DEU FMD 27(81) (2025) 457-465

Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi
Dokuz Eylul University Faculty of Engineering Journal of Science and Engineering
Elektronik/Online ISSN: 2547-958X

RESEARCH ARTICLE / ARAŞTIRMA MAKALESI

Prediction of Alzheimer's Diagnosis with Machine Learning and Innovative Feature Engineering

Makine Öğrenmesi ve Yenilikçi Özellik Mühendisliği Kullanılarak Alzheimer Tanısının Tahmini

Murat Emeç [1*], Gamze Gezgin [2]

[1]Department of Computer Science Application and Research Center, Istanbul University, Istanbul, TÜRKİYE
[2]Department of Software Engineering, Istanbul Aydın University, Istanbul, TÜRKİYE

*Corresponding Author / Sorumlu Yazar *: murat.emec@istanbul.edu.tr*

Abstract

Alzheimer's disease is a leading cause of dementia, presenting significant challenges to healthcare systems globally. Early diagnosis is essential for effective management and intervention, yet traditional diagnostic methods remain invasive, time-consuming, and costly. This study investigates the application of advanced machine learning models, emphasising the role of feature selection techniques, such as RFE (Recursive Feature Elimination) and hyperparameter optimisation, to enhance the early detection of Alzheimer's disease. Among the evaluated models, CatBoost with RFE achieved the highest performance, with an accuracy of 95.81% and an F1-score of 94.00%, demonstrating its robustness and reliability as a diagnostic tool. Random Forest and XGBoost models also showed strong results, particularly when combined with feature importance and RFE. The findings highlight the significant impact of feature engineering and hyperparameter tuning in improving model performance across key metrics, including accuracy, recall, precision, and F1-score. This research underscores the potential of integrating machine learning techniques into medical diagnostics, offering a non-invasive, cost-effective, and efficient approach to Alzheimer's disease prediction. The insights gained from this study lay the groundwork for future advancements in diagnostic models, aiming to improve early detection strategies and patient outcomes, ultimately contributing to the global effort to mitigate the impact of Alzheimer's disease on individuals and society.

Keywords: *Alzheimer's disease, Early Diagnosis, Machine Learning, Feature Engineering*

Öz

Alzheimer hastalığı, dünya çapında sağlık sistemleri için büyük sorunlar yaratır ve demansın önde gelen nedenlerindendir. Erken teşhis edilmesi, etkili yönetim ve müdahaleyi mümkün kılar. Bununla birlikte geleneksel tanı yöntemleri zaman alıcı, invaziv ve maliyetlidir Bu çalışma Alzheimer hastalığının erken tanısı için Recursive Feature Elimination(RFE) gibi özellik seçim teknikleri ve hiperparametre optimizasyonunun rolünü vurgulayarak, gelişmiş makine öğrenmesini modellerinin uygulanmasını araştırmaktadır. Değerlendirmeler sonucunda, RFE ile beraber kullanılan CatBoost modeli, %95,81 doğruluk ve %94,00 F1-skora ulaşarak sağlamlığı ve güvenilirliğiyle öne çıkan model olmuştur. Random Forest ve XGBoost modelleri de, özellik önemi ve RFE ile birleştirildiğinde güçlü sonuçlar elde edebilmiştir. Araştırma, doğruluk, duyarlılık, kesinlik ve F1-skor gibi temel metriklerde model performansını artırmada özellik mühendisliği ve hiperparametre ayarlarının öne çıkan etkisini vurgulamaktadır. Bu araştırma, tıbbi tanılarda makine öğrenimi tekniklerinin entegre edilmesi sonucunda invaziv olmayan, maliyet etkin ve verimli bir yaklaşım sunulabildiğini ortaya koymaktadır. Çalışma sonucunda elde edilen bilgiler, erken tanı yöntemlerini ve hasta sonuçlarını iyileştirmeyi hedefleyen teşhis modellerinde gelecekteki ilerlemelere zemin hazırlamaktadır. Bununla beraber alzheimer hastalığının bireyler ve toplum üzerindeki etkisini düşürmeye yönelik küresel çalışmalara katkıda bulunmaktadır.

Anahtar Kelimeler: *Alzheimer hastalığı, Erken Tanı, Makine Öğrenmesi, Özellik Mühendisliği*

1. Introduction

AD (Alzheimer's Disease) is a progressive neurodegenerative disorder and a significant global public health challenge. Characterised by cognitive decline, memory loss, and behavioural changes, AD affects millions of individuals worldwide, placing an immense burden on caregivers and healthcare systems. With the ageing global population, the prevalence of Alzheimer's disease is expected to rise dramatically, emphasising the urgent need for effective diagnostic and therapeutic strategies. This research directly addresses this critical need by exploring innovative approaches for early diagnosis.

Early detection of Alzheimer's disease is essential for managing symptoms, slowing disease progression, and improving patients' quality of life. However, traditional diagnostic methods, such as clinical assessments and neuroimaging, are often invasive, expensive, and time-consuming, limiting their accessibility and practicality. As a result, there is a growing demand for non-invasive, cost-effective, and efficient diagnostic tools that can facilitate early detection and intervention.

DOI: https://doi.org/10.21205/deufmd.2025278113

Geliş Tarihi / Received: 13.12.2024
Kabul Tarihi / Accepted: 02.01.2025

In recent years, machine learning has emerged as a transformative tool in medical diagnostics, offering the ability to analyse complex datasets and uncover patterns that may not be evident through conventional methods. Machine learning algorithms can process vast amounts of data, including genetic, biochemical, and clinical information, to accurately predict disease outcomes. This capability makes machine learning particularly well-suited for the early diagnosis of Alzheimer's disease, where subtle changes in biomarkers and clinical features can provide critical insights into disease progression [1]. The application of machine learning in Alzheimer's diagnosis represents a promising step forward in addressing this global health challenge. This study leverages advanced machine learning techniques to enhance the early diagnosis of Alzheimer's disease. Specifically, we focus on feature engineering, a process that involves selecting and transforming the most relevant data attributes to improve the predictive power of machine learning models. By refining input features, we aim to increase the accuracy, reliability, and interpretability of diagnostic predictions. Additionally, we employ hyperparameter optimisation, a critical process of fine-tuning model parameters to maximise performance metrics such as accuracy, recall, precision, and F1 score. These techniques ensure that the models can effectively distinguish between Alzheimer's and non-Alzheimer's cases. This research evaluated several machine learning models, including Random Forest, CatBoost, and XGBoost, to determine their effectiveness in predicting Alzheimer's disease. We compare the performance of these models with and without feature selection techniques such as Recursive Feature Elimination (RFE) and Feature Importance (FI), as well as with and without hyperparameter optimisation. This comprehensive analysis provides valuable insights into these models' diagnostic capabilities and highlights the impact of feature engineering and optimisation on their performance.

By integrating machine learning with innovative feature engineering and optimisation techniques, this study aims to inspire the development of more accurate, efficient, and accessible diagnostic tools for Alzheimer's disease. This research's findings can inform future studies and clinical practices, ultimately improving patient outcomes and reducing the societal burden of this debilitating disease. The promising results underscore the importance of continued exploration in this field, offering hope for more effective early detection strategies and better management of Alzheimer's disease.

2. Related Works

In recent years, the application of machine learning (ML) and deep learning (DL) techniques has not only revolutionised but also opened up a new frontier in the diagnosis and prediction of Alzheimer's Disease (AD). This significant leap forward in medical technology, which harnesses various data types, including neuroimaging, genetic, and clinical data, has the potential to enhance diagnostic accuracy and enable earlier disease detection significantly. This offers a promising and hopeful future for Alzheimer's research, inspiring the audience with the potential of machine learning in this field. Deep learning approaches, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have consistently demonstrated and delivered high accuracy in AD classification and in predicting the conversion from mild cognitive impairment (MCI) to AD. Studies have reported up to 96% accuracy for AD classification and 84.2% for MCI conversion prediction. Furthermore, combining traditional machine learning with deep learning, such as using stacked auto-encoders (SAE) for feature selection, can achieve even higher accuracies, reaching up to 98.8% for AD classification. These results underscore the reliability and confidence of these methods in Alzheimer's diagnosis. Integrating multiple data modalities, such as MRI, genetic data, and clinical tests, using deep learning models has outperformed single modality models in accuracy, precision, recall, and F1 scores [4, 7]. Multimodal neuroimaging and fluid biomarkers yield the best classification performance [2, 4]. Additionally, novel feature selection techniques, such as Neighbourhood Component Analysis and Correlation-based Filtration (NCA-F), have been developed to improve the accuracy of machine learning models for early AD detection [9]. Using methods like Random Survey Support Vector Machines (RS-SVM), Voxel-based feature detection frameworks have shown high prediction accuracy, particularly in distinguishing between AD and healthy controls. Various machine learning algorithms, including support vector machines (SVM), Random Forests, and ensemble methods, have been employed for AD prediction, with performance metrics varying widely, achieving AUC values from 0.59 to 0.98. However, the ensemble machine learning models, which combine multiple classifiers, have demonstrated improved robustness and accuracy in predicting AD, highlighting their potential and offering a hopeful future for AD prediction. Long short-term memory (LSTM) networks within RNNs have been used to predict biomarkers over time, achieving superior accuracy compared to other algorithms [5]. Pre-trained CNN models, such as ResNet50, have been effective for automatic feature extraction from MRI images, achieving high accuracy in AD diagnosis [8]. Machine learning and deep learning techniques have significantly advanced the prediction and diagnosis of Alzheimer's Disease. These models have achieved remarkable accuracy and robustness by integrating multimodal data and employing innovative feature selection and engineering methods. The high accuracy of these models instils confidence in the audience, reassuring them of the reliability of the research and the potential for early and precise AD detection. Additionally, it has extensively explored the application of machine learning (ML) and deep learning (DL) techniques for the diagnosis and prediction of Alzheimer's Disease (AD), highlighting their transformative potential in this field. Raza et al. (2019) [14] investigated classical and deep learning techniques, emphasising the importance of early diagnosis and the role of feature engineering in improving model performance. Cochrane et al. (2020) [15] proposed a practical solution for early diagnostics that excludes costly imaging, relying on cognitive tests and demographic data, and broadening access to non-invasive diagnostic methods. Chang et al. (2021) [16] reviewed clinical trials integrating machine learning with novel biomarkers, showcasing the potential of cutting-edge technologies in AD diagnosis. Similarly, Battineni et al. (2021) [17] demonstrated the effectiveness of multimodal machine learning algorithms, combining various data sources to enhance diagnostic accuracy. Li et al. (2021) [18] discussed the challenges and applications of ML in analysing multi-source data, reiterating the importance of feature engineering and hyperparameter optimisation. Aqeel et al. (2022) [19] introduced a Long Short-Term Memory (LSTM) framework, emphasising the refinement of clinical feature selection to improve predictive accuracy. Singh et al. (2022) [20] highlighted the potential of deep learning in automated AD detection using MRI images, reinforcing the importance of advanced algorithms for early diagnosis. Bucholc et al. (2023) [21] explored optimisation methods for dementia research, emphasising the role of ML in enhancing diagnostic accuracy. Kang et al. (2023) [22] presented an interpretable ML model utilising deep learning-based imaging biomarkers, achieving high accuracy and AUC metrics while emphasising explainability for clinical adoption. M. Abdelwahab et al. (2023)

[23] introduced microarray gene expression data for AD diagnostics, advocating for a multimodal approach combining genetic insights with neuroimaging. Finally, Ahmadi et al. (2024) [24] proposed an intensely supervised adaptable neural network for multitask feature extraction, integrating Raman spectroscopic examination of cerebrospinal fluid (CSF) to enable rapid and cost-effective early-stage AD detection. These studies collectively illustrate the evolving landscape of ML and DL applications in Alzheimer's diagnosis, emphasising the critical role of feature engineering, model optimisation, and multimodal data integration. The consistent identification of CatBoost as a high-performing algorithm across multiple studies further highlights its potential as a cornerstone in the future of Alzheimer's diagnostics.

3. Materials and Methods

This section summarises the dataset, methods, and analysis processes used in the prediction model. The primary aim of the study is to predict Alzheimer's disease using various techniques and to evaluate the accuracy of these predictions. Information about the dataset's sources, content, and significance is provided, along with detailed explanations of data processing, feature engineering, and the methods employed.

3.1. Materials

3.1.1. Dataset and Features

The dataset used in this study comprises a wide range of 33 features, including patients' demographic information, lifestyle factors, medical history, clinical measurements, and cognitive/functional assessments [25]. Of these features, 21 are integers, and 12 are floats. Upon verification, it was confirmed that all values are in the dataset. Demographic information includes age, Gender, Ethnicity, and education level. Lifestyle factors are assessed through BMI (Body Mass Index), Smoking, alcohol consumption, physical activity, and diet quality. Additionally, variables related to medical history, such as FamilyHistoryAlzheimers, CardiovascularDisease, and Diabetes, are included. Clinical measurements feature attributes like SystolicBP and CholesterolTotal. Variables such as MMSE (Mini-Mental State Examination), Functional Assessment, and ADL (Activities of Daily Living) are noteworthy in terms of cognitive and functional assessments. Finally, the dataset includes the Diagnosis feature, which indicates whether the patients have been diagnosed with Alzheimer's disease.

3.1.2. Statistical Analysis of the Dataset

This section focuses on the statistical analysis of the dataset used to predict Alzheimer's diagnosis.

At this stage, we will use a correlation matrix to gain a deeper understanding of the dataset's overall structure and examine the relationships between features. As a result, we aim to identify the factors influencing Alzheimer's diagnosis and understand how these factors are interrelated. This approach will assist in determining the variables of interest when developing future prediction models. Table 1 presents the statistical measures of the study participants, revealing that the average age of the 2,149 individuals was 74.91 years (±8.99). The ages ranged from 60 to 90 years, with a median age of 75 in the 50th percentile. In terms of Body Mass Index (BMI), the average BMI was 27.66 (±7.22), showing a wide distribution among individuals. The BMI values ranged from 15.01 to 39.99, with a median BMI of 27.82. These findings indicate that BMI values in the elderly population tend to be in the overweight and obesity range. These findings highlight that the study population, with an average age in the high-risk group for Alzheimer's and BMI values leaning toward overweight and obesity, reflects key risk factors. Higher BMI has been associated with cognitive decline, emphasising the importance of addressing modifiable factors to reduce Alzheimer's risk.

Table 1. Analysis of Age and BMI values.

	Age	BMI
count	2149	2149
mean	74.9087	27.6556
std	8.9902	7.2174
min	60	15.0088
25%	67	21.6114
50%	75	27.8239
75%	83	33.869
max	90	39.992

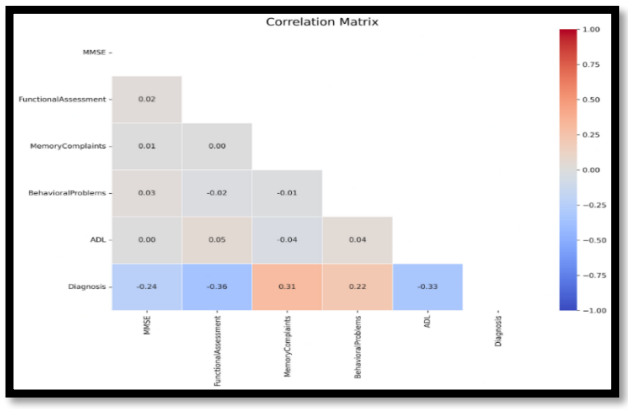

Figure 1. Correlation matrix of a dataset.

Figure 1 illustrates the correlation matrix, a powerful tool for understanding the relationships between various factors and Alzheimer's diagnosis. The matrix identifies and explains the significance of each correlation, focusing on relevant variables such as Memory Complaints, Behavioural Problems, Functional Assessment, ADL, and MMSE, which are crucial indicators of Alzheimer's disease. Positive correlations, like those with Memory Complaints and Behavioural Problems, are linked to an increased likelihood of diagnosis. In contrast, negative correlations with Functional Assessment, ADL, and MMSE highlight the impact of cognitive and functional declines. This analysis provides valuable insights into the progression of Alzheimer's, emphasising the importance of these factors in diagnosis and monitoring. A correlation matrix is scientifically sound, offering a quantitative basis for interpretation. However, the section could be enhanced by acknowledging potential limitations, such as confounding factors, and by including a brief explanation of the matrix construction for clarity. Overall, the section is well-structured and informative, contributing significantly to understanding the factors associated with Alzheimer's diagnosis. These findings underscore the complex interplay between cognitive, behavioural, and functional aspects in Alzheimer's disease. By identifying key correlations, the analysis highlights potential areas for early intervention and targeted management strategies. Understanding these relationships not only aids in diagnosis but also provides a

foundation for developing predictive models and personalised treatment approaches for Alzheimer's progression.

3.1.3. Feature Engineering

Feature engineering, a pivotal step in data analysis and modelling, significantly enhances the model's predictive accuracy. In our dataset, devoid of any categorical variables and comprising only float and int types, the 'Diagnosis' feature, the key variable we aim to predict, is earmarked as the target variable for the modelling process. We also eliminate the 'PatientID' and 'DoctorInCharge' features, as they do not contribute to the prediction and could potentially introduce noise or bias into the model.

Our dataset, a complete set with no missing values, simplifies the preprocessing stage. This absence of missing values and outliers revealed during dataset analysis indicates that the data is clean and ready for modelling without additional preprocessing steps to handle anomalies. To further enhance the model's performance, we consider the following feature engineering techniques:

- Normalisation and Standardisation: Although our dataset does not contain categorical variables, ensuring that all numerical features are on a similar scale is essential. Normalisation or standardisation can achieve this, which helps improve the convergence of gradient-based optimisation algorithms used in many machine learning models.
- Feature Interaction: We explore potential interactions between features that might capture complex relationships not evident from individual features alone. This involves creating new features by combining existing ones, such as calculating ratios or products of related features.
- Dimensionality Reduction: Principal Component Analysis (PCA) or Singular Value Decomposition (SVD) can reduce the dataset's dimensionality while retaining the most informative components. This can help reduce overfitting and improve model generalisation.
- Feature Selection: We apply feature selection methods with utmost care, ensuring that only the most relevant features are retained for the model. Techniques such as recursive feature elimination or feature importance scores from tree-based models are meticulously employed, guaranteeing the model's relevance and accuracy. This meticulous approach instils confidence in our model's robustness.
- Polynomial Features: Generating polynomial features can help capture non-linear relationships between variables. This involves creating new features by scaling up existing features or multiplying them.

By meticulously implementing these feature engineering strategies, we aim to significantly enhance the model's ability to learn from the data and improve its predictive performance, instilling confidence in the robustness of our approach.

3.1.4. Data Preprocessing

Data preprocessing forms the solid foundation of the model we train to accurately predict the diagnosis. In this stage, we meticulously analyse, clean, and prepare our dataset, ensuring it is in the best possible condition for model training. Below are the explanations of these steps:

- Handling Missing Values: The absence of missing data in our dataset means that missing data handling methods are unnecessary. This careful handling of missing values indicates that the dataset is clean and complete, which helps the machine-learning model produce reliable results. The integrity of the dataset ensures that the model can learn effectively without the risk of bias introduced by imputed values.
- Splitting the Dataset: Splitting the dataset into training and testing sets is critical in training and evaluating the model. This process, which involves dividing the data into a training set (80%) of 1719 samples and a test set (20%) of 430 samples, allows for an unbiased evaluation of the model's performance on unseen data.

3.1.5. Feature Engineering Processing

Feature engineering is a crucial step in building predictive models, as it involves selecting and transforming variables to enhance the model's performance.

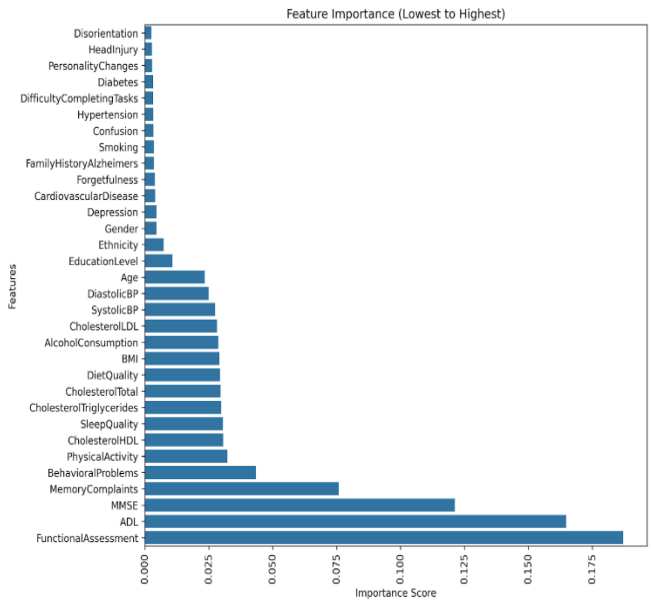

Figure 2. Feature importance values using Random Forest.

Figure 2 includes a Feature Importance graph based on the Random Forest model, which ranks the features according to their importance. In this graph, features are ordered from top to bottom in increasing order of importance, and the contribution of each feature to the model's performance is determined. In the feature importance ranking based on the Random Forest model, the features that provide the most significant contribution are Functional Assessment, ADL, and MMSE. These features play a critical role in the model's performance as they reflect individuals' levels of daily functionality and cognitive status. Notably, Functional Assessment and ADL impact the model's predictions most. Features of moderate importance include Memory Complaints, Behavioural Problems, Physical Activity, and Sleep Quality. These features indicate that physical health and quality of life also play an essential role in predicting cognitive disorders such as Alzheimer's. Features that contribute less include demographic and health-related variables such as

Age, Education Level, and Hypertension. This suggests that demographic characteristics and some health parameters have a lower determinative power in diagnosing the disease. The features with the lowest contribution are clinical symptoms such as Disorientation, Head Injury, and Personality Changes, indicating that the impact of these variables in the model is limited.

3.2. Methods

This study employed various ensemble models to ensure robust and reliable results. The models used in the study are as follows:

- **Random Forest Classifier:** This robust ensemble learning algorithm uses a forest structure composed of multiple decision trees. It classifies based on the majority vote of each tree's output, making it highly effective for classification tasks.
- **CatBoost Classifier:** A gradient boosting-based classification algorithm that works well with categorical data and reduces data preprocessing needs. It is known for its speed and high performance.
- **XGBoost Classifier:** An optimised version of the gradient boosting algorithm, XGBoost is highly effective in computational speed and model accuracy, making it a popular choice for large datasets.

In this study, the results of all methods are presented comparatively.

This analysis demonstrates which features should be considered more to optimise the model's performance.

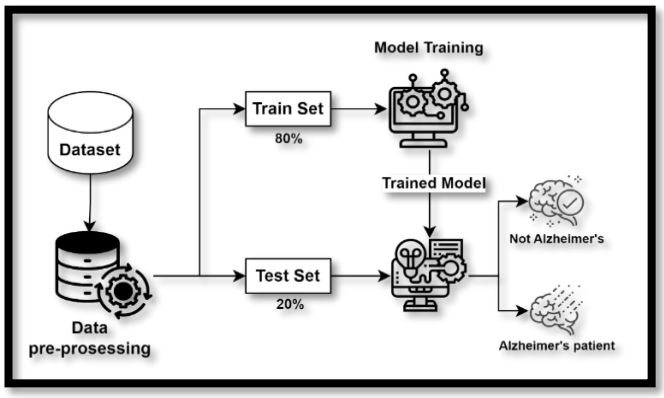

Figure 3. Proposed Machine Learning Architecture.

Figure 3 illustrates the proposed machine learning architecture for predicting Alzheimer's disease. Here's a detailed explanation of the process visualised in the diagram:

Dataset: The process begins with the dataset containing all the relevant data needed to train and test the model.

Data Pre-processing: The dataset undergoes pre-processing to clean and prepare it. This step ensures that the data is in a suitable format for model training and addresses any issues, such as missing values or scaling.

Splitting the Dataset: The pre-processed data is divided into two parts:

- **Train Set (80%):** This portion of the data is used to train the machine learning model. It helps the model learn patterns and relationships within the data.
- **Test Set (20%):** This portion is reserved for testing the model's performance. It objectively evaluates the model's generalisation ability to new, unseen data.

Model Training: The training set is fed into the model training process. Here, various machine learning algorithms (such as Random Forest, CatBoost, and XGBoost) are applied to develop a trained model capable of making predictions.

Trained Model: Once the model is trained, it can make predictions. The trained model is evaluated using the test set to ensure accuracy and reliability.

Prediction: Based on the input data, the trained model classifies individuals as either "Not Alzheimer's" or "Alzheimer's patient." This classification helps identify potential Alzheimer's cases.

This architecture provides a structured approach to developing a predictive model for Alzheimer's disease, ensuring that the model is both accurate and generalizable.

4. Results and Discussion

4.1. Our study result

This section presents the performance metrics of various machine learning models that predict Alzheimer's disease. The models were evaluated using a rigorous methodology that involved data preprocessing, model training, and evaluation. Table 2 summarises the results without hyperparameter optimisation, highlighting each model's accuracy, recall, F1 score, and precision. The models evaluated include Random Forest, CatBoost, and XGBoost, along with their variations using Feature Importance (FI) and Recursive Feature Elimination (RFE).

Table 2: Values without Hyperparameter Optimisation.

Models	Accuracy	Recall	F1-score	Precision
Random Forest	92.56	82.35	88.73	96.18
Random Forest+FI	94.88	88.89	92.52	96.45
Random Forest+RFE	94.88	88.89	92.52	96.45
XGBoost	94.88	89.54	92.57	95.80
XGBoost+FI	95.12	90.20	92.93	95.83
XGBoost+RFE	95.58	91.50	93.65	95.89
CatBoost	95.35	90.85	93.29	95.86
CatBoost+FI	95.35	90.85	93.29	95.86
CatBoost+RFE	95.81	92.16	94.00	95.92

The results indicate that the models perform exceptionally well, instilling confidence in their potential for accurate Alzheimer's disease prediction. CatBoost+RFE achieves the highest accuracy (95.81%) and F1-score (94.00%), suggesting feature selection techniques like RFE can significantly enhance model performance. This could potentially lead to earlier detection of Alzheimer's disease, which is crucial for improving patient outcomes.

Random Forest and XGBoost also show strong results, particularly when combined with feature importance and RFE, indicating that these techniques help refine the model's predictive capabilities. The recall values, which indicate the models' ability to identify actual positive cases, suggest that CatBoost models are particularly effective, which is crucial for early diagnosis of Alzheimer's disease. A high recall value means the model is good at identifying positive cases, which is critical in a disease like Alzheimer's, where early detection is key.

The analysis demonstrates that ensemble models, particularly when enhanced with feature selection methods, provide robust tools for predicting Alzheimer's disease. This robustness, which balances sensitivity and specificity, should reassure the audience about the reliability of the research and the potential for these models to be practical tools in clinical practice.

Table 3: Model with Hyperparameter Optimisation.

Models (HT)	Accuracy	Recall	F1-score	Precision
CatBoost	95.81	92.16	94.00	95.92
CatBoost+FI	95.58	91.50	93.65	95.89
CatBoost+RFE	95.58	91.50	93.65	95.89
XGBoost	95.58	91.50	93.65	95.89
XGBoost+FI	95.35	90.85	93.29	95.86
XGBoost+RFE	95.58	91.50	93.65	95.89
Random Forest	94.42	87.58	91.78	96.40
Random Forest+RFE	94.88	88.89	92.52	96.45
Random Forest+FI	95.12	89.54	92.88	96.48

The application of hyperparameter optimisation has led to improvements in model performance across most metrics. CatBoost continues to perform exceptionally well, achieving the highest accuracy (95.81%) and F1-score (94.00%), indicating its robustness in handling complex datasets.

Random Forest and XGBoost also show enhanced performance with hyperparameter tuning, particularly when combined with feature importance and RFE. These techniques help refine the model's predictive capabilities, as evidenced by the improved precision and recall scores. Overall, hyperparameter optimisation significantly enhances the models' ability to predict Alzheimer's disease, providing a more reliable and accurate tool for early diagnosis. The results underscore the importance of fine-tuning model parameters to achieve optimal performance, highlighting the significance of this aspect in the field of machine learning in healthcare. Based on the analysis of accuracy, recall, F1-score, and precision, we observe the following outcomes for the models used in predicting Alzheimer's disease:

- **Random Forest without Hyperparameter Optimisation:** The model starts with an accuracy of 92.56%, but a recall of 82.35% indicates a moderate ability to capture positive cases, with about 18% of true positives misclassified. The high precision is 96.18%, showing that most positive predictions are correct.
- **Random Forest with Hyperparameter Optimisation and Feature Importance (FI):** Accuracy improves to 94.88%, and recall increases to 88.89%, indicating better detection of positive classes and a reduction in false positives. Precision slightly improves to 96.45%, and the F1-score rises to 92.52%, reflecting a balanced enhancement in model performance.
- **CatBoost without Hyperparameter Optimisation:** This model starts with a high accuracy of 95.35% and a recall of 90.85%, indicating strong performance in identifying positive classes. Precision is also high at 95.86%.
- **CatBoost with Hyperparameter Optimisation and RFE:** Accuracy reaches 95.81%, with recall improving to 92.16%, showing even better identification of positive classes. Precision remains stable at 95.92%, and the F1-score increases to 94.00%, indicating an excellent balance between recall and precision.
- **XGBoost without Hyperparameter Optimisation:** The initial accuracy is 94.88%, with a recall of 89.54%, slightly lower than CatBoost in capturing positive classes. Precision is high at 95.80%.
- **XGBoost with Hyperparameter Optimisation and FI:** Accuracy rises to 95.12%, and recall improves to 90.20%, showing better capture of positive classes. However, using RFE, recall drops to 87.58%, and precision falls to 94.37%, indicating that RFE does not yield optimal results for XGBoost, resulting in a performance loss.

These results highlight the importance of hyperparameter tuning and feature selection in enhancing model performance, with CatBoost showing the most significant improvements.

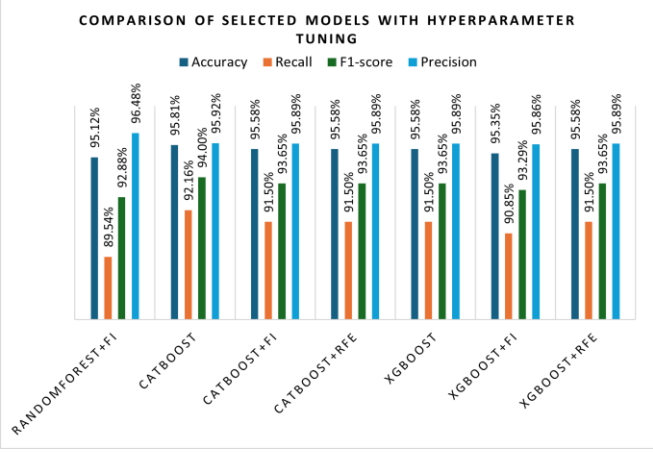

Figure 4. Comparison of selected models with Hyperparameter tuning.

Figure 4 illustrates training a model for Alzheimer's classification and compares the performance of various models after hyperparameter tuning. The first part of the figure shows the data pre-processing stage, where the dataset is split into a training set (80%) and a test set (20%). The model is trained on

the training set and then evaluated on the test set to classify individuals as either "Not Alzheimer's" or "Alzheimer's patient." The second part of the figure compares selected models, highlighting their accuracy, recall, F1 score, and precision. The Random Forest model with feature importance (FI) tuning achieves the highest accuracy of 96.48 %, demonstrating superior performance. Other models, such as CatBoost and XGBoost, also show competitive results, with CatBoost achieving a notable F1 score of 94.00 %. This comparison underscores the effectiveness of hyperparameter tuning in enhancing model performance across different metrics.

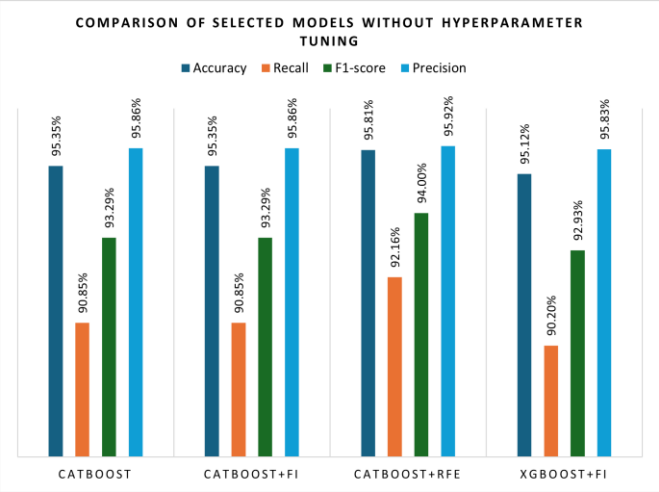

Figure 5. Comparison of selected models without Hyperparameter tuning.

Figure 5 compares selected models without hyperparameter tuning, focusing on their performance metrics: accuracy, recall, F1-score, and precision. The CatBoost model starts with a high accuracy of 95.35% and a recall of 90.85%, indicating initial solid performance in identifying positive cases. The precision is also high at 95.86%, reflecting accurate optimistic predictions. The CatBoost model with feature importance (FI) maintains similar performance, with slight variations in recall and precision. The XGBoost model shows an accuracy of 95.12% and a recall of 90.20%, with a precision of 95.83%, indicating competitive performance but a slightly lower recall than CatBoost. The figure highlights the models' effectiveness even without hyperparameter tuning, with CatBoost demonstrating the highest initial accuracy and recall.

4.2. Confidence Measures

To ensure the reliability of our model predictions, we incorporated confidence measures into our analysis. Confidence measures provide a quantitative assessment of the certainty associated with each prediction, which is particularly important in medical diagnostics, where decision-making requires high reliability.

By categorising predictions into three confidence levels [26]:

- Low [0-0.7]
- Medium [0.7-0.9]
- High [0.9-1.0]

We aimed to evaluate the distribution of prediction confidence and identify areas for potential improvement. This approach ensures a more transparent and interpretable diagnostic process, enabling clinicians to assess the reliability of model outputs better.

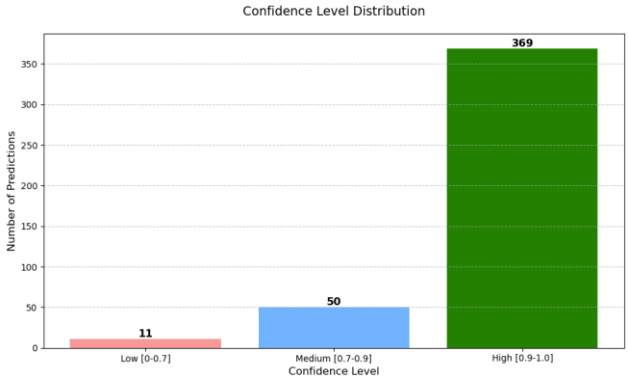

Figure 6. Confidence level distribution.

Figure 6 illustrates the distribution of predictions across the three confidence levels. The majority of predictions (369, **86.7%**) fall into the High confidence category, indicating that the model is remarkably accurate about most of its predictions. A smaller number of predictions (50, **11.8%**) are classified as Medium confidence, while only 11 (**2.6%**) 1 predictions fall into the Low confidence category.

This distribution demonstrates the robustness of the model, as the majority of predictions are made with high confidence. However, the presence of Low and Medium confidence predictions highlights the need for further refinement, particularly in cases where the model's certainty is lower. These cases may benefit from additional data or feature engineering to improve the model's performance in uncertain scenarios.

4.3. Comparison with previous studies

This section compares the proposed model's performance with previous studies. Table 4 summarises the accuracy of different models used in earlier research and highlights the proposed model's superiority.

Table 4: Comparison of the proposed model and the previous study.

Author(s)	Proposed Model	Accuracy (%)
[1]	Random Forest	86.80
[12]	Adaptive Neuro-fuzzy	84.00
[9]	Voting – Ensemble Model	93.92
[10]	Voting classifier	83.00
[13]	RNN	84.20
Our Model	CatBoost+RFE	95.81

The results presented in Table 4 show that the proposed model, CatBoost+RFE, achieved the highest accuracy with 95.81% accuracy, surpassing all other models in the list. This demonstrates the effectiveness of the proposed approach in addressing the problem and its robustness compared to previous methods. [12] The adaptive Neuro-fuzzy Inference System) The model with an accuracy of 84.00% is significantly lower than the proposed model, while [9] (Voting - Ensemble Model) with an accuracy of 93.92% is the closest to the proposed model, but with a difference of about 2.56%. [10] (Voting Classifier) was the lowest performing model with an accuracy of 83.00%, while [13] (RNN) performed similarly to [12] with an accuracy of 84.20%. These results show that the CatBoost+RFE model is more accurate and reliable than the other models. This improvement

can be achieved by combining CatBoost+RFE and effectively selecting and prioritising the most relevant features. This significant increase in accuracy demonstrates the potential of the proposed model for practical applications and emphasises the importance of feature selection and ensemble methods in achieving superior performance.

5. Conclusions

In this study, we explored the application of machine learning models for the early diagnosis of Alzheimer's disease, focusing on the impact of hyperparameter optimisation and feature selection techniques on model performance. Our findings demonstrate significant improvements in model accuracy, recall, precision, and F1-score when these techniques are applied. However, it's important to note that our study has limitations, such as using a specific dataset and the potential for overfitting. These limitations provide opportunities for future research to further enhance the performance of machine learning models in Alzheimer's disease prediction.

The increase in recall values across all models indicates a heightened ability to correctly identify positive cases, which is crucial in scenarios where early detection of Alzheimer's disease can lead to better patient outcomes. High precision values suggest that the models maintain or improve the accuracy of optimistic predictions, ensuring that most identified cases are true positives. The improvements in F1-score reflect a balanced enhancement in both recall and precision, underscoring the models' overall effectiveness. CatBoost exhibited the highest performance among the models tested, delivering the most balanced and superior results across recall, accuracy, and F1-score metrics. CatBoost achieved the best recall with hyperparameter optimisation, demonstrating its superior capability in capturing positive classes compared to other models. Although Random Forest and XGBoost also showed improvements with feature selection techniques, they did not surpass CatBoost's performance. These results emphasise the importance of considering multiple evaluation metrics, such as recall, precision, and F1-score, in addition to accuracy, when selecting models for medical diagnosis tasks. The ability to accurately identify positive cases is particularly vital in the context of Alzheimer's disease, where early intervention can significantly alter the disease's progression.

In conclusion, this study highlights the effectiveness of machine learning models in the early diagnosis of Alzheimer's disease and provides a valuable foundation for future research. By demonstrating the benefits of hyperparameter optimisation and feature selection, we offer insights into improving model performance and achieving more reliable diagnostic outcomes. Future work could explore further integrating additional data sources and advanced feature engineering techniques to enhance predictive accuracy and clinical applicability.

Ethics committee approval and conflict of interest statement

This article does not require approval from the ethics committee. This article has no conflicts of interest with any individual or institution

Author Contribution Statement

Author 1 (corresponding author) conducted the literature review, conceptualised the study, provided key insights into the theoretical framework, proposed the data set, and wrote the manuscript, focusing on presenting the results. Author 2 contributed to the literature review, created visual representations, added the study results, provided support for the data analysis processes, developed the code used in the study, and contributed to the overall organisation of the manuscript.

References

[1] Ünal, C., Gökşen, Y. 2024. Unlocking Neurological Mysteries: Machine Learning Approaches to Early Detection of Alzheimer's Disease, Güvenlik Bilimleri Dergisi, Vol. 13, no. 1, pp. 85-104.

[2] Jo, T., Nho, K., Saykin, A. 2019. Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data, Frontiers in Ageing Neuroscience, Vol. 11, DOI: 10.3389/fnagi.2019.00220.

[3] Alatrany, A., Hussain, A., Mustafina, J., Al-Jumeily, D. 2022. Machine Learning Approaches and Applications in Genome-Wide Association Study for Alzheimer's Disease: A Systematic Review, IEEE Access, DOI: 10.1109/ACCESS.2022.3182543.

[4] Venugopalan, J., Tong, L., Hassanzadeh, H., Wang, M. 2021. Multimodal Deep Learning Models for Early Detection of Alzheimer's Disease Stage, Scientific Reports, Vol. 11, DOI: 10.1038/s41598-020-74399-w.

[5] Aqeel, A., Hassan, A., Khan, M., Rehman, S., Tariq, U., Kadry, S., Majumdar, A., Thinnukool, O. 2022. A Long Short-Term Memory Biomarker-Based Prediction Framework for Alzheimer's Disease, Sensors, Vol. 22, DOI: 10.3390/s22041475.

[6] Bi, X., Li, S., Xiao, B., Li, Y., Wang, G., Ma, X. 2020. Computer-aided Alzheimer's disease diagnosis by an unsupervised deep learning technology, Neurocomputing, Vol. 392, pp. 296-304, DOI: 10.1016/J.NEUCOM.2018.11.111.

[7] Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M. 2015. Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease, IEEE Transactions on Biomedical Engineering, Vol. 62, pp. 1132-1140, DOI: 10.1109/TBME.2014.2372011.

[8] Alsaeed, D., Omar, S. 2022. Brain MRI Analysis for Alzheimer's Disease Diagnosis Using CNN-Based Feature Extraction and Machine Learning, Sensors, Vol. 22, DOI: 10.3390/s22082911.

[9] Irfan, M., Shahrestani, S., Elkhodr, M. 2023. Early Detection of Alzheimer's Disease Using Cognitive Features: A Voting-Based Ensemble Machine Learning Approach, IEEE Engineering Management Review, Vol. 51, pp. 16-25, DOI: 10.1109/EMR.2022.3230820.

[10] Kavitha, C., Mani, V., Srividhya, S., Khalaf, O., Romero, C. 2022. Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models, Frontiers in Public Health, Vol. 10, DOI: 10.3389/fpubh.2022.853294.

[11] Meng, X., Wu, Y., Liu, W., Wang, Y., Xu, Z., Jiao, Z. 2022. Research on Voxel-Based Features Detection and Analysis of Alzheimer's Disease Using Random Survey Support Vector Machine, Frontiers in Neuroinformatics, Vol. 16, DOI: 10.3389/fninf.2022.856295.

[12] Castellazzi, G., Cuzzoni, M.G., Cotta Ramusino, M., Martinelli, D. et al. 2020. A machine learning approach for the differential diagnosis of Alzheimer's and vascular dementia fed by MRI-selected features, Frontiers in Neuroinformatics, Vol. 14, p. 25.

[13] Jo, T., Nho, K., Saykin, A.J. 2019. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data, Frontiers in Ageing Neuroscience, Vol. 11, p. 220.

[14] Raza, M., Awais, M., Ellahi, W., Aslam, N., Nguyen, H.X., Le Minh, H. 2019. Diagnosis and Monitoring of Alzheimer's Patients Using Classical and Deep Learning Techniques.

[15] Cochrane, C., Castineira, D., Shiban, N., Protopapas, P. 2020. Application of Machine Learning to Predict the Risk of Alzheimer's Disease: An Accurate and Practical Solution for Early Diagnostics.

[16] Chang, C.H., Lin, C.H., Lane, H.Y. 2021. Machine Learning and Novel Biomarkers for the Diagnosis of Alzheimer's Disease.

[17] Battineni, G., Amran Hossain, M., Chintalapudi, N., Traini, E., Rao Dhulipalla, V., Ramasamy, M., Amenta, F. 2021. Improved Alzheimer's Disease Detection by MRI Using Multimodal Machine Learning Algorithms.

[18] Li, X., Qiu, Y., Zhou, J., Xie, Z. 2021. Applications and Challenges of Machine Learning Methods in Alzheimer's Disease Multi-Source Data Analysis.

[19] Aqeel, A., Hassan, A., Attique Khan, M., Rehman, S., Tariq, U., Kadry, S., Majumdar, A., Thinnukool, O. 2022. A Long Short-Term Memory Biomarker-Based Prediction Framework for Alzheimer's Disease.

[20] Singh, N.D.P., Soni, N., Kapoor, A. 2022. Automated Detection of Alzheimer's Disease Using MRI Images and Deep Neural Networks - A Review.

[21] Bucholc, M., James, C., Al Khleifat, A., Badhwar, A.P. et al. 2023. Artificial Intelligence for Dementia Research Methods Optimisation.

[22] Kang, W., Li, B., Papma, J.M., Jiskoot, L.C. et al. 2023. An Interpretable Machine Learning Model with Deep Learning-Based Imaging Biomarkers for Diagnosis of Alzheimer's Disease.

[23] Abdelwahab, M.M., Al-Karawi, K.A., Semary, H.E. 2023. Deep Learning-Based Prediction of Alzheimer's Disease Using Microarray Gene Expression Data.
[24] Ahmadi, M., Javaheri, D., Khajavi, M., Danesh, K., Hur, J. 2024. A Deeply Supervised Adaptable Neural Network for Diagnosis and Classification of Alzheimer's Severity Using Multitask Feature Extraction.
[25] El Kharoua, R. 2024. Alzheimer's Disease Dataset, Kagle, DOI: 10.34740/KAGGLE/DSV/8668279.
[26] Klasan, A., Neri, T., Sommer, C., Leie, M.A., Dworschak, P., Schofer, M.D., Heyse, T.J. 2019. Analysis of acetabular version: Retroversion prevalence, age, side and gender correlations, Journal of Orthopaedic Translation, Vol. 18, pp. 7-12, DOI: 10.1016/j.jot.2018.12.002.