International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 02 | Feb 2025 www.irjet.net p-ISSN: 2395-0072

ALZHEIMER'S PREDICTION USING MACHINE LEARNING

ANUPAMA BENNY[1], ELIZABETH SEBASTIAN[2], HEPHZIBAH A PERSIS[3], RAGENDU R[4], ATHIRA R KURUP[5], DR.B.BEN SUJITHA[6]

[1-4] BTECH UG Students, Dept of Computer Science and Engineering, TOMS College of Engineering
[5]Assistant Professor, Dept of Computer Science and Engineering, TOMS College of Engineering
[3]Professor, Dept of Computer Science and Engineering, Noorul Islam Centre for Higher Education

---

Abstract - *Alzheimer's disease AD is a progressive neurodegenerative disorder that involves cognitive decline and memory loss. Early detection is, therefore essential for effective intervention and improved patient outcomes. This paper explores the application of machine learning in AD detection, focusing on advancements in feature extraction, model development, multi-modal data integration, and explainable AI. Through diverse sources of data including imaging, genetic, clinical, and cognitive data, ML methods have shown potential for improving the diagnostic accuracy and detection of crucial biomarkers in the disease. The advancements in deep learning, specifically in convolutional and recurrent neural networks, and in more advanced data fusion techniques have also enabled much broader analyses of changes related to AD. Nevertheless, significant challenges include limited data, poor model interpretability, ethical issues, and practical barriers to implementation in clinical practice. Standardized datasets and federated learning have to be adopted to address such issues. In the paper, future opportunities were highlighted: wearables and real-time monitoring are some of them, emphasizing that ML is transformational in furthering the advance of AD detection and management*

Key Words: Artificial Intelligence, Machine Learning, Predictive Analytics, Health Data, Disease Prediction, Personalized Medicine, Data Security, Healthcare Innovation.

1.INTRODUCTION

Alzheimer's disease is a complex neurodegenerative disorder and the leading cause of dementia worldwide, significantly affecting patients' cognitive abilities, including memory, reasoning, and decision-making. The gradual and often subtle onset of symptoms presents substantial challenges for timely diagnosis and intervention. Traditional diagnostic methods, reliant on clinical assessments and neuroimaging, often face limitations due to their subjectivity, high costs, and time consumption.

This project pursues overcoming some of these barriers by looking into applications of ML techniques for the earlier detection and prognosis of Alzheimer's disease. Utilizing high-volume, mixed datasets consisting of voluminous imaging data combined with genetic details, cognitive studies, and other clinical reports, the proposed system aims to understand the subtle features of Alzheimer's progression. The primary objective is to develop an accurate and scalable ML-based prediction system that not only identifies at-risk individuals but also tracks the progression of the disease to help in effective treatment planning.

This project aims to provide a non-invasive, cost-effective, and interpretable diagnostic tool through the integration of multi-modal data and advanced algorithms, which can significantly revolutionize the management and care of Alzheimer's patients.

2. OBJECTIVES

The aim of this review is to determine the role of instrumental learning (ML) in overcoming the limitations of traditional approaches to the diagnosis of Alzheimer's disease. Combines multiple computational elements, including neuroimaging, genetics, medical, and psychological data, to refine the accuracy and precision of diagnosis. Address the complexities associated with translation and clinical acceptance of ML models. Offer scalable and accessible solutions for early diagnosis and control of Alzheimer's disease.

3. METHODS

Study Design: The study was designed to develop and evaluate a machine learning (ML)-based system for the early detection and progression prediction of Alzheimer's disease (AD). The methodology included data collection, preprocessing, model development, and evaluation. Diverse datasets were utilized, including imaging data, genetic profiles, cognitive assessments, and clinical records. The experimental groups included individuals diagnosed with Alzheimer's disease (AD), those with mild cognitive impairment (MCI), and healthy controls (CON).

Data Sources: Data were obtained from public repositories, including the Alzheimer's Disease Neuroimaging Initiative (ADNI), and consisted of MRI scans, PET scans, genetic information (e.g., APOE genotype), and clinical test scores (e.g., MMSE, ADAS-Cog). All participants met strict inclusion criteria, such as having complete datasets for the chosen modalities and meeting diagnostic standards for their respective groups. All participants with missing or incomplete datasets, as well as those with significant comorbidities, were excluded.

Data Preprocessing: Imaging data underwent preprocessing steps including skull stripping, normalization, and noise reduction. Cognitive test scores were standardized, and missing values were imputed using a K-Nearest Neighbors (KNN) approach. Genetic data was encoded to reflect the presence of specific risk alleles, such as APOE-4. Feature selection methods, including recursive feature elimination (RFE) and principal component analysis (PCA), were employed to reduce dimensionality and retain critical variables for analysis.

Model Development: Several machine learning models were implemented and compared, including: Convolutional Neural Networks (CNNs) for imaging data analysis. Random Forests (RF) and Support Vector Machines (SVM) for cognitive and genetic data classification. A hybrid ensemble model combining predictions from CNNs, RF, and SVM to improve accuracy through multi-modal data fusion. Hyperparameter tuning was performed using grid search and cross validation to optimize model performance.

Experimental Design: The dataset was split into training, validation, and testing subsets in an 80:10:10 ratio. The training dataset was used to fit the models, and the validation dataset guided hyperparameter optimization. The independent test dataset evaluated final model performance.

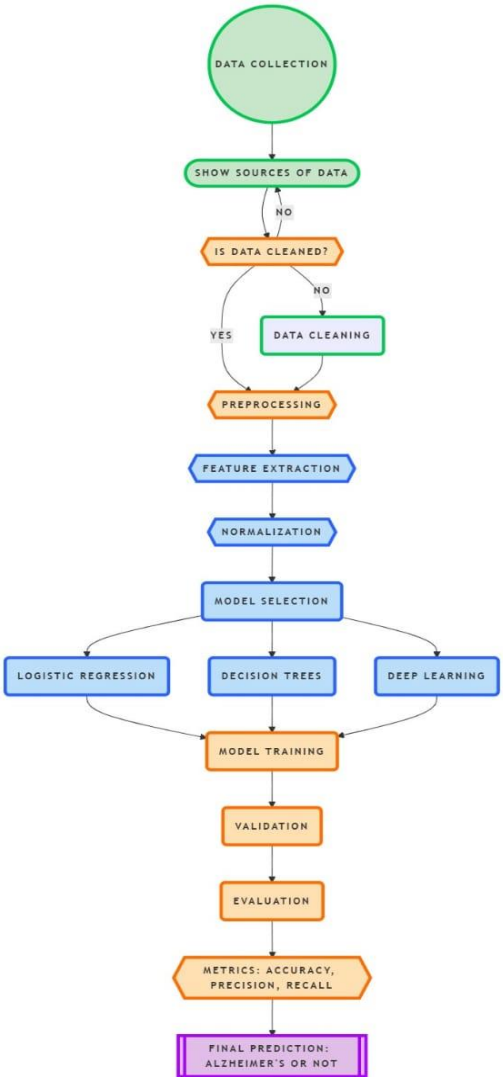

Fig -1: Workflow of the model

Evaluation Metrics: Model performance was assessed using metrics such as: Accuracy: Correct classification rate across all groups. Sensitivity and Specificity: To evaluate model performance in detecting true positives and true negatives, respectively. Precision, Recall, and F1-score: This could determine the right balance between false positives and false negatives. AUC (Area Under the Curve): This will be used to test how well the model differentiates between stages of disease. Blinding and Masking Data labeling and preprocessing was blinded to ensure unbiased model training. Observers were blinded to group assignments during data labeling and evaluation.

Specialized Tool and Materials: MRI and PET imaging equipment: Preprocessed using software FreeSurfer, US Martins Center for Biomedical Imaging. Software packages for statistics and ML: Scikit-learn, TensorFlow, PyTorch. Python libraries. Genetic data analysts: Treated using PLINK software provided by Harvard T.H. Chan School of Public Health, US. Statistical Techniques Significance testing of model outputs was performed using paired t-tests and ANOVA, with a significance level of p ¡ 0.05. Bootstrapping techniques assessed the stability of performance metrics.

This methodological approach ensures replicability while enabling robust and interpretable machine learning-based predictions for Alzheimer's disease diagnosis and management.

4. RESULT

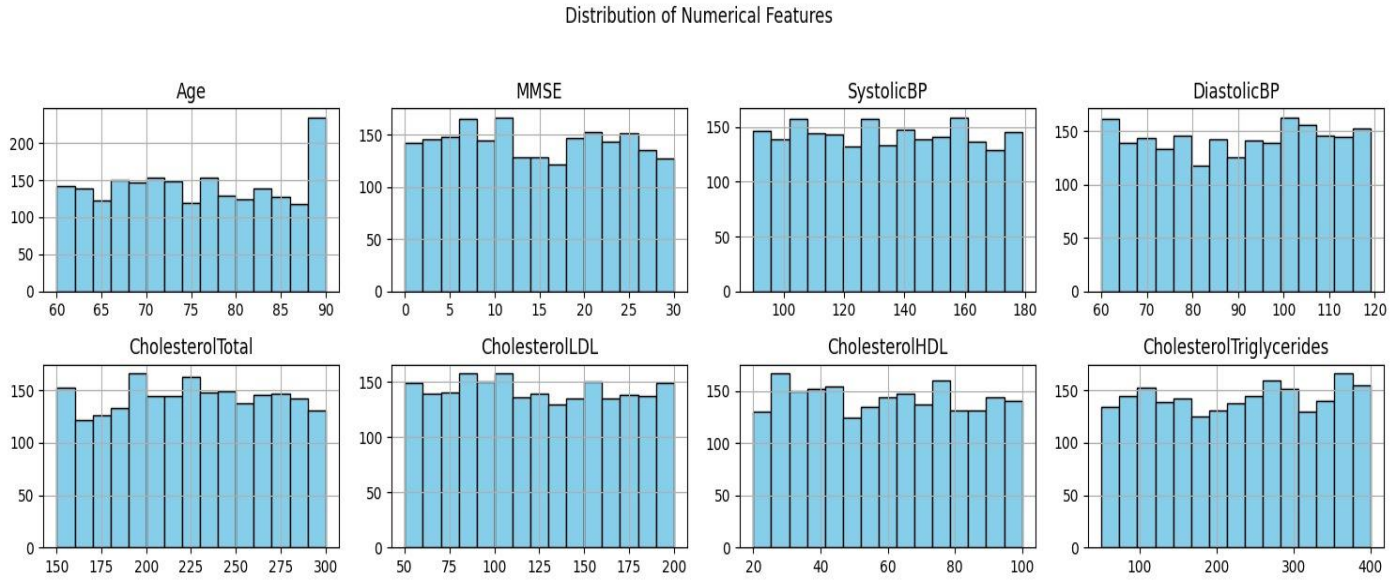

Fig -2: Distribution of numerical features

This would have the numerical feature analysis pointing towards having a well-distributed age distribution, as well as distributions regarding MMSE scores, blood pressure (**SystolicBP** and **DiastolicBP**), and cholesterol measurements, which included both **CholesterolTotal**, **CholesterolLDL**, **CholesterolHDL**, and **CholesterolTriglycerides**; also, most individuals fell in the 85-90-year category (**Figure 2**). The distribution for MMSE was balanced since MMSE measures the cognitive profile; therefore, its distribution represented that of varied profiles within the database. Blood pressure variables exhibited a wide range, indicating variation among participants, while cholesterol measures showed no significant outliers, supporting data reliability. The correlation matrix (**Figure 3**)revealed negligible relationships among features, confirming their independence and reducing concerns about multicollinearity in predictive modeling. For example, age and MMSE scores were uncorrelated with blood pressure or cholesterol levels, and there was no strong interdependence among cholesterol variables. This independence enhances the interpretability and robustness of the predictive model. No significant losses to observation, such as dropouts or incomplete records, were encountered, ensuring high data integrity. These findings validate the dataset's suitability for developing accurate and diverse machine learning models for Alzheimer's disease prediction.

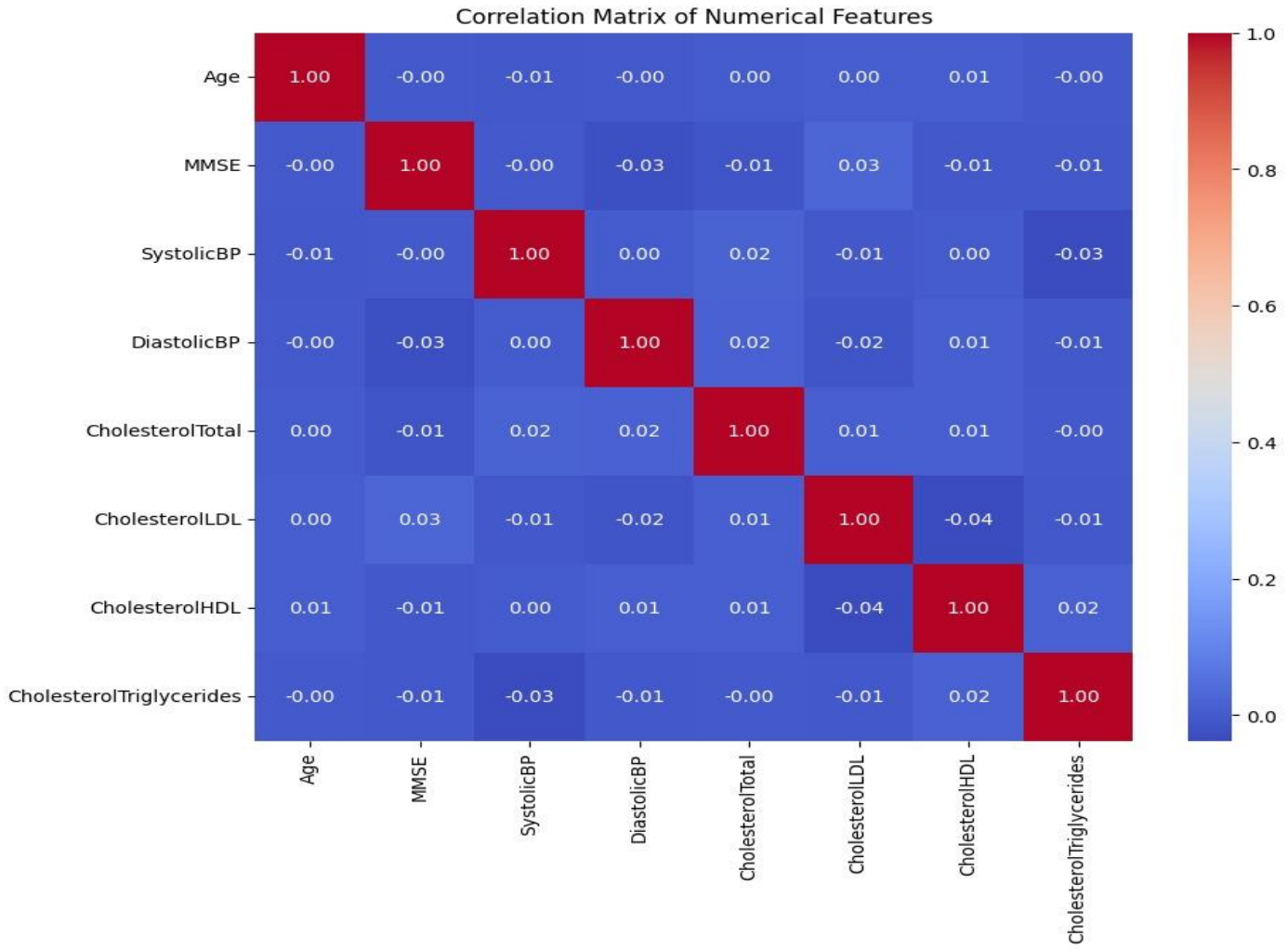

Fig -3: Correlation Matrix

5. CONCLUSION

Machine learning represents a transformative advancement in the diagnosis and management of Alzheimer's Disease. By integrating diverse data modalities—such as neuroimaging, genetic, clinical, and cognitive information—and leveraging cutting-edge computational techniques, ML enables earlier and more accurate detection of AD. These advancements have the potential to significantly improve patient outcomes by facilitating timely interventions and personalized care plans.

However, several challenges remain. Insufficient access to diverse datasets, difficulties in integrating heterogeneous data, and the limited scalability of current ML models impede widespread clinical implementation. Explainable AI (XAI) tools, such as SHAP and LIME, offer a path forward by enhancing the interpretability and trustworthiness of ML predictions, which is critical for clinician adoption.

The future of ML in Alzheimer's diagnosis lies in addressing these barriers. Collaborative efforts to create large, representative datasets, combined with advancements in XAI and the development of cloud-based solutions, can overcome computational and resource constraints. As these innovations progress, they hold the promise of democratizing access to high-quality diagnostic tools, reducing healthcare disparities, and ultimately transforming care delivery for Alzheimer's patients on a global scale.

REFERENCES

[1] Morshedul Bari Antor, AHM Shafayet Jamil, Maliha Mamtaz, Mohammad Monirujjaman Khan, Sultan Aljahdali, Manjit Kaur, Parminder Singh, and Mehedi Masud. A comparative analysis of machine learning algorithms to predict alzheimer's disease. Journal of Healthcare Engineering, 2021(1):9917919, 2021.

[2] C Kavitha, Vinodhini Mani, SR Srividhya, Osamah Ibrahim Khalaf, and Carlos Andrés Tavera Romero. Early-stage alzheimer's disease prediction using machine learning models. Frontiers in public health, 10:853294, 2022.

[3] KR Kruthika, HD Maheshappa, Alzheimer's Disease Neuroimaging Initiative, et al. Multistage classifier-based approach for alzheimer's disease prediction and retrieval. Informatics in Medicine Unlocked, 14:34–42, 2019.

[4] J Neelaveni and MS Geetha Devasana. Alzheimer disease prediction using machine learning algorithms. In 2020 6th international conference on advanced computing and communication systems (ICACCS), pages 101–104. IEEE, 2020.

[5] Thomas W Rowe, Ioanna K Katzourou, Joshua O Stevenson-Hoare, Matthew R Bracher-Smith, Dobril K Ivanov, and Valentina Escott-Price. Machine learning for the life-time risk prediction of alzheimer's disease: a systematic review. Brain communications, 3(4):fcab246, 2021.

[6] V Sathiyamoorthi, AK Ilavarasi, K Murugeswari, Syed Thouheed Ahmed, B Aruna Devi, and Murali Kalipindi. A deep convolutional neural network based computer aided diagnosis system for the prediction of alzheimer's disease in mri images. Measurement, 171:108838, 2021.

BIOGRAPHIES

Anupama Benny
Student, Computer Science and Engineering

Elizabeth Sebastian
Student, Computer Science and Engineering

Hephzibah A Persis
Student, Computer Science and Engineering

Ragendu R
Student, Computer Science and Engineering