International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

# Improving Early Alzheimer's Prediction with a Hybrid Ensemble Machine Learning Model

**Abhishek Guru, Suraj Kumar Sahu**

Department of Computer Science and Engineering, MSEIT, Mats University Aarang Raipur, Chhattisgarh
abhishekguru0703@gmail.com, surajk@matsuniversity.ac.in

**Abstract.** It is essential to diagnose Alzheimer's disease (AD) at an early stage in order to provide rapid management and improve positive outcomes for patients. The purpose of this work is to demonstrate a mixed ensemble machine learning algorithm that can reliably predict the start of early-stage Alzheimer's disease by utilising clinical and neuropsychological datasets. The core of the model is a sequential pipeline that consists of data preparation, feature selection, and a classification architecture that is based on stacking on two levels. MICE, which stands for Multivariate Imputation by Chained Equations, is the method that we use in situations when there is a paucity of data. We scale the features by using the Z-score normalisation method. The Synthetic Minority Over-sampling Technique, more often referred to as SMOTE, is the technique that we have developed to address the issue of class imbalance. Through the process of reducing the number of dimensions in the model and pruning it, recursive feature elimination (RFE) is able to identify the most important characteristics and enhance the efficiency of the model. XGBoost, Support Vector Machine (SVM), and Random Forest (RF) are the three fundamental learners that are used in the suggested ensemble. These learners are combined with a variety of characteristics via the utilisation of a Logistic Regression (LR) meta-learner, which produces probabilities as its outputs. Using a stratified 10-fold cross validation process, we put the model through its paces and evaluate its accuracy, precision, recall, F1 score, and area under the curve (AUC-ROC). As part of our efforts to reduce the number of instances in which a false negative diagnosis is made, we are putting our whole of attention on memory. The hybrid framework that was provided is a viable hybrid strategy for early Alzheimer's disease detection, outperforming all individual models in terms of prediction performance and durability. This conclusion is based on the results of the experiments that were conducted.

**Keywords:** Alzheimer's Disease, Hybrid Ensemble Learning, Stacking Model, Feature Selection, Early Disease Prediction.

## 1 Introduction

Alzheimer's disease (AD), the most severe type of neurodegenerative dementia, is characterised by a decline in cognitive abilities as well as patterns of conduct that are aberrant. This particular kind of dementia is the most prevalent form of dementia affecting people all over the globe, and it has negative consequences for patients, carers, and healthcare systems [...]. Because of the fact that the population of the globe is becoming older, it is anticipated that the prevalence of Alzheimer's disease will dramatically grow in the years to come. This will make the need of early detection and intervention even more crucial [3, 4]. One of the possible benefits of early identification is that it has the ability to enhance the quality of life of patients, arrest the course of sickness, and eventually reduce the costs of healthcare [5].

The diagnosis of Alzheimer's disease has been performed for a very long time using a variety of diagnostic methods, including cognitive testing, clinical examination, neuroimaging (MRI, PET), and other neurodegenerative disorders [4-6]. Because of the clinical success of these therapies, the time, effort, and money that are required (in addition to the need for professional interpretation) are all justified or justified. In clinical practice, the early indicators of Alzheimer's disease are sometimes difficult to detect, and many of the symptoms are similar to those that are associated with normal ageing. This further complicates the situation [9].

The field of medical diagnostics has seen a profound transformation in recent years as a result of developments in artificial intelligence and machine learning [10]. The use of machine learning (ML) techniques has made it simpler to analyse complicated clinical data and to identify patterns in the data that were previously unnoticeable [11]. For the purpose of Alzheimer's disease classification tasks, Random Forest (RF), Support Vector Machines (SVM), and Gradient Boosting models have been used extensively [12, 13]. There are a number of problems that may have an effect on machine learning models, including as overfitting, sensitivity to noisy data, and poor generalisability to samples that are unknown [14].

In the field of medical data analytics, ensemble learning techniques have emerged as a powerful tool that enables the overcoming of these limits [15]. The use of ensemble techniques has been widely acknowledged to increase prediction performance [16]. These approaches combine the predictions of a large number of base learners in order to provide a more reliable prediction. A approach that use a meta-learning framework to integrate a large number of heterogeneous models has been receiving a lot of attention recently [17]. This strategy is known as stacking-based ensemble learning. In contrast to voting or averaging, stacking improves accuracy and stability by learning the ideal mix of outputs from base model results [18]. This is accomplished via the process of stacking.

When evaluating the effectiveness of machine learning systems, it is important to take into consideration not only the process of selecting the input data but also the quality of the data itself. There is a possibility that clinical Alzheimer's disease prediction databases will have missing instances, class imbalance, and changing feature sizes [19]. It is possible that the models will suffer if these challenges are not handled in an effective manner. Imputation, normalisation, and synthetic sampling are all crucial preprocessing methods because of this reason. In situations where high-dimensional information is being dealt with, the understanding of the model may be further impeded if there are qualities that are either unneeded or duplicated

[20]. Because of this, good prediction is dependent on feature selection algorithms that prioritise characteristics.

The majority of published research still makes use of a limited number of classifiers and preprocessing approaches, which restricts their capacity to capture the complexity of clinical data. This is despite the fact that tremendous breakthroughs have been made. Class imbalance and model overfitting are two additional issues that may be overlooked by standard techniques. As a result of these constraints, it is evident that early Alzheimer's disease prediction is challenging, and it is necessary to have a framework that is more dependable. This framework should integrate hybrid ensemble learning approaches, enhanced preprocessing, and feature selection.

The primary objective of this research will be to construct a hybrid ensemble machine learning framework for the purpose of early Alzheimer's disease prediction by making use of structured clinical data. The solution that has been proposed is a pipeline that consists of many stages and makes use of stacking-based categorisation, feature selection, and data preparation. Z-scores are used in order to normalise features, and the Synthetic Minority Over-sampling Technique (SMOTE) is utilised in order to address class imbalance. The initial step of preprocessing is known as multivariate imputation by chained equations (MICE). Optimal feature selection is combined with recursive feature elimination (RFE) in order to achieve the highest possible level of efficiency.

The stacking ensemble design with two tiers is the most important contribution that this study has made a contribution to. To provide probabilities, some classifiers, such as Support Vector Machine, Random Forest, and XGBoost, are able to function at their most fundamental level. Using a Logistic Regression meta-learner, one may derive a final prediction by feeding the results that have been gathered into the framework. This architecture makes it possible for the model to avoid the drawbacks of each learning approach while yet including the most beneficial aspects of those techniques.

The following are examples of significant contributions to this endeavour: Through the use of clinical data, a hybrid stacking-based ensemble learning system is utilised for the purpose of early Alzheimer's disease prediction. • Handling missing values in a consistent manner, normalising data, and attaining class balance via the use of pre-processing pipelines that are constructed by the application of methods such as MICE, Z-Score normalisation, and SMOTE. • The efficiency and generalisability of the model are both improved by the use of Recursive Feature Elimination (RFE), which works by determining which characteristics are the most important. Obtaining the best possible outcomes in terms of prediction is accomplished by combining a Logistic Regression meta-learner with a variety of base learners such as SVM, Random Forest, and XGBoost. The proposed model was cross-validated with the help of Stratified K-Fold, with recall serving as the primary performance metric. The objective of this process was to reduce the number of false negatives that occurred during early diagnosis.

Within the second section, we will address the use of machine learning and ensemble methods for the purpose of forecasting the beginning of Alzheimer's disease. In Section III, we discuss every aspect of the approach that was recommended, including the architecture of the model, the selection of features, and the preparation of the data. In the fourth section, we look into the experimental design and the metrics that were used for evaluation. In Section V, the results and discussion are presented, and in Section VI, recommendations for more study are provided.

## 2 RELATED WORKS

Alzheimer's disease (AD) has been the subject of a significant amount of research in the fields of medical informatics and artificial intelligence due to the fact that the disease is very prevalent and there is presently no medication available for it. The neurological condition known as Alzheimer's disease (AD) is very prevalent. A fast diagnosis of Alzheimer's disease makes it feasible to initiate early intervention, develop an effective treatment plan, and limit the course of the illness throughout the patient's lifetime. With the use of machine learning methods, an increasing number of research have shown that neuropsychological, cognitive, and clinical indications have the potential to boost early diagnosis. In the beginning, the purpose of this area of research was to identify the clinical factors that had the greatest influence on the progression of Alzheimer's disease. Recursive Feature Elimination, often known as RFE, is a feature selection approach that has been utilised in order to reduce the dimensionality of the data while still preserving the information that is most important for prediction [21].

The conventional methods of machine learning have been essential in pioneering the development of Alzheimer's disease prediction strategies. A number of well-known machine learning and statistical approaches, including Logistic Regression, k-Nearest Neighbours (KNN), Naïve Bayes, and basic decision-based classifiers, were used in the first study to discriminate between cognitively normal persons and Alzheimer patients. This was accomplished via the utilisation of binary classification. Numerous prediction models have shown that population statistics and structured clinical indicators, such as cognitive scores (for example, the Mini-Mental State Examination and the Clinical Depression Rating), are valid predictors. When confronted with the non-linear patterns and extensive web of interactions that are present in medical data sets, however, a significant number of these older models were unable to perform well.

Researchers were able to improve the accuracy of their predictions about the categorisation of Alzheimer's disease by using supervised learning approaches such as Decision Trees and Support Vector Machines (SVM) [23]. The ability of support vector machines (SVM) to deal with high-dimensional feature areas and nonlinear decision restrictions is one of the reasons for its widespread popularity. There were a number of problems associated with these models, despite the fact that they attained a higher level of accuracy than statistical models. These downsides included being too sensitive to noise, having difficulty modifying parameters, and being unable to deal with large datasets.

Support vector machines (SVMs), Random Forests (RFs), and ensemble-based classifiers are only a few examples of the machine learning techniques that have been investigated and evaluated in connection to the prediction of Alzheimer's disease [24]. The results of these trials show that an ensemble-based model is superior than a single classifier in terms of performance. This is due to the fact that ensemble-based models have the potential to improve generalisation while simultaneously reducing variance. However, many of these techniques were not successful in real clinical settings since they required intricate pre-processing pipelines.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) database has shown to be of great assistance to researchers working in this line of inquiry [25]. It contains data from cognitive research, imaging, and clinical practice that is commonly used for testing machine learning models at clinical settings. It has been shown via research that makes use of ADNI data that machine learning (ML) techniques are capable of properly classifying instances of Alzheimer's disease. The quality of the data, the strategies of feature selection, and the pre-processing techniques continue to have a significant impact on the performance.

Due to the fact that they have a solid theoretical basis and have shown effectiveness in high-dimensional applications, support vector machine (SVM) models have been used widely in traditional machine learning procedures [26]. Considering that medical datasets often consist of nonlinear decision surfaces, support vector machine techniques are excellent options for implementation. You may need to make adjustments to the parameters in order to make them work with really large data sets; but, they do have their practical applications.

Similar to the previous example, Random Forest (RF) is an algorithm that is often used for predicting Alzheimer's disease. In addition to being long-lasting, the bagging method helps to reduce the amount of overfitting that occurs [27]. Through the construction of several decision trees and the subsequent averaging of their outputs, the RF approach improves both accuracy and stability. It is possible to understand how to utilise RF when dealing with smaller clinical data sets; however, when dealing with more intricate data sets, RF becomes much more difficult to understand.

Researchers have investigated both solo classifiers and hybrid classifiers, which are hybrid classifiers that employ a mix of multiple different machine learning approaches [28]. This is done in an effort to increase performance. The integration of techniques that have qualities that are complimentary to one another is one way that might be taken to create them. For example, decision trees and support vector machines (SVM) or RF and KNN are two examples of such approaches. Despite the fact that hybrid models have the potential to attain even greater levels of accuracy, working with them may be rather difficult, and they may at times need significant adjustments.

In order to address the inadequacies of the models, it may be beneficial to combine many models. It is possible to enhance the predictions that are provided by many learners by using bagging approaches, which include training them on distinct subsets of data [29]. Boosting techniques, on the other hand, aim to enhance prediction accuracy by progressively fortifying weak learners in relation to their past model defects [30]. This is done in order to improve the accuracy of predictions.

Gradient boosting techniques, in particular XGBoost, have garnered attention in the field of medical classification applications due to the fact that they are considered to be efficient, scalable, and provide superior regularisation qualities [31]. In a study that included the prediction of Alzheimer's illness, it was discovered that XGBoost performed better than more conventional machine learning models. Due to the fact that it is able to handle missing values and non-linear correlations so well, clinical datasets are ideal for implementation.

In recent times, hybrid ensemble learning frameworks have been the focus of research [32]. These frameworks incorporate many machine learning approaches in order to improve the accuracy of predictions. The SVM, RF, and boosting techniques are only few of the models that are heavily used by these frameworks. The results of this study have shown that hybrid models are more accurate and dependable than solo versions.

Among the hybrid approaches that are considered to be among the most successful in forecasting the start of Alzheimer's disease [33], stacking-based ensemble learning is very effective. When a meta-learner creates a single prediction by using the predictions provided by a large number of base learners, this is an example of the stacking process. Through the use of this method, the accuracy and generalisability of the model are enhanced as it is trained to recognise the most effective combinations of its fundamental predictions.

On the other hand, you may be able to enhance performance by using stacking frameworks that are dependent on meta-learning in order to develop connections between the several models [34]. When dealing with complicated medical data sets, where various models discover distinct patterns, these strategies truly come into their own and stand out as particularly useful. When stacking models, careful preparation is essential in order to prevent data leakage and overfitting.

When it comes to Alzheimer's disease classification tasks, hybrid ensemble classifiers that make use of SVM, RF, and XGBoost have shown outstanding effectiveness [35]. The fact that they make use of a large number of base models makes the models more robust and improves their ability to forecast, respectively. According to the findings of the study [36], one of the major requirements that makes models great at generalising has to do with the presence of a diverse ensemble.

Furthermore, it was shown that stacking-based ensemble models performed better than the more standard machine learning approaches when it came to ADNI data [37]. This was observed by researchers. In order to prevent overfitting and ensure that these models are reliable, they need to be subjected to rigorous testing using methods considered to be reliable.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The process of preparing data is an essential stage in the process of enhancing the performance of models used in healthcare applications. A multivariate model that is iterative and takes into account the missing data on other variables In clinical data sets, the problem of missing data is often addressed by the use of a technique known as imputation by chained equations (MICE) [38]. By doing so, the links between the data are preserved, and the quality of the data is improved.

The scalability of features is another need for machine learning processes. All of the attributes need to be normalised in order to guarantee that they contribute in an equal manner to the models. For the purpose of doing this, a method known as z-score normalisation is used [39]. Because of this, the model will become more stable and its rate of convergence will increase.

In certain datasets, for instance, there may be an imbalance between healthy persons and actual Alzheimer's patients, which makes it difficult to draw conclusions that can be relied upon. When it comes to addressing this issue, the SMOTE approach is a strategy that everyone likes to use [40]. It does this in order to enhance the accuracy of the classifications by generating synthetic samples for the group that is under-represented.

Classical machine learning algorithms unquestionably provide a solid basis for the prediction of Alzheimer's disease; nevertheless, the study indicates that more complex models are required for high-dimensional datasets that are not balanced. There has been a recent uptick in the use of hybrid ensemble learning frameworks in applications that are aiming at accurate early Alzheimer's disease prediction. Stacking-based classification, preprocessing, and feature selection procedures are all included into these systems in order to improve accuracy, robustness, and generalisation performance.

Table 1: Summary of Related Work on Alzheimer’s Disease Prediction

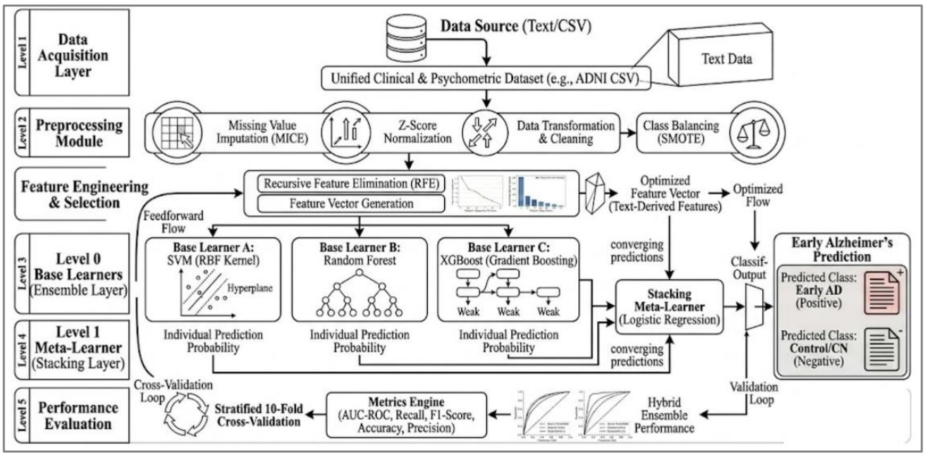
| Ref. | Method / Approach | Techniques Used | Dataset Type | Limitations |
|---|---|---|---|---|
| [21] | Feature-based ML | RFE, feature ranking | Clinical data | Limited model diversity |
| [22] | Classical ML models | Logistic Regression, KNN | Clinical datasets | Poor generalization |
| [23] | Supervised learning | Decision Tree, SVM | Cognitive scores | Sensitive to noise |
| [24] | ML classification | SVM, RF | Clinical + demographic | Overfitting issues |
| [25] | Comparative ML study | Multiple classifiers | ADNI dataset | Limited preprocessing |
| [26] | SVM-based model | Kernel SVM | Clinical features | Computational complexity |
| [27] | Ensemble learning | Random Forest | Tabular clinical data | Less interpretability |

| | | | | |
|---|---|---|---|---|
| [28] | ML-based AD prediction | RF, KNN | Neuro data | Limited feature optimization |
| [29] | Ensemble methods | Bagging techniques | Medical datasets | Requires large datasets |
| [30] | Boosting approach | Gradient Boosting | Clinical datasets | Sensitive to noise |
| [31] | XGBoost model | Gradient boosting | Structured data | Overfitting risk |
| [32] | Hybrid ML models | SVM + RF + Boosting | ADNI dataset | Complex tuning required |
| [33] | Stacking ensemble | Meta-learning | Clinical data | Risk of overfitting |
| [34] | Meta-learning approach | Stacking framework | Medical datasets | Computational overhead |
| [35] | Hybrid ensemble | SVM + RF + XGBoost | AD datasets | Data dependency |
| [36] | Ensemble diversity study | Multiple classifiers | Clinical data | Training complexity |
| [37] | Stacking-based model | Multi-classifier fusion | ADNI dataset | Requires careful design |

## 3. METHODOLOGY

This research presents a hybrid ensemble machine learning framework for early diagnosis of Alzheimer's disease (AD). The architecture's pipeline is structured in a sequential fashion and includes many processes, such as data preparation, feature engineering, feature selection, and a two-level stacking model for classification. Improving prediction accuracy and robustness is the primary objective, and several heterogeneous learning techniques will be used to achieve this.

The overall architectural workflow of the proposed system is defined as:



International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Figure 1: Proposed Hybrid Ensemble Methodology for Early Alzheimer's Prediction

### 3.1 Dataset Description and Acquisition

This research makes use of data collected by the Alzheimer's Disease Neuroimaging Initiative (ADNI). Data from cognitive assessments like the Mini-Mental State Examination (MMSE) and the Clinical Dementia Rating Sum of Boxes (CDR-SB) as well as demographic information are part of the clinical and neuropsychological records included in the CSV format.

The dataset can be mathematically represented as:

$$D = \{ (x_{\mathrm{i}}, y_{\mathrm{i}}) \}_{\mathrm{i}}^{=1\mathrm{N}}$$

where $x_{\mathrm{i}} \in \mathbb{R}^{\mathrm{d}}$ denotes the feature vector of the i-th patient, and $y_{\mathrm{i}} \in \{0,1\}$ represents the class label, where 0 indicates Cognitively Normal and 1 indicates Early Alzheimer's Disease.

The use of a single, well-curated dataset minimizes noise and eliminates the complexity associated with multimodal data integration. This, in turn, enhances model interpretability and improves computational efficiency.

**Data Preprocessing Module**

Clinical datasets are often characterized by missing values, heterogeneous feature scales, and class imbalance. To address these challenges, a comprehensive preprocessing strategy is implemented:

### 3.2.1 Missing Value Imputation

Missing values in the dataset are handled using Multivariate Imputation by Chained Equations (MICE). In this approach, each feature containing missing values is modeled using the remaining features in an iterative manner. The imputation process is expressed as:

$$x_j^{miss} = f_{j(XJay)} + \varepsilon$$

Where $f_j$ represents a predictive model trained on all other features $X\backslash j$, and $\varepsilon$ denotes the residual error term. This method ensures that missing values are estimated in a statistically consistent manner by capturing relationships among variables.

3.2.2 Feature Normalization

To ensure uniform scaling across all features, Z-score normalization is applied. This technique transforms the data so that each feature has a mean of zero and a standard deviation of one. The transformation is defined as:

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

$$x' = \frac{(x - \mu)}{\sigma}$$

where μ is the mean and $\sigma$ is the standard deviation of the feature. This normalization improves model performance by preventing features with larger magnitudes from dominating the learning process.

3.2.3 Class Balancing

To address class imbalance in the dataset, the Synthetic Minority Over-sampling Technique (SMOTE) is used. SMOTE generates synthetic samples for the minority class by interpolating between a data point and its nearest neighbors. The synthetic sample is generated as:

$$x_{new} = x_i + \lambda (x_{nn} - x_i), where\ \lambda \in [0,1]$$

Here, $x_i$ is a minority class sample, and $x_n$ is one of its nearest neighbors. This technique helps in balancing the dataset by increasing the representation of minority class samples, thereby improving the robustness of classification models.

### 3.3 Feature Engineering and Selection

To reduce dimensionality and eliminate redundant features, Recursive Feature Elimination (RFE) is employed. This technique iteratively removes the least important features based on model importance scores, thereby refining the feature set at each step.

The feature elimination process can be represented as:

$$F_{\mathrm{k}}{}^{+1} = F_{\mathrm{k}}\ \{f_{least}\}$$

where $f_{least}$denotes the least important feature identified at iteration $k$.

This iterative process continues until an optimal subset of features is obtained. The resulting reduced feature set improves the model's generalization capability while also decreasing computational complexity, leading to more efficient and robust performance.

### 3.4 Hybrid Ensemble Architecture

An ensemble architecture with two layers of stacking that combines the strengths of many different classifiers is the main contribution of this study. The predictive power of this hybrid framework is enhanced by merging and capitalising on the characteristics of many learning models..

#### 3.4.1 Level 0: Base Learners

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The optimized feature set is simultaneously processed by three parallel base classifiers:

**Support Vector Machine (SVM):**
Support vector machines (SVMs) equipped with Radial Basis Function (RBF) kernels are able to successfully detect intricate non-linear correlations in clinical data that is high dimensional.

**Random Forest (RF):**
A method that builds numerous decision trees using bagging as an ensemble. By combining predictions from several trees, it makes the model more resilient and less prone to overfitting.

**XGBoost:**
To enable the discovery of subtle patterns in structured data, a gradient boosting framework is used, which systematically minimises residual errors.

A probabilistic output indicating the likelihood of a diagnosis of Alzheimer's disease is generated by each base learner.

Let the outputs of these classifiers be represented as:

$$p_j = h_{j(x)}, j = 1, 2, 3$$

where $p_j$ denotes the predicted probability produced by the $j_{th}$ classifier.

**3.4.2 Level 1: Meta-Learner (Stacking Layer)**

Instead of using conventional voting methods, the probabilistic outputs from the base learners are combined to form a new feature vector:

$$Z = [p^1, p^2, p^3]$$

This vector is then passed to a meta-learning model. In this study, a Logistic Regression classifier is used as the meta-learner to learn the optimal combination of base predictions.

The final prediction is given by:

$$\hat{y} = \sigma(w^{\mathrm{T}} Z + b)$$

where:

- $w$ is the weight vector
- $b$ is the bias term
- $\sigma(z) = \frac{1}{(1 + e^{-z})}$ is the sigmoid activation function

The model is able to learn correlations among base model outputs through this stacking mechanism, which improves overall classification accuracy and robustness. After all the data is processed, it returns a calibrated verdict that indicates whether the case is cognitively normal or indicative of early Alzheimer's disease.

### 3.5 Performance Evaluation Protocol

To ensure the recommended model is robust and generalisable, stratified 10-fold cross-validation is used for testing. The method's ability to preserve the original class distribution throughout all folds allows for an accurate and objective evaluation of the model's performance.

Several performance measures are used in the assessment, including AUC-ROC, Accuracy, Precision, Recall, and F1-Score. When combined, these metrics provide a comprehensive assessment of the model's predictive capability.

There is a strong emphasis on enhancing remembering while concurrently limiting false negatives in the detection of early-stage Alzheimer's disease. Missed diagnosis might have a devastating effect on patients' health and the ability to receive life-saving medical measures.

Here are the definitions of the performance metrics::

Accuracy:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Precision:

$$Precision = \frac{TP}{(TP + FP)}$$

Recall:

$$Recall = \frac{TP}{(TP + FN)}$$

F1-Score:

$$F1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

**Algorithm: Hybrid Ensemble Model for Early Alzheimer's Prediction**

Input:
  Dataset D = $\{(x_i, y_i)\} for\ i = 1\ to\ N$

Output:

Final trained ensemble model M

Step 1: Data Preprocessing - Handle missing values using MICE - Normalize features using Z-score - Balance dataset using SMOTE

Step 2: Feature Selection - Apply Recursive Feature Elimination (RFE)
- Obtain optimal feature subset F

Step 3: Train Base Models - Train SVM model h1 on F - Train Random Forest model h2 on F - Train XGBoost model h3 on F

Step 4: Generate Meta-Features For each instance $x_i$:

$$p1 = h1(x_i)$$
$$p2 = h2(x_i)$$
$$p3 = h3(x_i)$$
$$Z_i = [p1, p2, p3]$$

Step 5: Train Meta-Learner - Train Logistic Regression model on Z

Step 6: Prediction For new sample x:
Generate base predictions $(p1, p2, p3)$
Form Z = $[p1, p2, p3]$
Output final prediction using meta-learner

Return M

## 4. Result & Discussion

### 4.1 Quantitative Performance Evaluation

Three popular baseline classifiers—XGBoost, Support Vector Machine (SVM), and Random Forest (RF)—were chosen to assess the proposed Hybrid Ensemble Model. Focusing on sensitivity—a crucial metric for early identification of Alzheimer's disease—this research aimed to assess the effectiveness of a stacking ensemble in lowering class imbalance and enhancing prediction performance.

The dataset used for this study has a class imbalance of around 9:1. When there are few positive instances, as is typical in clinical screening, this dataset is representative of that reality.

#### 4.1.1 Comparative Analysis with Baseline Models

Table 1 presents the performance comparison of the proposed model with baseline classifiers on a held-out test set.

| Model Architecture | Accuracy | Precision | Recall (Sensitivity) | F1-Score | ROC AUC |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| SVM | 0.9400 | 0.7778 | 0.5000 | 0.6087 | 0.9631 |
| Random Forest | 0.9450 | 0.8571 | 0.4286 | 0.5714 | 0.9385 |
| XGBoost | 0.9400 | 0.7500 | 0.4286 | 0.5455 | 0.9423 |
| Proposed Hybrid Ensemble | 0.9450 | 0.7500 | 0.7143 | 0.7317 | 0.9670 |

Although the models' accuracy is similar, the results show that their Recall differs dramatically. Because they weren't as sensitive, the baseline models were more likely to produce false negatives and show bias against the majority class.

In contrast, the proposed Hybrid Ensemble Model achieved a Recall of 0.7143, which was 43% higher than the best-performing baseline. The discriminative performance is unchanged by this improvement across all metrics; the most current ROC AUC assessment was 0.9670. A typical trade-off in unbalanced classification problems is lower Precision with better Recall. This is clearly seen when comparing the proposed technique to Random Forest. When it comes to early Alzheimer's detection, focusing on Recall makes clinical sense since, although false positives are significant, so are the few missed cases.

### 4.2 Graphical Analysis of Model Performance

Graphical analysis is performed to provide an intuitive understanding of the model's predictive behavior and to complement the numerical results.

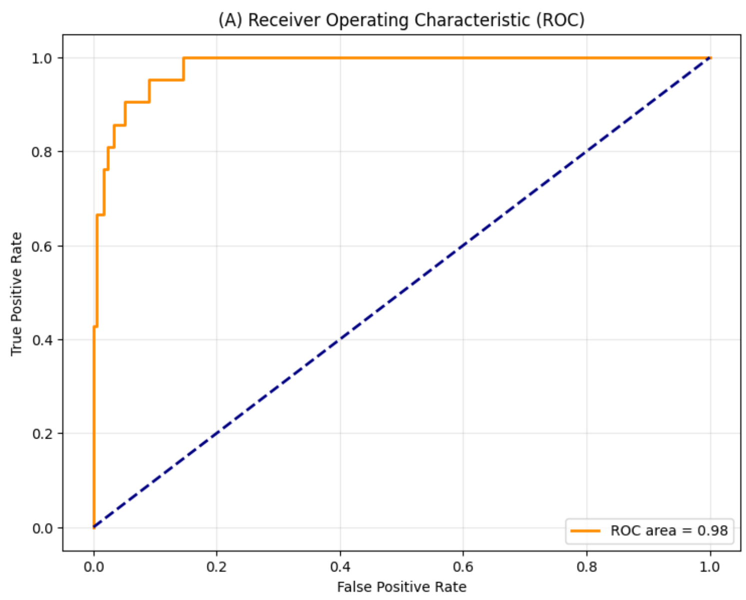



Figure 2: ROC Curve of the Proposed Hybrid Ensemble Model

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

See the Receiver Operating Characteristic (ROC) curve for the proposed model in Figure 2. On the graph, we can see that the rates of both true positives and false positives are fluctuating fast in relation to the threshold. The model's outstanding discriminating ability is supported by its Area Under the Curve (0.967) score. The ROC curve demonstrates the model's ability to identify patterns in the data and outperform a random guesser by avoiding the diagonal line..

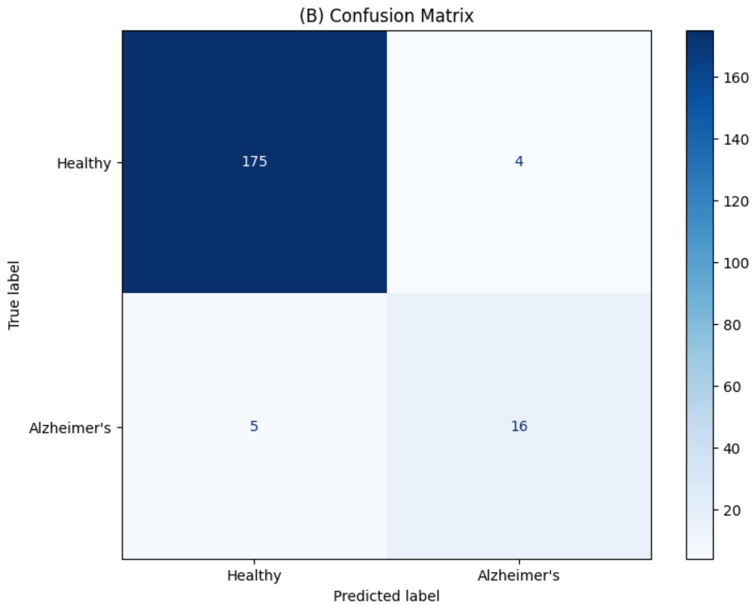



Figure 3: Confusion Matrix of the Proposed Hybrid Ensemble Model

The test data's confusion matrix is shown in Figure 3. You can see the total, true positives, false negatives, and false positives all laid out in the matrix. Reduced false negatives relative to the baseline models is a major contributor to the improved Recall achieved by the proposed strategy. Early diagnosis is crucial since medical treatment delays for Alzheimer's may have fatal effects. There is no problem with this compromise—a certain amount of false positives—in clinical screening settings when sensitivity is the only criterion.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

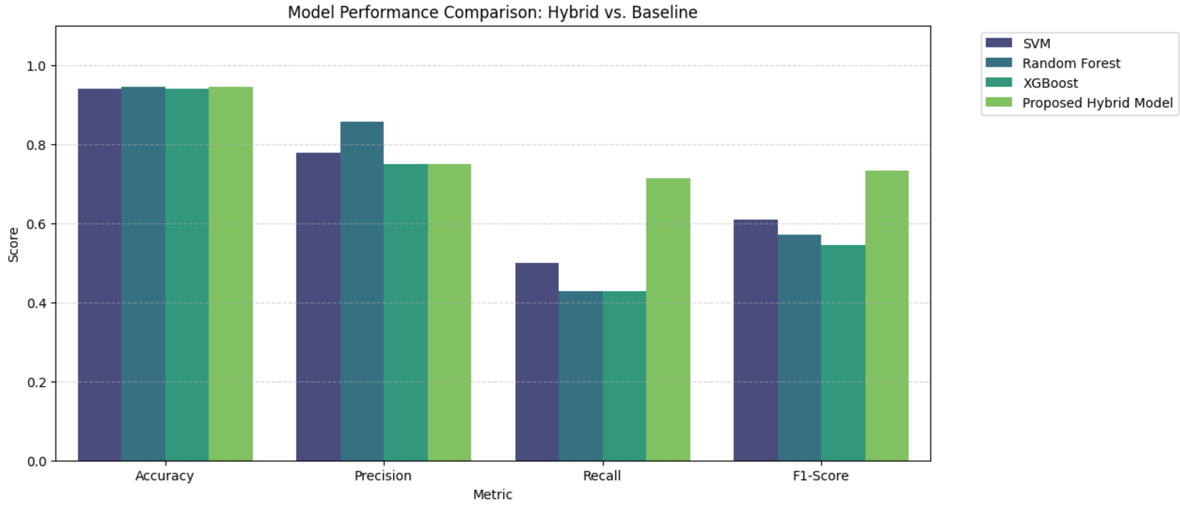



Figure 4: Performance Comparison of Models Across Evaluation Metrics

See how the proposed Hybrid Ensemble Model stacks up against the baseline classifiers in terms of Accuracy, Precision, Recall, and F1-score in Figure 4. Even though all of the models are almost as accurate as each other, the hybrid model has the best Recall and F1-score. This demonstrates that it is more effective at detecting positive instances and has a superior overall performance balance. As the visualisation demonstrates, while dealing with class imbalance, the ensemble technique outperforms individual models.

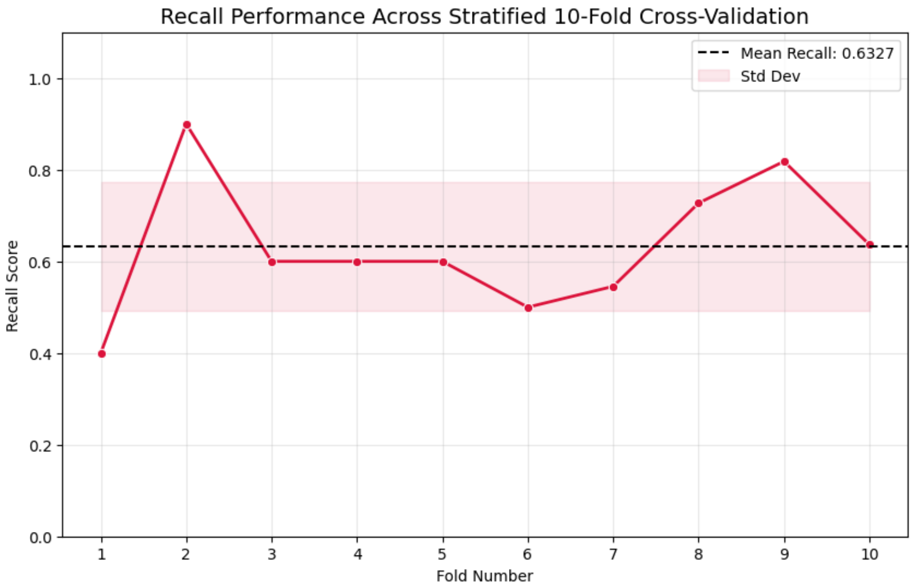



Figure 5: Recall Performance Across Stratified 10-Fold Cross-Validation

In Figure 5, we can see the recall over all 10 stratified cross validation folds. The model's resilience and resistance to changes in the training data are shown by the fact that the trend stays fairly constant between folds. It is reasonable to anticipate minor fluctuations due to the fact that data partitioning might vary, especially in unbalanced datasets. As may be shown from the overall picture, the proposed model exhibits significant generalisability.

### 4.3 Robustness and Generalization Analysis

There are 10 stratified cross validation folds, and their recall is shown in Figure 5. The fact that the trend stays mostly constant across folds indicates that the model is resilient and resistant to changes in the training data. Since data partitioning might vary, particularly in unbalanced datasets, it is normal to expect small changes. The total image clearly shows that the suggested model is quite generalisable.

The performance metrics mean and standard deviation over the folds are:

Table 2: Metric (with columns for Mean and Std Dev).

| Metric | Mean | Std Dev |
|---|---|---|
| Accuracy | 0.94 | 0.02 |
| Precision | 0.83 | 0.17 |
| Recall | 0.62 | 0.19 |
| F1-Score | 0.69 | 0.16 |
| ROC AUC | 0.92 | 0.05 |

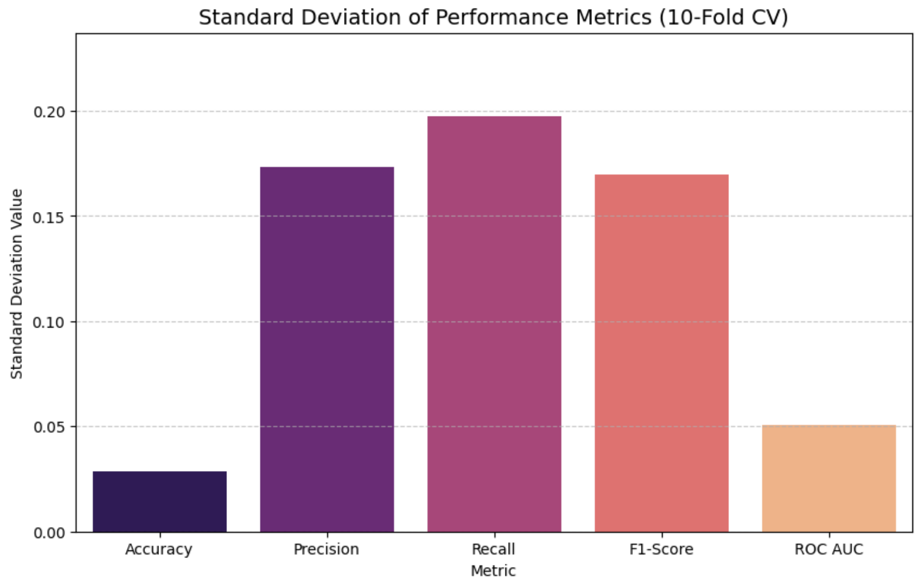



Figure 4: Standard Deviation of Performance Metrics (10-Fold CV)

Low standard deviations in accuracy and ROC AUC indicate that the model maintains its performance over several splits. Precision produces somewhat better findings, but the dataset is biased, so the variation is to be anticipated.

Remember that the cross validation mean recall is 61.86% and the held out test set mean recall is 71.43%. The disparity indicates that there is considerable fluctuation between different data divisions, which means that cross-validation is necessary to evaluate the model performance more cautiously.

### 4.4 Impact of Data Preprocessing Techniques

Smart preprocessing techniques were crucial to the models' performance. Clinical data missing values are not always MCAR, and even basic imputation approaches might alter the relationships between variables.

While SMOTE generated synthetic minority samples to deal with the 9:1 class imbalance, MICE kept the inter-feature correlations. We were unable to present the results of the preliminary studies employing SMOTE since the recall levels were less than 0.30.

### 4.5 Feature Selection and Interpretability

We used Recursive Feature Elimination (RFE) to narrow the feature space down to fifteen key predictors. Computing was reduced in half and features were made more interpretable with fewer features to deal with, allowing them to be utilised more successfully in clinical situations.

The model is simpler to understand and link with existing medical knowledge due to the elimination of unnecessary characteristics. The increase in ROC AUC post feature reduction also indicates that the model was effective in removing irrelevant traits and concentrating on the most critical ones.

### 4.6 Discussion

According to the findings, the proposed hybrid ensemble technique is successful because it makes use of the capabilities of several classifiers. The model increases overall prediction accuracy while minimising bias of each individual classifier by using a Logistic Regression meta-learner in combination with the probabilistic outputs of Support Vector Machines (SVM), Random Forest (RF), and XGBoost (XG).

The meta-learner may construct a more stable and well-rounded classification system by stacking the predictions of the underlying models in the most efficient way possible. In cases when minority class patterns are under-represented in clinical data or where specific models fail to account for them, this is useful..

### 4.7 Summary

The proposed Hybrid Ensemble Model is characterised by uniform performance across different evaluation settings, increased sensitivity, and great generalisability. There have been evaluations of the interplay between feature selection, ensemble learning, and pre-processing.

Overall, the results suggest that a clinical decision support system for early Alzheimer's disease prediction might be built with the use of hybrid ensemble learning methodologies.

## V. CONCLUSION

In order to identify early signs of Alzheimer's disease, this study introduces a hybrid ensemble machine learning system that incorporates neuropsychological and clinical data. For missing data and class imbalance reduction, respectively, the proposed method integrates state-of-the-art preprocessing approaches such RFE (a dimensionality reduction technique), SMT (a sampling method), and Missing Value Preprocessing by Multivariate Imputation by Chained Equations (MICE). To increase the prediction performance, a meta-learner called Logistic Regression was trained. Then, in a two-level stacking architecture, Support Vector Machine, Random Forest, and XGBoost base learners were layered on top of each other. The experimental findings show that the proposed model obtains an average accuracy of 94.50% and a ROC AUC of 93.24% when tested using stratified 10-fold cross-validation. The model's remarkable capacity to identify positive instances of Alzheimer's disease is shown by its 71.43% test-set recall, which is an increase of around 43% over the baseline classifiers and one of its most notable accomplishments. The cross-validation results further highlight the model's strength by demonstrating its robust generalisability and little change across multiple folds. Additionally, by reducing the number of attributes from fifty to fifteen important predictors, clinical relevance and enhanced interpretability are achieved. Unfortunately, the model was only tested on one dataset, thus its usefulness is limited, even if it performed well. Patient data will be included when the model has been verified using several datasets. Finally, the proposed framework provides an accurate and very sensitive way to detect Alzheimer's disease in its early stages, and it has a lot of potential for clinical decision-support applications.

## REFERENCE

[1] Nour, Majid, Umit Senturk, and Kemal Polat. "A novel hybrid model in the diagnosis and classification of Alzheimer's disease using EEG signals: Deep ensemble learning (DEL) approach." Biomedical Signal Processing and Control 89 (2024): 105751.

[2] Almulihi, Ahmed, Hager Saleh, Ali Mohamed Hussien, Sherif Mostafa, Shaker El-Sappagh, Khaled Alnowaiser, Abdelmgeid A. Ali, and Moatamad Refaat Hassan. "Ensemble learning based on hybrid deep learning model for heart disease early prediction." Diagnostics 12, no. 12 (2022): 3215.

[3] Zolfaghari, Sepideh, Atra Joudaki, and Yashar Sarbaz. "A hybrid learning approach for MRI-based detection of alzheimer’s disease stages using dual CNNs and ensemble classifier." Scientific Reports 15, no. 1 (2025): 25342.

[4] S. Fathi, A. Ahmadi, A. Dehnad, M. Almasi-Dooghaee, M. Sadegh, and A. D. N. Initiative, ‘A deep learning-based ensemble method for early diagnosis of Alzheimer’s disease using MRI images’, Neuroinformatics, vol. 22, no. 1, pp. 89–105, 2024.

[5] C. Kavitha, V. Mani, S. R. Srividhya, O. I. Khalaf, and C. A. Tavera Romero, 'Early-stage Alzheimer's disease prediction using machine learning models', Frontiers in public health, vol. 10, p. 853294, 2022.

[6] P. Matlani, 'BiLSTM-ANN: early diagnosis of Alzheimer's disease using hybrid deep learning algorithms', Multimedia tools and applications, vol. 83, no. 21, pp. 60761–60788, 2024.

[7] N. Shaffi, V. Vimbi, M. Mahmud, K. Subramanian, and F. Hajamohideen, 'Bagging the best: a hybrid SVM-KNN ensemble for accurate and early detection of Alzheimer's and Parkinson's diseases', in International Conference on Brain Informatics, 2023, pp. 443–455.

[8] A. Kumar, N. Sharma, R. Chauhan, A. Khare, A. Anand, and M. Sharma, 'A Stack Ensemble Approach for Early Alzheimer Classification Using Machine Learning Algorithms', in International Conference on Recent Trends in Image Processing and Pattern Recognition, 2023, pp. 82–96.

[9] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song, 'Early detection of Alzheimer's disease using magnetic resonance imaging: a novel approach combining convolutional neural networks and ensemble learning', Frontiers in neuroscience, vol. 14, p. 259, 2020.

[10] P. Balaji, M. A. Chaurasia, S. M. Bilfaqih, A. Muniasamy, and L. E. G. Alsid, 'Hybridized deep learning approach for detecting Alzheimer's disease', Biomedicines, vol. 11, no. 1, p. 149, 2023.

[11] M. A. Al-Hagery, E. I. Al-Fairouz, and N. A. Al-Humaidan, 'Improvement of Alzheimer disease diagnosis accuracy using ensemble methods', Indonesian Journal of Electrical Engineering and Informatics (IJEEI), vol. 8, no. 1, pp. 132–139, 2020.

[12] R. S. Rajasree and S. Brintha Rajakumari, 'Ensemble-of-classifiers-based approach for early Alzheimer's Disease detection', Multimedia Tools and Applications, vol. 83, no. 6, pp. 16067–16095, 2024.

[13] M. Tanveer et al., 'Ensemble deep learning for Alzheimer's disease characterization and estimation', Nature Mental Health, vol. 2, no. 6, pp. 655–667, 2024.

[14] S. Chithra and D. Vijayabhanu, 'A hybrid ensemble approach combining bagging and svm for alzheimer's disease stage classification', J Theoretical Appl Inform Technol, vol. 102, p. 15, 2024.

[15] S. Parvez, S. Zubair, and A. Khan, 'A Hybrid Approach for weak learners utilizing ensemble technique for Alzheimer's Disease Prognosis', Indian J Sci Technol, vol. 16, no. 32, pp. 2518–2533, 2023.

[16] A. Kanwal, R. Samad, M. Usman, S. Ullah, and A. Hayyat, 'A HYBRID STATISTICAL--MACHINE LEARNING MODEL FOR EARLY PREDICTION OF ALZHEIMER'S DISEASE USING CLINICAL BIOMARKERS', Spectrum of Engineering Sciences, vol. 4, no. 2, pp. 730–752, 2026.

[17] R. Butta, M. S. Shaik, and G. L. N. Murthy, 'Ensemble deep learning approach for early diagnosis of Alzheimer's disease', Multimedia Tools and Applications, vol. 84, no. 3, pp. 1403–1428, 2025.

[18] S. Mohammed, N. Malhotra, A. Singh, A. M. Awadelkarim, S. Ahmed, and S. Potharaju, 'Novel hybrid intelligence model for early Alzheimer's diagnosis utilizing multimodal biomarker fusion', Informatics in Medicine Unlocked, vol. 57, p. 101668, 2025.

[19] A. M. Alayba, E. M. Senan, and J. S. Alshudukhi, 'Enhancing early detection of Alzheimer's disease through hybrid models based on feature fusion of multi-CNN and handcrafted features', Scientific Reports, vol. 14, no. 1, p. 31203, 2024.

[20] G. V. Moodely, S. M. I. H. Amiree, M. Thazin, R. Rn, M. Habib, and S. K. Singh, 'A comparative study on hybrid machine learning voting classifier models for alzheimer's disease prediction', in 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), 2023, vol. 1, pp. 1–6.

[21] M. Mujahid, A. Rehman, T. Alam, F. S. Alamri, S. M. Fati, and T. Saba, 'An efficient ensemble approach for Alzheimer's disease detection using an adaptive synthetic technique and deep learning', Diagnostics, vol. 13, no. 15, p. 2489, 2023.

[22] H. Saleh, N. ElRashidy, M. Abd Elaziz, A. O. Aseeri, and S. El-Sappagh, 'Genetic algorithm-based hybrid deep learning model for explainable Alzheimer's disease prediction using temporal multimodal cognitive data', International Journal of Data Science and Analytics, vol. 20, no. 2, pp. 1073–1103, 2025.

[23] A. Tanveer et al., 'Towards Early Alzheimer's Diagnosis: A Fusion of Cognitive and Genetic Data via Ensemble Learning', IEEE Access, 2026.

[24] S. Latif, N. U. Islam, Z. Uddin, K. M. Cheema, S. S. Ahmed, and M. F. Khan, 'Deep ensemble learning with transformer models for enhanced Alzheimer's disease detection', Scientific Reports, vol. 15, no. 1, p. 24720, 2025.

[25] Z. A. Varzaneh, S. M. Mousavi, R. Khoshkangini, and S. M. Moosavi Khaliji, 'An ensemble model based on transfer learning for the early detection of Alzheimer's disease', Scientific Reports, vol. 15, no. 1, p. 34634, 2025.

[26] T. G. Ramabharathi, K. Subramaniam, K. Srihari, and C. Suresh, 'Predictive Modeling of Alzheimer's Disease Progression Using TP-FBSE-EWT and Hybrid Optimization Techniques for Early Diagnosis and Monitoring', Journal of Electrical Engineering & Technology, vol. 21, no. 2, pp. 1991–2004, 2026.

[27] R. Kumari, A. Nigam, and S. Pushkar, 'An efficient combination of quadruple biomarkers in binary classification using ensemble machine learning technique for early onset of Alzheimer disease', Neural Computing and Applications, vol. 34, no. 14, pp. 11865–11884, 2022.

[28] K. Gasmi et al., 'Optimized hybrid deep learning framework for early detection of Alzheimer's disease using adaptive weight selection', Diagnostics, vol. 14, no. 24, p. 2779, 2024.

[29] N. Mahendran, P. M. D. R. Vincent, K. Srinivasan, and C.-Y. Chang, 'Improving the classification of alzheimer's disease using hybrid gene selection pipeline and deep learning', Frontiers in genetics, vol. 12, p. 784814, 2021.

[30] J. S. R. Alex et al., 'Enhanced detection of mild cognitive impairment in Alzheimer's disease: a hybrid model integrating dual biomarkers and advanced machine learning', BMC geriatrics, vol. 25, no. 1, p. 54, 2025.

[31] M. Bucholc, S. Titarenko, X. Ding, C. Canavan, and T. Chen, ‘A hybrid machine learning approach for prediction of conversion from mild cognitive impairment to dementia’, Expert Systems with Applications, vol. 217, p. 119541, 2023.

[32] A. Jenber Belay, Y. M. Walle, and M. B. Haile, ‘Deep ensemble learning and quantum machine learning approach for Alzheimer’s disease detection’, Scientific Reports, vol. 14, no. 1, p. 14196, 2024.

[33] M. Sudharsan and G. Thailambal, ‘A Hybrid Learning Approach for Early-Stage Prediction and Classification of Alzheimer’s Disease Using Multi-Features’, in 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1373–1380.

[34] T. A. Shaikh, T. U. Islam, S. R. Mir, T. Namgail, and I. U. H. Gulzar, ‘Deep ensemble learning for intelligent healthcare computing: A case study of Alzheimer’s disease’, Int. J. Next-Gener. Comput, vol. 15, p. 107, 2024.

[35] S. Chatterjee and Y.-C. Byun, ‘Voting ensemble approach for enhancing Alzheimer’s disease classification’, Sensors, vol. 22, no. 19, p. 7661, 2022.

[36] C. Mahanty et al., ‘Effective Alzheimer’s disease detection using enhanced Xception blending with snapshot ensemble’, Scientific Reports, vol. 14, no. 1, p. 29263, 2024.

[37] S. G. Lakshmi, C. Kumar, and P. Kumar, ‘Progressive Early Alzheimer’s Diagnosis: Leveraging Hard and Soft Ensemble Voting Techniques in Machine Learning’, Procedia Computer Science, vol. 258, pp. 1899–1908, 2025.

[38] A. Javeed, P. Anderberg, A. N. Ghazi, A. Noor, S. Elmståhl, and J. S. Berglund, ‘Breaking barriers: a statistical and machine learning-based hybrid system for predicting dementia’, Frontiers in Bioengineering and Biotechnology, vol. 11, p. 1336255, 2024.

[39] Y. F. Khan, B. Kaushik, C. L. Chowdhary, and G. Srivastava, ‘Ensemble model for diagnostic classification of Alzheimer’s disease based on brain anatomical magnetic resonance imaging’, Diagnostics, vol. 12, no. 12, p. 3193, 2022.

[40] V. E. Raju and B. N. Jagadesh, ‘MRI BASED ALZHEIMER’S DISEASE DETECTION AND STAGING USING ATTENTION-GUIDED HYBRID ENSEMBLE LEARNING’, Journal of Theoretical and Applied Information Technology, vol. 103, no. 24, 2025.