Can machine learning models accurately detect early signs of neurological disorders like Alzheimer's disease using patient data?

By Chenwei Pan, Swati Dokania

Westmount Charter Mid-High School

Abstract

Keywords: Alzheimer's disease, machine learning, early detection, Random Forest, neurological disorders, predictive modeling, healthcare analytics, APOE-ε4, cognitive assessment

Alzheimer's disease (AD), a progressive neurological disorder causing memory loss and cognitive decline, poses a significant public health challenge, with millions affected globally and limited tools available for early diagnosis. Traditional methods often rely on symptom onset, which occurs late in the disease process. This project investigates whether machine learning (ML) models can detect early signs of Alzheimer's using patient data. Supervised learning algorithms, including Random Forest and Logistic Regression, were trained on publicly available datasets containing variables such as age, APOE-ε4 (Apolipoprotein E epsilon 4 allele), cognitive scores, and body mass index (BMI). Performance was assessed using accuracy, precision, recall, and F1-score. A Streamlit app was developed for real-time predictions. The Random Forest model achieved 92.1% accuracy, significantly outperforming Logistic Regression and proving more reliable than XGBoost, which suffered from severe overfitting issues. Feature importance analysis highlighted cognitive decline, depression, and genetic factors as key predictors. These results suggest that ML can play a valuable role in early intervention by identifying high-risk individuals sooner and supporting clinical decision-making.

1.0 Introduction

Alzheimer’s disease (AD) is a devastating neurological disorder that gradually erodes memory, thinking skills, and the ability to carry out daily tasks. It is the most common cause of dementia worldwide, affecting over 55 million people today, with this number expected to rise significantly as global populations age (World Health Organization, 2023).

1.1 Health Conditions and Risk Factors

Alzheimer’s disease is primarily characterized by the accumulation of abnormal proteins in the brain - beta-amyloid plaques and tau tangles - which disrupt normal communication and function of brain cells (Jack et al., 2018). While advancing age remains the strongest risk factor, genetic factors, such as the presence of the APOE-ε4 gene variant, significantly increase an individual’s risk of developing Alzheimer’s (Liu et al., 2013). Additionally, lifestyle factors including smoking, poor diet, lack of physical and mental activity, as well as medical conditions like hypertension, diabetes, and depression, have been linked to elevated disease risk (Livingston et al., 2020).

1.2 Traditional Detection Limitations

Conventional methods for Alzheimer’s detection often rely on neuroimaging techniques such as magnetic resonance imaging (MRI) and positron emission tomography (PET) scans. While these tools can identify structural and functional brain changes, they are costly, time-intensive, and typically performed only after noticeable symptoms arise, by which point significant, irreversible neurological damage has already occurred. This delay in detection greatly limits opportunities for early intervention and preventive strategies.

1.2 Machine Learning In Healthcare

Machine learning (ML) offers a promising path toward earlier, more accessible Alzheimer's detection by analyzing patterns in structured health data before severe symptoms develop. In healthcare, supervised learning algorithms are trained on labelled datasets to predict outcomes by identifying subtle, complex patterns that may be invisible to human observation.

The machine learning models explored in this project are Logistic Regression, Random Forest, and XGBoost (Extreme Gradient Boosting). Logistic Regression is a well-established classification method that estimates the probability of a binary outcome based on input features (Dawson, 2021). Random Forest is an ensemble learning method that constructs multiple decision trees, branch-like models that split data into smaller subsets based on feature values, and aggregates their predictions. This approach reduces the likelihood of overfitting, which occurs when a model performs exceptionally well on training data but fails to generalize to new, unseen data because it has "memorized" noise instead of learning underlying patterns (*What Is Random Forest? [Beginner's Guide + Examples]*, 2020). XGBoost, a more advanced gradient boosting framework, builds trees sequentially, with each tree correcting errors from the previous one, and applies regularization techniques to control further overfitting (*What Is XGBoost?*, 2019).

By training and testing these models on structured patient health data, including demographics, genetics, and cognitive scores, this study evaluates whether machine learning can match or surpass the predictive accuracy of traditional approaches while eliminating reliance on costly neuroimaging. In doing so, it aims to demonstrate the potential for low-cost, scalable screening tools that could enable earlier diagnosis and intervention for at-risk individuals.

2.0 Rationale

The progressive nature of Alzheimer's means that brain changes begin long before symptoms become noticeable, making early detection a critical yet challenging goal. The primary objective of this project is to determine whether machine learning can accurately detect early signs of Alzheimer's disease using patient health data. By focusing on structured data rather than medical imaging, the project aims to assess the feasibility of developing accessible, data-driven tools for earlier diagnosis. The study compares the predictive performance of Logistic Regression and Random Forest models on a labelled dataset that includes relevant patient features.

Beyond model evaluation, this project investigates which input features most significantly influence predictions. Understanding these features may improve the interpretability of model outputs and shed light on which health factors are most strongly associated with early signs of Alzheimer's, helping support clinicians in making more informed decisions during early diagnostic stages.

Hypothesis: Machine learning models, particularly ensemble methods like Random Forest, can accurately identify early signs of Alzheimer's disease using patient demographic, genetic, and health data, achieving clinically relevant accuracy rates above 90% while maintaining appropriate sensitivity for early intervention purposes.

3.0 Literature Review

This section examines current research on predictive models for Alzheimer's disease, feature selection approaches, and real-world clinical performance, helping identify gaps that this project addresses.

3.1 Predictive Models for Alzheimer's Disease

Recent research has focused on developing predictive models to assess Alzheimer's risk, particularly in individuals with mild cognitive impairment (MCI), who are at higher risk of progressing to dementia. A 2024 systematic review by Wang et al. examined 16 cohort studies involving 9,290 participants and found that machine learning models such as Random Forest and Support Vector Machines achieved an average area under the curve (AUC) of 0.87, suggesting strong predictive performance.

However, the review identified several limitations. Many models were developed using small sample sizes, lacked external validation, and failed to include essential methodological details. All models were rated as having either high or unclear risk of bias, with only two being externally validated, raising concerns about generalizability in clinical settings.

The most commonly used features included age, Mini-Mental State Examination (MMSE) scores, and the Functional Activities Questionnaire, reflecting a combination of demographic, cognitive, and functional data associated with Alzheimer's progression.

3.2 Model Performance in Clinical Settings

Licher et al. evaluated four existing dementia prediction models (CAIDE, BDSI, ANU-ADRI, and DRS) using real-world clinical data from the Rotterdam Study, following 6,667 individuals aged 55 and older over 75,581 person-years. The models showed varying discrimination ability: CAIDE performed weakest with a C-statistic of 0.55, while DRS performed best with 0.81. However, all models suffered from poor calibration, underestimating risk in low-risk groups and overestimating it in high-risk groups. Notably, using only age as a predictor gave results nearly identical to the full models.

This study highlights important limitations of current dementia prediction models in clinical settings, emphasizing the need for updated prediction tools that are both accurate and practical across diverse patient populations.

3.3 Study Objectives

Given the challenges identified in existing models, poor calibration, heavy reliance on age, and limited external validation, there is a clear need for improved predictive tools. This project aims to develop a machine learning model that addresses these limitations by leveraging modern algorithms and a comprehensive set of relevant features to create a more reliable and practical tool for early Alzheimer's detection.

4.0 Methodology

4.1 Data Collection

Data was sourced from a publicly available Alzheimer's prediction dataset hosted on Kaggle (https://www.kaggle.com/datasets/ankushpanday1/alzheimers-prediction-dataset-global/data). The dataset includes 24 features such as age, gender, BMI, cognitive test scores (e.g., MMSE), and APOE-ε4 allele presence. The target variable represents a binary Alzheimer's diagnosis, making it suitable for supervised classification tasks.

4.2 Dataset Features

The dataset includes comprehensive features representing biological, genetic, environmental, lifestyle, and socioeconomic factors that may influence Alzheimer's risk:

Demographic factors: Country, age, gender, education level, employment status, marital status, income level, urban vs rural living

Health indicators: BMI, diabetes, hypertension, cholesterol level, cognitive test score, depression level, sleep quality

Lifestyle factors: Physical activity level, smoking status, alcohol consumption, dietary habits, social engagement level, stress levels

Environmental factors: Air pollution exposure

Genetic factors: Family history of Alzheimer's, APOE-ε4 allele presence

4.3 Ethical Considerations

All data used in this research were obtained from publicly available datasets with appropriate consent and privacy protections already in place. The Kaggle dataset used contains de-identified patient information, ensuring participant confidentiality and anonymity. No personally identifiable information was accessed or utilized during the research process. The study adhered to ethical guidelines for secondary data analysis, recognizing that all original data collection procedures were conducted with proper informed consent from participants. Additionally, the research aims to benefit public health by advancing early detection capabilities for Alzheimer's disease, aligning with principles of beneficence and non-maleficence in medical research.

4.4 Development Challenges and Model Selection

The project involved extensive iterative development with significant technical challenges that ultimately informed the final model selection.

4.4.1 Model Performance and Overfitting Issues

Development began with Logistic Regression as a baseline due to its interpretability and simplicity. However, this model demonstrated poor recall performance, particularly struggling to identify true positive cases, critical in medical diagnosis, where missing positive cases have severe consequences.

Random Forest was implemented next, with extensive hyperparameter tuning using GridSearchCV optimizing n_estimators, max_depth, and min_samples_split. This transition marked a significant improvement, reaching 92.1% accuracy with better-balanced performance across precision and recall metrics.

XGBoost was also evaluated, initially showing impressive training metrics with over 92% accuracy. However, during deployment testing, XGBoost exhibited severe overfitting issues, predicting unrealistically high probabilities (99.97%) for obviously low-risk individuals, such as healthy 50-year-old women with no risk factors. This overfitting rendered the model clinically unusable despite impressive training metrics, as the predictions lacked the calibration necessary for real-world medical applications.

After extensive debugging and calibration attempts, Random Forest emerged as the optimal choice, demonstrating more realistic and clinically appropriate prediction behaviours while maintaining high accuracy.

4.4.2 Technical Implementation Challenges

File Size Constraints: GitHub's 25 MB file size restriction created deployment challenges. Models were compressed using pickle and gzip, achieving approximately 10:1 compression ratios. Feature selection techniques, including Recursive Feature Elimination (RFE) and SelectKBest, were employed to reduce model complexity while preserving predictive performance.

Categorical Encoding: The dataset's mix of numerical and categorical variables required careful preprocessing to ensure consistent encoding between training and prediction phases, particularly for medical yes/no variables, where incorrect encoding could invert risk factor relationships.

Web Application Development: The Streamlit application required extensive iteration on user interface design, input validation, error handling, and cross-platform compatibility to create an intuitive, professional interface suitable for medical applications.

4.5 Model Development

Data Preprocessing: Missing numerical values were imputed using the median. Categorical variables were encoded using label encoding or one-hot encoding based on cardinality. All features were standardized to ensure consistency across model inputs.

Addressing Class Imbalance: SMOTE (Synthetic Minority Oversampling Technique) was applied to generate synthetic examples for the minority class, significantly improving the model's sensitivity to detecting Alzheimer’s positive cases.

Final Model Selection: Based on a comprehensive evaluation, including real-world testing for clinical appropriateness, Random Forest was selected as the final model. The model was optimized for both performance and deployability, with hyperparameters tuned to balance accuracy with file size constraints.

Model Evaluation: Performance was assessed using accuracy, precision, recall, F1-score, ROC AUC, confusion matrices, and ROC curves. Clinical validation included testing realistic scenarios to ensure appropriate risk calibration.

4.6 Streamlit Application

A web-based application was built using Streamlit to make the predictive model accessible and interactive. The interface includes form fields with placeholder text, dropdown menus, tooltips, input validation, and responsive feedback elements. Despite GitHub's file size restrictions, the compressed Random Forest model was successfully integrated, resulting in a fast, accessible interface capable of delivering accurate Alzheimer's risk assessments in real time.

5.0 Findings and Discussion

5.1 Model Performance

Metric	Logistic Regression	Random Forest	XGBoost
Accuracy	89.0%	92.1%	92.23%*
Precision (Class 0)	0.87	0.92	0.91*
Recall (Class 0)	0.83	0.86	0.88*
F1-score (Class 0)	0.85	0.89	0.90*
Precision (Class 1)	0.90	0.92	0.93*
Recall (Class 1)	0.93	0.96	0.95*
F1-score (Class 1)	0.91	0.94	0.94*

Model Size	0.001 MB	11.27 MB	0.18 MB
Clinical Validation	Poor recall	Excellent	Severe overfitting

Table 1. A comparison of three different models across different factors

*XGBoost metrics reflect training performance; however, the model failed at predicting real-world scenarios due to overfitting.

Based on Table 1, Random Forest demonstrates superior overall performance across nearly all metrics, making it the clear optimal choice for this Alzheimer's prediction task. All three models achieve strong accuracy scores, with Random Forest (92.1%) and XGBoost (92.23%) showing nearly identical performance that significantly outperforms Logistic Regression (89.0%). However, the critical distinction emerges in clinical validation and real-world applicability.

Random Forest excels with the highest recall for Class 1 (0.96), which is crucial in medical diagnosis as it minimizes missed Alzheimer's cases—a factor of paramount importance since failing to identify true positive cases can have severe clinical consequences. The model also demonstrates strong precision across both classes, maintaining balanced performance that avoids excessive false positives while maximizing true positive detection. While XGBoost shows competitive training metrics with slightly better recall for Class 0 (0.88 vs 0.86), these advantages are rendered meaningless by severe overfitting that causes the model to fail in real-world scenarios.

Logistic Regression, despite being extremely lightweight at 0.001 MB, demonstrates poor recall performance that makes it unsuitable for medical screening applications. The model's underperformance across all metrics, particularly its inadequate sensitivity for detecting true Alzheimer's cases, eliminates it from consideration for clinical deployment where reliability is paramount.

The most significant finding is that only Random Forest achieves "excellent" clinical validation, maintaining realistic and reliable predictions when deployed. Although Random Forest requires more computational resources at 11.27 MB compared to XGBoost's 0.18 MB, this trade-off is justified by its proven real-world reliability and superior recall performance for detecting Alzheimer's cases. The combination of high accuracy, excellent sensitivity for positive case detection, and validated clinical performance establishes Random Forest as the only viable model for deployment in preliminary Alzheimer's screening applications.

5.2 Feature Importance Analysis

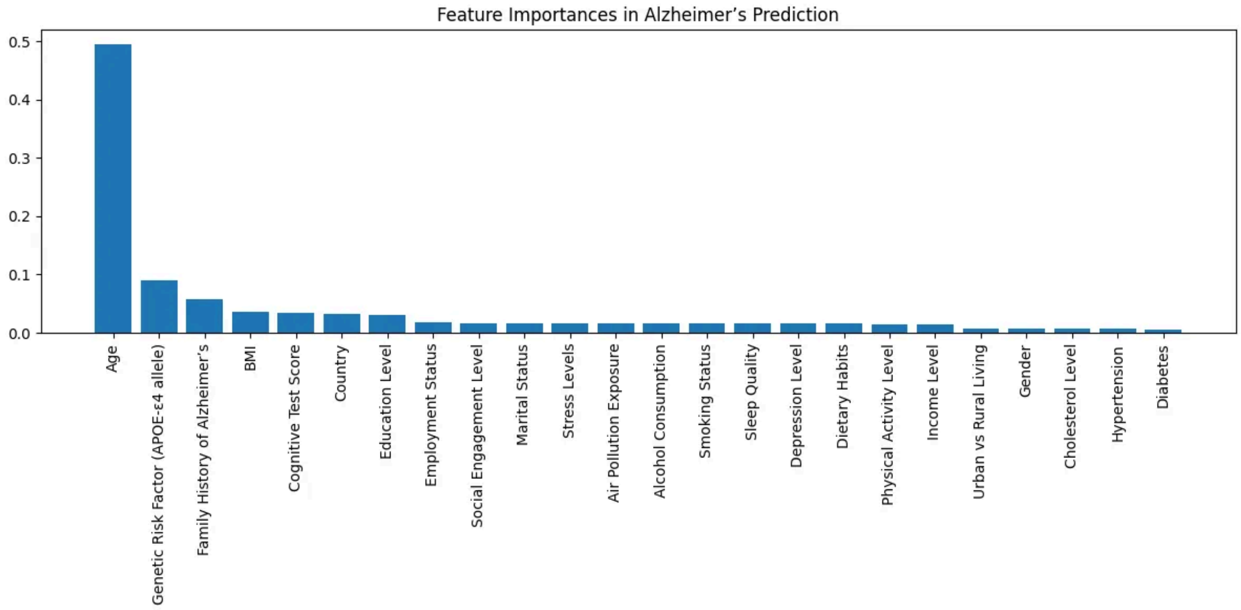

Figure 1. Most important factors for Alzheimer's Prediction

Figure 1 reveals that age is by far the most significant predictor of Alzheimer's disease risk, with an importance score of approximately 0.5. This finding strongly aligns with established medical knowledge, as age is the primary risk factor for Alzheimer's disease, with incidence rates doubling approximately every five years after age 65.

Following age, the most important predictive features were genetic risk factors, particularly the presence of the APOE-ε4 allele (importance ~0.09), and family history of Alzheimer's disease (importance ~0.06). Additional factors showing moderate importance included cognitive assessment scores such as MMSE (Mini-Mental State Examination), BMI measurements, country, and finally, education level.

The feature importance hierarchy demonstrates clinical validity, with the model correctly prioritizing age as the dominant risk factor while appropriately weighting genetic predisposition and family history as secondary but significant predictors. This alignment with established medical knowledge validates the model's clinical relevance and suggests it has successfully learned the underlying biological and clinical patterns associated with Alzheimer's disease risk.

5.3 Streamlit Application Performance

The deployed application successfully provides real-time risk assessments with appropriate validation and user-friendly interfaces. The Random Forest model's reliability and realistic probability outputs make it suitable for preliminary screening applications.

To evaluate the practical usability and effectiveness of the deployed application, a comprehensive user feedback survey was conducted. The survey, distributed through research networks and healthcare communities, gathered responses from users who tested the Alzheimer's risk assessment tool at https://alzheimersprediction.streamlit.app. The 8-section feedback form assessed multiple dimensions of the application, including user experience, assessment quality and accuracy, design and interface, content and features, medical and educational value, and technical performance.

The survey evaluated critical aspects such as navigation ease, completion rates for the 24-factor assessment, time requirements, question clarity, result confidence, visual design appeal, device compatibility, and the helpfulness of prevention recommendations. Participants were asked to rate their experience across various metrics and provide feedback on desired additional features, including more detailed risk factor explanations, result saving capabilities, progress tracking, population comparisons, and healthcare provider integration.

5.4 User Experience Analysis

Evaluation of five user feedback forms revealed an overwhelmingly positive reception. In terms of overall user experience, 80% of participants rated the app at the highest score of 5, while the remaining 20% rated it at 3, indicating a small margin for improvement in accommodating certain user needs. Ease of navigation scores followed a similar trend, with 60% awarding a score of 5, 20% a score of 4, and 20% a score of 3. These results suggest that while the layout and flow are generally effective, further refinements in button placement, progress indicators, or section organization could push satisfaction scores even higher.

5.5 Completion Metrics

The completion rate, with 60% of respondents answering all 24 factors and 40% skipping some, points to a partial engagement gap. Qualitative comments revealed that the primary challenge lay in self-assessing specific medical and psychological factors, particularly cognitive test scores and perceived stress levels. These findings highlight a key usability consideration: users without prior clinical context may find such self-assessment questions ambiguous or intimidating. Integrating tooltips, example scales, or optional built-in mini-assessments could address this gap and raise completion rates.

5.6 Efficiency and Time Performance

From a performance standpoint, the app demonstrated strong efficiency. All respondents completed the assessment in under five minutes, underscoring the success of the design in minimizing cognitive load and unnecessary complexity. The short completion time also suggests that the question flow was logical and the response inputs were optimized for speed, likely aided by the use of dropdowns and radio buttons rather than open-text fields for most items.

5.7 Question Clarity Assessment

The clarity of questions received high praise from most users, with 80% assigning the maximum score of 5. However, the outlier rating of 2, paired with feedback about certain technical terms, indicates that a small subset of users may require additional explanation for medically-oriented factors. Providing plain-language definitions alongside complex terms, as well as embedding "info" icons with short explanations, could enhance accessibility for all users.

5.8 Overall Performance Assessment

Collectively, these metrics indicate that the application is already delivering a smooth, fast, and user-friendly experience, achieving high scores in both usability and satisfaction. The main opportunity for improvement lies not in the technical performance, but in supporting user confidence when self-reporting subjective or clinical health data. Addressing this gap through embedded guidance and contextual help would likely yield higher completion rates, increased clarity scores, and overall stronger engagement.

5.9 Clinical Implications

The Random Forest model's success demonstrates that AI-driven predictive tools could enable earlier identification of Alzheimer's risk, giving healthcare providers more time to intervene before severe cognitive decline occurs. This shift toward early detection may encourage proactive health strategies among at-risk individuals and reduce long-term care costs.

From a system-wide perspective, early diagnosis through ML tools could ease healthcare infrastructure burden by allowing more efficient resource allocation and timely patient monitoring. The Streamlit app offers a potential starting point for such applications by enabling preliminary self-assessment, though it should complement rather than replace professional medical evaluation.

6.0 Limitations and Future Directions

6.1 Current Limitations

This project has several important limitations that must be acknowledged. The dataset was publicly available and limited in size, which may reduce generalizability across diverse populations and clinical settings. While SMOTE addressed class imbalance, synthetic sampling may introduce artificial patterns not observed in real clinical environments, potentially affecting model performance when deployed with actual patient data.

Additionally, inherent dataset bias could significantly impact model performance across diverse populations if certain demographic, ethnic, or socioeconomic groups are underrepresented. The model's heavy reliance on age as the primary predictor, while clinically valid, may limit its ability to identify younger individuals at risk due to genetic or lifestyle factors.

Interpretability remains a substantial challenge. Although Random Forest provides feature importance scores, the model's decision-making process remains difficult to fully explain to non-expert users, especially in clinical settings where transparency and trust are paramount. This black-box nature may limit adoption among healthcare professionals who require clear reasoning behind diagnostic recommendations.

6.2 Future Research Directions

Future work could significantly enhance predictive accuracy by incorporating neuroimaging data, such as MRI or PET scans, which provide direct visualization of brain structure and function. Alternative modelling approaches, including ensemble methods combining multiple algorithms or advanced deep learning architectures, could capture more nuanced patterns in patient data that single models might miss.

Collecting real-world data from local clinics or hospitals, with proper ethical approval and informed consent, would substantially enhance dataset relevance and diversity. Longitudinal studies tracking patient outcomes over extended periods would provide valuable insights into disease progression and model performance over time.

Integration of emerging biomarkers, such as blood-based amyloid and tau proteins, cerebrospinal fluid markers, or digital biomarkers from wearable devices, could provide additional predictive power while maintaining accessibility for routine screening.

6.3 Technical Enhancements

Priority should be given to enhancing the Streamlit application by integrating explainable AI tools like SHAP (SHapley Additive exPlanations) values to offer transparent reasoning behind each prediction. This would improve trust among users and clinicians by clearly showing which factors contributed most to each individual's risk assessment.

Implementation of uncertainty quantification techniques would provide confidence intervals for predictions, helping users understand the reliability of their risk assessment. Additionally, developing adaptive questionnaires that adjust based on user responses could improve completion rates and data quality.

6.4 Clinical Integration Pathways

Collaborating with healthcare professionals is essential to ensure the model aligns with established clinical standards and provides actionable insights that can be effectively integrated into existing diagnostic workflows. This includes developing clear protocols for when and how the tool should be used in clinical practice.

Expanding the dataset by gathering anonymized, diverse user data from multiple regions and healthcare systems would improve generalizability and move the project closer to real-world deployment. Establishing partnerships with medical institutions could facilitate prospective validation studies to assess the model's performance in actual clinical environments.

Future development should also focus on creating decision support tools for clinicians that combine ML predictions with clinical guidelines, rather than standalone diagnostic systems. This approach would leverage the strengths of both artificial intelligence and human expertise while maintaining appropriate oversight and accountability in medical decision-making.

7.0 Conclusion

This project demonstrates that machine learning models, particularly Random Forest, can accurately detect early signs of Alzheimer's disease using comprehensive patient data, including cognitive scores, genetic markers, and demographic information. The Random Forest model achieved 92.1% accuracy while maintaining realistic clinical behaviour and superior recall performance for detecting positive cases, unlike other models that suffered from overfitting issues or inadequate sensitivity.

The developed Streamlit application showcases how such predictive tools could be made accessible for practical use, enabling early intervention and better clinical decision-making through real-time risk predictions. User feedback indicates strong acceptance and usability, with the majority of users rating the application highly across multiple dimensions.

Feature importance analysis validated the clinical relevance of the model by correctly identifying age as the dominant risk factor while appropriately weighting genetic predisposition and family history as significant secondary predictors. This alignment with established medical knowledge provides confidence in the model's ability to capture meaningful biological and clinical patterns associated with Alzheimer's disease risk.

While results are promising, the research identified important limitations, including dataset constraints, potential bias issues, and interpretability challenges that must be addressed for successful clinical deployment. Future work involving larger and more diverse datasets, integration of neuroimaging and biomarker data, implementation of explainable AI techniques, and continued collaboration with healthcare professionals will be crucial for refining these tools to meet real-world diagnostic needs.

Machine learning holds strong potential to transform early Alzheimer's detection and improve patient outcomes through timely, data-driven care and prevention strategies. However, successful implementation will require careful attention to clinical validation, ethical considerations, and integration with existing healthcare workflows to ensure these tools enhance rather than replace human clinical judgment.

8.0 References

1. World. (2025, March 31). *Dementia*. Who.int; World Health Organization: WHO. https://www.who.int/news-room/fact-sheets/detail/dementia
2. Gotzner, N., Solt, S., & Benz, A. (2018). Scalar Diversity, Negative Strengthening, and Adjectival Semantics. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01659
3. Liu, C.-C., Takahisa Kanekiyo, Xu, H., & Bu, G. (2013). Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature Reviews Neurology*, *9*(2), 106–118. https://doi.org/10.1038/nrneurol.2012.263
4. Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Jiska Cohen-Mansfield, Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Adesola Ogunniyi, & Orgeta, V. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*, *396*(10248), 413–446. https://doi.org/10.1016/s0140-6736(20)30367-6
5. Dawson, C. (2021, February 11). *A Guide to Logistic Regression for Beginners - Chris Dawson - Medium*. Medium. https://dawsonc96.medium.com/a-guide-to-logistic-regression-for-beginners-c53632fea4e4
6. *What is Random Forest? [Beginner's Guide + Examples]*. (2020, October 21). CareerFoundry. https://careerfoundry.com/en/blog/data-analytics/what-is-random-forest/
7. *What is XGBoost?* (2019). NVIDIA Data Science Glossary. https://www.nvidia.com/en-us/glossary/xgboost/

8. Jo, T., Nho, K., & Saykin, A. J. (2019). Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Frontiers in Aging Neuroscience*, *11*. https://doi.org/10.3389/fnagi.2019.00220
9. Lu, D., Karteek Popuri, Ding, G. W., Rakesh Balachandar, Beg, M. F., Weiner, M., Aisen, P., Petersen, R., Jack, C., Jagust, W., Trojanowki, J., Toga, A., Beckett, L., Green, R., Saykin, A., Morris, J., Shaw, L., Kaye, J., Quinn, J., & Silbert, L. (2018). Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-22871-z
10. Wang, X., Zhou, S., Ye, N., Li, Y., Zhou, P., Chen, G., & Hu, H. (2024). Predictive models of Alzheimer's disease dementia risk in older adults with mild cognitive impairment: a systematic review and critical appraisal. *BMC Geriatrics*, *24*(1). https://doi.org/10.1186/s12877-024-05044-8
11. Licher, S., Yilmaz, P., Leening, M. J. G., Wolters, F. J., Vernooij, M. W., Stephan, B. C. M., Ikram, M. K., & Ikram, M. A. (2018). External validation of four dementia prediction models for use in the general community-dwelling population: a comparative analysis from the Rotterdam Study. *European Journal of Epidemiology*, *33*(7), 645–655. https://doi.org/10.1007/s10654-018-0403-y

9.0 Glossary

Accuracy: The overall percentage of correct predictions across both classes in a classification model. In medical contexts, this represents how often the model correctly identifies both positive and negative cases.

AD (Alzheimer's Disease): A progressive neurological disorder causing memory loss and cognitive decline.

AI (Artificial Intelligence): The simulation of human intelligence processes by machines, especially computer systems.

ANU-ADRI (Australian National University Alzheimer's Disease Risk Index): A model for predicting Alzheimer's risk.

APOE-ε4 (Apolipoprotein E epsilon 4 allele): A genetic variant associated with increased risk of Alzheimer's disease.

AUC (Area Under the Curve): A performance metric for classification models, showing the ability to distinguish between classes.

BDSI (Brief Dementia Screening Indicator): A clinical tool for screening dementia risk.

Beta-amyloid plaques: Abnormal protein deposits in the brain that disrupt normal brain cell communication and function, characteristic of Alzheimer's disease.

BMI (Body Mass Index): A measure of body fat based on weight and height.

C-statistic: A measure of discrimination performance for prediction models, equivalent to AUC for binary outcomes.

CAIDE (Cardiovascular Risk Factors, Aging and Dementia): A dementia risk prediction model.

Calibration: The degree to which predicted probabilities match actual observed frequencies; well-calibrated models produce realistic risk estimates.

Class 0: In this Alzheimer's prediction context, it represents patients without Alzheimer's disease (negative cases).

Class 1: In this Alzheimer's prediction context, it represents patients with Alzheimer's disease (positive cases).

CNN (Convolutional Neural Network): A deep learning architecture commonly used for image analysis.

Cognitive reserve: The brain's resilience to neuropathological damage through efficient networks and cognitive processes, often built through education and mental stimulation.

DRS (Dementia Risk Score): A composite score estimating dementia risk.

F1-score: A metric that combines precision and recall into a single score for each class, providing a balanced measure that accounts for both false positives and false negatives. This is particularly useful for evaluating model performance on imbalanced datasets.

Feature importance: A measure of how much each input variable contributes to a machine learning model's predictions.

GridSearchCV: A method for systematically testing different combinations of hyperparameters to find the optimal model configuration.

Hyperparameter tuning: The process of optimizing model parameters that are set before training to improve performance.

Logistic Regression: A statistical model used for binary classification tasks that estimates probabilities.

MCI (Mild Cognitive Impairment): A condition involving cognitive decline greater than expected for age but not severe enough to significantly interfere with daily life.

ML (Machine Learning): A branch of AI that uses algorithms to learn patterns from data and make predictions.

MMSE (Mini-Mental State Examination): A widely used test that assesses cognitive function and screens for cognitive impairment.

Overfitting: When a model learns training data too specifically, it performs well on training data but poorly on new, unseen data.

Precision: The proportion of positive predictions that are actually correct (true positives / [true positives + false positives]). For Class 0 (no Alzheimer's), this represents the proportion of negative predictions that were actually correct, while for Class 1 (Alzheimer's positive), this represents the proportion of positive predictions that were actually correct.

Random Forest: An ensemble machine learning method that builds multiple decision trees and combines their results for improved accuracy and reduced overfitting.

Recall (Sensitivity): The proportion of actual positive cases that are correctly identified (true positives / [true positives + false negatives]). Recall for Class 0 measures how well the model identifies true negative cases (correctly identifying people without Alzheimer's), while Recall for Class 1 measures the model's ability to correctly identify true positive cases (people who actually have Alzheimer's) - this is particularly critical in medical diagnosis, where missing positive cases can have severe consequences.

RFE (Recursive Feature Elimination): A feature selection technique that recursively removes features and builds the model on the remaining attributes.

ROC AUC (Receiver Operating Characteristic Area Under the Curve): A measure of a model's ability to distinguish between positive and negative classes across all classification thresholds.

SelectKBest: A feature selection method that selects the k highest-scoring features based on statistical tests.

SHAP (SHapley Additive exPlanations): A method for interpreting machine learning model predictions by showing feature contributions.

SMOTE (Synthetic Minority Oversampling Technique): A method for addressing class imbalance by creating synthetic examples of minority class data.

Streamlit: A Python framework for building and deploying web applications, particularly useful for machine learning demos.

Supervised learning: A machine learning approach where algorithms learn from labelled training data to make predictions on new data.

Tau tangles: Twisted protein fibres inside brain cells that are another hallmark of Alzheimer's disease, alongside beta-amyloid plaques.

XGBoost (Extreme Gradient Boosting): An advanced gradient boosting framework that builds trees sequentially, where each tree learns from previous errors.