EARLY STAGE OF ALZHEIMER'S DISEASE DETECTION USING MACHINE LEARNING

Hemavathi Patil[1], Asha Rani[2], Sowmya Bellur[3], Tanuja[4], Vaishnavi D Patil[5]

[1,2,3,4,5]Department Of CSE (Data Science), Guru Nanak Dev Engineering College Bidar, Karnataka, India.

hbpatilbidar@gmail.com ,ranis61685@gmail.com , sowmyabellur04@gmail.com, tanujachendapure@gmail.com , vaishnavipatil0582@gmail.com

Abstract

Alzheimer's disease is a progressive brain disorder that mainly affects elderly people, gradually impairing memory, reasoning ability, and behavior. As the global population continues to age, cases of this condition are increasing, making it a serious concern for healthcare systems. Because of this rise, there is a growing need for early and accurate detection methods. Recent advancements in machine learning have made it possible to analyze large medical datasets more effectively and identify patterns that may not be obvious through manual observation. This study focuses on using such techniques to support early detection of Alzheimer's disease so that patients can receive timely medical attention. For this purpose, multiple machine learning models were implemented, including Decision Tree, Random Forest, Support Vector Machine (SVM), Gradient Boosting, and Voting Classifier. The OASIS dataset was used for training and testing, and the models were evaluated using standard performance metrics to compare their effectiveness. The experimental results show that Random Forest performed the most consistently, achieving accuracy in the mid-80% range compared to other models. This indicates that ensemble-based methods can be particularly useful for this type of medical prediction task. Overall, the study highlights how machine learning can support early diagnosis of Alzheimer's disease. Detecting the condition at an early stage can help slow its progression and improve the quality of life for affected individuals.

Keywords- Alzheimer's disease, machine learning, early diagnosis, dementia prediction, MRI analysis, feature selection, brain imaging.

1. Introduction

Alzheimer's disease is a brain disorder that slowly affects memory, thinking ability, and behavior. It usually starts with small memory problems, like forgetting recent events or names, but over time, it can become severe and interfere with daily life. Many people mistake these early signs as normal aging, which often delays proper diagnosis. Because of this, identifying the disease at an early stage becomes very important.

Alzheimer's disease is becoming more common worldwide, especially as people are living longer than before. With increasing age, a larger population is now at risk of developing this condition. This situation affects not only the patients themselves but also places emotional pressure and financial burden on their families as well as healthcare services. Since a complete cure for Alzheimer's disease has not yet been found, identifying the disease at an early stage becomes very important. Early diagnosis can help slow its progression and support patients in maintaining better daily functioning and quality of life.

In the traditional approach, diagnosing Alzheimer's disease involves several medical examinations, specialist consultation, and the use of advanced brain imaging techniques. This process often takes a lot of time and can be costly for patients. To address these limitations, researchers are increasingly exploring machine learning methods to support early detection. These models are capable of processing large volumes of medical data, discovering patterns that are not easily noticeable, and generating reliable predictions, which can assist healthcare professionals in improving diagnostic decisions [1], [2].

This process makes use of various types of data, such as MRI brain images, patient clinical records, and cognitive assessment results like the MMSE score. These different inputs help machine learning models better understand brain health and identify early indicators of dementia. Several research studies have also found that integrating medical data with machine learning techniques can improve the accuracy of disease prediction [3].

Modern technology, particularly machine learning, plays an important role in this area. Machine learning, a branch of artificial intelligence, enables computers to learn from data and make informed predictions. Rather than depending only on human analysis, these systems can efficiently process large volumes of medical information and detect patterns that may not be easily noticed by humans. As a result, it helps improve diagnostic accuracy and reduces the likelihood of errors.

In this work, the objective is to design a system based on machine learning that can detect Alzheimer's disease at an early stage. By working with structured medical data and applying different algorithms, the system classifies individuals as either demented or non-demented. This method helps improve prediction accuracy while also assisting healthcare professionals in making quicker and more dependable decisions.

2. Related Work

Many researchers are currently using machine learning and deep learning techniques to support the early detection of Alzheimer's disease, particularly through the analysis of brain imaging datasets such as OASIS and ADNI. These approaches have gained popularity because they are capable of identifying complex patterns that are often difficult to detect through traditional methods. In addition, similar techniques are also being applied to other neurological conditions, including brain tumors, epilepsy, and Parkinson's disease. This highlights the versatility and effectiveness of these technologies in medical applications.

The extracted features were then classified using two machine learning algorithms, namely K-Nearest Neighbor (KNN) and Extreme Learning Machine (ELM). In the case of binary classification, the method achieved an accuracy of 96.88% with KNN and 98.97% with ELM. These results were obtained despite the relatively small dataset and the overall complexity of the process. [10]

Chaves et al. proposed a method for discovering association rules from SPECT imaging data to support the early detection of Alzheimer's disease. The study included 97 participants, out of which 43 were healthy individuals and 54 were diagnosed with Alzheimer's disease. The proposed technique achieved an accuracy of 95.87%, with a sensitivity of 100% and a specificity of 92.86%. It also demonstrated faster processing and required fewer computational resources compared to PCA-SVM and GMM-SVM methods. Nevertheless, the dependence on unconfirmed pathologically data could impair the consistency and accuracy of the results. In 2011, Chaves et al. continued working on the earlier research by recognizing perfusion patterns in the SPECT images for diagnosing Alzheimer's disease. This algorithm provided 94.87% accuracy, 91.07% sensitivity, and 100% specificity. Yet, the confidence in the classification results was affected by drawbacks, including incomplete data and databases that had no pathological confirmation. [3]

The deep learning methodology adopted by Martinez-Murcia et al. consisted of convolutional autoencoders that were used to identify features from MRI images. These features included symptoms of cognitive decline and signs of neurodegenerative changes. After extracting these features, the researchers applied both regression and classification methods using imaging-based measures such as MMSE and ADAS-Cog (C11) scores. This approach achieved more than 80% accuracy in predicting Alzheimer's disease. The method used by Prajapati et al. was a deep neural network–based model for binary classification of Alzheimer's disease patients. This network included several hidden layers, with activation functions for better learning. Their findings showed that deep learning techniques significantly improved the classification accuracy. [1]

Turrisi et al. (2024) reviewed contemporary deep learning methods for MRI data. Previous studies had used CNNs, adversarial learning, and GANs for predicting Alzheimer's disease progression and identifying unusual brain connections. They were effective in improving the classification rate and aiding in the automatic recognition of various stages of Alzheimer's disease. Turrisi et al. (2024) emphasized that both the dataset size and model architecture play critical roles in influencing the effectiveness of models. Earlier studies utilized datasets containing between 170 and 1662 subjects, making direct and fair comparisons challenging. In addition, variations in evaluation metrics and the division of model training and validation data resulted in inconsistent performance outcomes. [9]

3. Methodology

The proposed system focuses on detecting Alzheimer's disease as early as possible, especially when the first signs of cognitive decline begin to appearIt uses machine learning models to make early detection more efficient and accurate. The entire process follows a systematic sequence, beginning with data collection, followed by data cleaning and preprocessing. After that, the data is explored, important features are selected, and the model is developed. Finally, the performance of the model is evaluated and validated to ensure its reliability.

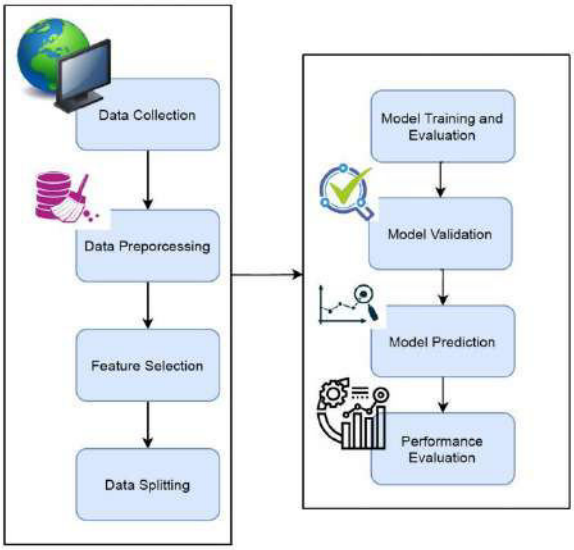

Fig1. Proposed Workflow [1]

3.1 Data Collection

In this study, the dataset is taken from the Open Access Series of Imaging Studies (OASIS). It includes long-term MRI scans along with medical records of older adults. The data set consists of about 150 people ranging from ages 60 to 96, including dementia and non-dementia samples. In this study, several important factors were considered, including age, gender, education level, and socio-economic status. Clinical assessment scores such as MMSE and CDR were also included, along with brain-related measurements like eTIV, nWBV, and ASF.

3.2 Data Preprocessing

The raw dataset may include missing entries, duplicated records, and inconsistencies that can reduce the effectiveness of the predictive model. Hence, preprocessing is performed as an important step before model training.

In this stage, missing values in variables such as SES are handled using the median imputation method. Categorical data is converted into numerical form through techniques like label encoding. To improve data quality, duplicate entries and irrelevant records are removed from the dataset. In addition, numerical features are normalized so that all variables are brought to a common scale, which helps improve model performance.[6]

3.3 Exploring The Data

Exploring the data (EDA) will be carried out to analyze the connection between the independent factors and whether the patient is diagnosed with dementia. Statistical tools are used alongside graphical representation to analyze the data distribution.

The analysis mainly focuses on comparisons based on:

- **age**
- **gender**
- **MMSE score**
- **education level**
- **brain volume parameters**

This stage helps in recognizing meaningful patterns, such as decreased brain volume and lower MMSE scores among demented patients.[8]

3.4 Selecting important features

Selecting important features s used to figure out which ones matter most relevant predictors and remove less significant variables from the dataset.

The major selected features include:

- **Age**
- **MMSE**
- **CDR**
- **EDUC**
- **SES**
- **eTIV**
- **nWBV**
- **ASF**

We use techniques like correlation analysis, information gain, and chi-square tests to figure out which features really matter. This helps the model perform better while keeping things simple and not too heavy on computation.[3]

3.5 Data Splitting

The dataset is split into two sections: a training set and a testing set. About 80% of the data is used to train the model, while the remaining 20% is reserved for evaluating how well the model performs on unseen data.

An **80:20 ratio** is adopted, where:

- **80% of the samples is allocated for training**
- **20% of the samples is reserved for testing**

This method helps the model learns from sufficient data while also being evaluated on previously unseen samples.

3.6 Model Development

We use several machine learning methods to sort patients as either demented or non-demented.

A range of models was tested in this work.

- **Decision Tree**
- **Random Forest**
- **Support Vector Machine (SVM)**
- **XGBoost**
- **Voting Classifier**

Each model is learned using the training data and then fine-tuned to improve its classification performance.[1]

3.7 Model Validation

To minimize overfitting and enhance generalization capability, k-fold cross-validation is employed.

In this analysis, both five-fold and ten-fold cross-validation methods are used. To begin with, the dataset is divided into ten equal sections.

After this, the model is trained and evaluated in rotation. In each iteration, one portion of the data is used for testing, while the remaining nine parts are used for training. [7]

3.8 Performance Evaluation

We checked how well the models performed using common measures like:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-score**

These metrics help evaluate and compare the performance of different machine learning algorithms.

The results indicate that the Random Forest model gave the best accuracy at 86.92%, while XGBoost followed closely with an accuracy of 85.92%.

3.9 Final Prediction

The classifier that performs the best is selected as the final model for predicting early-stage Alzheimer's disease.

The final system predicts whether a patient belongs to:

- **non-demented**
- **early-stage demented**

This facilitates early detection and supports timely medical intervention and treatment planning.[3]

4. Results and Discussion

This section presents the results obtained from applying machine learning models for the early detection of Alzheimer's disease. It examines the performance of each model in detail and compares their effectiveness. The findings help in understanding how accurately the models can differentiate between individuals with dementia and those without it.

Results

The suggested machine learning framework was successfully evaluated using the **OASIS dataset**, which includes clinical information and MRI-based patient features. The system efficiently processed the input data, identified relevant patterns, and accurately classified the patient records.

Model Performance Evaluation

In this study, we implemented and compared several machine learning models, including Decision Tree, Random Forest, Support Vector Machine (SVM), XGBoost, and a Voting Classifier. After testing them, Random Forest turned out to perform the best, achieving achieving an accuracy of 86.92%, which was better than the other models.

Classification Accuracy

The developed system showed effective performance in differentiating demented and non-demented cases. In addition, both XGBoost and the Voting Classifier showed strong performance, achieving accuracies above 85%. This further supports the effectiveness of ensemble learning techniques when applied to this healthcare dataset.

Feature-Based Prediction

The prediction model effectively utilized significant features such as:

- **Age**
- **MMSE score**
- **CDR score**
- **Education level**
- **Brain volume measurements (nWBV)**

These features contributed significantly in enhancing prediction accuracy and supported the recognition of early symptoms of Alzheimer's disease.[10]

Discussion

The results indicate that the machine learning-based prediction model is effective in detecting early-stage Alzheimer's disease.

Performance Efficiency

The classifier was able to give accurate predictions with an extremely low error rate. Using preprocessing methods, feature selection, and cross-validation played an important role in making the model more robust and preventing overfitting.

Accuracy and Reliability

Among all the classifiers evaluated, Random Forest proved to be the most consistent and reliable model. This is mainly because it brings together multiple decision trees that helps make its predictions more stable and accurate. We also noticed that ensemble models like Random Forest and XGBoost performed better overall and gave more consistent results compared to single models like Decision Tree and SVM.

Clinical Significance

The results indicate a strong relationship between lower MMSE scores and reduced brain size, which is closely associated with the presence of dementia.The above statement corresponds to the features observed in Alzheimer's disease patients.

System Challenges

There were a few limitations in the proposed system, including:

- **limited dataset size**
- **presence of missing values**
- **dependence on selected features**
- **lower generalization for external datasets**

These limitations may influence the model's performance when applied to larger, real-world medical datasets.[8]

Future Improvements

The proposed system can be further improved by:

- **implementing advanced neural network methods**
- **using larger MRI datasets**
- **applying CNN-based image models**
- **enhancing multiclass stage classification**

These improvements can further improve the precision of early diagnosis and provide better support for clinical decision-making.

5. Conclusion

This study introduces a machine learning-based approach for detecting Alzheimer's disease at an early stage using the OASIS dataset. The main idea is to catch the disease as early as possible by looking at key factors like age, MMSE score, CDR score, education level, and normalized whole-brain volume. Instead of relying on a single step, we followed a clear and structured process that included preparing the data, selecting the most useful features, building different models, and carefully evaluating their performance.

To identify the most effective method, several machine learning models were implemented, including Decision Tree, Random Forest, Support Vector Machine (SVM), XGBoost, and a Voting Classifier. The performance of each model was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. After comparing the results, Random Forest emerged as the best-performing model, achieving an accuracy of 86.92%. XGBoost also showed strong performance with an accuracy of 85.92%, making it the second-best model.

These results clearly show that machine learning can be very effective in predicting Alzheimer's disease, especially when the right features and models are used. More importantly, such systems can support doctors by providing quick and reliable predictions, helping them make better decisions during diagnosis.

Looking ahead, there is still scope for improvement. The system can be further improved by using larger and more diverse MRI datasets, adopting advanced deep learning methods such as Convolutional Neural Networks (CNNs), and integrating multiple models to enhance overall performance. With further development, this kind of approach has strong potential to become a valuable tool in real-world healthcare settings.

References

1. C. Kavitha, V. Mani, S. R. Srividhya, O. I. Khalaf, and C. A. Tavera Romero, "Early-stage Alzheimer's disease prediction using machine learning models," *Frontiers in Public Health*, vol. 10, art. no. 853294, Mar. 2022.

2. V Patil, M. Madgi, and A. Kiran, "Early prediction of Alzheimer's disease using convolutional neural network: a review," *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, vol. 58, no. 130, 2022.

3. A.Khan and M. Usman, "Early diagnosis of Alzheimer's disease using machine learning techniques: a review paper," in *Proc. 7th Int. Joint Conf. Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) – Vol. 1: KDIR*, Lisbon, Portugal, 2015, pp. 380–387.

4. C. S. Eke, E. Jammeh, X. Li, C. Carroll, S. Pearson, and E. Ifeachor, "Early detection of Alzheimer's disease with blood plasma proteins using support vector machines," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 218–227, Jan. 2021.

5. S. Liu, A. V. Masurkar, H. Rusinek, J. Chen, B. Zhang, W. Zhu, C. Fernandez-Granda, and N. Razavian, "Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs," Scientific Reports, vol. 12, no. 17106, 2022.

6. A. M. Ahmed, F. A. Hanandeh, A. M. Soliman, A. M. Gaber, A. Sharif, M. O. Al Enany, and M. M. Khattab, "Early Alzheimer's Disease Detection with Multiple Machine Learning Models," Research Square, Preprint, Dec. 30, 2024.

7. P. K. Mandal and R. V. Mahto, "Deep Multi-Branch CNN Architecture for Early Alzheimer's Detection from Brain MRIs," Sensors, vol. 23, no. 8192, Sep. 2023.

8. S. W. Park, N. Y. Yeo, J. Lee, S.-H. Lee, J. Byun, D. Y. Park, S. Yum, J.-K. Kim, G. Byeon, Y. Kim, and J.-W. Jang, "Machine learning application for classification of Alzheimer's disease stages using 18F-flortaucipir positron emission tomography," BioMedical Engineering OnLine, vol. 22, no. 40, 2023.

9. R. Turrisi, A. Verri, and A. Barla, "Deep learning-based Alzheimer's disease detection: reproducibility and the effect of modeling choices," Frontiers in Computational Neuroscience, vol. 18, Sep. 2024.

10. A. A. Abd El-Latif, S. A. Chelloug, M. Alabdulhafith, and M. Hammad, "Accurate Detection of Alzheimer's Disease Using Lightweight Deep Learning Model on MRI Data," Diagnostics, vol. 13, no. 1216, Mar. 2023.

11. G. M. McKhann et al., "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," Alzheimer's Dementia, vol. 7, no. 3, pp. 263–269, 2011.

12. Sharma, S.; Gupta, S.; Gupta, D.; Altameem, A.; Saudagar, A.K.J.; Poonia, R.C.; Nayak, S.R. HTLML: Hybrid AI Based Model for Detection of Alzheimer's Disease. Diagnostics 2022, 12, 1833.

13. Mehmood A, Yang S, Feng Z, Wang M, Ahmad AS, Khan R, et al. A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images. Neuroscience. 2021.

14. Mohammed, B.A.; Senan, E.M.; Rassem, T.H.; Makbol, N.M.; Alanazi, A.A.; Al-Mekhlafi, Z.G.; Ghaleb, F.A. Multi-Method Analysis of Medical Records and MRI Images for Early Diagnosis of Dementia and Alzheimer's Disease Based on Deep Learning and Hybrid Methods. Electronics 2021, 10, 2860.

15. Murugan, S.; Venkatesan, C.; Sumithra, M.G.; Gao, X.Z.; Elakkiya, B.; Akila, M.; Manoharan, S. DEMNET: A Deep Learning Model for Early Diagnosis of Alzheimer Diseases and Dementia from MR Images. IEEE Access 2021.

16. Pirrone, D.; Weitschek, E.; Di Paolo, P.; De Salvo, S.; De Cola, M.C. EEG Signal Processing and Supervised Machine Learning to Early Diagnose Alzheimer's Disease. Appl. Sci. 2022

17. Veeramuthu, A., S. Meenakshi, et al. (2014). A New Approach for Alzheimer's Disease Diagnosis by using Association Rule over PET Images. International Journal of Computer Applications 91(9), 9-14.

18. Bangyal, W.H.; Rehman, N.U.; Nawaz, A.; Nisar, K.; Ibrahim, A.A.A.; Shakir, R.; Rawat, D.B. Constructing Domain Ontology for Alzheimer Disease Using Deep Learning Based Approach. Electronics 2022.

19. Huanhuan Ji, Zhenbing Liu, Wei Qi Yan, Reinhard Klette. Early Diagnosis of Alzheimer's Disease Using Deep Learning, ICCCV 2019: In Proceedings of the 2nd International Conference on Control and Computer Vision. 2019.

20. Bae, J.B.; Lee, S.; Jung, W.; Park, S.; Kim, W.; Oh, H.; Kim, K.W. Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. Sci. Rep. 2020, 10, 22252. [CrossRef] 12. Ebrahimi, A.; Luo, S. Convolutional neural networks for Alzheimer's disease detection on MRI images. J. Med. Imaging 2021,