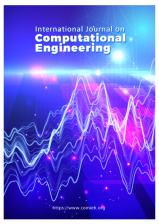
comien

International Journal on Computational Engineering

Journal Homepage www.comien.org/index.php/comien



AI-Based Doctor Agent for Early Detection of Alzheimer's Applying Hybrid Machine-Learning Techniques

Debasish Saha Roy [a], Suhodip Koley [b,*], Pronay Pal [c]

[a] *Computer Sc. & Engg., JIS College of Engineering, Kalyani, Nadia, India*

[b] *Computer Application, JIS College of Engineering, Kalyani, Nadia, India*

[c] *Information Technology, JIS College of Engineering, Kalyani, Nadia, India*

Corresponding author: [*]subhodip.koley@jiscollege.ac.in

***Abstract*— Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by memory loss, cognitive decline, and irreversible brain death, rendering it a global health concern. Even though early detection using machine learning (ML) has gotten better, there are still problems with expensive neuroimaging, limited multimodal data integration, small homogeneous cohorts, and models that are hard to understand. This study proposes an AI-driven "doctor" agent that forecasts Alzheimer's disease state, progression risk, and personalized therapy recommendations, grounded in standard clinical and cognitive indicators. We used Mutual Information-based feature selection and an 80/20 stratified methodology to test seven hybrid ensemble models on 2,149 Kaggle patients. The second hybrid model had an accuracy of 95.81% and an AUC of 0.9492, which was better than both classical and deep-learning baselines. All hybrid models did better than individual learners, showing that they could work well with a wide range of clinical data. With alzheime.pkl, it was possible to make real-time predictions with confidence ratings and comments on diagnoses that were like those made by doctors. Real-time diagnostic results that can be understood, along with confidence ratings and automated "doctor's notes," help research move from the lab to the clinic. Multimodal fusion, external validation, calibrated probability calculation, and explainable decision outputs make Alzheimer's screening fairer, faster, and clearer. The study shortens the time it takes to make a diagnosis and the amount of work that doctors must do, which is good for healthcare systems. Longitudinal, multi-omics, and wearable sensor data will be combined with privacy-preserving federated learning and explainable-AI dashboards to help doctors understand things better and use them in the real world.**

***Keywords*—Alzheimer's disease prediction; hybrid machine-learning ensemble; soft-voting classifier; AI-Driven doctor agent system.**

Manuscript received 6 Apr. 2025; revised 25 Jul. 2025; accepted 19 Aug. 2025. Date of publication 30 Sep. 2025.

I. INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by the accumulation of amyloid-β plaques and tau tangles, leading to the degeneration of cholinergic neurons, especially in the hippocampus. This degeneration causes memory problems, confusion, and a decline in thinking skills. The pathophysiology encompasses synaptic dysfunction, neuroinflammation, oxidative stress, and vascular abnormalities. Alzheimer's disease (AD) accounts for 60–70% of the 57 million [1] dementia cases worldwide, affecting 34–40 million [2] individuals in 2021, with prevalence continuing to rise. Alzheimer's disease and other dementias kill about 1.8 million [3] people each year, making them the seventh leading cause of death in the world. Although incurable, early detection and timely interventions—including pharmacological treatment, cognitive exercises, and lifestyle modifications—can significantly slow disease progression and improve quality of life. This study is driven by the imperative to improve early detection and clinical management of Alzheimer's disease, a leading cause of dementia worldwide. Conventional diagnostic methods heavily depend on costly imaging technologies, specialized analysis, and the protracted recognition of symptoms, often resulting in delayed interventions and irreversible brain damage. This study aims to develop a cost-efficient, AI-driven diagnostic instrument that forecasts the risk of Alzheimer's disease using conventional clinical data. This method aims to make screening faster, easier to get to, and easier to understand, which will reduce diagnostic delays and improve timely, personalized patient care.

The literature review shows that there have been significant improvements in the use of machine learning and deep learning methods to find Alzheimer's disease using MRI, PET, genetic, and cognitive data. However, most studies use only one type of input, small and similar datasets, and don't have external validation, which makes it hard to generalize their results. Many models function as "black boxes," offering little clinical interpretability, and only a few are appropriate for application in real-world healthcare settings. Inconsistent preprocessing, lack of probability calibration, and poor integration of multimodal features are other problems that make it hard to get reliable, clear, and scalable early diagnosis of Alzheimer's disease.

The goal of the method is to create an AI-based "doctor" agent that can diagnose Alzheimer's disease and predict its risk. We got a Kaggle dataset [4] with 2,149 patient records that had 35 clinical and demographic features. We first preprocessed the data by dealing with missing values, getting rid of columns that weren't needed, and normalizing features. Then we used histograms, bar charts, and pie charts to show the data. We used Mutual Information (MI) to choose the features, and we split the dataset into 80% training and 20% testing sets. We used the AUC-ROC curve to compare the performance of seven hybrid models that combine ANN, RF, LR, SVM, GBC, KNN, GNB, XGB, SGB, CART, LightGBM, XGBoost, CatBoost, and SGD. These models are called first, second, third, fourth, fifth, sixth, and seventh.

The results section looks at how well seven hybrid machine-learning ensembles did at predicting Alzheimer's disease. The second hybrid did better than any other single or ensemble baseline, with an accuracy of 95.81% and an AUC of 0.9492. More models showed significant results, which proved that the hybrid method works. The final model was saved as an alzheime.pkl file so that it could be used for real-time diagnosis. It made reports that could be understood and included confidence scores.

The trained hybrid model for real-time prediction uses information from patients, such as their demographics, lifestyle, and cognitive scores. It uses standardized preprocessing to turn these inputs into outputs that show Alzheimer's status, confidence scores, and autogenerated "doctor's notes." This report gives personalized suggestions for early care. The model fixes problems with earlier research by making multimodal fusion easier, giving clear decision outputs, and allowing for deployable inference. This links experimental AI research with real-world clinical use.

A. Literature Review

"Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models" uses good ways to sort data to find early signs of Alzheimer's. The Decision Tree, Random Forest, Support Vector Machine, XGBoost, and Voting classifiers are tested on the OASIS and Kaggle MRI datasets, which have 150 patient samples (ages 60 to 96). After preprocessing and selecting features using correlation, information gain, and chi-square methods, Random Forest had the highest accuracy (86.92%), followed by XGBoost (85.92%). The study's novel methodology for early-stage Alzheimer's diagnosis employs multiple machine learning models and optimal feature selection to enhance predictive accuracy and clinical relevance [3]. The study "Alzheimer's Disease Prediction Using Machine Learning Techniques and Principal Component Analysis (PCA)" utilizes structural MRI (sMRI) data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), consisting of 214 individuals aged 65 to 96, to enhance early diagnosis of Alzheimer's disease. After selecting features based on PCA, it uses SVM, IVM, and RELM to lower the number of dimensions. Research shows that RELM did better than SVM and IVM in both binary and multiclass classification, with an accuracy of about 81.03%. Combining PCA for reducing dimensions with RELM for fast, accurate, and less expensive Alzheimer's detection works well for small datasets with high-dimensional sMRI features [5]. “A Machine Learning Model for Alzheimer’s Disease Prediction” makes it easier to find Alzheimer's disease early. The Decision Tree (DT), Extreme Gradient Boosting (XGB), and Random Forest (RF) studies used the Kaggle Open Access Series of Imaging Studies (OASIS) dataset. To make the classes equal, SMOTE was used. On the balanced dataset, SMOTE-RF was 94.03% accurate, which was better than other methods. Combining data balancing with ensemble learning makes real-world clinical prediction more sensitive and more reliable for classification [6]. "Alzheimer's Disease Prediction Using Three Machine Learning Methods" integrates gene expression data from GEO datasets GSE63060 and GSE63061, totaling 569 samples (16,382 genes), to forecast early Alzheimer's disease. We used Information Gain to choose features and SVM, NB, and K-Nearest Neighbors to group 44 genes. With an accuracy of 96.6%, SVM did better than other models. It is new to use feature-selection-driven gene expression analysis, dimensionality reduction, and blood-based biomarkers to make early-stage AD predictions more accurate [7]. The focus of "Machine Learning–Based Multi-Modal Prediction of Future Decline Toward Alzheimer's Disease: An Empirical Study" is to find high-risk AD patients early. It uses clinical, cognitive, CSF, MRI, and PET data from 1,404 ADNI subjects to predict the change from CN to MCI and from MCI to AD. Author made three deep learning models: LSM, NSM, and NMM. The NMM model had the highest AUC of 0.90 and the best long-term accuracy for CN-to-MCI. The research is unique because it uses missing data, multiple time horizons, and multimodal fusion to make strong, flexible predictions about AD risk and improve clinical forecasting [8]. An Explainable and Efficient Deep Learning Framework for EEG-Based Diagnosis of Alzheimer's Disease and Frontotemporal Dementia meets the need for a clear and correct diagnosis of dementia. Author used a hybrid model that combines a Temporal Convolutional Network (TCN) and a Long Short-Term Memory (LSTM) model with Modified Relative Band Power (RBP) feature extraction on an EEG dataset of 88 people (36 with AD, 29 healthy, and 23 with FTD) from AHEPA University Hospital. The framework's 99.7% binary and 80.34% multi-class classification accuracy beat all the other models that were already available. Lightweight deep learning and SHAP-based explainable AI enhance model interpretability, diagnostic precision, and real-time clinical utilization [9]. The article, "Machine Learning Approaches for Predicting Progression to Alzheimer's Disease in Patients with Mild Cognitive Impairment," utilizes the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset comprising 819 subjects, including 402 individuals

with mild cognitive impairment (MCI), to forecast the progression of Alzheimer's disease. SHAP, XGBoost importance, and Gini index were used to test Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Decision Tree, XGBoost, Gradient Boosting, and Artificial Neural Network. Shap + XGBoost + NEAR-MISS sampling had the highest accuracy (76.13%, AUC = 0.84), which was better than any other model before it. Interpretable SHAP-based feature selection and sample optimization enhance biomarker identification and predictive reliability for real-time clinical Alzheimer's disease progression forecasting [10]. Utilization of MRI scans from an e-Health clinical dataset to improve the early detection of Alzheimer's disease. We tried out the manual feature extraction (LBP), CV2-based feature vectors, and CNN plus SVM deep feature extraction methods. The best accuracy was 75% for SVM, Logistic Regression, Random Forest, and XGBoost. Comparisons of handmade, traditional ML, and deep-learning pipelines demonstrate improved diagnostic accuracy and plausible future applications [11]. The large NACC dataset (169,408 records, 1,024 characteristics) makes it easier to diagnose Alzheimer's disease early and in a way that makes sense. After reducing the number of features, this algorithm uses SVM, RF, KNN, and NB to sort NC, MCI, and AD. With SVM, the F1-scores went up to 88%. Combining explainable-AI rule extraction methods like CAR, SIRUS, SHAP, and LIME to help doctors make decisions is new [12]. The OASIS MRI dataset, which has 416 pictures, is used to make it easier to find Alzheimer's disease. It looks at how well SVM, KNN, Decision Tree, Random Forest, and Logistic Regression work for binary classification. The Random Forest model is the most accurate, with an accuracy rate of 94%. Ensemble-based ML surpasses traditional classifiers in the early diagnosis of Alzheimer's by offering a lightweight, clinically adaptable framework [13].

Despite notable progress across these ten studies, several key gaps remain. Most research relies on limited or homogeneous datasets (OASIS, ADNI, or small GEO cohorts), restricting generalizability and external validation. Imaging-based approaches face high computational costs and data scarcity, while gene-expression studies suffer from small sample sizes and class imbalance. Many models prioritize accuracy but lack interpretability or clinical explain ability. Few frameworks integrate longitudinal, multimodal, or real-time data, and cross-cohort standardization is minimal—hindering reproducibility, scalability, and clinical deployment of Alzheimer's predictive AI systems.

II. Material and Method

The objective of this paper is to design and implement an AI-driven specialist doctor agent for Alzheimer's disease, as illustrated in Figure 1, by evaluating seven hybrid machine-learning ensembles, selecting and serializing the top performer for real-time inference. The system will enable automatic diagnosis, future risk estimation, comprehensive patient report generation with clinician notes, and result interpretation, thereby accelerating diagnostic workflows, enhancing accuracy, supporting clinical decision-making, and ultimately improving patient management through an integrated end-to-end AI solution.

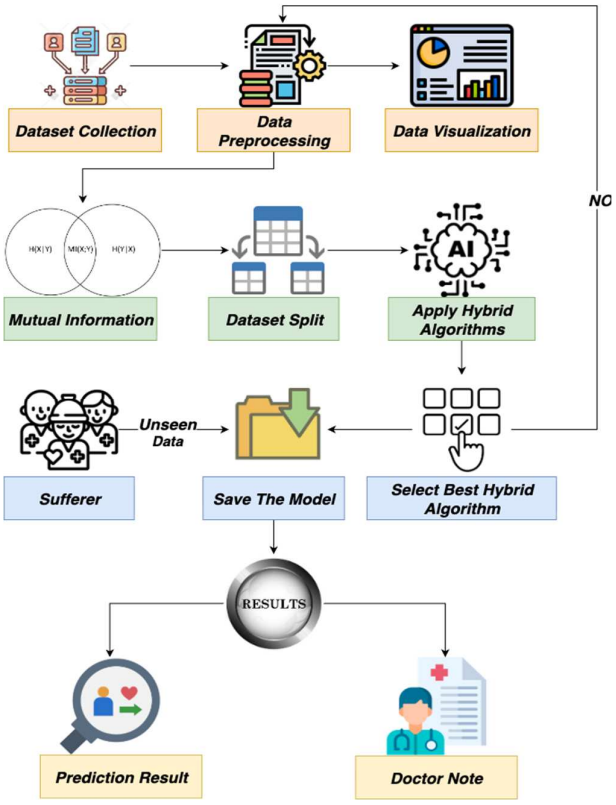


Fig. 1 Diagram of the full process flow

There are 2,149 records and 35 variables in the Kaggle Alzheimer's dataset [4]. Along with a binary label that reads "Diagnosis," these variables include demographics, lifestyle factors, medical history, clinical measures, and cognitive scores. We start by using data.info() to look at the schema and non-null counts. Next, we use data.isnull().sum() to determine the missing values for every column. Features with a lot of null values are eliminated. The mode is imputed into categorical fields, and the mean or median values are used to fill in the remaining numerical fields. To stop data leakage, unique identifiers and columns specific to a particular clinician have been removed. The key continuous variables are visualized using plt.hist(), the categorical distributions are visualized using plt.bar() or sns.countplot(), and the balance between the NC, MCI, and AD classes is evaluated using plt.pie(). Mutual_info_classif, a k-nearest-neighbors method appropriate for mixed data, is used to estimate Mutual Information during the feature selection process. Out of 33 features, 18 are kept using this method if their MI scores are equal to or higher than 0.002. For a fair and impartial assessment, we use an 80/20 stratified train-test split.

The first hybrid model incorporates Random Forest, Logistic Regression, and an Artificial Neural Network (ANN) in a soft-voting ensemble. Using the Keras Sequential model, the artificial neural network (ANN) is built with dense layers, ReLU activation functions, and a softmax output layer at the end. The RandomForestClassifier and LogisticRegression tools from Scikit-learn are used to create RF and LR. The fit() method is used to independently train each of the three models on the 80% stratified training set. The final class selection is determined during the inference phase by averaging and aggregating the predict_proba() function from each model. With a 94.88% accuracy rate and an AUC of roughly 0.9483 for Alzheimer's prediction, the complementary design improves robustness.

The second hybrid model uses a soft-voting technique that gives each component equal weights to integrate Support Vector Machine (SVM), Random Forest, and Gradient Boosting Classifier. SVC is used to implement SVM with kernel='rbf' and probability=True. RandomForestClassifier is used for Random Forest, and GradientBoostingClassifier is used for Gradient Boosting. The fit() function is used to independently train each model on the 80% stratified training dataset. The final prediction is generated during the inference phase by averaging the outputs from predict_proba(). This ensemble predicts Alzheimer's disease with an accuracy of 95.81% and an AUC of 0.9492 by utilizing the sequential error-correction of GBC, the robustness of RF, and the margin optimization capabilities of SVM.

The third hybrid model uses a soft-voting technique to combine KNN, ANN, and Gaussian Naïve Bayes. The KNeighborsClassifier() function is used to implement KNN. The compile() and fit() methods are used after the ANN is built using Keras Sequential, which includes dense layers with ReLU activations and a softmax output. The GaussianNB() class is used to instantiate GNB. The stratified dataset, which makes up 80% of the data, is used to train each model separately using the fit() function. Each of the three models uses the predict_proba() function to create probabilities during the prediction phase. These probabilities are then averaged to produce the final class outcome. With an accuracy of 85.81% and an AUC of 0.915, this system combines probabilistic modeling, deep nonlinear pattern extraction, and local similarity learning.

The fourth hybrid model uses a soft-voting ensemble method to combine Logistic Regression, XGBoost, and Stochastic Gradient Boosting. You can use LogisticRegression() to run Logistic Regression, XGBClassifier() to run XGBoost, and Scikit-learn's GradientBoostingClassifier() to run Stochastic Gradient Boosting (SGB). The fit() function is used to train each model on its own on the 80% stratified training dataset. In the inference phase, the final prediction is made by averaging the predict_proba() function from each of the three models. This combination brings together the easy-to-understand nature of logistic regression, the regularized tree learning of XGBoost, and the iterative error-correction of stochastic gradient boosting. It works very well, with an accuracy of 95.35% and an AUC of 0.954 when predicting Alzheimer's disease.

The fifth hybrid model combines CART, SVM, and Random Forest by using a soft-voting mechanism. The DecisionTreeClassifier() function is used by the CART model, the SVC(kernel='rbf', probability=True) function is used by the SVM model, and the RandomForestClassifier() function is used by the RF model. Each model undergoes independent training utilizing the fit() function on the stratified dataset comprising 80% of the data. During the prediction phase, all models use the predict_proba() function to give probability outputs. These outputs are then averaged to choose the final class. CART gives clear decision paths, SVM handles high-dimensional boundaries well, and RF makes things more stable by using bagging techniques. The complementary design can predict Alzheimer's disease with 94.42% accuracy and an AUC of 0.9471.

The sixth hybrid model uses a soft-voting ensemble approach to combine LightGBM, XGBoost, and CatBoost. The LGBMClassifier() function lets you use LightGBM, the XGBClassifier() function lets you use XGBoost, and the CatBoostClassifier() function lets you use CatBoost. Using the fit() function, each model gets its own training on the 80% stratified training set. During the inference phase, the three algorithms produce probability outputs using the predict_proba() function. These outputs are then averaged to establish the final class. LightGBM makes training fast and memory-efficient. XGBoost has strong regularization features. CatBoost handles categorical features well and deals with missing values. This group of models has a 95.35% accuracy rate and an AUC of 0.947 for predicting Alzheimer's disease.

The seventh hybrid model uses soft voting to combine Gaussian Naïve Bayes, Logistic Regression, and an SGD-based linear classifier. GaussianNB is created using GaussianNB(), Logistic Regression is established with LogisticRegression(), and the SGD model is defined with SGDClassifier(loss='log'). The fit() function trains each of the three models separately using the 80% stratified training set. The predict_proba() function gives each classifier a set of probability outputs during the inference phase. The final prediction is then the average of these outputs. The Gaussian Naive Bayes (GNB) model works well with small-sample probabilistic patterns, and the Logistic Regression (LR) model gives weights that can be understood. Stochastic

Gradient Descent (SGD) makes sure that the process can be scaled up and that it converges quickly. When used together, these methods can predict Alzheimer's disease with 83.95% accuracy and an AUC of 0.901.

III. Result and Discussion

The primary aim of this research project is to develop an artificial intelligence-driven "doctor" agent capable of early diagnosis of Alzheimer's disease, risk prediction, and automated clinical reporting. Seven hybrid machine-learning models were evaluated to ascertain the most reliable ensemble for deployment. Table 1 shows that the second hybrid model, which used SVM, RF, and GBC, had the highest accuracy of 95.81% and had good precision, recall, and F1-Score values. Fig. 3 shows the trends for accuracy and precision, and Fig. 2 shows the ROC–AUC curves. Fig. 2 shows that the fourth hybrid had the highest AUC of 0.954. Because of how well it worked overall, the second hybrid model was saved as alzheime.pkl so it could be used in real-time clinical settings.

TABLE I
Accuracy, Precision, Recall, and F1-Score comparison of every Hybrid Model

Sl.No	Model Name	Accuracy	Precision	Recall	F1-Score
1	First Hybrid Model	94.88%	94.12%	98.19%	96.11%
2	Second Hybrid Model	95.81%	95.76%	97.83%	96.79%
3	Third Hybrid Model	85.81%	89.42%	88.45%	88.93%
4	Fourth Hybrid Model	95.35%	95.09%	97.83%	96.44%
5	Fifth Hybrid Model	94.42%	94.08%	97.47%	95.74%
6	Sixth Hybrid Model	95.35%	95.09%	97.83%	96.44%
7	Seven Hybrid Model	83.95%	84.90%	91.34%	88.00%

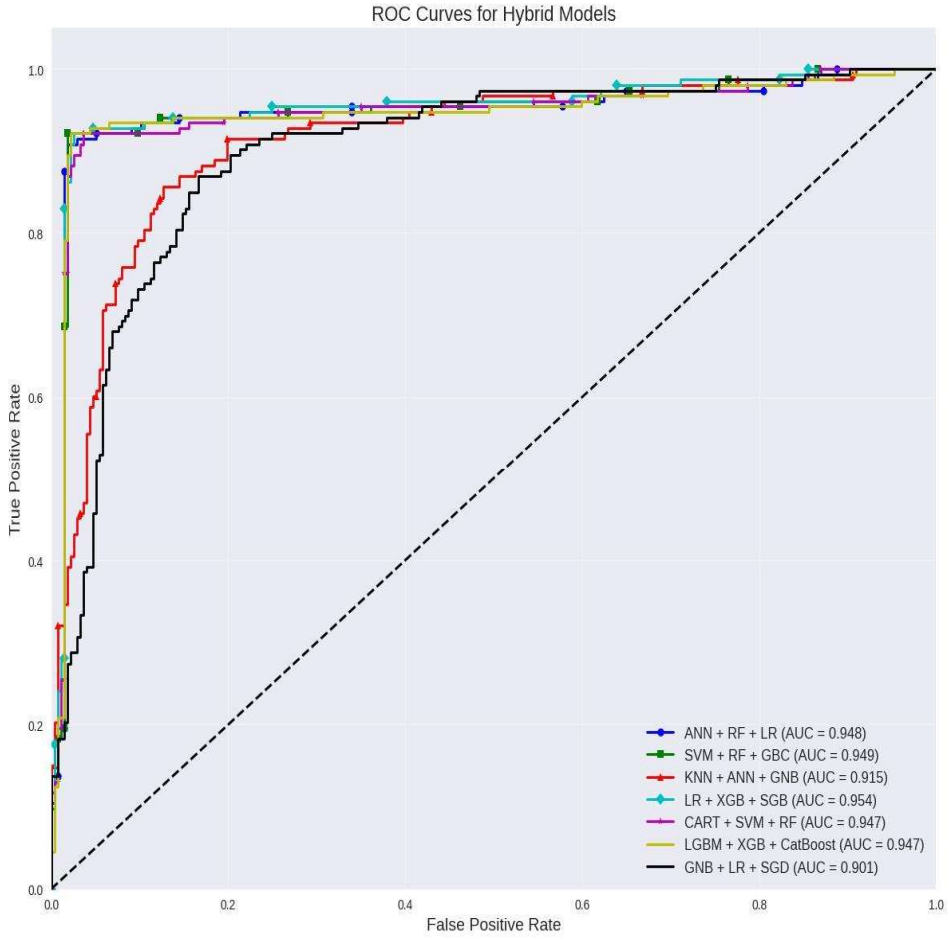


Fig. 2 The ROC–AUC curves

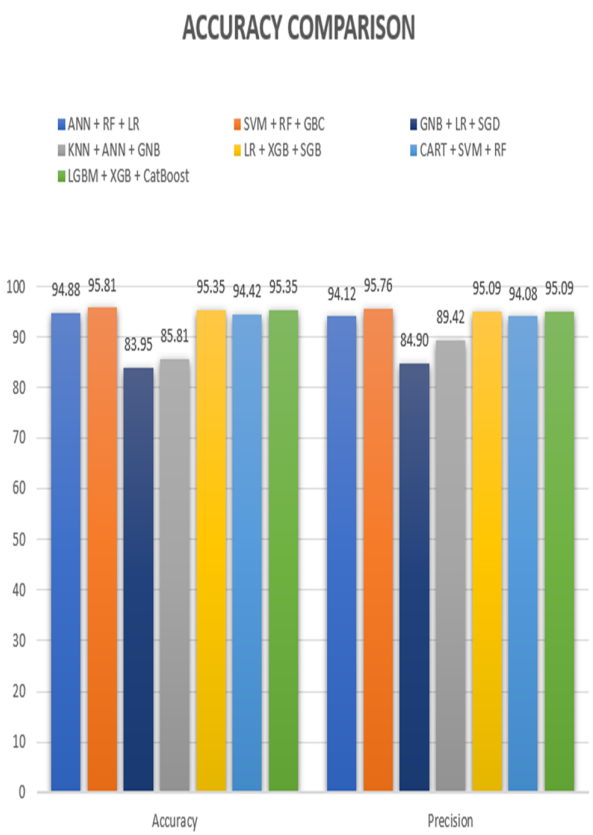


Fig. 3 Accuracy and precision Comparison of the Hybrid Model

The deployed alzheime.pkl model uses 18 clinical and demographic inputs, such as gender, education level, alcohol use, physical activity, sleep quality, family history of Alzheimer's, depression, head injury, diastolic blood pressure, cholesterol HDL, mini-mental state examination (MMSE), functional assessment, memory complaints, behavioral problems, activities of daily living (ADL), disorientation, personality changes, and difficulty completing tasks. The system uses these features to make a full Alzheimer's Diagnostic Report that includes the predicted class (0 or 1), the confidence level, the diagnosis stage, and an automated note from a doctor. A thorough probability analysis is provided to improve comprehension for both patients and healthcare professionals. Figures 4 and 5 show how the model sorted people into Alzheimer's Disease (AD) and non-AD groups. This is compared to expert medical assessments, which show a strong agreement with clinical judgment.

This model is recommended for clinical decision support, especially in outpatient memory clinics, community health centers, and telemedicine platforms to make it easier to screen for Alzheimer's disease early on. This method is good for doing a full risk assessment of a large population, talking to caregivers, and keeping an eye on things in digital health systems so that interventions can happen quickly.

The discussion is divided into six sections, including a comparison of real-time predictions from our model with doctor-verified diagnoses, an explanation of how the system fills in previous research gaps, a performance comparison with previous studies, a rationale for the choice of Mutual Information for feature selection, a reason for favoring hybrid models over single algorithms, and a brief summary of study limitations and novelty.

A. Doctor and our Model Comparison through real data:

1) Case Study 1:

The National Institute of Aging [14] provided a clinical profile for Alzheimer's without the disease, consisting of 18 characteristics: 1 for gender and 3 for education level. Drinking: 6 grams a day; Exercising: 7.7 hours a week; Quality of sleep: 9 out of 10. A family is a group of people who are related by blood, marriage, or adoption. They usually have shared responsibilities, support, and social ties. No history of Alzheimer's, depression, or head injury. The person's diastolic blood pressure was 71 mmHg, their high-density lipoprotein cholesterol was 43 mg/dL, and their mini-mental state examination score was 19. Functional Assessment: 6; Memory Complaints: 0; Behavioral Problems: 1; Activities of Daily Living (ADL): 9.1; Disorientation: 0; Personality Changes: 1; Difficulty Completing Tasks: 0. After standardization, the alzheime.pkl model used its pre-fitted scaler to analyze the whole vector and run the predict() and predict_proba() functions. The system's output was Class 0, which means that there is no dementia. This is exactly what the medical expert said. Fig. 4 shows the diagnostic report, which includes a clinical format with a confidence level, stage interpretation, and a recommendation. This shows that the clinician's judgment and the model's prediction are very similar.

```
🔎 Please enter values for the following features:

Enter value for 'Gender': 1
Enter value for 'EducationLevel': 3
Enter value for 'AlcoholConsumption': 6
Enter value for 'PhysicalActivity': 7.7
Enter value for 'SleepQuality': 9
Enter value for 'FamilyHistoryAlzheimers': 0
Enter value for 'Depression': 0
Enter value for 'HeadInjury': 0
Enter value for 'DiastolicBP': 71
Enter value for 'CholesterolHDL': 43
Enter value for 'MMSE': 19
Enter value for 'FunctionalAssessment': 6
Enter value for 'MemoryComplaints': 0
Enter value for 'BehavioralProblems': 1
Enter value for 'ADL': 9.1
Enter value for 'Disorientation': 0
Enter value for 'PersonalityChanges': 1
Enter value for 'DifficultyCompletingTasks': 0

==================================================
        🧠 Alzheimer's Diagnostic Report
==================================================
Prediction Result     : Class 0
Confidence Level      : 95.37%
Diagnosis Stage       : No signs of dementia

→ Doctor's Note       : 👍 Continue regular cognitive screening annually.

→ Full Probability Breakdown:
   Class 0 (No signs of dementia): 95.37%
   Class 1 (Very mild cognitive impairment): 4.63%
==================================================
```

Fig. 4 Model and Doctor both are predict Alzheimer's Not Detected

2) Case Study 2:

The Alzheimer's Association [15] provided a clinically validated profile indicative of Alzheimer's disease, consisting of 18 distinct characteristics: Gender: 1; Level of Education: 0; Drinking: 4.2 grams per day; Exercise: 6.4 hours per week; Quality of Sleep: 9.55 out of 10. No family history of Alzheimer's, depression, or head injury. The person's diastolic blood pressure is 93 mmHg, their high-density lipoprotein cholesterol is 76.4 mg/dL, their mini-mental state examination score is 9, their functional assessment score is 7, their memory complaints are 0, their behavioral problems are 1, their activities of daily living score is 1.69, their disorientation is 0, their personality changes are 0, and their difficulty completing tasks is 0. After standardization, the entire vector was processed by the alzheime.pkl hybrid model, which used the pre-fitted scaler and ran the functions predict() and predict_proba(). The system recognized Class 1, which is the same as Alzheimer's disease, and it matched the doctor's note perfectly. Fig. 5 shows a diagnostic report that includes a confidence score, the stage of the disease, and an AI-generated doctor's note that suggests cognitive therapy and planning for caregivers.

Fig. 5 Model and Doctor both are predict Alzheimer's Detected

B. How model solved previous research gaps:

The suggested hybrid model effectively addresses several major limitations found in earlier studies by improving generalization, aligning with clinical practices, and making results easier to understand. Prior studies often relied on limited or singular datasets, lacked external validation, and offered constrained comparisons with actual clinical evaluations. Our method tests the trained model on both the Kaggle dataset [4] and real national and international patient profiles, which makes it easy to compare it directly to expert diagnoses. This empirical validation demonstrates that the model consistently correlates with physician-annotated outcomes for both Alzheimer's-positive and Alzheimer's-negative cases. The model improves on earlier research by using Mutual Information-based feature selection, which reduces noise, makes decisions clearer, and makes predictions more stable by using hybrid ensemble learning instead of just one algorithm.

C. Comparison between Other Research Paper and Our Model

As Table 2 shows, prior studies reported accuracies of 73–92.13 %, while the best single model (SVM) reached 90.07 %. Our hybrid ensemble (SVM + RF + GBC) achieves 95.81 % accuracy, outstripping every earlier approach and establishing the highest reported performance for Alzheimer's prediction.

TABLE II
COMPARISON RESEARCH ARTICLE

Sl. No.	Name of the Algorithm	Accuracy
1	Random Forest	86.92% [3]
2	SVM	81.03% [5]
3	SMOTE-RF	94.03% [6]
4	NMM	90% [8]
5	TCN+LSTM	80.34% [9]
6	Shap + XGBoost + NEAR-MISS	76.13% [10]
7	SVM	75% [11]
8	Random Forest	94% [13]
9	Second Hybrid Model (Combined of SVM, RF, GBC)	95.81% [Our model]

D. Why MI was chosen as a feature selection

Here selected Mutual Information (MI) because it can find both linear and non-linear relationships between features and the target, unlike correlation-based methods. It handles different types of data, cuts down on unnecessary or duplicate variables, and makes the model more stable. This makes sure that feature inputs are clearer and more useful for making accurate predictions about Alzheimer's.

E. Why use a hybrid approach instead of a single machine learning algorithm

A hybrid approach is preferable to a single machine-learning algorithm because Alzheimer's data has many different patterns, such as linear, non-linear, probabilistic, and interaction-based elements, that a single model can't fully capture. The ensemble combines different algorithms to take advantage of their strengths, such as being easy to understand, strong, able to learn non-linearly, and able to fix mistakes. This method reduces overfitting, improves generalization across different patient profiles, and consistently achieves better accuracy, reliability, and clinical alignment than any one classifier.

F. Novelty and Limitation

This paper introduces a new AI-powered "doctor" agent that uses a seven-hybrid ensemble framework to make diagnoses more reliable. It uses MI-based feature selection to find more accurate biomarkers and makes reports like a doctor with confidence scoring. These are features that have rarely been used in previous Alzheimer's research. The system shows external validity because its predictions match up with

real clinical cases in the US and around the world. As a prototype, it has some flaws: the autogenerated doctor's note isn't always completely accurate, confidence thresholds might misclassify borderline cases (like 50% predictions), and clinical recommendations can vary from country to country and doctor to doctor. More fine-tuning and a lot of clinical calibration are needed.

IV. Conclusion

This research presents a robust AI-driven hybrid ensemble framework for the detection of Alzheimer's disease, featuring seven tri-model combinations that exhibit enhanced performance relative to standalone algorithms. The evaluated hybrids exhibited accuracies ranging from 83.95% to 95.81%, with the optimal model (the second hybrid model) selected for implementation. Table 1 shows a summary of the F1-scores, accuracy, precision, and recall. Figure 2 shows how well the ROC–AUC works, and Figure 3 shows how the accuracy and precision trends compare. The deployed alzheime.pkl model efficiently processes real-time clinical inputs from both national and international sources, producing diagnostic classifications, confidence levels, stages, and physician-style notes. The combination of MI-based feature selection and hybrid voting fixes problems with interpretability, generalization, and external clinical validation, making it a reliable way to screen for Alzheimer's disease early in real-world healthcare settings.

In the future, research will combine longitudinal, multi-omics, wearable sensor, and 3D Alzheimer's imaging data with federated learning and explainable AI dashboards to make it easier to use electronic health records in real time. To improve clinical decision support, the system will make results that are specific to each stage, as well as personalized medication lists and treatment schedules.

Acknowledgment

We would like to express our deep gratitude to JIS College of Engineering and the Research & Development (R&D) team for giving me the chance, the tools, and the constant support I needed to do this research.

References

[1] "Dementia," World Health Organization. Accessed: Nov. 17, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia.

[2] "Dementia," Wikipedia, Jul. 11, 2025. Accessed: Nov. 17, 2025. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Dementia&oldid=1300017519

[3] C. Kavitha, V. Mani, S. R. Srividhya, O. I. Khalaf, and C. A. T. Romero, "Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models," *Front. Public Health*, vol. 10, Mar. 2022, doi: 10.3389/fpubh.2022.853294.

[4] "Alzheimer's Disease Dataset," Kaggle. Accessed: Nov. 17, 2025. [Online]. Available: https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset.

[5] M. Sudharsan and G. Thailambal, "Alzheimer's disease prediction using machine learning techniques and principal component analysis (PCA)," *Mater. Today: Proc.*, vol. 81, pp. 182–190, Jan. 2023, doi:10.1016/j.matpr.2021.03.061.

[6] P. Rani, R. Lamba, R. K. Sachdeva, K. Kumar, and C. Iwendi, "A machine learning model for Alzheimer's disease prediction," *IET Cyber-Phys. Syst.: Theory Appl.*, vol. 9, no. 2, pp. 125–134, 2024, doi:10.1049/cps2.12090.

[7] S. T. Ahmed and S. M. Kadhem, "Alzheimer's disease prediction using three machine learning methods," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 3, pp. 1689–1697, Sep. 2022, doi:10.11591/ijeecs.v27.i3.pp1689-1697.

[8] B. K. Karaman, E. C. Mormino, M. R. Sabuncu, and Alzheimer's Disease Neuroimaging Initiative, "Machine learning based multi-modal prediction of future decline toward Alzheimer's disease: An empirical study," *PLoS One*, vol. 17, no. 11, Nov. 2022, doi:10.1371/journal.pone.0277322.

[9] W. Khan et al., "An explainable and efficient deep learning framework for EEG-based diagnosis of Alzheimer's disease and frontotemporal dementia," *Front. Med.*, vol. 12, Jul. 2025, doi:10.3389/fmed.2025.1590201.

[10] F. Gelir, T. Akan, S. Alp et al., "Machine Learning Approaches for Predicting Progression to Alzheimer's Disease in Patients with Mild Cognitive Impairment," *J. Med. Biol. Eng.*, vol. 45, pp. 63–83, 2025, doi: 10.1007/s40846-024-00918-z.

[11] N. P. D, H. M. V, D. C, S. S, and A. K. S, "Alzheimer's Disease Prediction Using Machine Learning Methodologies," in Proc. 2022 Int. Conf. Comput. Commun. Informat. (ICCCI), Jan. 2022, pp. 1–6, doi: 10.1109/ICCCI54379.2022.9740942.

[12] A. S. Alatrany, W. Khan, A. Hussain, et al., "An explainable machine learning approach for Alzheimer's disease classification," *Sci. Rep.*, vol. 14, 2024, Art. no. 2637, doi: 10.1038/s41598-024-51985-w.

[13] A. N. Upadhyaya et al., "Alzheimer Disease Classification Model Using Machine Learning Techniques," *Int. J. Basic Appl. Sci.*, vol. 14, no. 3, pp. 396–405, Jul. 2025, doi: 10.14419/rcsfqr61.

[14] "Alzheimer's Disease Fact Sheet," National Institute on Aging. Accessed: Nov. 17, 2025. [Online]. Available: https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet.

[15] "Medical Tests for Diagnosing Alzheimer's & Dementia," alz.org. Accessed: Nov. 17, 2025. [Online]. Available: https://www.alz.org/alzheimers-dementia/diagnosis/medical_tests