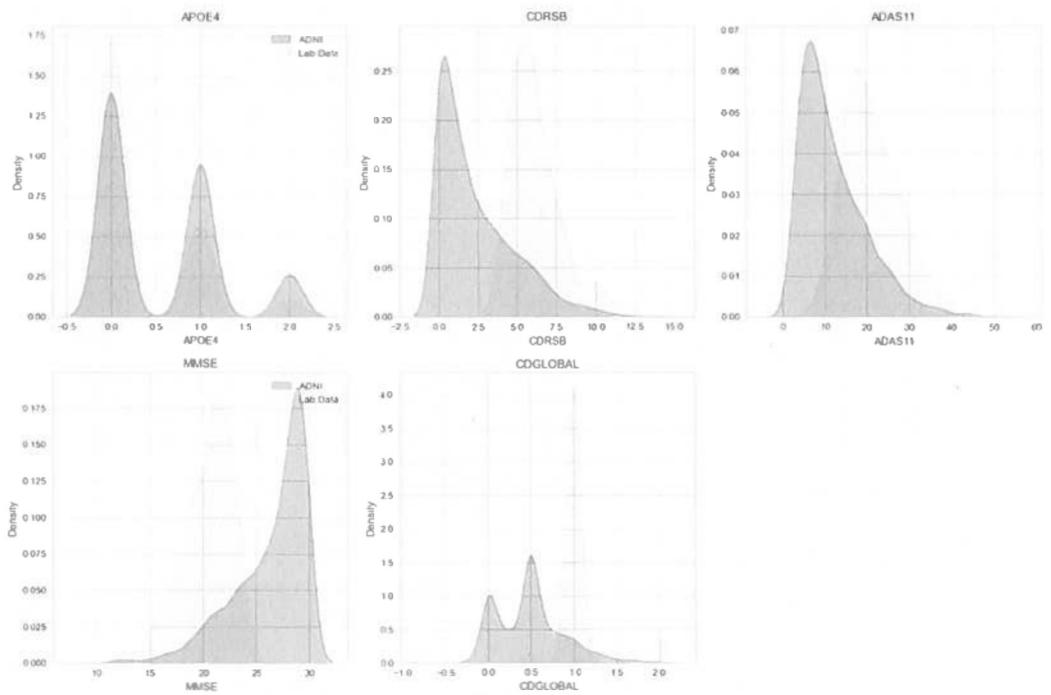


Fig. 7. Distributions of the same features in training (ADNI) and testing (Lab Data) for 7-feature model.

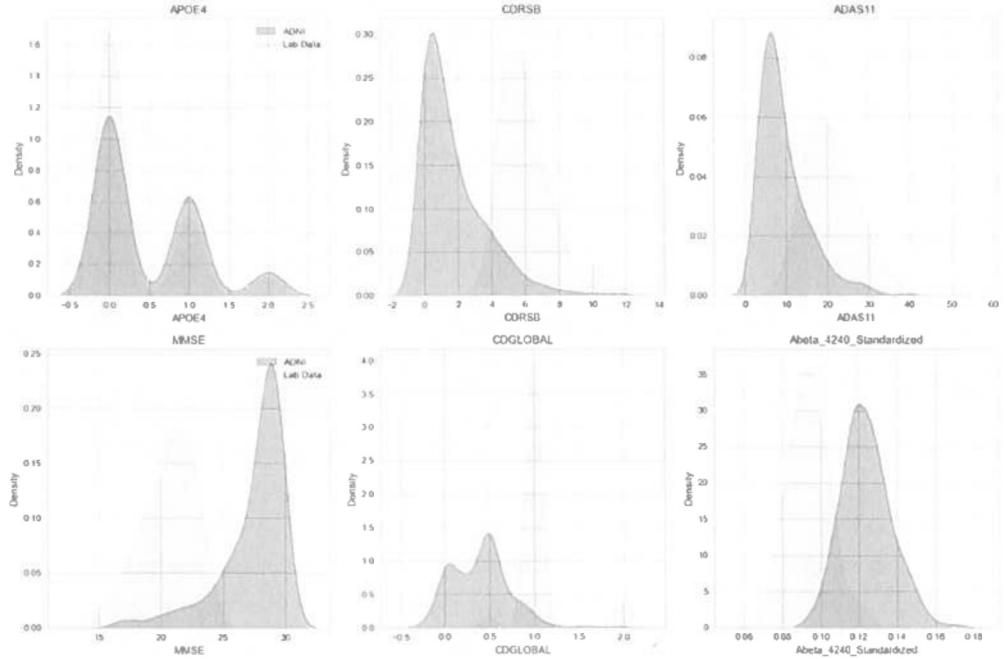


Fig. 8. Distributions of the same features in training (ADNI) and testing (Lab Data) for 8-feature model.

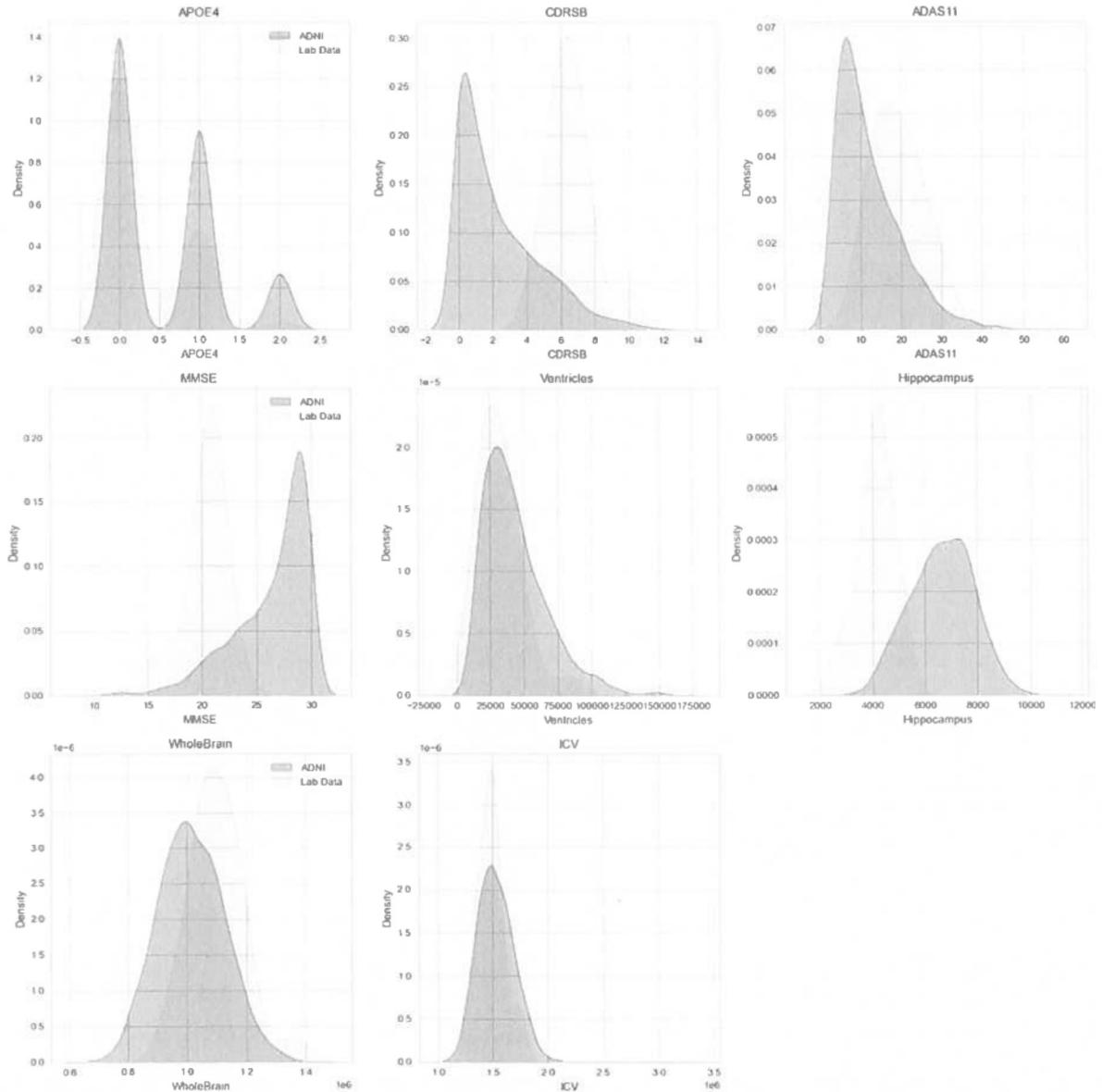


Fig. 9. Distributions of the same features in training (ADNI) and testing (Lab Data) for 10-feature model.

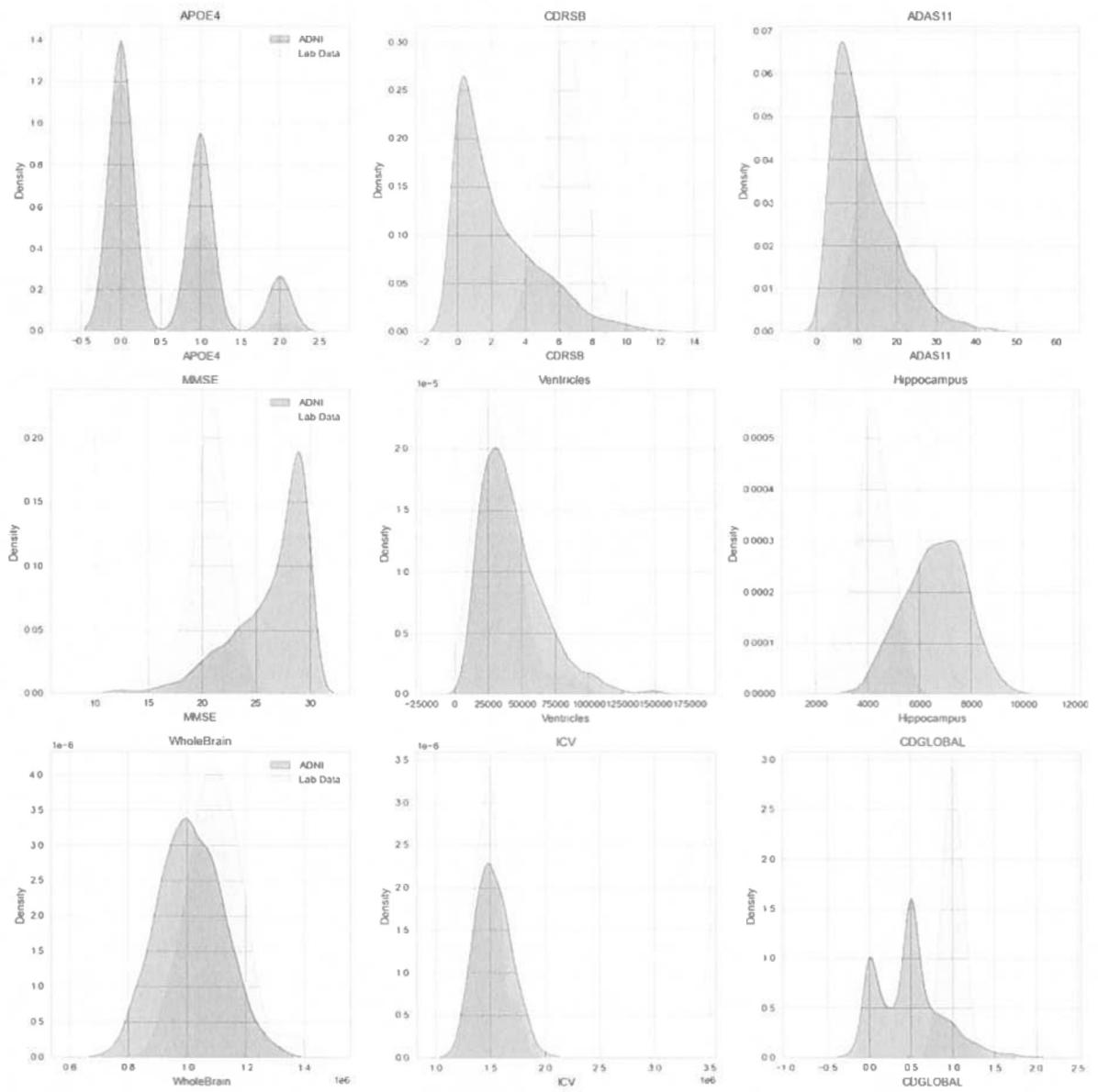


Fig. 10. Distributions of the same features in training (ADNI) and testing (Lab Data) for 11-feature model.

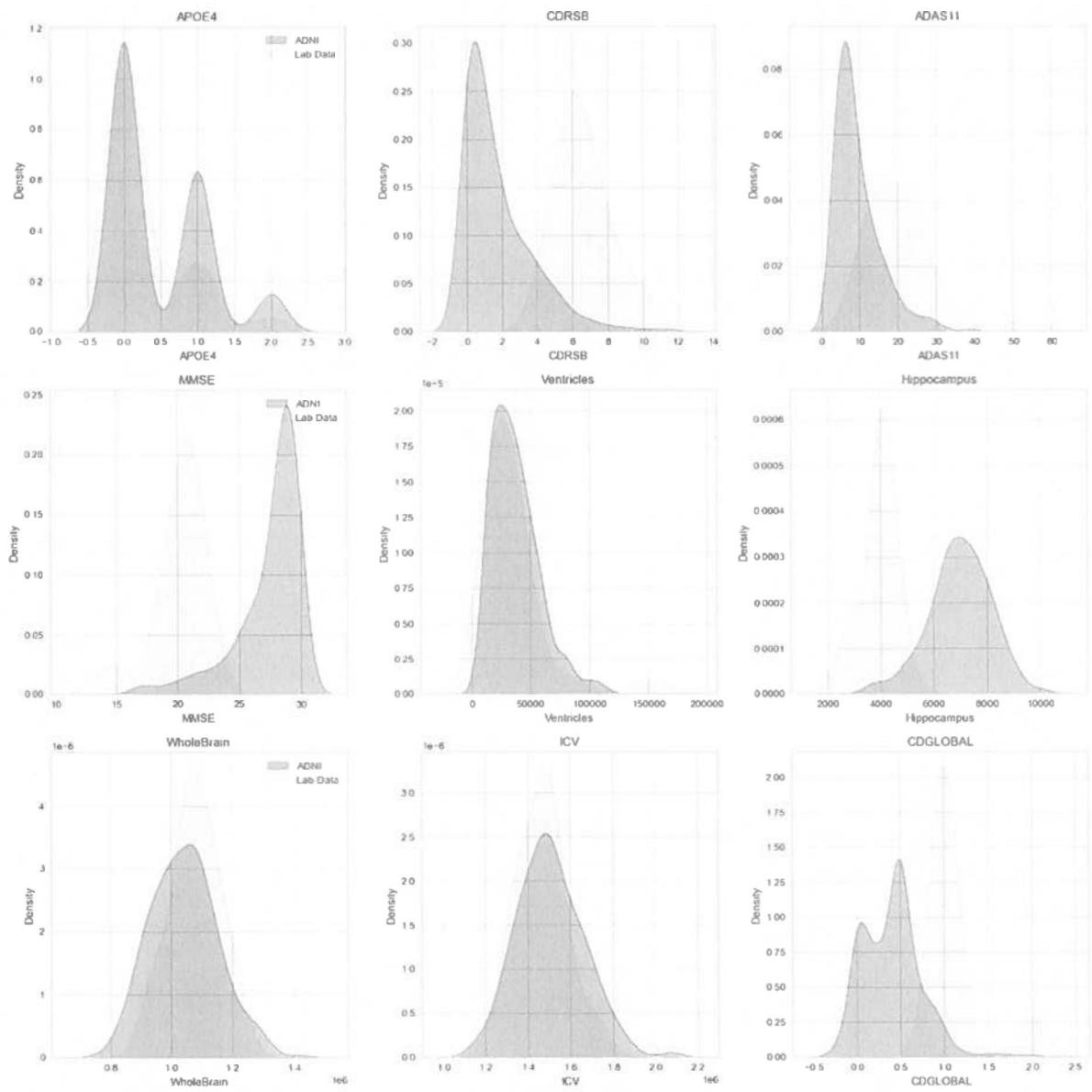


Fig. 11. Distributions of the same features in training (ADNI) and testing (Lab Data) for 12-feature model.

4. DISCUSSION

4.1 Significance

The replication and application of the multimodal deep learning model for Alzheimer's disease prediction in a real-world clinical study setting marks a significant advancement in the field of neurodegenerative disease diagnostics. Our study successfully identified common significant features for the AD class, which are pivotal in the accurate classification and progression tracking of the disease. The model's ability to achieve relatively high accuracy of prediction with a limited set of biomarkers underscores its efficiency and potential for clinical utility. The implications of these findings are manifold.

Firstly, the model's predictive power, harnessed from a small yet robust feature set, demonstrates that a streamlined approach to AD diagnosis is feasible without compromising diagnostic integrity. This is particularly relevant in clinical settings where extensive testing is often resource-intensive and may not be readily available.

Secondly, the promising results obtained from the model's performance offer a foundation for further refinement and development. The integration of additional biomarkers and clinical data could enhance the model's sensitivity and specificity, thereby improving the precision of AD diagnoses. Moreover, the model's cost-effectiveness and time efficiency present a compelling case for its adoption in routine clinical practice, especially in the preliminary screening of AD in diverse patient populations.

4.2 Interpretations

The difference in distribution (Fig.1) is explained by the fact that ADNI data contains 3 groups of subjects: cognitively normal, mild cognitive impairment, and Alzheimer's disease, however, Lab Data contains only, based

on physician's diagnosis, AD subjects. A shift in distribution makes sense. For example, the higher CDRSB and ADAS11 scores, the worse cognitive impairment is. In contrast, the higher MMSE scores, the better cognitive functions are.

The predictive outcomes of our model, which predominantly identified patients as having mild cognitive impairment rather than Alzheimer's disease, warrant a nuanced interpretation. Two primary factors may explain this result.

Firstly, the model's reliance on cognitive scores and genetics as the main predictive features, while MRI modalities did not emerge as significant for the AD class, could be instrumental in this pattern. Cognitive assessments alone, without the support of more definitive biomarkers, may not possess the specificity required to distinguish AD from other forms of dementia, such as vascular dementia. This is due to the symptomatic commonalities that AD shares with other cognitive disorders, which neuropsychological tests alone may not adequately differentiate.

Secondly, the possibility of diagnostic inaccuracies during the study cannot be entirely dismissed. This conjecture is substantiated by the discovery of significant deviations from the established protocol and violations of Good Clinical Practice, which led to the exclusion of data from 15 sites. Such procedural lapses could have inadvertently skewed the diagnosis towards MCI, reflecting a systemic issue rather than a limitation of the model itself.

Thirdly, experiments from the second part underline the importance of usage of concentrations of A β and CDR global scores because accuracy is increased when utilizing both of features. However, there is a big limitation of a model that uses 8 features due to low number of subjects. We need to find data for higher number of subjects to build better predictions. Moreover, adding grid search optimization to 6-, 7-, and 11-feature models gives different prediction results compared with non-

optimized same models although accuracy haven't changed significantly. This finding leads to conclusion that our diagnostic AI model doesn't work correctly, and we have to figure out what causes such significant shift in predictions when using optimization.

These interpretations underscore the complexity of diagnosing AD and the imperative for multi-faceted diagnostic approaches. They also highlight the critical need for stringent adherence to protocols in clinical research to ensure the integrity and accuracy of the findings.

4.3 Connections

Our investigation has reinforced the significance of cognitive scores as reliable biomarkers for Alzheimer's disease prediction. The Mini-Mental State Examination (MMSE), Clinical Dementia Rating Sum of Boxes (CDRSB), and Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog 11) have emerged as pivotal in diagnosing AD. Conversely, magnetic resonance imaging (MRI) results did not exhibit substantial importance in predicting AD within our experiments. Additionally, the role of the APOE ε4 genotype remains inconclusive, warranting further investigation to elucidate its diagnostic value.

5. CONCLUSION

5.1 Limitations

Our study acknowledges several limitations. The reliance solely on cognitive tests and the APOE ε4 gene marker does not provide a comprehensive means to differentiate AD from other dementia forms. The absence of cerebrospinal fluid (CSF) biomarkers, specifically amyloid-beta and tau protein concentrations, is a notable gap, as their inclusion could significantly enhance diagnostic accuracy.

5.2 Implications

The implications of our findings suggest that a broader spectrum of biomarkers is necessary to improve diagnostic precision. Incorporating additional genetic, biochemical, medical history, and neuroimaging markers could provide a more holistic view of the disease, leading to improved patient outcomes.

5.3 Future Research

Future research will focus on elucidating the main pathophysiological mechanisms of AD, particularly the accumulation of A β plaques and tau protein aggregation. We aim to establish a conversion factor for these proteins from plasma biospecimens to CSF, enhancing the model's predictive capability. Furthermore, we plan to integrate medical history data to distinguish AD from vascular dementia effectively. Lastly, an expert in AD diagnosis will be sought to validate the model's predictions, ensuring the reliability of our approach in diagnosing patients from study data.

6. REFERENCES/BIBLIOGRAPHY

1. World Health Organization. (2019). Dementia. In *International statistical classification of diseases and related health problems* (11th ed.). <https://icd.who.int/browse10/2019/en#/F00-F09>
2. Geldmacher, D. S., & Whitehouse, P. J. (1996). *Evaluation of Dementia*. In New England Journal of Medicine (Vol. 335, Issue 5, pp. 330–336). Massachusetts Medical Society. <https://doi.org/10.1056/nejm199608013350507>
3. Blennow, K., De Leon, M. J., & Zetterberg, H. (2006, July 1). *Alzheimer's disease*. The Lancet. [https://doi.org/10.1016/s0140-6736\(06\)69113-7](https://doi.org/10.1016/s0140-6736(06)69113-7)
4. Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., Cummings, J. L., & Van Der Flier, W. M. (2021, April 1). *Alzheimer's disease*. The Lancet. [https://doi.org/10.1016/s0140-6736\(20\)32205-4](https://doi.org/10.1016/s0140-6736(20)32205-4)
5. Hendrie, H. C. (1998, March 1). *Epidemiology of Dementia and Alzheimer's Disease*. The American Journal of Geriatric Psychiatry. <https://doi.org/10.1097/00019442-199821001-00002>
6. [Epidemiology of Alzheimer's disease: latest trends]. (2020, February 1). PubMed.
7. Meek, P. D., McKeithan, E. K., & Schumock, G. T. (1998, March 4). *Economic Considerations in Alzheimer's Disease*. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. <https://doi.org/10.1002/j.1875-9114.1998.tb03880.x>
8. Passeri, E., Elkhoury, K., Morsink, M., Broersen, K., Linder, M., Tamayol, A., Malaplate, C., Yen, F. T., & Arab-Tehrany, E. (2022, November 12). *Alzheimer's Disease: Treatment Strategies and Their Limitations*.

International Journal of Molecular Sciences.

<https://doi.org/10.3390/ijms232213954>

9. McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011, April 22). *The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease*. Alzheimer's & Dementia.
<https://doi.org/10.1016/j.jalz.2011.03.005>
10. Balsis, S., Benge, J. F., Lowe, D. A., Geraci, L., & Doody, R. S. (2015, October 3). *How Do Scores on the ADAS-Cog, MMSE, and CDR-SOB Correspond?* The Clinical Neuropsychologist.
<https://doi.org/10.1080/13854046.2015.1119312>
11. Walhovd, K. B., Fjell, A. M., Brewer, J. B., McEvoy, L. K., Fennema-Notestine, C., Hagler, D. J., Jennings, R. G., Karow, D. S., & Dale, A. M. (2010, January 14). *Combining MR Imaging, Positron-Emission Tomography, and CSF Biomarkers in the Diagnosis and Prognosis of Alzheimer Disease*. American Journal of Neuroradiology.
<https://doi.org/10.3174/ajnr.a1809>
12. Kalaria, R. N. (2002, November 1). *Similarities between Alzheimer's disease and vascular dementia*. Journal of the Neurological Sciences.
[https://doi.org/10.1016/s0022-510x\(02\)00256-3](https://doi.org/10.1016/s0022-510x(02)00256-3)
13. Jellinger, K. A. (2007, January 1). *The enigma of mixed dementia*. Alzheimer's & Dementia. <https://doi.org/10.1016/j.jalz.2006.09.002>
14. Hurtado, M. O., Köhler, I., & De Lange, E. C. M. (2018, September 1). *Next-generation biomarker discovery in Alzheimer's disease using metabolomics –*

from animal to human studies. Bioanalysis. <https://doi.org/10.4155/bio-2018-0135>

15. Bai, B., Vanderwall, D., Li, Y., Wang, X., Suresh, P., Wang, H., Dey, K., Chen, P., Yang, K., & Peng, J. (2021, August 12). *Proteomic landscape of Alzheimer's Disease: novel insights into pathogenesis and biomarker discovery.* Molecular Neurodegeneration. <https://doi.org/10.1186/s13024-021-00474-z>
16. Chang, C., Lin, C., & Lane, H. Y. (2021, March 9). *Machine Learning and Novel Biomarkers for the Diagnosis of Alzheimer's Disease.* International Journal of Molecular Sciences. <https://doi.org/10.3390/ijms22052761>
17. Zetterberg, H., & Blennow, K. (2021, February 19). *Moving fluid biomarkers for Alzheimer's disease from research tools to routine clinical diagnostics.* Molecular Neurodegeneration. <https://doi.org/10.1186/s13024-021-00430-x>
18. Budelier, M. M., & Bateman, R. J. (2019, December 30). *Biomarkers of Alzheimer Disease.* The Journal of Applied Laboratory Medicine. <https://doi.org/10.1373/jalm.2019.030080>
19. Aerqin, Q., Wang, Z. T., Wu, K., He, X., Dong, Q., & Yu, J. (2022, November 8). *Omics-based biomarkers discovery for Alzheimer's disease.* Cellular and Molecular Life Sciences. <https://doi.org/10.1007/s00018-022-04614-6>
20. Badhwar, A., McFall, G. P., Sapkota, S., Black, S. E., Chertkow, H., Duchesne, S., Masellis, M., Liang, L., Dixon, R. A., & Bellec, P. (2019, December 13). *A multiomics approach to heterogeneity in Alzheimer's disease: focused review and roadmap.* Brain. <https://doi.org/10.1093/brain/awz384>
21. Tan, M. S., Cheah, P., Chin, A., Looi, L., & Chang, S. W. (2021, December 1). *A review on omics-based biomarkers discovery for Alzheimer's disease from the bioinformatics perspectives: Statistical approach vs machine*

- learning approach.* Computers in Biology and Medicine.
<https://doi.org/10.1016/j.combiomed.2021.104947>
- 22.Xie, L., He, B., Varathan, P., Nho, K., Risacher, S. L., Saykin, A. J., Salama, P., & Yan, J. (2021, May 10). *Integrative-omics for discovery of network-level disease biomarkers: a case study in Alzheimer's disease.* Briefings in Bioinformatics. <https://doi.org/10.1093/bib/bbab121>
- 23.Wong-Lin, K., McClean, P., McCombe, N., Kaur, D., Sánchez-Bornot, J. M., Gillespie, P., Todd, S., Finn, D. P., Joshi, A., Kane, J., & McGuinness, B. (2020, December 1). *Shaping a data-driven era in dementia care pathway through computational neurology approaches.* BMC Medicine. <https://doi.org/10.1186/s12916-020-01841-1>
- 24.Senanarong, V. (2021). *Application of Artificial Intelligence-Machine Learning in Dementia.* <http://dx.doi.org/10.31031/sbb.2021.05.000608>
- 25.Karako, K., Song, P., & Yu, C. (2023, February 28). *Recent deep learning models for dementia as point-of-care testing: Potential for early detection.* Intractable & Rare Diseases Research. <https://doi.org/10.5582/irdr.2023.01015>
- 26.Skolariki, K., & Exarchos, T. P. (2020, January 1). *Computational Approaches Applied in the Field of Neuroscience.* Advances in Experimental Medicine and Biology. https://doi.org/10.1007/978-3-030-32622-7_17
- 27.Wong-Lin, K., McClean, P., McCombe, N., Kaur, D., Sánchez-Bornot, J. M., Gillespie, P., Todd, S., Finn, D. P., Joshi, A., Kane, J., & McGuinness, B. (2020, December 1). *Shaping a data-driven era in dementia care pathway through computational neurology approaches.* BMC Medicine. <https://doi.org/10.1186/s12916-020-01841-1>
- 28.El-Sappagh, S., Alonso, J. M., Islam, S. M. R., Sultan, A., & Kwak, K. S. (2021, January 29). *A multilayer multimodal detection and prediction model*

- based on explainable artificial intelligence for Alzheimer's disease.* Scientific Reports. <https://doi.org/10.1038/s41598-021-82098-3>
29. Reading, C. L., Ahlem, C. N., & Murphy, M. (2021, August 1). *NM101 Phase III study of NE3107 in Alzheimer's disease: rationale, design and therapeutic modulation of neuroinflammation and insulin resistance.* Neurodegenerative Disease Management. <https://doi.org/10.2217/nmt-2021-0022>