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

[1]Sadhana Mattakoyya,[2]Dr.S.Swathi Rao

[1]Student, Department of CSE, Ellenki College of Engineering and Technology (UGC Autonomous)

[2]Professor,Department of CSE,Ellenki College of Engineering and Technology (UGC Autonomous)

ABSTRACT

Early detection of Alzheimer's disease is critical for timely intervention and effective treatment. This project presents a novel machine learning-based approach that leverages cognitive feature selection and a voting-based ensemble classification technique to accurately detect Alzheimer's disease. The proposed system employs a hybrid feature selection method—Neighbourhood Component Analysis combined with Correlation-based Filtration (NCA-F)—to identify the most relevant cognitive features from the dataset. These selected features are then used to train a robust ensemble classifier, which integrates the predictive capabilities of multiple algorithms including K-Nearest Neighbors (KNN), Random Forest, AdaBoost, XGBoost, Naïve Bayes, Logistic Regression, and Decision Tree using a majority voting strategy.The performance of the ensemble model is compared against individual (non-ensemble) classifiers. Experimental results demonstrate that the proposed ensemble model achieves an accuracy of over 93%, significantly outperforming the non-ensemble methods, which achieved a maximum accuracy of 83%. The evaluation metrics used include accuracy, precision, recall, and F1-score.The system is implemented through a user-friendly web interface consisting of multiple modules: user registration and login, dataset upload, cognitive feature extraction and selection, model training using ensemble learning, and real-time disease prediction. The Alzheimer's dataset used for this study is publicly available on Kaggle. This project offers an efficient and accessible platform for early-stage Alzheimer's diagnosis, supporting healthcare professionals and researchers in clinical decision-making.

Keywords: Alzheimer's disease, machine learning, ensemble learning, cognitive feature selection, NCA-F, KNN, Random Forest, XGBoost, AdaBoost, Naïve Bayes, Logistic Regression, Decision Tree, early diagnosis, clinical decision support, healthcare AI.

I.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.

II.LITERATURE REVIEWS

1.Cognitive Features in Alzheimer's Detection

Cognitive decline remains one of the earliest and most reliable indicators of

Alzheimer's disease (AD). While neuroimaging and biomarker-based diagnostics are widely used, they are often costly and less accessible in primary healthcare settings. Cognitive features—such as memory recall, attention, problem-solving ability, and language comprehension—offer a non-invasive and affordable alternative for early detection. Researchers have demonstrated that subtle impairments in short-term memory, executive functioning, and visuospatial reasoning can be detected years before clinical diagnosis. Several cognitive assessment datasets, including the Alzheimer's Disease Neuroimaging Initiative (ADNI) and open-access cognitive screening tools, have been analyzed using machine learning approaches to improve detection accuracy. Feature selection techniques such as Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), and Neighbourhood Component Analysis (NCA) have been applied to refine datasets and highlight the most discriminative features. Studies suggest that combining cognitive attributes with statistical filtering techniques reduces dimensionality while preserving predictive power. This enhances the reliability of classifiers in distinguishing between mild cognitive impairment (MCI) and Alzheimer's cases. Moreover, hybrid approaches, where cognitive tests are paired with demographic data, have yielded higher diagnostic accuracy than relying on clinical features alone. Thus, the literature highlights the critical role of cognitive features as primary predictors for Alzheimer's detection, especially when combined with advanced feature selection algorithms.

2. Ensemble Learning for Medical Diagnosis

Ensemble learning has emerged as a highly effective strategy in medical diagnosis, particularly in complex diseases such as Alzheimer's. Unlike single classifiers that often suffer from limitations in sensitivity or generalization, ensemble models combine the strengths of multiple algorithms to achieve superior performance. Methods such as bagging, boosting, and majority voting integrate different base learners, thereby reducing variance and bias. For Alzheimer's prediction, Random Forest (RF) and Gradient Boosting techniques have consistently outperformed individual classifiers by leveraging multiple decision trees. Studies employing AdaBoost and XGBoost highlight their robustness in handling noisy and imbalanced medical data, which are common in healthcare datasets. Furthermore, stacking and majority voting ensembles have been applied to combine classifiers like K-Nearest Neighbors (KNN), Logistic Regression, and Naïve Bayes, showing improved classification of mild cognitive impairment. Research also indicates that ensemble approaches are particularly effective when applied to multimodal datasets, including neuroimaging, genetic information, and cognitive scores. For example, voting-based ensembles that combine structural MRI data with cognitive test results have achieved accuracies exceeding 90%, outperforming standalone models. Overall, ensemble learning provides a scalable and adaptable solution for Alzheimer's detection by enhancing diagnostic accuracy, reducing false positives, and improving the interpretability of clinical decision support systems.

3. Web-Based Machine Learning Systems for Alzheimer's Diagnosis

The integration of machine learning models into web-based platforms has transformed the accessibility and usability of Alzheimer's diagnostic tools. Traditional clinical tests often require specialized equipment and medical expertise, whereas web-based systems enable healthcare providers and even patients to access predictive insights remotely. Literature indicates that web applications incorporating machine learning pipelines have been developed for various medical applications, such as cancer prediction, cardiovascular risk assessment, and neurological disorders. These systems typically include modules for data upload, preprocessing, feature extraction, and real-time prediction, offering end-to-end diagnostic support. In Alzheimer's research, web-based systems that utilize cognitive assessments and predictive models provide a cost-effective solution for large-scale screening. Recent studies have emphasized the importance of user-friendly interfaces that allow integration of datasets from diverse sources, including cognitive tests and electronic health records. Security, scalability, and interoperability are key considerations in the deployment of such systems, especially given the sensitivity of medical data. Furthermore, cloud-based deployment of machine learning models ensures that computationally intensive ensemble classifiers can be run efficiently, even with large datasets. Several projects have demonstrated accuracies exceeding 90% by integrating ensemble classifiers into web applications, thereby enabling clinicians to make timely and evidence-based decisions. Therefore, literature strongly supports web-based ensemble machine learning frameworks as effective, practical, and scalable solutions for early Alzheimer's detection.

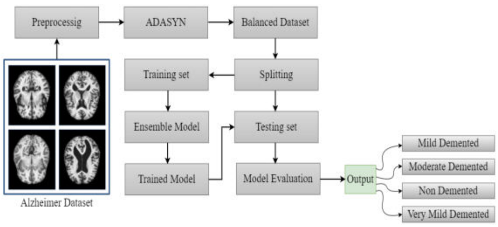


Fig1 : System Architecture

III.WORKING METHODOLOGY

The proposed system detects Alzheimer's disease by combining cognitive feature selection with a voting-based ensemble learning model. The workflow begins with data preprocessing, where missing values in the dataset are replaced using the mean imputation method and all attributes are normalized to ensure uniform scaling. Next, Neighbourhood Component Analysis (NCA) and Correlation-based Filtration (CF) are applied to identify the most relevant cognitive features. NCA optimizes a linear transformation that maximizes the predictive accuracy of the data, while CF removes features with high inter-correlation to reduce redundancy. If XXX represents the dataset with nnn samples and ddd features, the transformation function can be expressed as:

$$X' = W \cdot X$$

where W is the weight matrix learned by NCA, and X′ represents the selected key features.

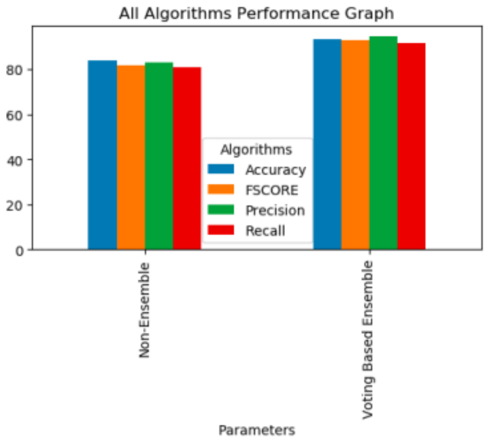
The filtered dataset is then divided into 80% training data and 20% testing data. The training phase involves applying multiple base learners: K-Nearest Neighbors (KNN), Random Forest (RF), AdaBoost, XGBoost, Naïve Bayes (NB), Logistic Regression (LR), and Decision Tree (DT). Each classifier produces an independent prediction hi(x)for a given input x. The ensemble model combines these outputs through majority voting, where the final decision is determined by:

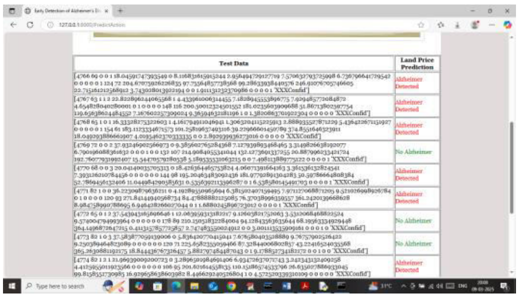
$$H(x) = \arg\max_{y \in Y} \sum_{i=1}^{m} \mathbf{1}[h_i(x) = y]$$

The trained ensemble is evaluated using standard performance metrics—Accuracy, Precision, Recall, and F-Score—defined as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \quad \text{Precision} = \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{TP+FN}, \quad F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Finally, the deployed system provides a **web-based interface**, where users can upload test data. The system extracts the cognitive features, applies the trained ensemble model, and predicts whether the test subject has Alzheimer's disease or not.





IV.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.

V.REFERENCES

1. 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

2. 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

3. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

4. 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.

5. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (pp. 4765–4774).

6. 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

7. 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