

A Systematic Review on Machine Learning-Based Predictive Modeling for Early Risk Detection of Alzheimer's Disease

Tanmay Kasbe¹, Sailesh Iyer²

¹ Postdoctoral Researcher, Lincoln University College, 47301, Petaling Jaya, Malaysia.

² Supervisor, Lincoln University College, 47301, Petaling Jaya, Malaysia & Principal, Narnarayan Shastri Institute of Technology-IFSCS (Affiliated to NFSU)

Email ID : tanmay.kasbe@gmail.com, drsaileshiyer@gmail.com

Abstract: Alzheimer's disease (AD) is a progressive neurological condition that, especially in older adults, causes memory loss, cognitive decline, and a lower quality of life. For prompt intervention and better illness treatment, early identification of those who are at risk is essential. The capacity of machine learning (ML) approaches to evaluate complex medical datasets and facilitate early Alzheimer's disease prediction has drawn a lot of attention in recent years. Examining and summarizing previous studies on machine learning-based predictive models created for Alzheimer's disease early risk identification is the goal of this systematic review. The review examines research using machine learning algorithms like Support Vector Machines, Random Forest, Decision Trees, Neural Networks, and Deep Learning techniques that has been published in significant scientific databases. This analysis also covers frequently used datasets, such as clinical records, neuroimaging, cognitive assessment scores, and genetic data. When compared to conventional statistical techniques, the results show that machine learning models can greatly improve the precision and effectiveness of early Alzheimer's risk prediction. However, issues including model interpretability, data heterogeneity, and limited dataset availability continue to be major worries. This work offers a thorough summary of recent developments and suggests future lines of inquiry for creating more dependable and comprehensible machine learning models for early Alzheimer's disease risk assessment.

Keywords: Machine Learning (ML); Early Disease Prediction; Predictive Modeling in Healthcare; Disease Detection; Clinical Data Analysis; Deep Learning in Healthcare; Medical Data Mining; Artificial Intelligence in Healthcare.

Introduction

One of the leading causes of cognition in the world is Alzheimer's disease (AD), a progressive neurological condition. It is typified by a slow decline in memory, cognitive function, and behavioral abilities, which eventually impairs a person's capacity to carry out daily tasks. Global health statistics state that as the population ages, the number of persons with Alzheimer's disease is rising quickly, posing a serious threat to public health. In order to improve patient care, enable prompt medical intervention, and reduce the illness's course, early identification of Alzheimer's disease is essential. However, conventional diagnostic

methods frequently depend on clinical evaluations, cognitive testing, and neuroimaging methods, which can be costly, time-consuming, and occasionally inadequate for early identification.

[1].

Machine learning (ML) and artificial intelligence (AI) techniques have become viable tools for improving Alzheimer's disease detection and prediction with the development of computer technology. Large amounts of complicated medical data can be analyzed by machine learning algorithms, which might then find hidden patterns that traditional clinical analysis might miss. By learning from datasets like neuroimaging data, clinical records, cognitive test scores, and genetic information, these methods enable automated disease prediction. Because of its potential to aid in early identification and clinical decision-making in the diagnosis of Alzheimer's disease, machine learning-based predictive models have drawn a lot of attention in recent years.

[2].

Related work

Tanveer et al. [3] presented a comprehensive review of machine learning techniques used for the diagnosis of Alzheimer's disease (AD). The study analyzed approximately 165 research papers published between 2005 and 2019, focusing on the application of machine learning algorithms for automated AD detection. The authors categorized existing approaches into three major groups: Support Vector Machines (SVM), Artificial Neural Networks (ANN), and deep learning and ensemble-based methods. These techniques were widely applied to medical datasets, particularly neuroimaging data such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Diffusion Tensor Imaging (DTI), which are commonly used for identifying structural and functional brain abnormalities related to Alzheimer's disease.

Raza et al. [6] reviewed recent advancements in deep learning techniques for the early diagnosis of Alzheimer's disease using multimodal neuroimaging data. The study examined the role of deep learning models, particularly convolutional neural networks (CNNs), in analyzing brain imaging data such as MRI and PET scans. The authors discussed how multimodal data integration can improve the accuracy of Alzheimer's disease detection by combining structural, functional, and metabolic brain information. The review highlighted several deep learning architectures used for feature extraction and disease classification. Furthermore, the study addressed key challenges in implementing deep learning models, including limited dataset availability, high computational requirements, and issues related to model interpretability. The authors suggested that future research should focus on developing explainable artificial intelligence models and improving data-sharing frameworks to enhance the reliability and clinical adoption of deep learning-based diagnostic systems.

Arya et al. [7] conducted a systematic review of machine learning and deep learning techniques used for the effective diagnosis of Alzheimer's disease. The study analyzed a wide range of research papers focusing on computational methods applied to neuroimaging and clinical datasets. The authors categorized the reviewed studies based on machine learning algorithms, datasets, and evaluation

techniques used for Alzheimer's disease detection. Their findings indicated that support vector machines, random forests, and convolutional neural networks are among the most frequently used algorithms in this research domain. The review also highlighted the growing importance of deep learning approaches for automatically extracting complex features from large medical datasets. Additionally, the authors emphasized the need for improved feature selection methods and standardized evaluation protocols. The study provides a comprehensive overview of current research trends and identifies opportunities for developing more accurate and robust machine learning models.

Mirzaei and Adeli [9] explored machine learning techniques for diagnosing Alzheimer's disease, mild cognitive impairment (MCI), and other types of dementia. The study reviewed various computational methods used for analyzing biomedical signals and neuroimaging data. The authors focused on the application of machine learning algorithms for detecting neurological disorders and differentiating between different stages of cognitive impairment. The research highlighted the effectiveness of machine learning-based diagnostic systems in identifying subtle changes in brain structure and function associated with Alzheimer's disease. Additionally, the study emphasized the importance of integrating multiple data sources, including imaging data and clinical information, to improve classification performance. Despite promising results, the authors noted challenges related to data heterogeneity and limited availability of standardized datasets. The study concluded that machine learning techniques have strong potential to enhance the accuracy and efficiency of neurological disorder diagnosis.

Nayaki et al. [10] investigated the application of machine learning models for Alzheimer's disease detection using medical datasets. The study compared several machine learning classification algorithms to evaluate their effectiveness in identifying Alzheimer's disease. The authors implemented different models and analyzed their performance using evaluation metrics such as accuracy, precision, and recall. The results indicated that machine learning techniques can successfully detect patterns in medical data and support early diagnosis of Alzheimer's disease. The study also highlighted the importance of feature selection techniques in improving the performance of predictive models. By selecting the most relevant features from clinical and cognitive datasets, the researchers were able to enhance classification accuracy. However, the authors noted that larger datasets and more advanced models are required to further improve prediction reliability and ensure practical applicability in healthcare environments.

Shuja et al. [11] explored the use of deep learning techniques for the diagnosis of Alzheimer's disease. The study focused on applying deep neural network models to analyze neuroimaging data and detect early signs of Alzheimer's disease. Deep learning algorithms, particularly convolutional neural networks, were used to automatically extract features from brain imaging datasets. The research demonstrated that deep learning models can effectively identify structural brain changes associated with Alzheimer's disease and achieve high classification accuracy. The authors also discussed the advantages of deep learning methods over traditional machine learning approaches, particularly in handling large and complex medical datasets. Despite these benefits, the study highlighted several challenges, including the need for large annotated datasets and high computational resources. The findings suggest that deep learning-based diagnostic systems hold significant potential for improving early detection of Alzheimer's disease.

Literature Gap

Although significant progress has been made in applying machine learning and deep learning techniques for the diagnosis and prediction of Alzheimer’s disease, several research gaps still exist. **First**, many existing studies focus on achieving high classification accuracy using specific datasets; however, the generalizability of these models across different datasets and clinical environments remains limited. Variations in data collection methods, imaging protocols, and patient demographics often lead to reduced model performance when applied to external datasets. **Second**, most research relies heavily on single-modality datasets, particularly neuroimaging data such as MRI or PET scans. While these datasets provide valuable structural and functional information about the brain, relying on a single data source may limit the effectiveness of predictive models. Integrating multimodal data sources, including clinical records, cognitive test scores, genetic information, and neuroimaging data, is still an underexplored area that could significantly improve prediction accuracy. Another important limitation is the lack of model interpretability and explainability. Many advanced machine learning and deep learning models, particularly deep neural networks, function as “black-box” systems, making it difficult for healthcare professionals to understand how predictions are generated. This lack of transparency can reduce the trust and adoption of such models in real clinical environments.

Furthermore, several studies are conducted using relatively small and imbalanced datasets, which may lead to overfitting and biased predictions. The availability of large, standardized, and publicly accessible datasets remains a major challenge in Alzheimer’s disease research. Additionally, there is limited research on real-time clinical deployment and validation of machine learning models, which is necessary to ensure their reliability in practical healthcare settings. Finally, many review and survey papers focus primarily on summarizing existing algorithms rather than systematically analyzing model performance, datasets, and evaluation methods. Therefore, there is a need for a comprehensive systematic review that critically evaluates machine learning-based predictive models, identifies methodological trends, and highlights future research directions for improving early risk detection of Alzheimer’s disease.

Table 1. Comparison study of previous work with limitations and key contribution

Ref	Author & Year	Study Type	Data Type / Dataset	ML / DL Techniques	Key Contribution	Limitations
1	Lodha et al., 2018	Research Paper	Clinical dataset	Machine Learning classifiers	Developed ML-based diagnostic system for AD detection	Limited dataset and feature diversity
2	Aghdam et al., 2025	Survey	Neuroimaging datasets	Various ML models	Evaluated reproducibility and generalizability of	Lack of standardized datasets

					ML models for AD diagnosis	
3	Tanveer et al., 2020	Review Paper	Multiple datasets	SVM, Decision Trees, Neural Networks	Comprehensive review of ML techniques used for AD diagnosis	Focused mainly on traditional ML methods
4	Diogo et al., 2022	Research Study	Multi-diagnostic clinical and imaging data	Machine Learning models	Proposed a generalizable ML framework for early AD diagnosis	Complexity in multi-dataset integration
5	Jumaili & Sonuç, 2025	Research Paper	Clinical and neuroimaging data	Deep Ensemble Learning	Developed ensemble ML model improving prediction accuracy	High computational complexity
6	Raza et al., 2025	Review Study	Multimodal neuroimaging	Deep Learning models (CNN, etc.)	Analyzed advancements in DL for early AD detection	Data availability and interpretability challenges
7	Arya et al., 2023	Systematic Review	Multiple datasets	ML and Deep Learning techniques	Comprehensive review of ML/DL approaches for AD diagnosis	Limited comparison of algorithm performance
8	Dara et al., 2023	Survey	Neuroimaging and clinical datasets	ML algorithms	Survey on ML-based AD diagnosis methods	Need for improved model generalization
9	Mirzaei & Adeli, 2022	Research Review	Biomedical signals and imaging	ML algorithms	ML techniques for diagnosing AD, MCI, and dementia	Dataset heterogeneity
10	Nayaki et al., 2024	Research Paper	Medical datasets	ML classification models	Compared ML algorithms for AD detection	Limited scalability for large datasets

11	Shuja et al., 2023	Conference Paper	Neuroimaging datasets	Deep Learning techniques	Applied DL models for automated AD diagnosis	Requires large training datasets
----	--------------------	------------------	-----------------------	--------------------------	--	----------------------------------

Conclusions

This systematic review analyzed recent research on the application of machine learning and deep learning techniques for the early detection and prediction of Alzheimer’s disease. The reviewed studies demonstrate that machine learning algorithms such as Support Vector Machines, Random Forest, Decision Trees, Neural Networks, and deep learning models can effectively analyze medical and neuroimaging datasets to improve diagnostic accuracy. These approaches enable automated identification of disease-related patterns, supporting early diagnosis and assisting healthcare professionals in clinical decision-making. However, several challenges remain, including limited dataset availability, data heterogeneity, lack of model interpretability, and insufficient validation across diverse populations. Addressing these limitations through the integration of multimodal data, development of explainable models, and use of larger standardized datasets can significantly enhance the reliability and clinical applicability of machine learning-based diagnostic systems. Overall, machine learning techniques show strong potential for improving early risk prediction and supporting effective management of Alzheimer’s disease.

References

1. P. Lodha, A. Talele and K. Degaonkar, "Diagnosis of Alzheimer's Disease Using Machine Learning," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-4. doi: 10.1109/ICCUBEA.2018.8697386.
2. Aghdam, Maryam Akhavan et al. "Machine-learning models for Alzheimer's disease diagnosis using neuroimaging data: survey, reproducibility, and generalizability evaluation." Brain informatics vol. 12,1 8. 21 Mar. 2025, doi:10.1186/s40708-025-00252-3.
3. M. Tanveer et al., "Machine Learning Techniques for the Diagnosis of Alzheimer’s Disease," ACM Trans. Multimedia Comput. Commun. Appl., vol. 16, no. 1s, pp. 1–35, Jan. 2020, doi: 10.1145/3344998.
4. V. S. Diogo, H. A. Ferreira, and D. Prata, "Early diagnosis of Alzheimer’s disease using machine learning: a multi-diagnostic, generalizable approach," Alz Res Therapy, vol. 14, no. 1, Aug. 2022, doi: 10.1186/s13195-022-01047-y.
5. M. L. F. Jumaili and E. Sonuç, "ML-Driven Alzheimer’s disease prediction: A deep ensemble modeling approach," SLAS Technology, vol. 32, p. 100298, Jun. 2025, doi: 10.1016/j.slast.2025.100298.

6. M. L. Raza et al., "Advancements in deep learning for early diagnosis of Alzheimer's disease using multimodal neuroimaging: challenges and future directions," *Front. Neuroinform.*, vol. 19, May 2025, doi: 10.3389/fninf.2025.1557177.
7. A. D. Arya et al., "A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease," *Brain Inf.*, vol. 10, no. 1, Jul. 2023, doi: 10.1186/s40708-023-00195-7.
8. O. A. Dara, J. M. Lopez-Guede, H. I. Raheem, J. Rahebi, E. Zulueta, and U. Fernandez-Gamiz, "Alzheimer's Disease Diagnosis Using Machine Learning: A Survey," *Applied Sciences*, vol. 13, no. 14, p. 8298, Jul. 2023, doi: 10.3390/app13148298.
9. G. Mirzaei and H. Adeli, "Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia," *Biomedical Signal Processing and Control*, vol. 72, p. 103293, Feb. 2022, doi: 10.1016/j.bspc.2021.103293.
10. A. P. Nayaki et al., "Alzheimer's Disease Detection using Machine Learning Models," 2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT). IEEE, pp. 1–8, Jul. 24, 2024. doi: 10.1109/iceeict61591.2024.10718398.
11. MM. Shuja, S. J. Sidiq, K. Dutt, D. Kaur, M. Kumar, and K. Bajaj, "Use of Deep Learning Techniques in Alzheimer's Disease Diagnosis," 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI). IEEE, pp. 1–7, Dec. 29, 2023. doi: 10.1109/icaiihi57871.2023.10489461.