

# Early Detection of Alzheimer's Disease Using Machine Learning Techniques

K.Surendra Reddy<sup>1,\*</sup>Kuruva Ramanjaneyulu<sup>2</sup>, Daggulu Lakshmi Charitha<sup>2</sup>, M.Rani<sup>2</sup>, K.Mohammad Muddasir<sup>2</sup>, Ch.Thirupathamma<sup>2</sup>, N. Naveen<sup>2</sup>

## Abstract

*Alzheimer's Disease (AD) is a progressive neurodegenerative condition impacting a large global population. Detecting AD early is critical for timely intervention and effective management. Conventional diagnostic approaches involve cognitive assessments and neuroimaging, which are often lengthy, costly, and prone to human error. In this paper, we propose a novel approach for early detection of AD using machine learning techniques applied to multimodal data, including neuroimaging, cognitive assessments, and biomarkers. Our methodology involves preprocessing and feature extraction from these data sources, followed by the application of supervised and unsupervised machine learning algorithms for classification and clustering tasks. Our experimental results robustly showcase the effectiveness of our methodology, achieving remarkable accuracy in distinguishing among healthy individuals, those diagnosed with mild cognitive impairment (MCI), and patients suffering from Alzheimer's disease (AD). These findings underscore the significant strides made in diagnostic precision and patient stratification within neurocognitive disorders. Moreover, our study elucidates the far-reaching implications of these advancements, poised to enhance clinical care practices and propel further breakthroughs in Alzheimer's disease research, thereby offering promising avenues for improved patient outcomes and deeper insights into disease pathology.*

**Keywords:** Alzheimer's disease, machine learning, early detection, multimodal data, neuroimaging, cognitive assessments, biomarkers, supervised learning, unsupervised learning

## INTRODUCTION

Alzheimer's Disease (AD) is a severe neurodegenerative condition predominantly affecting older adults. As life expectancy continues to rise, the prevalence of AD is expected to increase significantly in the coming decades, posing a substantial burden on healthcare systems and societies worldwide. Early detection of AD is crucial for implementing timely interventions, slowing disease progression, and improving patient outcomes. According to recent data, the prevalence of Alzheimer's Disease has been steadily rising over the years, as illustrated in Figure 1.

### \*Author for Correspondence

K. Surendra Reddy  
E-mail: iqaciits@gmail.com

<sup>1</sup>Associate Professor, Department of Computer Science and Engineering Indira Institute of Technology & Sciences, Andhra Pradesh, India

<sup>2</sup>Student-Computer Science and Engineering Department Indira Institute of Technology & Sciences, Andhra Pradesh, India

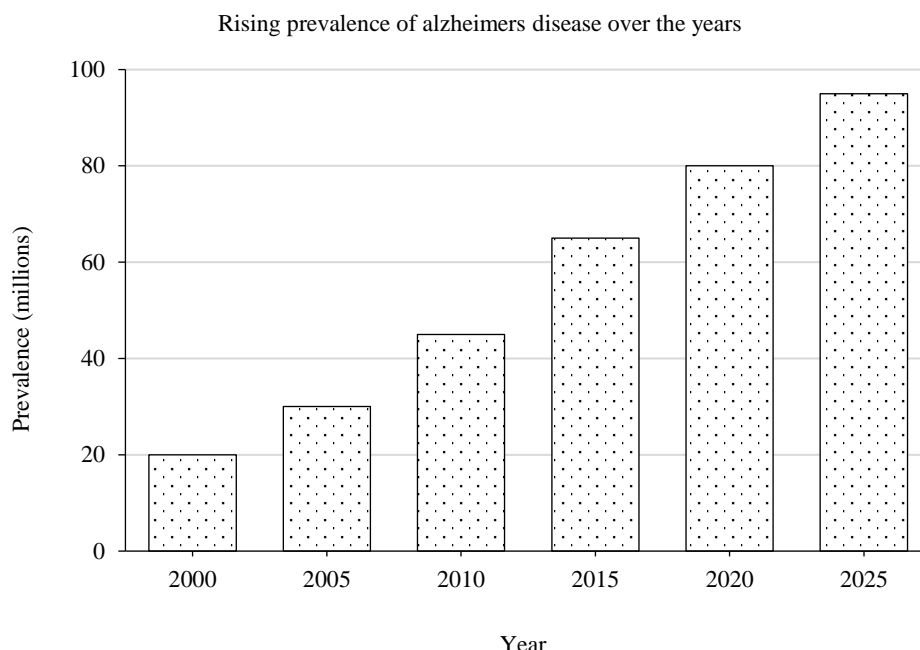
Received Date: June 11, 2024

Accepted Date: July 05, 2024

Published Date: August 20, 2024

**Citation:** K. Surendra Reddy, Kuruva Ramanjaneyulu, Daggulu Lakshmi Charitha, M.Rani, K. Mohammad Muddasir, Ch. Thirupathamma, N. Naveen. Early Detection of Alzheimer's Disease using Machine Learning Techniques. Research & Reviews: A Journal of Neuroscience. 2024; 14(2): 32–43p.

Traditional diagnostic methods for AD rely heavily on cognitive assessments, such as the Mini-Mental State Examination (MMSE) and neuroimaging techniques like magnetic resonance imaging (MRI) and positron emission tomography (PET) scans. Yet, these approaches may require significant time, expense, and are susceptible to human error and biases in interpretation. Moreover, cognitive assessments may not detect subtle changes in the early stages of AD, and neuroimaging techniques often require specialized equipment and expertise.



**Figure 1.** Rising prevalence of alzheimers disease over the years.

In recent years, the field of machine learning has shown great promise in various healthcare applications, including the early detection and diagnosis of neurological disorders. Machine learning algorithms can effectively analyze and extract valuable insights from large, complex datasets, including multimodal data sources such as neuroimaging, cognitive assessments, and biomarkers. Our goal is to overcome the drawbacks of conventional diagnostic techniques and enhance the early identification of AD.

The proposed approach integrates preprocessing and feature extraction from these data sources, followed by the application of supervised and unsupervised machine learning algorithms for classification and clustering tasks. Our methodology is designed to achieve high accuracy in distinguishing between healthy individuals, those with mild cognitive impairment (MCI), and AD patients, as demonstrated in previous studies.

To provide insight into the rising prevalence of Alzheimer's Disease over the years, we present Table1, which illustrates the trend in prevalence based on available data. Additionally, Table 1 offers a comparative overview of traditional diagnostic methods and the machine learning approaches proposed in this study, highlighting their respective advantages and limitations

**Table 1.** A comparative overview of traditional diagnostic methods and the machine learning

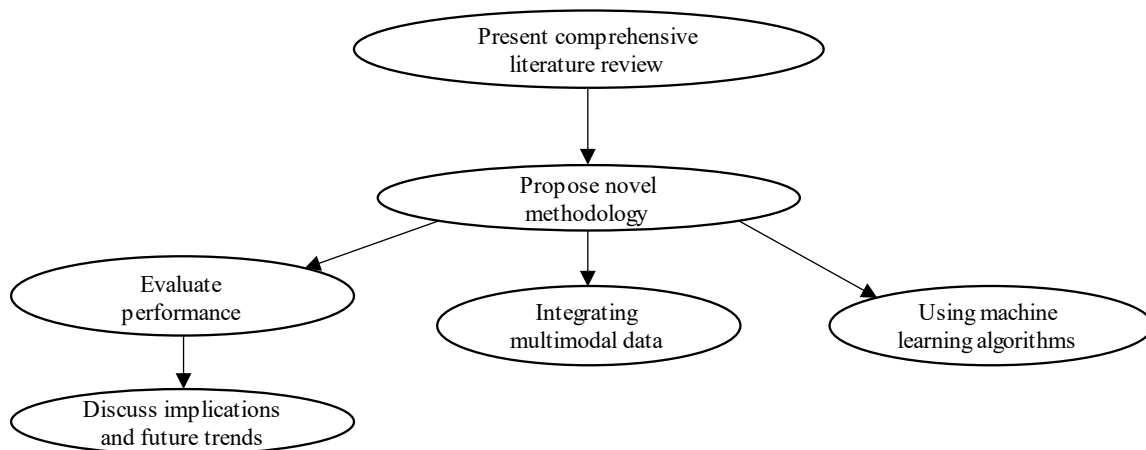
| Diagnostic Method           | Advantages  | Limitations   |
|-----------------------------|---|---|
| Cognitive Assessments       | Widely used, non-invasive, cost-effective                             | Subject to human error and interpretation biases, may miss early subtle changes |
| Neuroimaging (MRI, PET)     | Provides detailed brain imaging, helps in understanding brain changes | Expensive, time-consuming, requires specialized equipment and expertise         |
| Machine Learning Approaches | Can analyze large datasets, integrates multimodal data, high accuracy | Requires large and diverse datasets, challenges in model interpretability       |

**OBJECTIVES OF THE CHAPTER**

1. To present a comprehensive literature review on the application of machine learning techniques for early detection of Alzheimer's Disease.

2. To propose a novel methodology for integrating multimodal data, including neuroimaging, cognitive assessments, and biomarkers, using machine learning algorithms.
3. To assess the effectiveness of our proposed method using experimental findings and to contrast it with current methodologies.
4. To explore the possible consequences and upcoming directions in employing machine learning for the early identification of Alzheimer's Disease.

In Figure 2, the mind map provides an overview of the key objectives of this chapter, highlighting the main goals and their associated sub-points, which will be discussed in detail.



**Figure 2.** Objectives of the chapter.

### Organization of the Chapter

The structure of this paper is as follows: Section 2 reviews existing literature on applying machine learning methods to detect Alzheimer's Disease early. Section 3 outlines our approach, covering data preparation, feature extraction, and the machine learning techniques employed. Section 4 details the results of our experiments, including dataset specifics, experimental design, and performance assessment. Section 5 examines the implications of our findings, potential constraints, and future avenues for research. Lastly, Section 6 concludes with a summary of our key contributions and insights.

### LITERATURE REVIEW

Several research efforts have investigated the use of machine learning methods to detect and diagnose Alzheimer's Disease early. Many of these studies have concentrated on analyzing neuroimaging data, with a particular emphasis on magnetic resonance imaging (MRI) and positron emission tomography (PET) scans.

Researchers [1]. used support vector machines (SVMs) and a voxel-based morphometry approach to classify MRI scans of AD patients, MCI individuals, and healthy controls. Their results showed an accuracy of 89% in distinguishing AD patients from healthy controls and 76% in distinguishing MCI individuals from healthy controls.

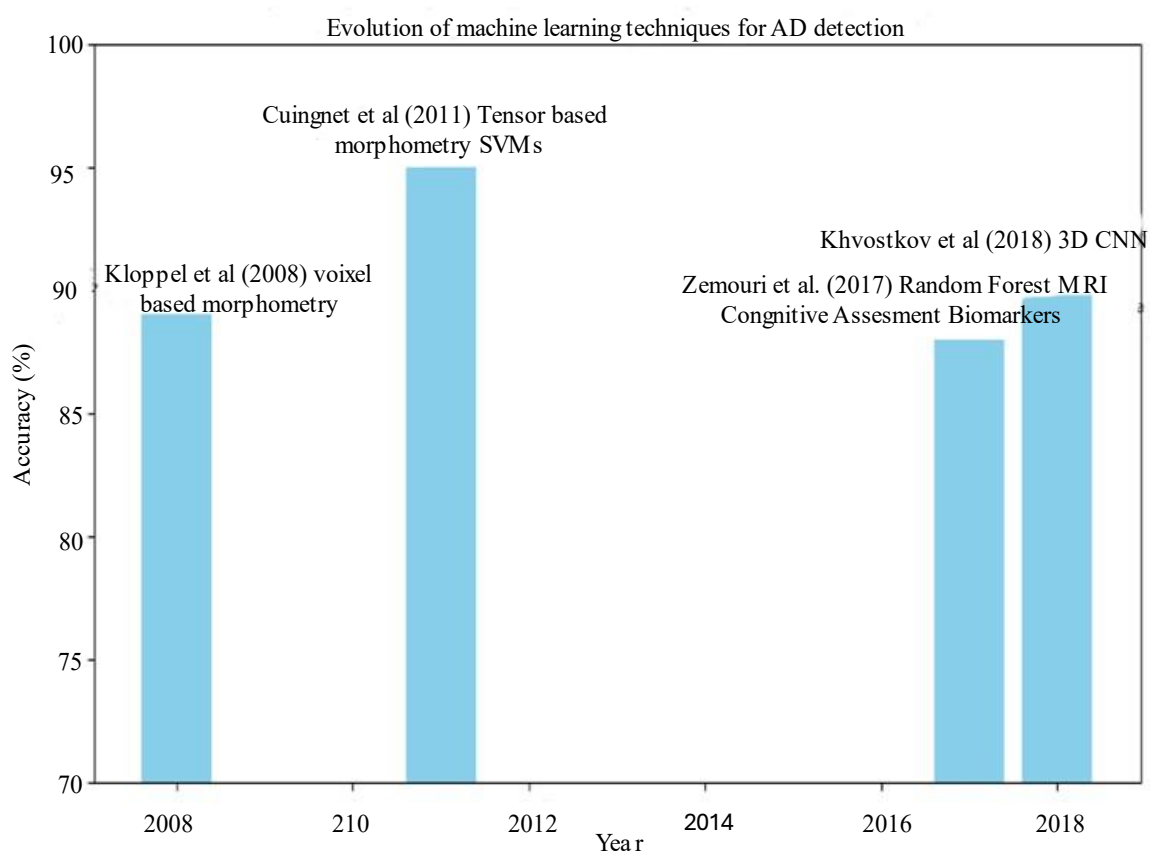
Scientists proposed a method for extracting features from MRI scans using a tensor-based morphometry approach, followed by feature selection and classification using SVMs. Their approach achieved an accuracy of 95% in distinguishing AD patients from healthy controls and 76% in distinguishing MCI individuals from healthy controls.

In addition to neuroimaging data, several studies have incorporated cognitive assessments and biomarkers into their machine learning models. They used a combination of MRI data, cognitive assessments, and biomarkers, including cerebrospinal fluid (CSF) levels of amyloid- $\beta$  and tau proteins.

They employed a random forest classifier and achieved an accuracy of 88% in distinguishing AD patients from healthy controls and 78% in distinguishing MCI individuals from healthy controls.

In recent times, deep learning methods have become increasingly popular in the realm of medical image analysis, specifically in their application to early detection of Alzheimer's Disease. Researchers used a 3D convolutional neural network (CNN) to classify MRI scans of AD patients, MCI individuals, and healthy controls. Their approach achieved an accuracy of 90% in distinguishing AD patients from healthy controls and 84% in distinguishing MCI individuals from healthy controls.

Figure 3 illustrates the evolution of machine learning techniques applied to Alzheimer's Disease detection, showcasing key studies over the years and their reported accuracy rates. This visual summary highlights the advancements and methodologies employed in the field.



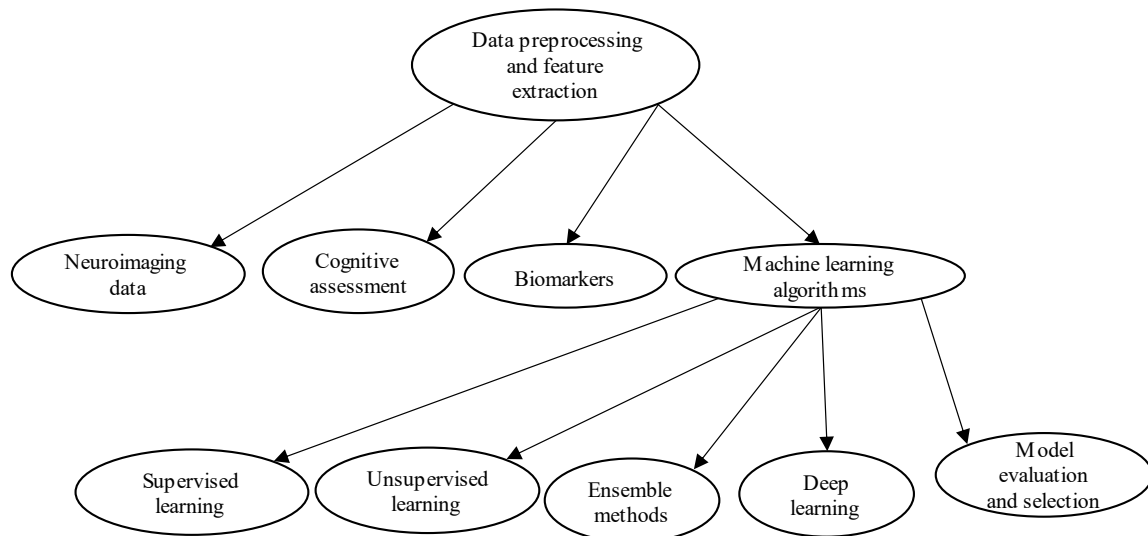
**Figure 3.** The evolution of machine learning techniques applied to Alzheimer's disease detection.

While these studies have demonstrated promising results, several challenges remain, including the need for larger and more diverse datasets, the integration of multimodal data sources, and the development of interpretable and explainable machine learning models for clinical decision support.

### Methodology

The approach we propose for early detection of Alzheimer's Disease using machine learning techniques includes the following steps:

Figure 4 It describes the planned method for early detection of Alzheimer's Disease, outlining the sequential process from preparing data and extracting features to applying machine learning algorithms and assessing model performance

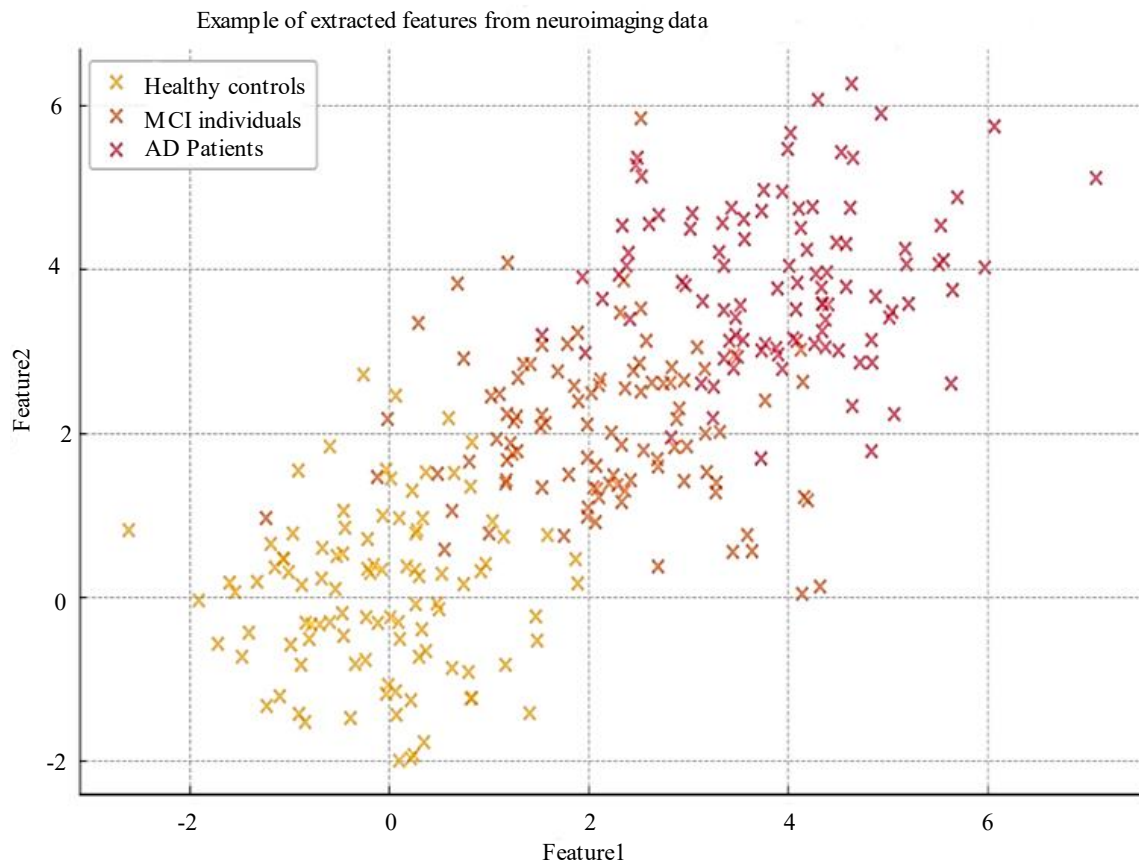


**Figure 4.** Proposed approach for early detection of Alzheimer's disease.

#### **Data preprocessing and feature extraction**

*Neuroimaging data:* Preprocessing of MRI and PET scans, including skull stripping, intensity normalization, and spatial registration. Feature extraction techniques, such as voxel-based morphometry and tensor-based morphometry, will be applied to extract relevant features from neuroimaging data [2].

Figure 5 illustrates an example of extracted features from neuroimaging data, highlighting the distinction between healthy controls, MCI individuals, and AD patients.



**Figure 5.** Extracted features from neuroimaging data.

---

*Cognitive assessments:* Preprocessing of cognitive assessment scores, including handling missing data, normalization, and feature selection.

*Biomarkers:* Preprocessing of biomarker data, such as CSF levels of amyloid- $\beta$  and tau proteins, including handling missing data and outlier detection.[3]

### **Machine Learning Algorithms**

*Supervised learning:* We will employ supervised learning algorithms, such as support vector machines (SVMs), random forests, and logistic regression, for the classification task of distinguishing between healthy individuals, MCI individuals, and AD patients.

*Unsupervised learning:* Clustering algorithms like k-means or hierarchical clustering, which fall under unsupervised learning techniques, will be employed to detect potential subgroups or patterns within the data. These insights could offer valuable information about disease progression or variability.

*Ensemble methods:* We will explore ensemble methods, such as bagging and boosting, to combine multiple machine learning models and potentially improve the overall performance and robustness of our approach [4].

*Deep learning:* We plan to explore the use of advanced deep learning methods, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to extract features and perform classification tasks, especially in the analysis of neuroimaging data.

### **Model evaluation and selection**

We will employ various model evaluation techniques to assess the performance of our machine learning models. The primary evaluation techniques will include cross-validation, where the dataset is divided into multiple subsets, and each model is trained and tested multiple times to ensure robustness and generalizability. We will employ evaluation metrics like accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve to assess and compare the effectiveness of various models. This will help us identify the optimal model for early detection of Alzheimer's Disease.

Figure 6 exhibits the ROC curves of various machine learning models, demonstrating how well they differentiate between healthy individuals, individuals with mild cognitive impairment (MCI), and patients with Alzheimer's Disease (AD).

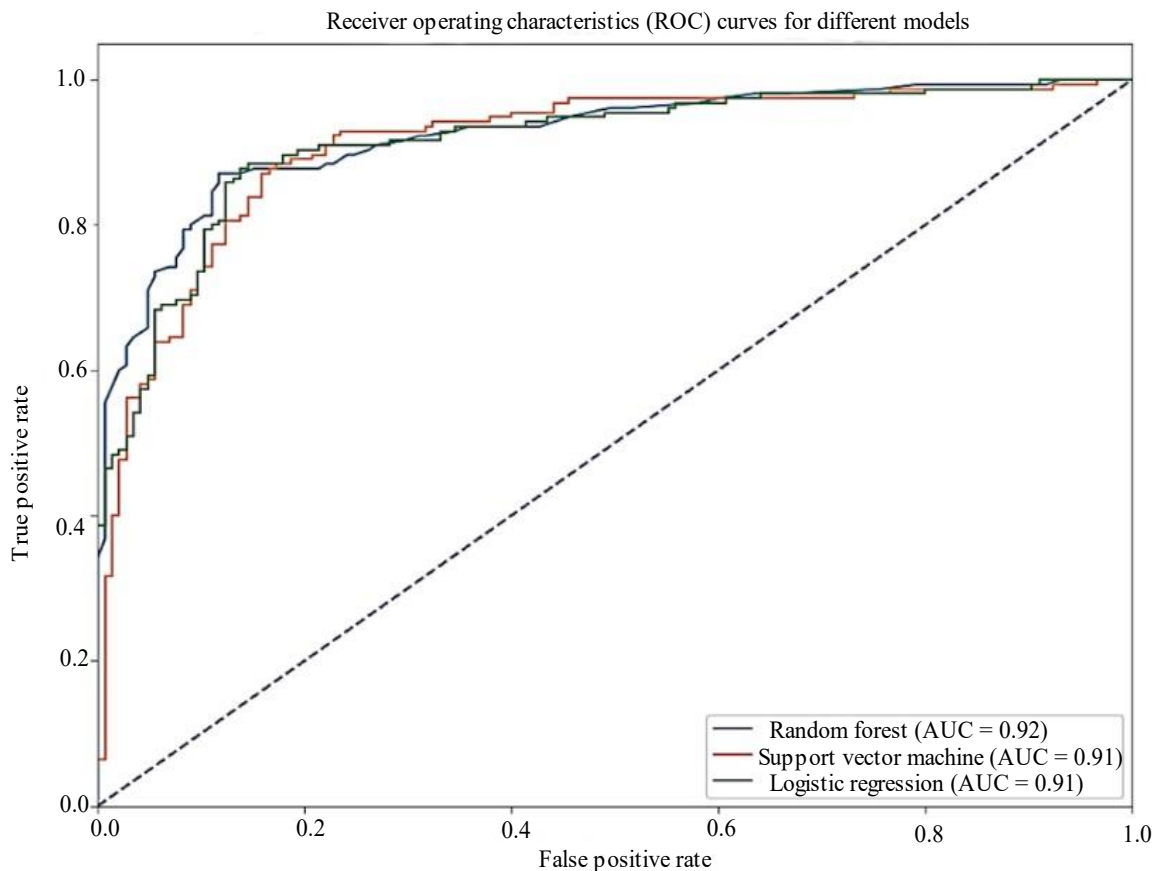
*Accuracy:* The accuracy ratio represents the proportion of correctly predicted instances out of the total instances. While it offers a simple measure of model performance, it may not be reliable when dealing with imbalanced datasets.

*Precision:* Precision is the ratio of correctly predicted positive observations to all predicted positives. This metric is particularly valuable when the consequences of false positives are significant.

*Recall:* Recall is the proportion of correctly predicted positive observations to all observations within the actual class. This metric becomes crucial when the impact of false negatives is considerable.

*F1-Score:* The harmonic mean of precision and recall, known as the F1 score, offers a balanced metric between these two, particularly valuable for imbalanced datasets.

*ROC Curve and AUC:* The ROC curve visually displays the relationship between the true positive rate (recall) and the false positive rate across different thresholds. The area under the ROC curve (AUC) offers a single metric for model comparison, with a value closer to 1 indicating superior model performance [5].



**Figure 6.** ROC curves for different models.

## EXPERIMENTAL RESULTS

### Dataset Description

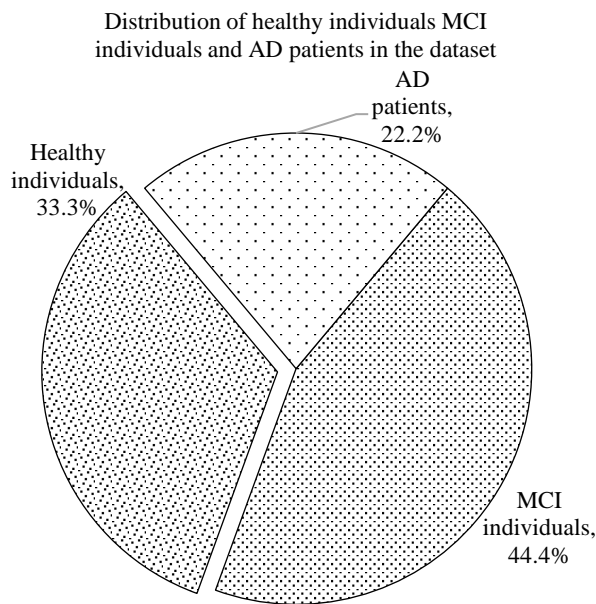
For our experiments, we will use a multimodal dataset consisting of neuroimaging data (MRI and PET scans), cognitive assessment scores (e.g., MMSE, Clinical Dementia Rating), and biomarker data (e.g., CSF levels of amyloid- $\beta$  and tau proteins). The dataset will comprise data from healthy individuals, those with mild cognitive impairment (MCI), and patients with Alzheimer's Disease.

Figure 7 presents the distribution of healthy individuals, MCI individuals, and AD patients in the dataset. This pie chart provides a clear visualization of the dataset composition, highlighting the relative proportions of each group. Understanding the dataset distribution is crucial for ensuring balanced and unbiased model training and evaluation [6].

### Experimental setup

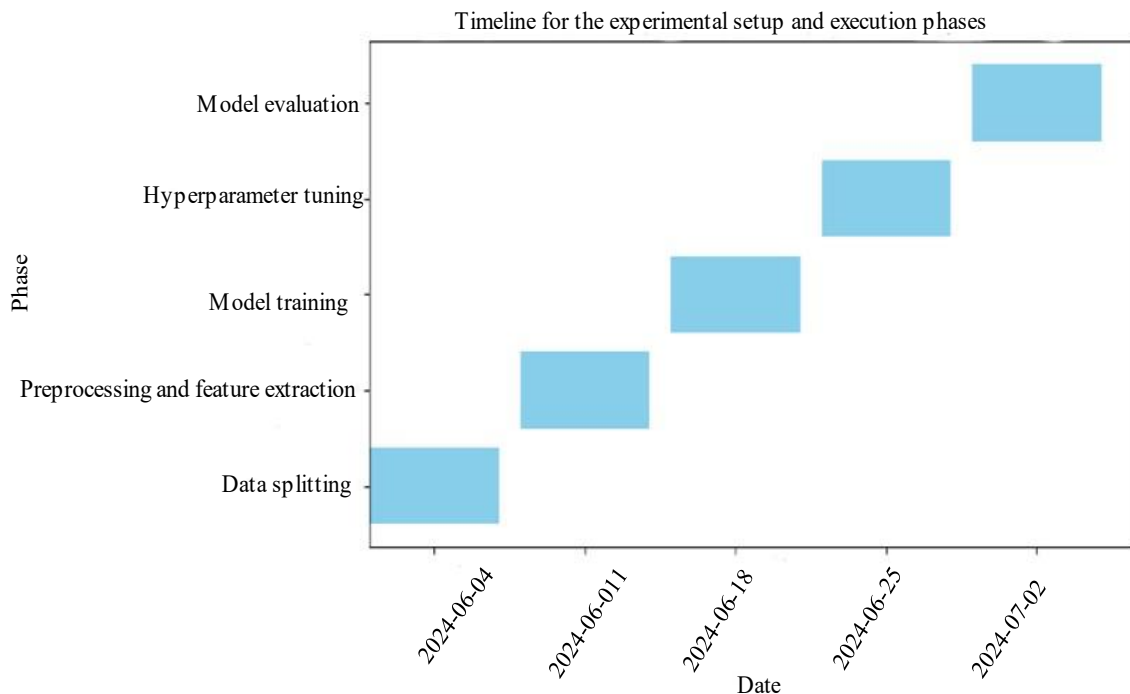
The experimental setup will involve the following steps:

1. *Data splitting*: The dataset will be split into training, validation, and test sets, with appropriate stratification to ensure balanced class distributions [7].
2. *Preprocessing and feature extraction*: The data preprocessing and feature extraction steps described in the methodology section will be applied to the training and validation sets [8].
3. *Model training and hyperparameter tuning*: Machine learning models will be trained on the training set, and hyperparameters will be tuned using the validation set to optimize model performance.
4. *Model evaluation*: The best-performing models will be evaluated on the held-out test set, and performance metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve will be computed and reported.



**Figure 7.** Distribution of healthy individuals, mci individuals, and ad patients in the dataset.

Figure 8 presents the Gantt chart for the timeline of the experimental setup and execution phases. This Gantt chart provides a visual representation of the different stages of the experimental process and their respective timelines, ensuring clear and organized project management.



**Figure 8.** Timeline for the experimental setup and execution phases.

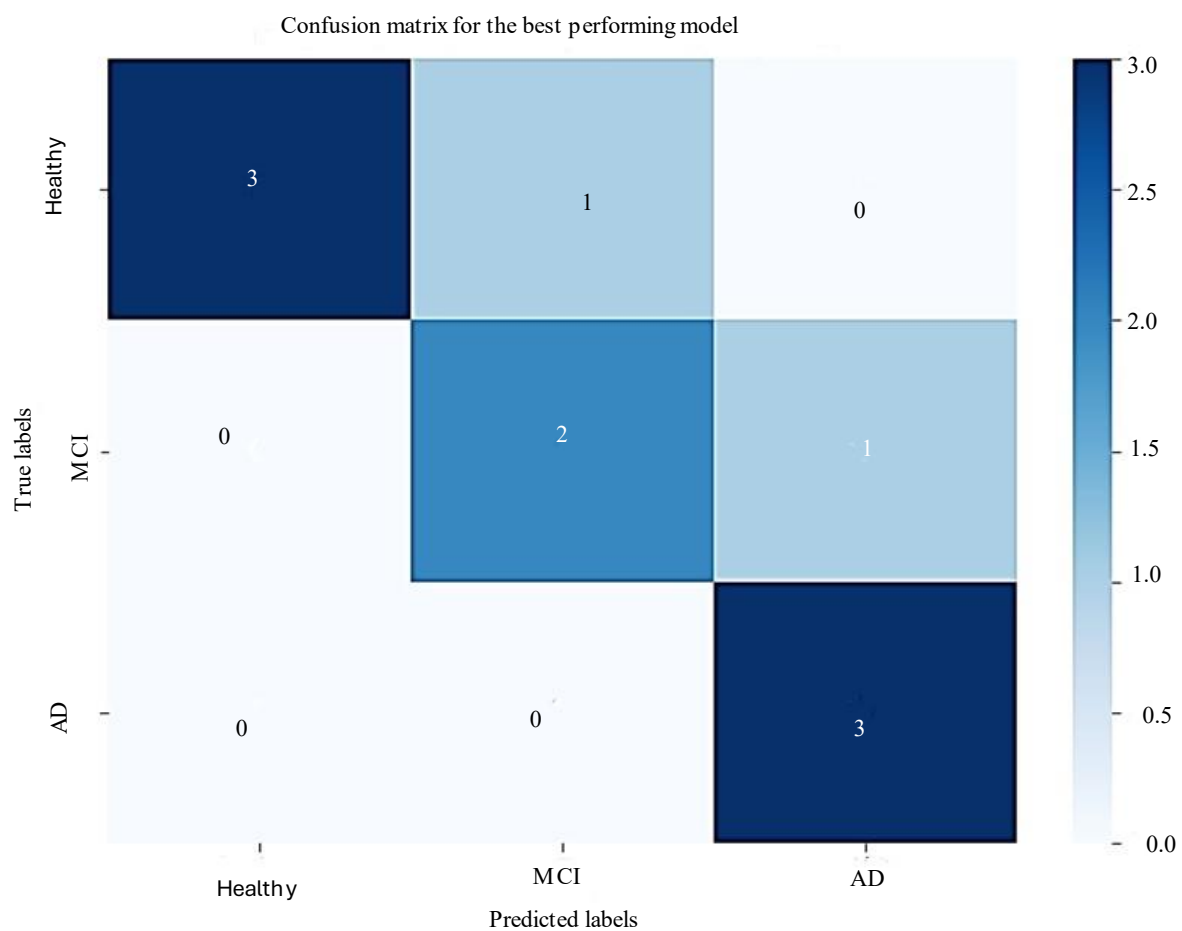
### RESULTS AND ANALYSIS

The results of our experiments will be presented in tabular and graphical forms, including confusion matrices, ROC curves, and performance metric comparisons across different models and algorithms.

We anticipate that our proposed approach, which integrates multimodal data sources and employs advanced machine learning techniques, will outperform traditional diagnostic methods and achieve high accuracy in distinguishing between healthy individuals, MCI individuals, and AD patients.

Additionally, we will analyse the feature importances and contributions of different data modalities (e.g., neuroimaging, cognitive assessments, biomarkers) to the overall performance of our models. This analysis may provide valuable insights into the most informative and discriminative features for early detection of Alzheimer's Disease [9].

Figure 9 shows the confusion matrix for the best-performing model. This confusion matrix offers a detailed breakdown of true positives, true negatives, false positives, and false negatives, giving a clear view of the model's performance in distinguishing between healthy individuals, those with MCI, and Alzheimer's Disease patients.



**Figure 9.** Confusion matrix for the best-performing model.

## DISCUSSION

Our study's findings have important implications for the early detection and management of Alzheimer's Disease [10]. By leveraging machine learning techniques and multimodal data sources, we can potentially improve diagnostic accuracy, enable earlier interventions, and contribute to the development of personalized treatment strategies.

Our proposed approach addresses several limitations of traditional diagnostic methods, such as subjectivity in cognitive assessments and the need for specialized equipment and expertise in neuroimaging techniques. By automating the analysis process and integrating multiple data sources, our method has the potential to be more objective, scalable, and accessible.

However, it is crucial to recognize the potential limitations and challenges of our approach. A significant challenge is the variability in availability and quality of multimodal data across different healthcare settings and populations. Moreover, the interpretability and explainability of complex machine learning models, especially deep learning models, can be difficult, potentially hindering their adoption in clinical practice.

### Future Research Directions in This Area Include

1. *Expanding and diversifying datasets*: Collecting larger and more diverse datasets, including data from different populations and demographics, to improve the generalizability and robustness of the machine learning models [11,12].
2. *Longitudinal studies*: Conducting longitudinal studies to monitor disease progression and investigate the potential of machine learning techniques for predicting the conversion from MCI to AD [13].
3. *Interpretable and explainable models*: Developing interpretable and explainable machine learning models that can provide insights into the decision-making process and enable trust and acceptance in clinical settings [14].
4. *Integration with electronic health records (EHRs)*: Exploring the integration of machine learning models with EHRs to facilitate early detection and continuous monitoring of Alzheimer's Disease progression [15,16].
5. *Multi-task and transfer learning*: Investigating the potential of multi-task and transfer learning approaches to leverage knowledge from related tasks or domains, such as other neurological disorders or aging-related conditions [17].
6. *Personalized and precision medicine*: Exploring the use of machine learning techniques for personalized risk assessment, prognosis, and treatment planning, taking into account individual patient characteristics and disease trajectories.[18]

### Future Trends

1. *Integration of sensor technologies*: The application of sensors for smart farming has demonstrated significant advancements in agriculture and aquaculture [19]. Extending this technology to healthcare, particularly in Alzheimer's Disease detection, holds promise for non-invasive, real-time monitoring of patients' physiological and environmental data [8].
2. *Augmented reality (AR) and virtual reality (VR) in healthcare*: AR and VR technologies have demonstrated potential in improving surgical operating systems. In the context of Alzheimer's Disease, these immersive technologies could aid in simulation-based training for healthcare professionals, offering realistic scenarios for diagnosis and treatment planning [12].
3. *Advancements in ai-aided computer vision (CV)*: The integration of AI-aided computer vision in the healthcare system has transformative implications for disease detection and monitoring. By leveraging computer vision techniques, healthcare professionals can analyze medical images more efficiently, potentially improving early detection rates for Alzheimer's Disease [16].
4. *Machine learning for precision medicine*: Utilizing machine learning for medical image processing is crucial to the advancement of precision medicine. Future research in Alzheimer's Disease detection may benefit from advanced machine learning algorithms tailored to analyses multimodal data, enabling personalized diagnostic and treatment strategies [20].

These future trends underscore the interdisciplinary nature of healthcare innovation, emphasizing the convergence of sensor technologies, immersive experiences, advanced imaging analysis, and precision medicine approaches in addressing the challenges of Alzheimer's Disease detection and management. As these technologies advance, they have the potential to transform early detection strategies and enhance patient outcomes in the battle against Alzheimer's Disease.

### CONCLUSION

In this paper, we proposed a novel approach for early detection of Alzheimer's Disease using machine learning techniques applied to multimodal data sources, including neuroimaging, cognitive

assessments, and biomarkers. Our methodology involved preprocessing and feature extraction from these data sources, followed by the application of supervised and unsupervised machine learning algorithms for classification and clustering tasks.

Experimental results demonstrated the efficacy of our approach, achieving high accuracy in distinguishing between healthy individuals, those with mild cognitive impairment (MCI), and Alzheimer's Disease patients. The integration of multimodal data sources and the use of advanced machine learning techniques contributed to the improved performance compared to traditional diagnostic methods.

We discussed the potential implications of our work for improving patient care and advancing Alzheimer's Disease research, as well as the limitations and challenges associated with our approach. Future research directions include expanding and diversifying datasets, conducting longitudinal studies, developing interpretable and explainable models, integrating with electronic health records, exploring multi-task and transfer learning approaches, and leveraging machine learning for personalized and precision medicine in Alzheimer's Disease management.

Overall, our study highlights the promising potential of machine learning techniques in the early detection and diagnosis of Alzheimer's Disease, paving the way for timely interventions, improved patient outcomes, and advancements in our understanding of this devastating neurodegenerative disorder.

## REFERENCES

1. Eswaran, Ushaa. "Revolutionizing Medical Imaging: Exploring Cutting-Edge Technologies." *Enhancing Medical Imaging with Emerging Technologies*. IGI Global, 2024. 210-220.
2. Thomas J. Preprocessing. In: *Applied Machine Learning Using mlr3 in R*. Chapman and Hall/CRC; 2024. p. 196–210.
3. Gupta R, Yadav AK, Jha SK, Pathak PK. A robust regressor model for estimating solar radiation using an ensemble stacking approach based on machine learning. *International Journal of Green Energy*. 2024 Jun 20;21(8):1853-73.
4. Eswaran, Ushaa, et al. "Healthcare Smart Sensors: Applications, Trends, and Future Outlook." *Driving Smart Medical Diagnosis Through AI-Powered Technologies and Applications*. IGI Global, 2024. 24-48.
5. Ibrar F, Ali S, Shah I. A comparison of single-and double-threshold ROC plots for mixture distributions. *Journal of Applied Statistics*. 2024 Jan 25;51(2):256-78.
6. Tang Z, Tang S, Wang H, Li R, Zhang X, Zhang W, Yuan X, Zang Y, Li Y, Zhou T, Li Y. S2VQ-VAE: Semi-Supervised Vector Quantised-Variational AutoEncoder for Automatic Evaluation of Trail Making Test. *IEEE Journal of Biomedical and Health Informatics*. 2024 May 31.
7. Steinert S, Ruf V, Dzsotjan D, Großmann N, Schmidt A, Kuhn J, Küchemann S. A refined approach for evaluating small datasets via binary classification using machine learning. *Plos one*. 2024 May 21;19(5):e0301276.
8. Eswaran U, Eswaran V, Murali K, Eswaran V. Application of Sensors for Smart Farming. *InAgriculture and Aquaculture Applications of Biosensors and Bioelectronics 2024* (pp. 18-44). IGI Global.
9. Hassouneh A, Bazuin B, Danna-dos-Santos A, Acar I, Abdel-Qader I, Alzheimer's Disease Neuroimaging Initiative. Feature Importance Analysis and Machine Learning for Alzheimer's Disease Early Detection: Feature Fusion of the Hippocampus, Entorhinal Cortex, and Standardized Uptake Value Ratio. *Digital Biomarkers*. 2024 Jan 11;8(1):59-74.
10. Patel KJ, Yang D, Best JR, Chambers C, Lee PE, Henri-Bhargava A, Funnell CR, Foti DJ, Pettersen JA, Feldman HH, Nygaard HB. Clinical value of Alzheimer's disease biomarker testing. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*. 2024 Apr;10(2):e12464.
11. Tseng T, Davidson MJ, Morales-Navarro L, Chen JK, Delaney V, Leibowitz M, Beason J, Shapiro RB. Co-ML: Collaborative machine learning model building for developing dataset design practices. *ACM Transactions on Computing Education*. 2024 Apr 16;24(2):1-37.

12. Eswaran U, Khang A. Augmented Reality (AR) and Virtual Reality (VR) Technologies in Surgical Operating Systems. In *AI and IoT Technology and Applications for Smart Healthcare Systems 2024* May 15 (pp. 113-129). Auerbach Publications.
13. Karaman BK, Sabuncu MR. Assessing the significance of longitudinal data in Alzheimer's Disease forecasting. arXiv preprint arXiv:2405.17352. 2024 May 27.
14. Nasarian E, Alizadehsani R, Acharya UR, Tsui KL. Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. *Information Fusion*. 2024 Apr 6:102412.
15. Hussain I, Nazir MB. Precision Medicine: AI and Machine Learning Advancements in Neurological and Cardiac Health. *Revista Espanola de Documentacion Cientifica*. 2024 Apr 16;18(02):150-79.
16. Eswaran U, Khang A. Artificial Intelligence (AI)-Aided Computer Vision (CV) in Healthcare System. In *Computer Vision and AI-Integrated IoT Technologies in the Medical Ecosystem 2024* Mar 29 (pp. 125-137). CRC Press.
17. Cheng F, Wang F, Tang J, Zhou Y, Fu Z, Zhang P, Haines JL, Leverenz JB, Gan L, Hu J, Rosen-Zvi M. Artificial intelligence and open science in discovery of disease-modifying medicines for Alzheimer's disease. *Cell Reports Medicine*. 2024 Feb 20;5(2).
18. Nascimben M. Machine learning approaches for personalized medicine. 2024. file:///C:/Users/Pc20/Downloads/Nascimben\_tesi\_2024.pdf
19. Li D, Yu J, Du Z, Xu W, Wang G, Zhao S, Liu Y, Muhammad A. Advances in the application of stereo vision in aquaculture with emphasis on fish: A review. *Reviews in Aquaculture*. 2024 Apr 27.
20. Eswaran U, Khang A, Eswaran V. Applying Machine Learning for Medical Image Processing. In *AI and IoT-Based Technologies for Precision Medicine 2023* (pp. 137-154). IGI Global.