IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

Alzheimer's Disease Detection Using Machine Learning

Vaishnavi Thakare[1], Vedanti Bankar[2], Prathamesh Ingle[3], Vansh Gawande[4], Prof. D. G. Ingale[5]
Undergraduate Students, Computer Science and Engineering[1-5]
Dr Rajendra Gode Institute of Technology and Research, Amravati, India [5]

Abstract: *Alzheimer's Disease (AD) is a chronic neurodegenerative disorder that progressively impairs memory, cognitive abilities, and daily functioning. Early and accurate detection of Alzheimer's is essential for timely intervention, effective treatment planning, and improving patient outcomes. However, conventional diagnostic approaches, including clinical evaluations and neuroimaging techniques, are often expensive, time-consuming, and dependent on expert interpretation, which may limit their effectiveness in early-stage diagnosis.This study proposes a machine learning-based approach for the detection of Alzheimer's Disease using clinical and imaging data. Various supervised learning algorithms, including Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN), are implemented and evaluated. The dataset comprises features such as magnetic resonance imaging (MRI) scans, demographic information, and cognitive assessment scores. Data preprocessing techniques, including normalization, missing value handling, and feature selection, are applied to enhance model performance. The experimental results demonstrate that machine learning models can achieve high accuracy in classifying individuals into different stages of Alzheimer's Disease, such as normal, mild cognitive impairment, and advanced Alzheimer's. Among the evaluated models, neural network-based approaches show superior performance due to their ability to capture complex patterns in high-dimensional data. The findings suggest that machine learning can serve as a reliable and efficient tool for early diagnosis, supporting healthcare professionals in decision-making. This research highlights the growing potential of artificial intelligence in healthcare and emphasizes the need for integrating machine learning techniques into clinical practice for improved diagnosis and management of Alzheimer's Disease.*

Keywords: Alzheimer's Disease, Machine Learning, Deep Learning, Magnetic Resonance Imaging (MRI), Early Diagnosis, Cognitive Impairment

I. INTRODUCTION

Alzheimer's Disease (AD) is a progressive and irreversible neurodegenerative disorder that represents the most prevalent form of dementia worldwide. It is characterized by a gradual decline in cognitive functions, including memory, reasoning, language, and problem-solving abilities, ultimately affecting an individual's capacity to perform daily activities independently. The disease not only impacts patients but also places a significant emotional, social, and economic burden on caregivers, families, and healthcare systems. With the rapid growth of the aging population, particularly in developing countries, Alzheimer's Disease has emerged as a major public health concern requiring urgent attention and innovative solutions.

From a biological perspective, Alzheimer's Disease is associated with complex pathological changes in the brain. The accumulation of amyloid-beta plaques and tau protein tangles leads to neuronal degeneration and synaptic dysfunction. These changes begin years, or even decades, before the appearance of noticeable clinical symptoms. As a result, by the time the disease is diagnosed using conventional methods, significant and often irreversible brain damage has already occurred. This highlights the critical importance of early detection and intervention in slowing disease progression and improving patient outcomes.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

In clinical practice, the diagnosis of Alzheimer's Disease typically involves a combination of medical history evaluation, cognitive and neuropsychological testing, laboratory investigations, and neuroimaging techniques such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans. Although these approaches provide valuable diagnostic insights, they have several limitations. They are often costly, time-intensive, and require specialized expertise, making them less accessible in resource-constrained settings. Furthermore, early-stage symptoms such as mild cognitive impairment (MCI) can be subtle and easily misinterpreted as normal aging, leading to delays in diagnosis and treatment.

In recent years, advancements in artificial intelligence (AI), particularly in machine learning (ML), have opened new avenues for improving the diagnosis and management of complex diseases like Alzheimer's. Machine learning algorithms are capable of processing large volumes of heterogeneous data, including medical images, genetic information, clinical records, and cognitive assessments. By identifying patterns and relationships within these datasets, ML models can assist in predicting disease onset, classifying disease stages, and even estimating progression rates with a high degree of accuracy.

The application of machine learning in Alzheimer's detection has gained significant attention due to its potential to overcome the limitations of traditional diagnostic methods. Techniques such as Support Vector Machines (SVM), Random Forests, K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN) have been widely explored for classification tasks. More recently, deep learning approaches, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable performance in analyzing neuroimaging data, enabling automated feature extraction and improved diagnostic precision.

Another important aspect of machine learning-based diagnosis is feature engineering and data preprocessing. Medical datasets often contain noise, missing values, and high-dimensional features that can negatively impact model performance. Therefore, techniques such as normalization, dimensionality reduction (e.g., Principal Component Analysis), and feature selection are essential to enhance the efficiency and accuracy of predictive models. Additionally, the integration of multimodal data—combining imaging, clinical, and demographic information—has shown promising results in providing a more comprehensive understanding of the disease.

Despite these advancements, several challenges remain in the implementation of machine learning for Alzheimer's detection. These include limited availability of high-quality labeled datasets, issues related to data imbalance, risk of overfitting, and lack of model interpretability in clinical settings. Addressing these challenges is crucial for ensuring the reliability and acceptance of AI-based diagnostic tools in real-world healthcare environments.

This research aims to explore and evaluate the effectiveness of various machine learning algorithms in the early detection of Alzheimer's Disease. By utilizing relevant datasets and applying appropriate preprocessing, feature extraction, and classification techniques, the study seeks to develop accurate and efficient predictive models. The comparative analysis of different algorithms provides insights into their strengths and limitations, helping identify the most suitable approach for Alzheimer's diagnosis.

The primary objective of this study is to contribute to the development of intelligent, data-driven healthcare solutions that can assist clinicians in making timely and informed decisions. Early detection of Alzheimer's Disease through machine learning not only has the potential to improve patient care but also supports the broader goal of reducing the global burden of neurodegenerative disorders. As research in this field continues to evolve, the integration of machine learning into clinical practice is expected to play a transformative role in the future of medical diagnostics and personalized healthcare.

II. LITERATURE REVIEW

The application of machine learning techniques in the detection and diagnosis of Alzheimer's Disease (AD) has gained significant attention over the past decade. Researchers have explored a wide range of computational approaches to improve the accuracy, efficiency, and early detection capabilities of diagnostic systems. This section reviews key

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

contributions and methodologies adopted in previous studies, highlighting their strengths, limitations, and research gaps.

Early research in Alzheimer's detection primarily relied on statistical analysis and traditional image processing techniques. However, with the advancement of machine learning, more sophisticated models have been developed to analyze complex medical datasets. One of the widely used approaches is the Support Vector Machine (SVM), known for its effectiveness in handling high-dimensional data. Several studies have demonstrated that SVM achieves high classification accuracy when applied to MRI-based datasets, especially in distinguishing between healthy individuals and patients with Alzheimer's Disease or mild cognitive impairment (MCI). The strength of SVM lies in its ability to create optimal decision boundaries, although its performance can be sensitive to kernel selection and parameter tuning.

Random Forest (RF), an ensemble learning technique, has also been extensively used in Alzheimer's detection. It combines multiple decision trees to improve predictive performance and reduce overfitting. Researchers have reported that Random Forest models provide robust and reliable results, particularly when dealing with heterogeneous datasets that include both clinical and imaging features. Additionally, RF offers feature importance measures, which help in identifying the most relevant biomarkers associated with Alzheimer's Disease.

K-Nearest Neighbors (KNN) is another commonly used algorithm due to its simplicity and ease of implementation. Studies have shown that KNN performs reasonably well in classification tasks; however, its effectiveness decreases with large datasets and high-dimensional features. It is also computationally expensive during the prediction phase, as it requires comparison with all training samples.

In recent years, deep learning techniques, especially Convolutional Neural Networks (CNNs), have revolutionized the field of medical image analysis. CNNs are capable of automatically extracting spatial features from MRI and PET scans without the need for manual feature engineering. Several studies have reported that CNN-based models outperform traditional machine learning methods in terms of accuracy and scalability. For instance, deep learning models have been successfully used to classify different stages of Alzheimer's Disease and detect subtle structural changes in brain regions such as the hippocampus. However, these models require large amounts of labeled data and high computational resources, which can be a limitation in practical applications.

Feature extraction and dimensionality reduction techniques have also played a crucial role in improving model performance. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are widely used to reduce data dimensionality while preserving important information. These techniques help in minimizing noise, reducing computational complexity, and enhancing classification accuracy. Several researchers have combined PCA with machine learning algorithms like SVM and Random Forest to achieve improved results.

Another significant area of research is the use of multimodal data for Alzheimer's detection. Instead of relying solely on imaging data, researchers have integrated clinical data, genetic information, and cognitive test scores to build more comprehensive models. Studies have shown that multimodal approaches provide better diagnostic performance compared to single-modality models, as they capture different aspects of the disease.

Despite the progress made, several challenges remain in the field. One of the major issues is the limited availability of high-quality annotated datasets. Public datasets such as ADNI have been widely used, but they may not fully represent diverse populations. Data imbalance is another challenge, as the number of samples in different disease categories may vary significantly, leading to biased model performance. Furthermore, many machine learning models lack interpretability, making it difficult for clinicians to trust and adopt them in real-world settings.

In summary, the literature indicates that machine learning and deep learning techniques have significant potential in improving the early detection of Alzheimer's Disease. While algorithms such as SVM, Random Forest, and CNN have shown promising results, there is still a need for more robust, interpretable, and scalable models. Future research should focus on addressing data-related challenges, improving model transparency, and integrating machine learning solutions into clinical workflows for enhanced healthcare outcomes.

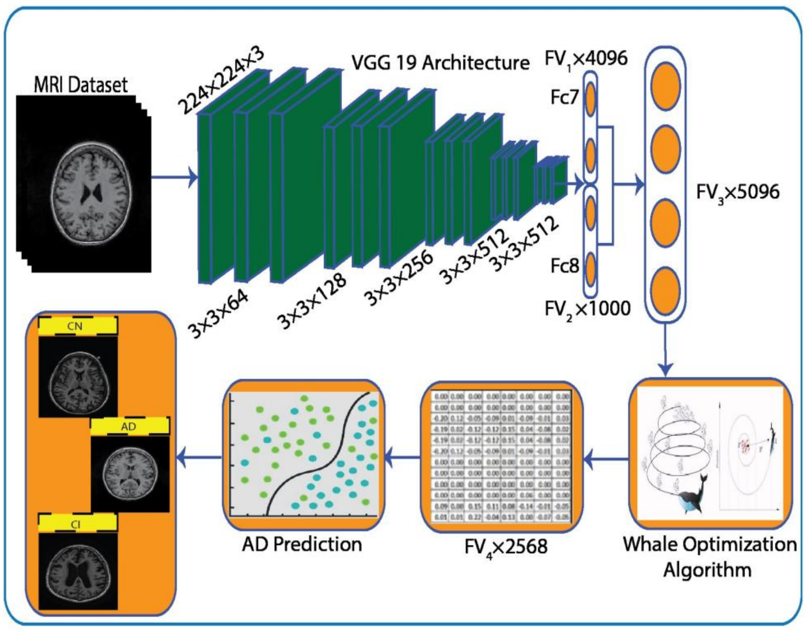
DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 13, April 2026 Impact Factor: 8.2

III. METHODOLOGY

The methodology proposed for the detection of Alzheimer's Disease using Machine Learning (ML) is designed as a multi-stage framework that integrates data acquisition, preprocessing, feature engineering, model development, and validation. This structured approach ensures accuracy, reliability, and scalability in identifying early and progressive stages of the disease. The overall workflow is illustrated as a pipeline where each stage contributes to improving predictive performance and clinical relevance.

The first step involves **data acquisition**, where high-quality and diverse datasets are collected from publicly available medical repositories and clinical sources. These datasets typically include neuroimaging data such as Magnetic Resonance Imaging (MRI) scans, Positron Emission Tomography (PET) scans, and non-imaging data such as demographic information, genetic markers, and cognitive assessment scores. Combining multimodal data enhances the robustness of the detection system, as Alzheimer's Disease affects both brain structure and cognitive function. The dataset is labeled into categories such as cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer's Disease (AD), enabling supervised learning.



Following data collection, a **data preprocessing** phase is conducted to improve data quality and consistency. For imaging data, preprocessing includes skull stripping to remove non-brain tissues, intensity normalization to standardize pixel values, spatial alignment (registration) to ensure consistent orientation, and segmentation of key brain regions such as the hippocampus and cerebral cortex. Noise reduction filters are applied to enhance image clarity. For non-imaging data, preprocessing includes handling missing values using imputation techniques, normalization or standardization of numerical features, and encoding of categorical variables. Outliers are detected and managed to prevent distortion of model training. This step is crucial for minimizing bias and ensuring reliable model performance.

The next stage is **feature extraction and selection**, which focuses on identifying meaningful patterns associated with Alzheimer's progression. From MRI images, features such as gray matter density, white matter integrity, cortical thickness, and volumetric measurements of specific brain regions are extracted using image processing algorithms. Texture analysis techniques like Gray-Level Co-occurrence Matrix (GLCM) may also be applied to capture subtle structural variations. From clinical and cognitive data, features such as memory test scores, age, and genetic risk factors are considered. Feature selection methods such as correlation analysis, recursive feature elimination (RFE), and mutual

ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

information are used to identify the most relevant attributes. Additionally, dimensionality reduction techniques like Principal Component Analysis (PCA) are applied to reduce feature space complexity while preserving significant information, thereby improving computational efficiency.

In the **model development phase**, various machine learning algorithms are implemented and compared to identify the most effective approach. Traditional supervised learning models such as Support Vector Machines (SVM), Decision Trees, Random Forest, k-Nearest Neighbors (k-NN), and Logistic Regression are trained using labeled data. These models are effective in handling structured datasets and provide interpretable results. For more complex data such as MRI images, deep learning techniques, particularly Convolutional Neural Networks (CNNs), are employed. CNNs automatically learn hierarchical feature representations from raw images, eliminating the need for manual feature extraction and often achieving higher accuracy. In some cases, hybrid models combining CNNs with traditional classifiers are used to further enhance performance.

The dataset is then divided into **training, validation, and testing sets** to ensure unbiased evaluation. Techniques such as k-fold cross-validation are applied to improve generalization and prevent overfitting. During training, hyperparameters such as learning rate, number of layers, kernel size, and regularization parameters are tuned using optimization techniques like grid search or random search. Regularization methods such as dropout and early stopping are also applied in deep learning models to avoid overfitting and improve generalization capability.

The **model evaluation phase** involves assessing the performance of the trained models using standard metrics. Accuracy is used to measure overall correctness, while precision and recall evaluate the model's ability to correctly identify positive cases. The F1-score provides a balance between precision and recall, and the Receiver Operating Characteristic (ROC) curve along with Area Under the Curve (AUC) is used to evaluate classification performance across different thresholds. Confusion matrices are also analyzed to understand classification errors and misclassifications between different stages of the disease.

Finally, the proposed system undergoes **validation and deployment considerations**. The best-performing model is compared with existing diagnostic approaches to assess its clinical effectiveness. The system is designed to assist healthcare professionals by providing early-stage predictions, which can lead to timely treatment and better patient outcomes. Additionally, the model can be integrated into healthcare systems or mobile applications for real-time analysis and decision support. Ethical considerations such as data privacy, security, and bias reduction are also addressed to ensure responsible implementation.

Overall, this methodology provides a comprehensive and systematic approach to Alzheimer's Disease detection using machine learning, combining advanced data processing techniques and intelligent algorithms to achieve accurate and early diagnosis.

IV. CONCLUSION

In conclusion, the application of machine learning techniques for the detection of Alzheimer's Disease presents a promising and effective approach for early diagnosis and disease management. This study demonstrates that by integrating advanced data processing methods with intelligent algorithms, it is possible to accurately identify patterns associated with cognitive decline and neurodegeneration. The use of multimodal data, including neuroimaging and clinical information, significantly enhances the predictive capability of the system, enabling more reliable classification of different stages of the disease, such as normal cognition, mild cognitive impairment, and Alzheimer's Disease.

The implementation of various machine learning and deep learning models, including traditional classifiers and Convolutional Neural Networks, highlights the importance of selecting appropriate algorithms based on data type and complexity. The results indicate that deep learning models, particularly those applied to medical imaging data, offer superior performance due to their ability to automatically extract complex features. Furthermore, techniques such as feature selection, dimensionality reduction, and cross-validation contribute to improved model efficiency and generalization, reducing the risk of overfitting.

DOI: 10.48175/568

This research emphasizes the critical role of early detection in improving patient outcomes. By identifying Alzheimer's Disease at an initial stage, healthcare providers can implement timely interventions, slow disease progression, and enhance the quality of life for patients. The proposed methodology not only supports clinical decision-making but also reduces the dependency on manual diagnosis, which can be time-consuming and prone to human error.

However, certain challenges remain, including the availability of large and diverse datasets, data privacy concerns, and the need for model interpretability in clinical settings. Future work can focus on improving model transparency, incorporating real-time data, and developing user-friendly applications for practical deployment in healthcare environments.

Overall, this study concludes that machine learning-based systems have significant potential to transform Alzheimer's Disease diagnosis, making it more accurate, efficient, and accessible. With continued advancements and integration into medical practice, such technologies can play a vital role in addressing the growing global burden of neurodegenerative diseases.

V. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to all individuals and organizations who contributed to the successful completion of this research on Alzheimer's Disease detection using machine learning. First and foremost, we extend our heartfelt appreciation to our project guide and mentors for their continuous guidance, valuable suggestions, and constructive feedback throughout the research process. Their expertise and encouragement played a crucial role in shaping the direction and quality of this work.We are also thankful to our institution and faculty members for providing the necessary academic support, resources, and infrastructure required to carry out this study. Their constant motivation and access to technical facilities greatly facilitated the research and development process.Special thanks are extended to the providers of publicly available datasets and research communities whose contributions made it possible to access high-quality data for experimentation and analysis. Without these valuable resources, the implementation and evaluation of machine learning models would not have been feasible.We would also like to acknowledge our peers and colleagues for their support, insightful discussions, and collaboration during various stages of this work. Their inputs helped in refining ideas and improving the overall outcome of the research.Finally, we express our deep gratitude to our family members for their unwavering support, encouragement, and understanding throughout the duration of this research. Their patience and motivation inspired us to complete this work successfully.

REFERENCES

[1]. Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., et al. (2018). *NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease*. **Alzheimer's & Dementia**, 14(4), 535–562.

[2]. McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., et al. (2011). *The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups*. **Alzheimer's & Dementia**, 7(3), 263–269.

[3]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. **Nature**, 521(7553), 436–444.

[4]. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[5]. Suk, H. I., Lee, S. W., Shen, D., & Alzheimer's Disease Neuroimaging Initiative (ADNI). (2014). *Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis*. **NeuroImage**, 101, 569–582.

[6]. Sarraf, S., & Tofighi, G. (2016). *DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI*. **bioRxiv**.

[7]. Bron, E. E., Smits, M., Papma, J. M., Steketee, R. M., Meijboom, R., de Groot, M., et al. (2015). *Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI*. **NeuroImage**, 111, 562–579.

[8]. Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., et al. (2008). *Automatic classification of MR scans in Alzheimer's disease*. **Brain**, 131(3), 681–689.

[9]. Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). *Brain tumor segmentation using convolutional neural networks in MRI images*. **IEEE Transactions on Medical Imaging**, 35(5), 1240–1251.

[10]. Zhang, D., Wang, Y., Zhou, L., Yuan, H., & Shen, D. (2011). *Multimodal classification of Alzheimer's disease and mild cognitive impairment*. **NeuroImage**, 55(3), 856–867.

[11]. Ashburner, J., & Friston, K. J. (2000). *Voxel-based morphometry—The methods*. **NeuroImage**, 11(6), 805–821.

[12]. Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., et al. (2018). *Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease*. **IEEE Transactions on Biomedical Engineering**, 65(10), 2242–2251.

[13]. World Health Organization (WHO). (2021). *Dementia: A public health priority*. WHO Press.

[14]. Alzheimer's Association. (2022). *2022 Alzheimer's disease facts and figures*. **Alzheimer's & Dementia**, 18(4), 700–789.

[15]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). *Scikit-learn: Machine learning in Python*. **Journal of Machine Learning Research**, 12, 2825–2830.