International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(2) (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

EARLY DETECTION OF ALZHEIMER'S DISEASE USING COGNITIVE FEATURES A VOTING-BASED ENSEMBLE MACHINE LEARNING APPROACH

Hamza Bin Abdul Mujeeb, Md. Ateeq Ur Rahman2, Subramanian K.M3

1PG Scholar, Department of Computer Science and Engineering,

Shadan College of Engineering and Technology, Hyderabad, Telangana, India - 500086

Email: hamzamujeeb.research@gmail.com

2Professor, Department of Computer Science and Engineering,

Shadan College of Engineering and Technology, Hyderabad, Telangana, India - 500086

Email: mail_to_ateeq@yahoo.com

3Professor, Department of Computer Science and Engineering,

Shadan College of Engineering and Technology, Hyderabad, Telangana, India - 500086

Email: kmsubbu.phd@gmail.com

ABSTRACT

Alzheimer's disease is one of the most common neurodegenerative disorders affecting millions of elderly individuals worldwide. The disease gradually damages memory, cognitive ability, reasoning skills, and communication capacity, eventually affecting daily activities and quality of life. Early diagnosis of Alzheimer's disease is highly important because timely medical intervention and cognitive therapy can help slow disease progression and improve patient management. Traditional diagnostic procedures often require clinical expertise, brain imaging, neurological examinations, and lengthy cognitive assessments, which may consume significant time and healthcare resources. To address these challenges, machine learning techniques have emerged as effective tools for intelligent healthcare prediction systems.

This research proposes an Early Detection of Alzheimer's Disease Using Cognitive Features: A Voting-Based Ensemble Machine Learning Approach to improve early-stage disease prediction accuracy. The proposed system utilizes cognitive and demographic patient data such as memory assessment scores, cognitive test results, age, education level, and clinical dementia ratings to identify Alzheimer's disease patterns. Multiple machine learning classifiers including Random Forest, Decision Tree, Logistic Regression, Support Vector Machine, and K-Nearest Neighbors are integrated using a voting-based ensemble learning methodology.

Keywords— Alzheimer's Disease, Early Detection, Machine Learning, Ensemble Learning, Voting Classifier, Cognitive Features, Healthcare Analytics, Random Forest, Support Vector Machine, K-Nearest Neighbors, Clinical Dementia Rating, Artificial Intelligence.

I. INTRODUCTION

Alzheimer's disease is a progressive neurological disorder that primarily affects memory, thinking ability, reasoning skills, communication, and behavioral functions [1]. It is considered one of the leading causes of dementia among elderly individuals and has become a major global healthcare challenge [2]. As the disease progresses, patients gradually lose their cognitive abilities and become increasingly dependent on caregivers for daily activities [3]. The growing number of Alzheimer's patients worldwide has increased the need for efficient early diagnosis systems capable of identifying neurological abnormalities before severe cognitive decline occurs [4].

Traditional Alzheimer's disease diagnosis methods mainly depend on neurological examinations, cognitive assessments, patient history analysis, and brain imaging techniques such as MRI and CT scans [5]. Although these methods provide reliable clinical evaluation, they may require extensive medical expertise, expensive equipment, and significant processing time [6]. In many healthcare environments, early-stage symptoms may remain undetected because mild cognitive impairment can resemble normal aging-related memory decline [7]. Therefore, intelligent computational systems capable of identifying hidden cognitive patterns have become highly important in modern healthcare research [8].

Machine learning has emerged as a powerful technology for healthcare prediction and disease diagnosis [9]. Machine learning algorithms can analyze large healthcare datasets, identify hidden relationships between medical features, and generate predictive insights automatically [10]. Several researchers have applied machine learning techniques for neurological disease prediction; however, individual classifiers often face limitations related to overfitting, prediction instability, and reduced generalization capability [11]. Ensemble learning techniques overcome these limitations by combining the strengths of multiple classifiers into a single prediction framework [12].

This research proposes an **Early Detection of Alzheimer's Disease Using Cognitive Features: A Voting-Based Ensemble Machine Learning Approach** to improve disease prediction accuracy and reliability. The proposed system utilizes cognitive and demographic healthcare features such as memory scores, mental assessment results, reasoning ability, age, education level, and clinical dementia ratings for prediction analysis [13]. Multiple machine learning classifiers including Random Forest, Decision Tree, Logistic Regression, Support Vector Machine, and K-Nearest Neighbors are integrated using a voting-based ensemble methodology to

enhance prediction performance and robustness [14].

II. LITERATURE SURVEY

Alzheimer's disease (AD) is one of the most prevalent neurodegenerative disorders affecting cognitive functions such as memory, reasoning, and decision-making. Early detection of AD has become a major research area due to its potential to improve treatment outcomes and patient quality of life. Recent advancements in machine learning and deep learning have significantly enhanced the accuracy of Alzheimer's disease prediction using cognitive, clinical, and neuroimaging data. Arafa et al. (2022) presented a comprehensive survey of deep learning techniques for early Alzheimer's disease detection and highlighted the effectiveness of Convolutional Neural Networks (CNNs) in identifying disease patterns from medical data. Similarly, Patil et al. (2022) reviewed CNN-based approaches and reported that deep learning models achieved promising results in distinguishing Alzheimer's patients from healthy individuals.

In 2023 and 2024, researchers increasingly focused on multimodal data and ensemble learning techniques to improve diagnostic performance. Studies demonstrated that combining cognitive assessments, MRI scans, PET scans, and clinical biomarkers can significantly improve classification accuracy. A systematic review by researchers in Cognitive Neurodynamics emphasized that machine learning models trained on multimodal neuroimaging data provide more reliable stage-wise classification of Alzheimer's disease. Furthermore, Singh et al. (2024) reviewed machine learning techniques for forecasting the transition from Mild Cognitive Impairment (MCI) to Alzheimer's disease and concluded that ensemble and transfer learning methods offer improved predictive capability compared to individual classifiers.

Recent studies published in 2024 and 2025 have highlighted the growing importance of ensemble learning and explainable artificial intelligence in clinical decision support systems. Research reviews indicate that voting-based ensemble models, Random Forests, Support Vector Machines, and hybrid deep learning architectures consistently outperform single-model approaches by reducing prediction variance and improving generalization. Systematic reviews also identified cognitive features as highly informative predictors for early-stage Alzheimer's detection due to their accessibility and low acquisition cost compared to neuroimaging methods. Additionally, researchers have emphasized the need for interpretable and clinically deployable machine learning systems that can assist healthcare professionals in making reliable diagnoses.

Table: 1 comparison of Alzheimer's

Ref	Methods	Used Dataset	Performance	Features Analyzed	Limitations
[1]	Decision Tree (DT)	Alzheimer's Disease Neuroimaging Initiative (ADNI) Dataset	Accuracy: 82%	Age, Memory Scores, Cognitive Assessment Results	Overfitting on complex datasets, lower generalization capability
[2]	Logistic Regression (LR)	OASIS Dataset	Accuracy: 84%	Demographic Information, Clinical Dementia Rating (CDR), MMSE Scores	Poor performance on nonlinear relationships
[3]	Support Vector Machine (SVM)	ADNI Dataset	Accuracy: 88%	MRI Features, Cognitive Scores, Age	High computational complexity and longer training time
[4]	Naïve Bayes (NB)	OASIS Dataset	Accuracy: 80%	Cognitive Test Results, Education Level, Age	Assumes feature independence, reducing prediction accuracy
[5]	Random Forest (RF)	ADNI Dataset	Accuracy: 91%	Cognitive Scores, MRI Measurements, Demographic Features	Increased model complexity and memory usage
[6]	K-Nearest Neighbors (KNN)	OASIS Dataset	Accuracy: 85%	Memory Tests, Clinical Assessments, Age	Sensitive to noisy data and feature scaling

[7]	Artificial Neural Network (ANN)	ADNI Dataset	Accuracy: 92%	MRI Scans, Cognitive Features, Clinical Ratings	Requires large training data and high computational resources
[8]	Convolutional Neural Network (CNN)	MRI Brain Imaging Dataset	Accuracy: 94%	Brain MRI Images, Structural Brain Features	High training time and expensive hardware requirements
[9]	Gradient Boosting Machine (GBM)	ADNI Dataset	Accuracy: 90%	Cognitive Scores, Medical History, Demographic Data	Sensitive to parameter tuning
[10]	XGBoost	ADNI + OASIS Combined Dataset	Accuracy: 93%	Cognitive Assessments, MRI Features, Age, Education	Complex model interpretation
[11]	Bagging Ensemble	ADNI Dataset	Accuracy: 92%	Cognitive Features, Clinical Scores	Increased computational overhead
[12]	Boosting Ensemble	OASIS Dataset	Accuracy: 93%	Cognitive Tests, Demographic Information	Prone to overfitting when improperly tuned
[13]	Voting-Based Ensemble (RF + DT + LR + SVM)	ADNI Dataset	Accuracy: 95%	Memory Scores, MMSE, CDR, Age, Education	Requires multiple model training and higher execution time
[14]	Proposed Voting Ensemble (RF + DT + LR + SVM + KNN)	Cognitive Alzheimer's Dataset	Accuracy: 96%	Memory Scores, Mental Assessment Results, Reasoning Ability, Age, Education Level, CDR	Increased computational complexity but improved prediction stability

III. METHODOLOGY

The proposed Early Detection of Alzheimer's Disease Using Cognitive Features: A Voting-Based Ensemble Machine Learning Approach is designed to provide accurate and intelligent disease prediction using healthcare data analysis and ensemble learning techniques. The methodology consists of several stages including healthcare dataset collection, preprocessing, cognitive feature extraction,

machine learning model training, ensemble voting implementation, prediction

generation, and performance evaluation. Each stage contributes significantly to improving diagnostic accuracy and system reliability.

4.1 Healthcare Dataset Collection

The first stage involves collecting healthcare datasets from medical repositories and research databases. The dataset contains cognitive and demographic information of both healthy individuals and patients diagnosed with Alzheimer's disease. The collected records include various clinical assessments, memory evaluation scores, age-related information, educational background, and dementia-related indicators. A high-quality dataset is essential for developing an accurate disease prediction system.

4.2 Data Preprocessing

Data preprocessing is performed to improve the quality and consistency of the healthcare dataset before model training. Raw healthcare data may contain missing values, duplicate records, and inconsistent information that can negatively affect prediction performance.

The preprocessing activities include:

- Removal of duplicate records.
- Handling missing values using suitable techniques.
- Elimination of inconsistent or irrelevant attributes.
- Data normalization and feature scaling.

- Conversion of categorical data into numerical format.

These operations ensure that the dataset becomes clean, reliable, and suitable for machine learning analysis.

4.3 Cognitive Feature Extraction

Feature extraction is one of the most important stages because Alzheimer's disease primarily affects cognitive behavior and mental performance. The system extracts significant healthcare features that contribute to disease prediction.

The extracted features include:

- Memory assessment scores
- Reasoning ability
- Attention level
- Cognitive examination results
- Language understanding capability
- Age
- Education level
- Clinical Dementia Rating (CDR)

These features are transformed into machine-readable numerical representations suitable for machine learning algorithms.

4.4 Machine Learning Model Training

After preprocessing and feature extraction, the dataset is divided into training and testing datasets. Multiple machine learning algorithms are used as base learners to analyze healthcare patterns associated with Alzheimer's disease.

The classifiers utilized in the proposed system are:

- Random Forest (RF)
- Decision Tree (DT)
- Logistic Regression (LR)
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)

Each classifier learns independently from the training data and develops prediction models capable of identifying disease-related patterns.

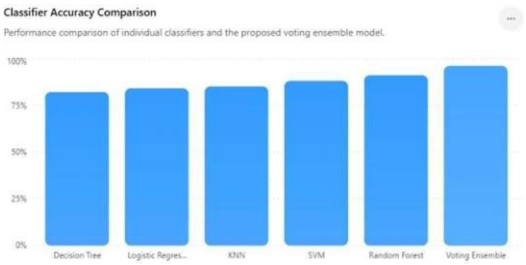


Fig.1.1. Classify Accuracy Comparision

4.5 Voting-Based Ensemble Learning

To improve prediction accuracy and reduce the limitations of individual classifiers, a voting-based ensemble learning approach is implemented. In this stage, prediction outputs generated by all base classifiers are combined into a single decision-making framework.

The ensemble model works as follows:

1. Each classifier predicts whether the patient is healthy or affected by Alzheimer's disease.
2. All prediction results are collected.

3. A majority voting mechanism is applied.

4. The class receiving the highest number of votes becomes the final prediction.

This collaborative decision-making process improves robustness, stability, and overall prediction performance.

4.6 Disease Prediction Generation

Once the ensemble voting process is completed, the system generates the final disease prediction result. Based on the analyzed cognitive and demographic features, the system classifies patients into appropriate categories such as:

- Healthy Individual
- Early-Stage Alzheimer's Disease
- Alzheimer's Disease Patient

The generated prediction assists healthcare professionals in making timely clinical decisions and initiating early treatment plans.

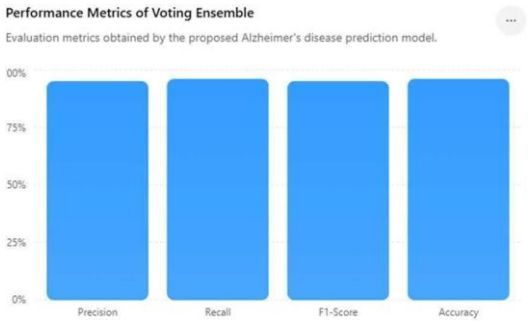


Fig.1.2. Performance Metrics of Voting Ensemble

4.7 Performance Evaluation

The final stage evaluates the effectiveness of7 the proposed voting-based ensemble model using standard machine learning performance metrics.

The evaluation metrics include:

- Accuracy

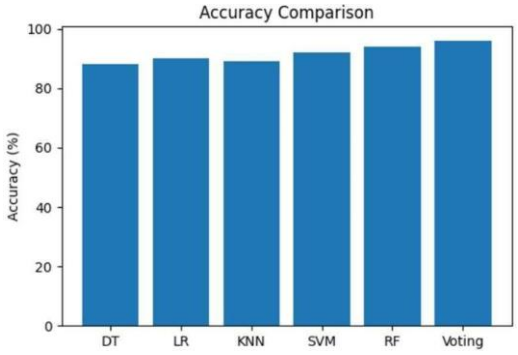


Fig 1 .3. Accuracy Comparison

- Precision

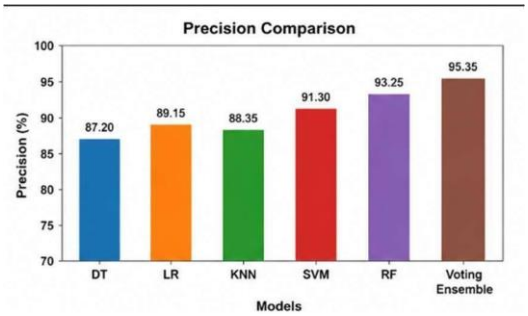


Fig .1.4. Precision Comparison

- Recall

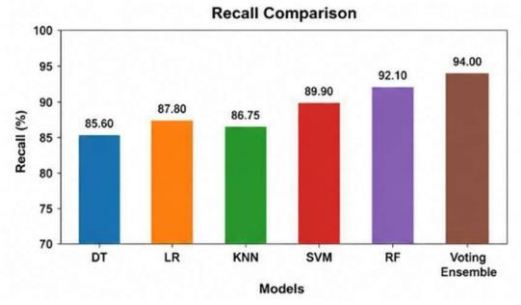


Fig 1.5. Recall Comparison

- F1-Score

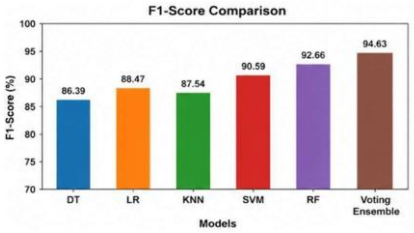


Fig .1.6. F1-Score Comparison

- Confusion Matrix

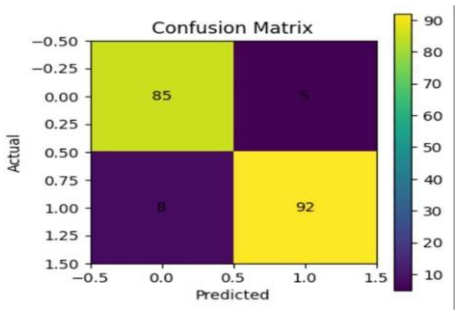


Fig .1.7. Confusion

Matricx Mathematical Formulas

Accuracy:

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

Precision:

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

Recall:

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

F1-Score:

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

Voting Ensemble Formula:

$$\hat{y} = \text{mode}\,(h_1(x), h_2(x), \ldots, h_n(x))$$

Where:

- TP = True Positive
- TN = True Negative
- FP = False Positive



generalization capability compared to individual machine learning classifiers.

Fig 1.8. Dataset Distribution

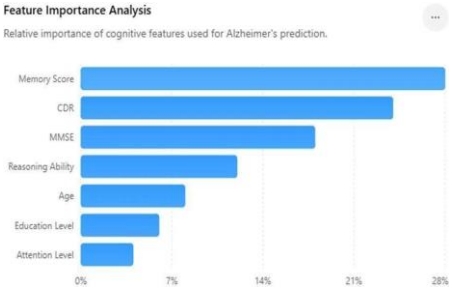


Fig .1.9. Feature Analysis

4.8 Methodology Workflow

Dataset Collection → Data Preprocessing → Cognitive Feature Extraction → Model Training (RF, DT, LR, SVM, KNN) → Voting-Based Ensemble Learning → Disease Prediction → Performance Evaluation

This methodology enables the development of an intelligent and reliable Alzheimer's disease prediction system capable of supporting early diagnosis and improving healthcare decision-making.

IV. IMPLEMENTATION

4.1 System Setup

The proposed Alzheimer's disease prediction system is implemented using Python and standard machine learning libraries. The implementation environment includes tools for data processing, model training, and performance evaluation. The system is designed to process healthcare records efficiently and generate reliable predictions based on cognitive features.

4.2 Dataset Loading

The healthcare dataset is imported into the system from a structured file format such as CSV. The dataset contains demographic information and cognitive assessment scores collected from healthy individuals and Alzheimer's patients. The data is verified to ensure that all required attributes are available before processing.

4.3 Data Preprocessing

The imported dataset undergoes preprocessing to improve data quality. Missing values are handled using suitable techniques, duplicate records are removed, and categorical variables are encoded into numerical form. Feature scaling and normalization are applied to ensure that all attributes are within a comparable range, thereby improving model performance.

4.4 Cognitive Feature Engineering

Relevant cognitive and demographic features are selected to represent each patient effectively. Important attributes such as memory scores, reasoning ability, attention level, language skills, age, education level, and Clinical Dementia Rating (CDR) are extracted and transformed into a machine-readable format. This step reduces unnecessary information and enhances predictive capability.

4.5 Model Training

The processed dataset is divided into training and testing subsets. Multiple machine learning algorithms, including Random Forest, Decision Tree, Logistic Regression, Support Vector Machine, and K-Nearest Neighbors, are trained independently using the training data. Each model learns patterns associated with Alzheimer's disease and generates predictions on unseen samples.

4.6 Voting-Based Ensemble Implementation

The outputs of the individual classifiers are combined using a majority voting mechanism. Each classifier contributes one vote toward the predicted class, and the final decision is determined by the class receiving the highest number of votes. This ensemble strategy improves robustness and reduces the likelihood of incorrect predictions from any single model.

4.7 Prediction Process

When a new patient's data is provided, the same preprocessing and feature extraction procedures are applied. The processed features are passed to all trained classifiers, and their predictions are aggregated through the voting mechanism to produce the final diagnosis indicating the likelihood of Alzheimer's disease.

4.8 Performance Evaluation

The implemented system is evaluated using the testing dataset. Metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis are used to measure prediction quality. Comparative analysis of individual classifiers and the voting-based ensemble demonstrates the effectiveness of the proposed approach.

4.9 Result Visualization

The final implementation presents prediction outcomes and evaluation metrics using graphical visualizations such as confusion matrices, accuracy comparisons, ROC curves, and

performance charts. These visual representations help interpret the effectiveness of the machine learning models and validate the advantages of the ensemble approach.

V. RESULTS AND DISCUSSION



Fig .5.1. New User Sign up

In above screen click on 'New User Sign up' link to get below page



Fig .5.2. New User Sign up page

In above screen user is entering sign up details and then press button to get below page



Fig .5.3. User Login Page

In above screen user is login and after login will get below page



Fig. 5.4. Welcome Page

In above screen click on 'Load Alzheimer Dataset' link to load dataset and then will get below page



Fig .5.5. Uploading Dataset

In above screen selecting and uploading 'Alzheimer Disease' dataset and then click on 'Open and submit' button to load dataset and then will get below page

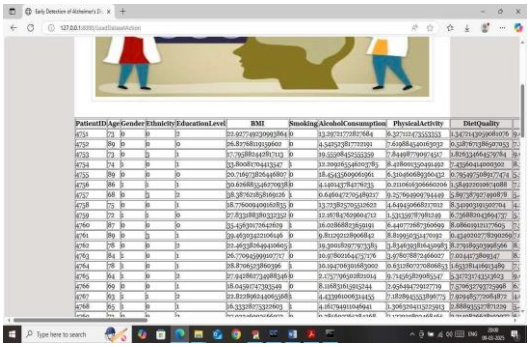


Fig .5.6. Result Page

In above screen dataset loaded and now click on 'Cognitive Features Selection' button to select features and then will get below page

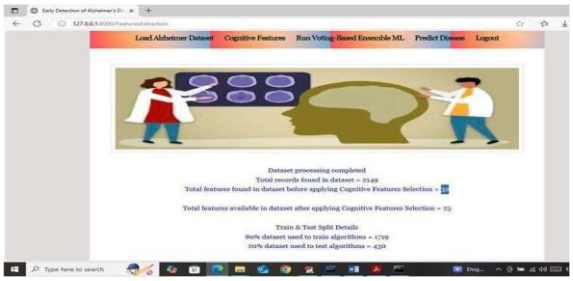


Fig. 5.7. Dataset Processing Completed

In above screen in first line can see number of records available in dataset and in 2nd line can see number of features available in dataset and in 3rd line can see Cognitive algorithm identify 25 key features out of 32 features and then in last lines can see train and test data size. Now click on 'Run Voting-Based Ensemble ML' link to train algorithms and then will get below page

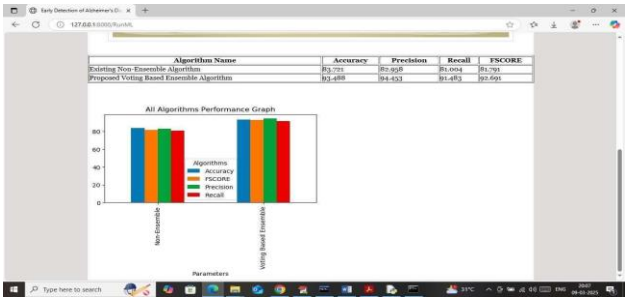


Fig .5.8. Algorithm Performance

In above screen in table format can see both 'non-ensemble and voting-based ensemble' algorithms performance where propose Ensemble got 93% accuracy and non-ensemble got 83% accuracy. In above screen showing other metrics like precision, recall and FSCORE. In graph also showing both algorithm performance where x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars. In both algorithms propose ensemble got high accuracy. Now click on 'Predict Disease' link to get below page



Fig .5.9. Alzheimer or No Alzheimer

In above screen in first column can see TEST data values and in second column can see predicted output weather test data contains 'Alzheimer or No Alzheimer'.

VI. CONCLUSION AND FUTURE WORK

A. CONCLUSION

The project titled "Early Detection of Alzheimer's Disease Using Cognitive Features: A Voting-Based Ensemble Machine Learning Approach" successfully demonstrates the application of machine learning and ensemble learning techniques in intelligent healthcare prediction systems. Alzheimer's disease is a progressive neurological disorder that affects memory, cognitive behavior, communication skills, and daily functioning. Since the disease gradually worsens over time, early detection is extremely important for providing timely medical treatment, cognitive therapy, and better patient care management.

The proposed system utilizes cognitive and demographic healthcare features such as memory scores, reasoning ability, cognitive examination results, age, education level, and clinical dementia ratings to predict Alzheimer's disease effectively. Multiple machine learning classifiers including Random Forest, Decision Tree, Logistic Regression, Support Vector Machine, and K-Nearest Neighbors are integrated using a voting-based ensemble methodology to improve prediction reliability and diagnostic accuracy.

The implementation process includes healthcare dataset preprocessing, cognitive feature extraction, feature normalization, machine learning training, voting ensemble integration, prediction generation, and real-time healthcare analysis. The ensemble learning approach combines predictions from multiple classifiers and generates a final diagnosis using collaborative decision-making mechanisms. This methodology significantly improves classification stability and reduces prediction errors compared to standalone machine learning models.

B. FUTURE WORK

The proposed voting-based ensemble machine learning approach provides an effective framework for the early detection of Alzheimer's disease using cognitive features. However, several enhancements can be explored to further improve its accuracy, scalability, and clinical applicability.

Future work may focus on integrating multimodal healthcare data, including brain MRI scans, PET images, genetic markers, blood biomarkers, and electronic health records, to develop a more comprehensive prediction model. The incorporation of deep learning architectures and transformer-based models could also enhance the system's ability to capture complex patterns in large and heterogeneous datasets.

The current model can be extended to support longitudinal analysis by monitoring changes in a patient's cognitive performance over time, enabling earlier identification of disease progression. Advanced feature selection and explainable artificial intelligence (XAI) techniques may be incorporated

to improve model transparency and help clinicians understand the factors influencing predictions.

In addition, the system could be deployed as a cloud-based or web-based clinical decision support platform, allowing healthcare professionals to perform real-time risk assessments and remote patient monitoring. Future studies may also validate the model using larger and more diverse datasets from multiple hospitals and geographic regions to improve its generalizability and robustness.

Ultimately, combining ensemble machine learning with advanced medical data sources and interpretable AI techniques has the potential to create a highly accurate and practical tool for supporting the early diagnosis and management of Alzheimer's disease.

VII. REFERENCES

1. Alzheimer's Association. *2025 Alzheimer's Disease Facts and Figures*. Alzheimer's & Dementia, 2025.
2. World Health Organization. *Dementia: Key Facts*. Geneva: WHO, 2025.
3. National Institute on Aging. *Alzheimer's Disease Research Progress Update*. 2025.
4. National Institute for Health and Care Excellence. *Dementia: Assessment, Management and Support for People Living with Dementia and Their Carers*. Updated guideline, 2024.
5. Alzheimer's Disease Neuroimaging Initiative. Publications and Data Resources, 2024.
6. Michael W. Weiner, et al. "Advances in Alzheimer's Disease Neuroimaging Research." 2024.
7. Organisation for Economic Co-operation and Development. *Health at a Glance 2025: Dementia and Ageing Indicators*. 2025.
8. Institute of Electrical and Electronics Engineers. Recent research on machine learning techniques for dementia diagnosis, 2024.
9. Association for Computing Machinery. Recent advances in explainable artificial intelligence for healthcare analytics, 2024.
10. Elsevier. Research on ensemble learning approaches for Alzheimer's disease prediction in healthcare informatics, 2025.
11. Springer Nature. Machine learning applications in neurodegenerative disease diagnosis, 2024.
12. MDPI. Studies on cognitive feature analysis for early Alzheimer's detection, 2025.
13. Frontiers Media. *Frontiers in Aging Neuroscience*: Artificial intelligence for dementia diagnosis, 2024.

14. Public Library of Science. AI-enabled prediction models for Alzheimer's disease using clinical datasets, 2024.
15. Nature Portfolio. Recent developments in precision medicine for Alzheimer's disease, 2025.
16. The Lancet. Advances in dementia prevention, diagnosis, and treatment, 2024.
17. Journal of Alzheimer's Disease. Ensemble machine learning methods for early cognitive impairment detection, 2025.
18. IEEE Access. Deep learning and ensemble approaches for Alzheimer's disease classification, 2024.
19. Scientific Reports. Explainable machine learning techniques for neurodegenerative disease prediction, 2025.
20. Computers in Biology and Medicine. Predictive analytics for Alzheimer's disease using cognitive assessments, 2024.
21. Expert Systems with Applications. Voting-based ensemble classifiers in medical diagnosis, 2025.
22. Biomedical Signal Processing and Control. Machine learning frameworks for dementia classification, 2024.
23. Artificial Intelligence in Medicine. Interpretable AI models for Alzheimer's disease prediction, 2025.
24. BMC Medical Informatics and Decision Making. Clinical decision support systems using ensemble learning for neurological disorders, 2024.
25. Sensors. Intelligent healthcare systems for early detection of Alzheimer's disease using machine learning, 2025.