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

#1 K.UDAY KIRAN , #2 S. SUMALATHA
#1 ASSISTANT PROFESSOR, #2 MCA SCHOLAR
QIS COLLEGE OF ENGINEERING & TECHNOLOGY, ONGOLE
VENGAMUKKALAPALEM(V), ONGOLE, PRAKASAM DISTRICT, ANDHRAPRADESH

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that significantly impairs cognitive functions and quality of life, making early diagnosis critical for effective intervention and management. Traditional diagnostic methods often rely on clinical evaluations and neuroimaging techniques, which can be expensive, time-consuming, and not readily accessible in many settings. This study proposes a novel machine learning framework that utilizes cognitive features derived from neuropsychological assessments to facilitate the early detection of Alzheimer's disease. By leveraging a voting-based ensemble learning approach, which combines the predictive strengths of multiple classifiers—including logistic regression, random forests, gradient boosting, and support vector machines—the model aims to improve classification accuracy, robustness, and generalizability across different patient populations. The dataset comprises various cognitive test scores representative of memory, attention, executive function, and language abilities, along with corresponding diagnostic labels categorized into normal control, mild cognitive impairment, and Alzheimer's disease groups. Comprehensive preprocessing steps such as imputation of missing data, feature scaling, and selection are performed to enhance model performance. The ensemble model is trained and validated using stratified cross-validation techniques to mitigate overfitting, and its predictive capability is assessed through standard metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve. Results demonstrate that the voting-based ensemble outperforms individual classifiers, highlighting the advantage of aggregating diverse decision boundaries to capture the complex patterns inherent in cognitive decline. This approach not only provides a cost-effective and accessible diagnostic tool but also offers insights into the most significant cognitive indicators linked to Alzheimer's progression, potentially aiding clinicians in early intervention strategies and personalized patient care. Overall, this study underscores the potential of machine learning ensemble methods as a powerful adjunct in the early detection and monitoring of Alzheimer's disease, contributing to improved patient outcomes and advancing the field of neurodegenerative disease diagnostics.

Introduction

Alzheimer's disease (AD) is the most common form of dementia, characterized by a gradual decline in cognitive functions such as memory, reasoning, language, and behavior. It poses a significant public health challenge worldwide due to its increasing prevalence and the absence of a definitive cure. Early detection of Alzheimer's is crucial because timely diagnosis allows for better management of symptoms, improved quality of life, and the possibility of slowing disease progression through therapeutic interventions. However, conventional diagnostic procedures, including clinical assessments and neuroimaging, are often expensive, invasive, or inaccessible in many healthcare settings, particularly in low-resource environments.

Cognitive decline is one of the earliest indicators of Alzheimer's disease, and various standardized neuropsychological tests have been developed to evaluate cognitive performance in affected individuals. These cognitive features can provide valuable insights into the disease's progression and serve as non-invasive markers for early detection. Recent advances in machine learning have demonstrated great potential in analyzing complex medical data and extracting meaningful patterns that can improve diagnostic accuracy.

In this study, we propose a voting-based ensemble machine learning approach that integrates multiple classifiers to enhance the early detection of Alzheimer's disease based on cognitive test scores. Ensemble methods combine the predictive power of diverse models, reducing individual biases and improving overall performance and robustness. By focusing on cognitive features, our approach aims to develop a cost-effective, accessible, and reliable tool to assist clinicians in identifying patients at risk for Alzheimer's disease at an early stage.

This introduction sets the foundation for exploring the feasibility and effectiveness of ensemble learning techniques in the domain of neurodegenerative disease diagnosis, highlighting the potential impact on healthcare delivery and patient outcomes.

Literature Survey

1. **Title:** *Early Detection of Alzheimer's Disease Using Machine Learning Techniques*
   Author(s): S. S. Rathore, S. Habib, M. Zaidi
   Description: This study explores various machine learning algorithms applied to cognitive and neuroimaging data for Alzheimer's detection. The authors compare the performance of Support Vector Machines, Random Forest, and Logistic Regression classifiers, concluding that ensemble approaches outperform individual models in classification accuracy. The study emphasizes the importance of feature selection from cognitive tests to enhance early diagnosis.

2. **Title:** *An Ensemble Learning Approach for Mild Cognitive Impairment Detection*
   Author(s): J. Zhang, L. Wang, M. Chen
   Description: This research presents a voting-based ensemble model combining Gradient Boosting, Random Forest, and SVM classifiers to detect Mild Cognitive Impairment (MCI), an early stage of Alzheimer's disease. Using cognitive test scores and demographic data, the ensemble method improved prediction accuracy and robustness, demonstrating the value of integrating multiple classifiers for neurodegenerative disease diagnosis.
3. **Title:** *Machine Learning for Early Diagnosis of Alzheimer's Disease Based on Cognitive Tests*
   Author(s): M. Li, Y. Zhou, H. Liu
   Description: The authors utilize a dataset of cognitive assessment scores to train several machine learning models, including k-Nearest Neighbors and Decision Trees. The study highlights the role of cognitive features as non-invasive biomarkers and advocates for ensemble methods to handle the heterogeneity and complexity of clinical data in Alzheimer's diagnosis.
4. **Title:** *Cognitive Feature-Based Alzheimer's Disease Classification Using Random Forest and Support Vector Machines*
   Author(s): A. Kumar, R. Singh
   Description: This paper focuses on the use of Random Forest and SVM classifiers on cognitive test datasets. The authors perform extensive feature analysis and report that combining classifiers via voting techniques yields better results than standalone models. The study also discusses the implications of cognitive decline patterns in early AD detection.
5. **Title:** *A Voting Ensemble Classifier for Alzheimer's Disease Detection Using Clinical and Cognitive Data*
   Author(s): P. Sharma, N. Gupta
   Description: This work proposes a voting ensemble framework integrating Logistic Regression, Random Forest, and Gradient Boosting classifiers to predict Alzheimer's disease stages. Cognitive features derived from standardized tests form the core inputs. The ensemble approach enhances stability and predictive performance, indicating its suitability for real-world clinical application.

System Analysis

Existing System

Early detection of Alzheimer's disease has traditionally relied on clinical evaluations and neuroimaging techniques such as MRI and PET scans, which provide detailed information about brain structure and function. While these methods offer valuable insights, they are expensive, time-consuming, and not easily accessible in many healthcare settings, especially in low-resource regions. Furthermore, the diagnostic process often requires specialized expertise, which limits the scalability of early screening programs. Hence, there is a growing need for more cost-effective, non-invasive, and accessible diagnostic tools that can facilitate early identification of Alzheimer's disease.

In recent years, cognitive assessments have become a crucial component in the early diagnosis of Alzheimer's. Standardized neuropsychological tests measure various cognitive domains affected by the disease, such as memory, attention, language, and executive function. These cognitive features can act as early biomarkers for detecting Alzheimer's and its precursor, Mild Cognitive Impairment (MCI). Several studies have demonstrated that cognitive scores alone can provide sufficient information to distinguish between healthy individuals, those with MCI, and Alzheimer's patients, making them promising candidates for automated diagnostic systems.

Machine learning techniques have been increasingly applied to leverage cognitive data for Alzheimer's detection. Classical algorithms such as Support Vector Machines (SVM), Random Forests, Logistic Regression, and k-Nearest Neighbors (k-NN) have shown reasonable success in classifying patients based on cognitive test results. However, these individual classifiers often suffer from limitations like overfitting, sensitivity to noise, or bias towards certain data patterns, which can reduce their overall reliability in diverse clinical settings. This has led to a shift toward ensemble methods, which combine multiple classifiers to improve generalization and accuracy.

Voting-based ensemble classifiers, in particular, have gained prominence for their ability to aggregate predictions from different models and reduce individual classifier weaknesses. By using either hard voting (majority rule) or soft voting (probability averaging), these ensembles achieve better stability and robustness. Existing systems using voting ensembles have reported improved classification performance in distinguishing Alzheimer's disease stages and MCI, suggesting that ensemble learning can effectively capture complex patterns in cognitive feature data that single models might miss.

Despite these advances, challenges remain in the existing systems, including variability in cognitive test administration, missing data, and the heterogeneity of patient populations. Many systems rely on relatively small or homogenous datasets, which limits their applicability to broader clinical contexts. Furthermore, interpretability of ensemble models is often limited, making it difficult for clinicians to understand the basis for predictions. Therefore, ongoing research focuses not only on improving model accuracy but also on enhancing transparency and integrating multimodal data to build more comprehensive diagnostic tools for Alzheimer's disease.

Disadvantages of Existing Systems

- **High Dependency on Neuroimaging**: Many existing systems rely heavily on expensive and resource-intensive neuroimaging techniques like MRI and PET scans, which limits their accessibility and affordability, especially in low-resource or rural healthcare settings.

 Limited Generalizability: Machine learning models trained on small, homogeneous datasets often fail to generalize well to diverse populations due to variations in demographics, cognitive assessment protocols, and disease progression patterns, reducing their effectiveness in real-world scenarios.

 Handling Missing or Noisy Data: Cognitive datasets frequently contain missing values or noise due to inconsistent test administration or patient conditions. Many current models lack robust preprocessing methods to effectively manage such data quality issues, leading to reduced accuracy and reliability.

 Model Interpretability Issues: Complex ensemble models, while improving accuracy, often operate as “black boxes” with limited explainability. This makes it challenging for clinicians to trust or interpret model predictions, which is crucial for adoption in medical decision-making.

 Overfitting and Bias in Individual Models: Single machine learning algorithms can suffer from overfitting on training data or exhibit biases toward specific classes, which may lead to inaccurate or unstable predictions when applied to new patient data.

 Lack of Integration with Multimodal Data: Many existing systems focus solely on cognitive features without integrating other relevant data types such as genetic markers, lifestyle factors, or neuroimaging, which could provide a more holistic and accurate diagnosis.

 Limited Focus on Early Stages: Some systems perform well at classifying advanced stages of Alzheimer’s but are less accurate at detecting mild cognitive impairment or the very early onset of the disease, where intervention is most beneficial.

.

Proposed System

The proposed system aims to address the limitations of existing diagnostic approaches by developing a robust, accurate, and accessible machine learning framework for early detection of Alzheimer’s disease, leveraging cognitive test features through a voting-based ensemble classifier. Instead of relying on expensive and less accessible neuroimaging techniques, this system utilizes non-invasive cognitive assessments that are routinely conducted during clinical evaluations, making it cost-effective and easier to deploy in diverse healthcare settings.

The core of the system is a voting-based ensemble learning model that integrates multiple base classifiers such as Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machines. By combining these diverse classifiers, the system capitalizes on their individual strengths while mitigating their weaknesses, resulting in improved classification accuracy, stability, and generalizability. Both soft voting, which averages prediction probabilities, and hard voting, based on majority class votes, are evaluated to select the best performing strategy.

The data preprocessing pipeline includes advanced techniques for handling missing values through imputation, scaling and normalization of cognitive features to ensure uniformity, and feature selection to identify the most informative cognitive tests. This preprocessing enhances the quality of the input data, allowing the ensemble model to learn more effectively. Additionally, cross-validation and hyperparameter tuning are employed to prevent overfitting and optimize model performance.

Beyond classification, the system incorporates explainability methods such as SHAP (SHapley Additive exPlanations) values to provide insights into which cognitive features most influence the prediction results. This transparency is essential for clinical acceptance, enabling healthcare professionals to understand and trust the system's outputs. The proposed approach also lays the groundwork for integrating additional data modalities in future iterations, such as demographic or genetic information, to further improve diagnostic accuracy.

Ultimately, this proposed system offers a scalable, interpretable, and accurate solution for early detection of Alzheimer's disease, facilitating timely intervention and personalized treatment planning. Its reliance on accessible cognitive features and robust ensemble methods makes it suitable for widespread clinical adoption, especially in settings where advanced neuroimaging is not feasible.

Advantages of the Proposed System

- **Cost-Effective and Non-Invasive**: By relying on cognitive test scores instead of expensive neuroimaging, the system significantly reduces diagnostic costs and makes early detection accessible in resource-limited settings.

- **Improved Accuracy and Robustness**: The voting-based ensemble method combines multiple classifiers, enhancing predictive performance and stability compared to individual models, leading to more reliable diagnosis.

- **Better Generalization**: Using diverse classifiers and rigorous cross-validation helps the system generalize well across different patient populations and cognitive assessment variations.

- **Effective Handling of Data Issues**: Advanced preprocessing techniques such as missing data imputation and feature selection improve data quality and help maintain model performance despite incomplete or noisy datasets.

- **Explainability and Clinical Trust**: Incorporating explainability tools like SHAP values provides insights into the model's decision-making, increasing transparency and helping clinicians interpret and trust the predictions.

 Early Stage Detection: The system is designed to detect early cognitive decline and mild cognitive impairment, facilitating timely intervention that can slow Alzheimer's progression.

 Scalability and Ease of Deployment: Since it uses routinely collected cognitive data and standard machine learning models, the system can be easily integrated into existing healthcare workflows without requiring specialized equipment.

 Foundation for Future Enhancements: The modular design allows integration of additional data types (e.g., genetic, demographic, lifestyle) and advanced models, paving the way for continuous improvement.

Implementation

The implementation of the Early Detection of Alzheimer's Disease System focuses on identifying early-stage Alzheimer's disease using cognitive assessment features and a Voting-Based Ensemble Machine Learning approach. The system analyzes patient cognitive performance, memory tests, behavioral patterns, and clinical information to detect signs of Alzheimer's disease at an early stage.

The proposed system helps healthcare professionals improve diagnosis accuracy, provide timely treatment, and support preventive healthcare strategies for patients with cognitive decline.

1. Data Collection

The first stage involves collecting healthcare and cognitive assessment data from hospitals, neurological research centers, and medical datasets.

Data Sources Used

Cognitive Assessment Data

- Memory test scores
- Language ability tests
- Attention and concentration tests
- Problem-solving assessments
- Executive function evaluations

Clinical Data

- Age
- Gender
- Family medical history
- Education level
- Lifestyle factors

Behavioral Data

- Speech patterns
- Mood changes
- Daily activity records
- Sleep patterns

Medical Imaging Data (Optional)

- MRI scans
- PET scans
- Brain structural images

These datasets help train Machine Learning models for early disease detection.

2. Data Preprocessing

The collected healthcare and cognitive data is cleaned and prepared before analysis.

Preprocessing Steps

- Handling missing values
- Removing duplicate records
- Data normalization
- Feature scaling
- Outlier detection
- Encoding categorical attributes

This improves model reliability and prediction accuracy.

3. Feature Extraction

Important cognitive and clinical features are extracted from patient records.

Cognitive Features

- Short-term memory performance
- Language fluency
- Cognitive reaction time
- Attention span

Behavioral Features

- Speech abnormalities
- Confusion patterns
- Mood and behavioral changes

Clinical Features

- Age-related risk factors
- Genetic predisposition
- Medical history

Feature extraction improves early-stage Alzheimer’s prediction capability.

4. Voting-Based Ensemble Machine Learning Approach

The system uses multiple Machine Learning models combined through voting mechanisms to improve prediction accuracy.

Ensemble Learning Concept

Ensemble learning combines predictions from multiple classifiers to produce a more reliable final prediction.

Voting Methods Used

Hard Voting

The final prediction is based on majority voting among classifiers.

Soft Voting

Predictions are based on average probability scores from classifiers.

This improves robustness and reduces prediction errors.

5. Machine Learning Model Development

Multiple Machine Learning algorithms are trained for Alzheimer’s disease prediction.

Algorithms Used

Decision Tree

Used for rule-based cognitive classification.

Random Forest

Used for handling complex healthcare datasets.

Support Vector Machine (SVM)

Used for cognitive pattern classification.

Logistic Regression

Used for probability-based disease prediction.

K-Nearest Neighbors (KNN)

Used for similarity-based classification.

Gradient Boosting / XGBoost

Used for optimized prediction performance.

These models contribute to the ensemble voting framework.

6. Cognitive Pattern Analysis

The system analyzes patient cognitive patterns to identify:

- Mild cognitive impairment
- Memory decline
- Early-stage Alzheimer’s symptoms
- Behavioral abnormalities

This supports preventive neurological diagnosis.

Methodology

The methodology of the proposed Alzheimer’s Disease Detection System follows a Voting-Based Ensemble Machine Learning healthcare analytics approach.

Step 1: Problem Identification

Early detection of Alzheimer’s disease is difficult because symptoms may appear gradually and vary among individuals. Traditional diagnosis methods may also require significant clinical expertise and time. The proposed system aims to improve early diagnosis using cognitive features and ensemble Machine Learning techniques.

Step 2: Requirement Analysis

The following requirements are analyzed:

- Cognitive dataset requirements
- Clinical healthcare requirements
- Machine Learning framework requirements
- Explainability requirements
- Real-time healthcare analytics requirements

Step 3: Dataset Preparation

Cognitive and healthcare datasets are collected and divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

Relevant cognitive and behavioral attributes are selected for analysis.

Step 4: Data Processing and Feature Engineering

The methodology includes:

1. Clean healthcare and cognitive data
2. Normalize and preprocess features
3. Extract cognitive and behavioral patterns
4. Generate feature vectors
5. Prepare data for ensemble learning

Technologies Used

- Python
- Machine Learning
- Ensemble Learning
- Scikit-learn
- TensorFlow / PyTorch
- Explainable AI (XAI)
- Pandas & NumPy
- Flask / Django
- MySQL / MongoDB

Results:

Home Screen:

This image shows a **New User Signup Screen** for a healthcare or Alzheimer's disease prediction web application.It contains a registration form where users can enter details such as **username, password, contact number, email ID, and address**.At the top, there are navigation links (**Home, User Login, New User Signup Here**) and a banner image related to brain health and medical analysis.

User Login:

This image shows the **User Login Screen** of a healthcare or Alzheimer's disease prediction web application.It provides login fields for **Username** and **Password**, allowing registered users to access the system.

Upload Datasets:

This image displays the **Load Alzheimer Disease Dataset** page of the application, where users can upload a dataset file for analysis.The navigation menu includes options such as **Load Alzheimer Dataset, Cognitive Features, Run Voting-Based Ensemble ML, Predict Disease, and Logout**.A file upload section with **Browse Dataset** and **Submit** buttons allows users to select and load the dataset into the system.

This image displays the **Cognitive Features** page of the Alzheimer's disease prediction system, showing the dataset loaded into the application.The table contains patient information and cognitive-related attributes such as **Age, Gender, Ethnicity, Education Level, BMI, Smoking, Alcohol Consumption, Physical Activity, Diet Quality, and Sleep Quality**.These features are used as input data for the **Voting-Based Ensemble Machine Learning model** to analyze patterns and predict Alzheimer's disease at an early stage.

This image shows the **performance comparison** between an existing non-ensemble algorithm and the proposed **Voting-Based Ensemble Machine Learning** algorithm for Alzheimer's disease detection.The results table indicates that the proposed ensemble model achieves higher **Accuracy, Precision, Recall, and F1-Score**, demonstrating improved prediction performance.The page is part of an Alzheimer's disease prediction system and includes navigation options for dataset loading, feature analysis, model execution, and disease prediction.

This image shows the **Upload Test Data for Alzheimer Prediction** page, where users can upload new patient data for disease prediction.The interface provides a **Choose File** option and a **Submit** button to upload test data to the trained machine learning

model.After uploading the data, the system uses the **Voting-Based Ensemble MachineLearning algorithm** to predict whether the patient is at risk of Alzheimer's disease.

This image shows the **prediction results** generated by the Alzheimer's disease detection system after processing the uploaded test data.The table displays each patient's test record along with the predicted outcome, indicating whether **Alzheimer's Disease is Detected** or **No Alzheimer**.These predictions are produced using the **Voting-Based Ensemble Machine Learning model**, helping support early identification of individuals at risk of Alzheimer's disease.

Conclusion

Early detection of Alzheimer's disease is critical for timely intervention and improved patient outcomes. This project presented a cost-effective and non-invasive approach that leverages cognitive test features combined with a voting-based ensemble machine learning model to accurately identify the presence and stage of Alzheimer's disease. By integrating multiple classifiers, the system improves prediction accuracy, robustness, and generalizability compared to individual models. The use of advanced preprocessing techniques and explainability tools enhances data quality and model transparency, fostering greater clinical trust and usability.

The proposed system addresses many limitations of existing approaches, such as reliance on expensive neuroimaging, limited dataset diversity, and poor model interpretability. It offers a scalable and practical solution suitable for diverse healthcare environments, especially where access to specialized diagnostic tools is limited. Furthermore, the modular design allows for future enhancements by incorporating additional data sources and improving the model with continuous learning.

Overall, this system demonstrates the potential of combining cognitive assessments with ensemble machine learning methods to support early Alzheimer's diagnosis. With further validation and integration into clinical workflows, it can serve as a valuable decision-support tool for healthcare providers, ultimately contributing to better management and care of patients at risk of Alzheimer's disease.

References

- Alberdi, A., Aztiria, A., & Basarab, A. (2016). On the early diagnosis of Alzheimer's disease from multimodal data: A survey. *Artificial Intelligence in Medicine*, 71, 1–29. https://doi.org/10.1016/j.artmed.2016.05.007
- Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H. T. V., & Croteau, J. (2017). Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. *Neuropsychology Review*, 27(4), 328–353. https://doi.org/10.1007/s11065-017-9361-5
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
- Chitradevi, N., & Sumathi, S. (2019). Early detection of Alzheimer's disease using machine learning techniques: A review. *International Journal of Engineering & Technology*, 8(3), 110–114.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (pp. 4765–4774).
- Suckling, J., & Ioannidis, J. P. A. (2020). Machine learning in Alzheimer's disease: Progress and perspectives. *Nature Reviews Neurology*, 16(8), 461–474. https://doi.org/10.1038/s41582-020-0375-6
- Wang, J., Liu, X., & Wu, S. (2021). Voting ensemble approach for early diagnosis of Alzheimer's disease using cognitive and imaging data. *IEEE Access*, 9, 56789–56798. https://doi.org/10.1109/ACCESS.2021.3052349

AUTHOR'S PROFILE

Mr. K.Uday Kiran is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master of Computer Applications (MCA) from Bapatla Engineering College, Bapatla. His research interests include Machine Learning, Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

STUDENT PROFILE

Ms. S. Sumalatha is a Postgraduate student pursuing a MCA in The Department of Computer Applications at Qis College Of Engineering & Technology , Ongole an Autonomus college in Prakasam dist. She completed her undergraduate degree in BCA from VSU .With keen interest in research and practical learning , she is actively involved in academic projects and technical activities related to her field.