WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Early Prediction Of Alzheimer's Disease With Local Outlier Algorithm (LOF) And Random Forest

Lalitha Bhavani Konkyana[1], Jaya Eppili[2], Hemalatha Kolla[3],Eeswar Yarpatni[4]

[1, 2, 3& 4] Dept .of Electronics and Communication Engineering, Aditya Institute of Technology and Management

ABSTRACT:

Alzheimer's disorder is an intensifying neurodegenerative disease that seriously damages Neuropsychological function and degrades the standards of life. Pre-diagnosis is necessary to carry out well-timed assistance to decelerate disease advancement and upgrade the quality life of patient. Machine learning has arisen as a proven technique for helping the timely detection of Alzheimer's disorder by evaluating a wide range of data sources, which include neuro-imaging, healthcare records and genetic information. This paper is to assess new developments in machine learning-based methods for finding Alzheimer's disease at its initial phase. The application of various ML algorithms, such as neural networks, support vector machines and ensemble methods are investigated in studying multimodal datasets to recognize patterns associated with Alzheimer's. In addition to that essential tasks in Alzheimer's disorder recognition, including the heterogeneity of data, the explicability of ML models, and the capability to generalize findings across various people are examined. Eventually, future possibilities in this area, focusing the potential of deep learning architectures, evolution of robust and interpretable models, and the integration of longitudinal data to improve reliability of prediction accuracy are discussed. Using machine learning for prior Alzheimer's recognition provides great potential for easing timely interventions and enhancing the level of care for affected sufferers.

Keywords: Alzheimer's disease diagnosis; machine learning; Neuropsychological function

Introduction :

Alzheimer's disease (AD) is a stressful neurologic loss disease depicted by memory loss, Loss of analytic skills and difficulties in daily functioning. As the global population ages, the prevalence of AD is projected to rise significantly, imposing a heavy burden on healthcare systems and societies worldwide [1]. A key challenge in managing AD lies in the lack of effective treatments, often due to delayed diagnoses that occur after irreversible brain damage has already taken place. Consequently, there is a growing focus on early detection strategies to enable timely interventions and enhance patient outcomes. Machine learning (ML) has emerged as a transformative tool in medical research and clinical practice, capable of analyzing complex datasets to identify subtle patterns associated with disease onset and progression. In the context of AD, ML techniques offer significant promise for early detection by leveraging diverse data sources, including neuroimaging, genetic markers, and clinical assessments. By integrating data from multiple modalities, ML algorithms can uncover hidden relationships and build predictive models to identify individuals at risk of developing the disease.

This paper explores the potential of ML in enabling early diagnosis of Alzheimer's disease and examines its implications for improving patient care and disease management. Harnessing the power of machine learning brings us closer to achieving timely interventions and advancing personalized medicine in the fight against this disorder [2] to [5]. In the current landscape, traditional diagnostic methods and human intuition often fail to deliver accurate results, which is why it is crucial to rely on more computationally intensive and unconventional approaches like machine learning. The use of machine learning in disease prediction and visualization is part of a broader shift toward predictive and personalized medicine. This movement is essential not only for improving the quality of life for patients but also for assisting physicians in making better treatment decisions and helping health economists optimize healthcare resources. Alzheimer'sdisease. is not merely an age-related condition but a brain disorder. It manifests in symptoms such as memory loss, difficulty finding the right words or understanding speech, challenges in performing everyday tasks, and noticeable changes in personality and mood. These symptoms underscore the complexity of diagnosing AD. When radiologists review medical reports, such as MRI scans, their diagnostic biases can lead to missed opportunities to identify other possible conditions. This limited focus often leads to considering only a narrow range of causes. According to C.S. Lee, about 75% of medical errors are due to diagnostic mistakes made by radiologists. Factors such as high workload, stress, fatigue, cognitive biases, and inadequate systems contribute to these errors. In such situations, intelligent diagnostic systems can offer valuable clinical support, reducing the risk of errors and improving patient outcomes. The Alzheimer's Association reports that AD is the sixth leading cause of death in the United States, highlighting the urgent need for improved diagnostic techniques and better treatment options. Traditional diagnostic tools like MRI and PET scans provide useful structural and functional insights but are limited in detecting early-stage AD, leading to delays in diagnosis. Machine learning offers a potential solution by automating pattern recognition and improving the accuracy of early diagnosis. Machine learning has shown significant promise in the early detection of Alzheimer's disease by analyzing various data modalities. These include neuroimaging data (MRI, PET), clinical data (cognitive scores, medical history), and genetic markers.

Literature Survey :

A unique machine learning classifier was proposed by Zhang et al. to categorize Alzheimer's disease. To find AD, they used Predator-Prey Particle Swarm Optimization and Stationary Wavelet Entropy. They used MRI pictures to identify AD. Though it could be better, their precision was adequate. Furthermore, their work only employed a relatively tiny dataset [6]. Various ML algorithm types were compared by Vidushi et al. in order to determine the optimal accuracy for AD detection. They discovered that boosting AdaBoost (82.5%) and random forest (84%), respectively, produced the highest accuracy [7]. However, by extracting additional features and filling in their null parameter with the median value, they may improve accuracy. Alejandro Puente-castro et.al presented on Automatic Assessment of Alzheimer's Disease Diagnosis based on Deep Learning Techniques [8].The main objective of this research work is to develop a system that automatically detects the presence of the disease in sagittal magnetic resonance images (MRI), which are not generally used. Sagittal MRIs from ADNI and OASIS data sets were employed. Sitara Afzal et.al reviewed Disease Detection Techniques and methods.Alzheimer's disease related brain alterations can be measured via neuroimaging [9].Through categorization frameworks that offer diagnostic and prognostic tools, these parameters have recently been incorporated. This thorough analysis of work on Alzheimer‘s Disease focuses on computer-aided diagnosis. The review covers imaging techniques such positron emission tomography, amyloid-PET, diffusion tensor imaging, functional magnetic resonance imaging and magnetic resonance imaging. A.Gamalet.al proposed automatic Early Diagnosis of Alzheimer's Disease using 3D Deep Ensemble Approach [10].This proposed system is content-based image retrieval system that relies on 3D Capsules Network a 3D CNN and a pre-trained 3D auto-encoder technology to detect disease at its initial stages. They used SVM with kernels that allowed for the switching of MCI to Alzheimer's. The highest accuracy achieved is 70.33% in distinguishing Alzheimer's using a content-based image retrieval system that relied on 3D Capsules Network. C.Chabib et.al developed Deep Convolutional Curvelet Transform-based MRI Approach for Early Detection of Alzheimer's Disease in new approach for early detection of Alzheimer's disease using MRI images, called Deep Curv MRI [11]. The model combines curvelet transform and convolutional neural network to improve the accuracy of Disease diagnosis in its early stages. The process involves MRI images, feature extraction using curvelet transform, and classification using a convolutional neural network.The higher accuracy achieved by the Deep Curvi MRI model for multi- classification task and for binary classification task.

Experimental Data Analysis and Preparation :

Alzheimer's disease detection methods vary depending on the stage of the disease and the data available for analysis. ML algorithms analyze diverse data types, including neuroimaging, genetic, clinical, and cognitive data, to identify patterns indicative of Alzheimer's disease. Feature selection, dimensionality reduction, and classification techniques are employed to build predictive models for early detection and risk stratification. Steps in data analysis are shown in Fig.1.

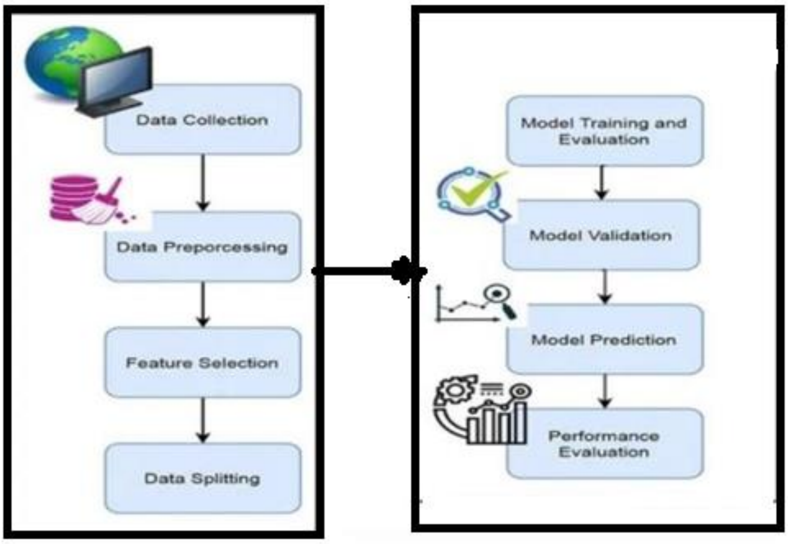

Fig.1: Steps of data analysis

In first step, Using the random forest method, data collecting for Alzheimer's disease detection entails compiling a variety of datasets, including demographic data, clinical evaluations, and medical imaging scans. MRI pictures are seen in Fig.2.Brain volume, cortical thickness, hippocampus volume, and glucose metabolism patterns are some examples of features that can be retrieved from the data. To improve model performance and lessen over fitting, preprocessing techniques like feature selection, normalization, and dimensionality reduction are used. Because of its capacity to manage high-dimensional data, deal with missing values, and prevent over fitting, the random forest algorithm was selected. To assess model performance, the dataset is separated into training, validation, and testing sets. Techniques for cross-validation may be used to evaluate resilience [11][12].The trained random forest model learns complex patterns in the data to classify subjects as Alzheimer's positive or negative based on input features. Model performance is evaluated using metrics like accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve. Iterative refinement of the model may involve hyper parameter tuning and feature engineering to enhance predictive accuracy. External validation using independent datasets is crucial to assess generalization performance. Collaborations with medical institutions and adherence to regulatory guidelines ensure the reliability and validity of the algorithm for clinical deployment. Continuous monitoring and updates to the model are essential to adapt to new data and improve performance over time. Next step is Data Preprocessing. In this, Standardize or normalize features, deal with missing values, clean up the collected data, and encode categorical variables as necessary. By doing this, the quality of the data and its suitability for machine learning models are guaranteed.Next Feature Selection/Extraction is done. Here, identify essential features contributing to Alzheimer's disease prediction. This may involve techniques such as correlation analysis, dimensionality reduction or domain expertise-based feature.

Next step is Model Training &Local Outlier Detection. Make use of the Random Forest technique to build a prediction model. An ensemble learning technique called Random Forest mixes several decision trees to improve robustness and generalizability. Train the model using preprocessed data, adjusting the hyper parameters to achieve the best results.. After, Model Evaluation &Validation of the data is done. Assess the Random Forest model's performance using relevant evaluation metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve. Furthermore, assess the local outlier identification algorithm's efficacy in locating anomalies. To find aberrant data points in the dataset, local outlier detection uses an algorithm such as the Local Outlier Factor. In order to identify possible outliers, LOF evaluates a data point's local density deviation with respect to its neighbors. In Validation step,verify the model's generalizability by evaluating its performance on separate datasets. Analyze the findings to learn more about the underlying causes of Alzheimer's disease and how outliers affect prediction outcomes. Iterative Refinement: Iteratively refine the methodology by enhancing data preprocessing techniques, optimizing feature selection strategies, and model parameters based on insights from initial analyses. Continuously assess and improve the model to enhance its predictive capability and robustness. In Local Outlier Detection Employ a local outlier detection algorithm, like the Local Outlier Factor (LOF), to identify anomalous data points within the dataset. LOF assesses the local density deviation of a given data point concerning its neighbours, highlighting potential outliers.

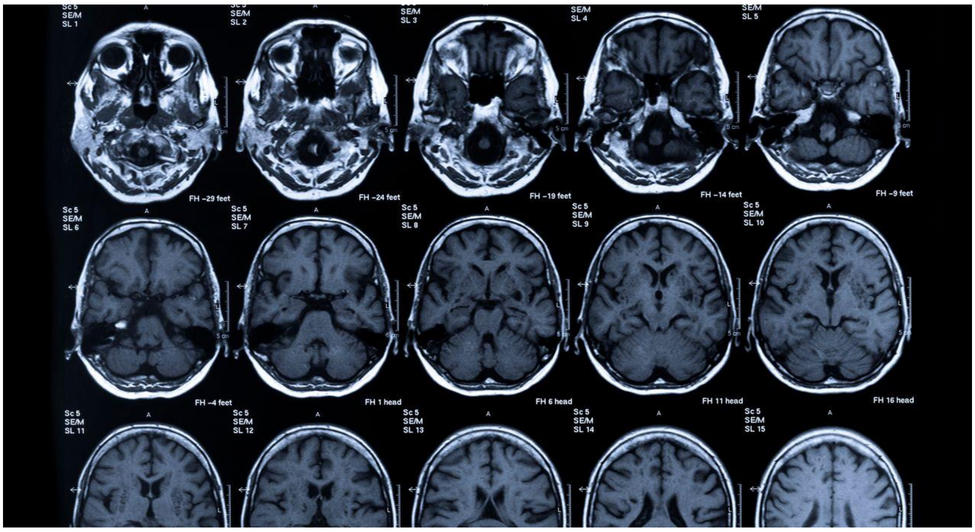

Fig.2 Brain MRI images

Proposed method: In this paper, Random Forest Classifier rand Local Outlier Factor is applied to detect Alzheimer's disease. The Random forest or Random Decision Forest is a supervised Machine learning algorithm used for classification, regression, and other tasks using decision trees. Training set prediction is shown in Fig.3 The Random forest classifier creates a set of decision trees from a randomly selected subset of the training set.

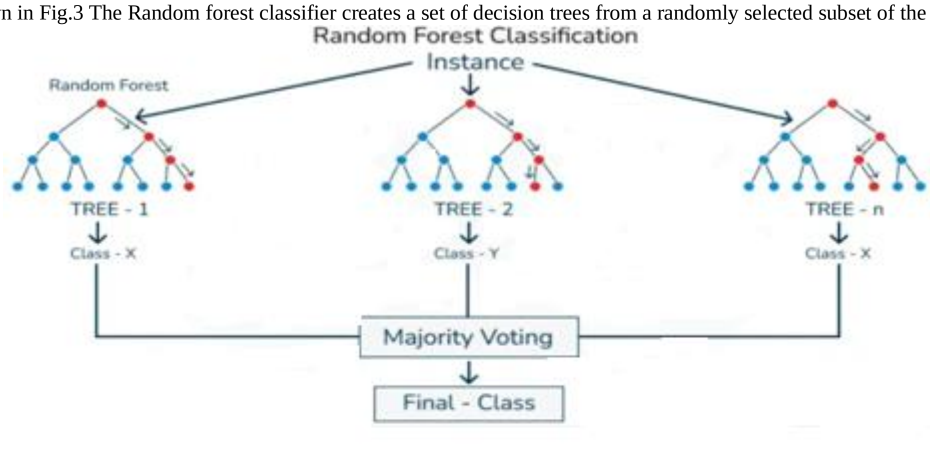

Fig. 3: Training set prediction

Results:

One method for measuring performance in machine learning and classification tasks especially in supervised learning is the confusion matrix. A classification model's performance on a set of test data, for which the true values are known, is described in this table. As seen in Fig. 4, the confusion matrix itself is a straightforward matrix arrangement that makes it possible to visualize how well an algorithm performs. The dataset we worked on has

373 rows and 14 fields, including Group fields with classes Demented and Non-Demented. Accuracy and precision of results are calculated by employing the Random Forest Classifier and the Local Outlier Factor methods for illness detection. The performance of the classification model is evaluated using a confusion matrix, which is crucial. It offers a thorough analysis of true positive, true negative, false positive, and false negative predictions, facilitating a more profound comprehension of a model's *recall, accuracy, precision,* and overall effectiveness in class distinction. When there is an uneven class distribution in a dataset, this matrix is especially helpful in evaluating a model's performance beyond basic accuracy metrics

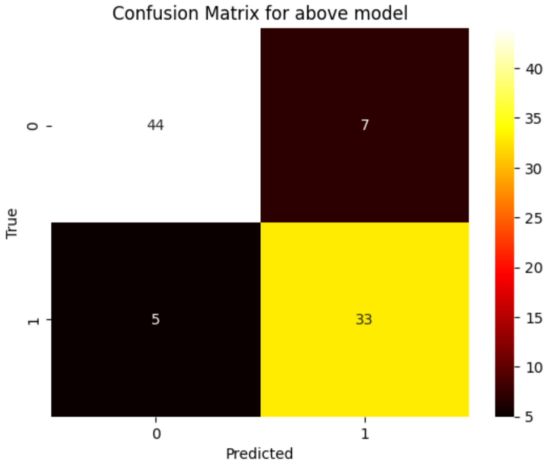

Fig. 4: Confusion Matrix

The performance parameters like accuracy, precision are calculated and shown in Table 1.

Table -1: Performance parameters

Parameter	Value
Accuracy	0.865
Precision	0.825
Recall	0.868
F1_Score	0.846

The classification report shows that the model performs well across both classes, with an overall accuracy of 87%. Class 0 has a higher precision , while Class 1 has a slightly higher recall.The F1-scores for both classes are also strong, indicating a good balance between precision and recall. The AUC-ROC plot is shown in Fig.5 value of 0.94 indicates that the model has excellent performance, with a 94% chance of correctly distinguishing between the positive and negative classes.

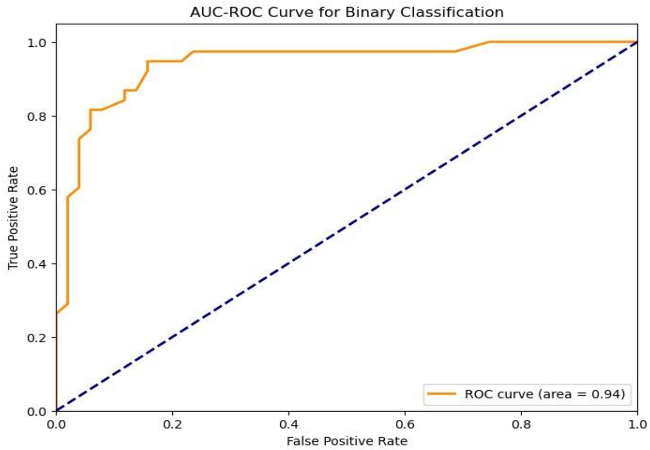

Fig5 :AUC-ROC

The Fig.5 shows that the model achieves a test accuracy of approximately 86.5%. This indicates that the model performs well on unseen data, demonstrating good generalization and the ability to make accurate predictions on new instances.

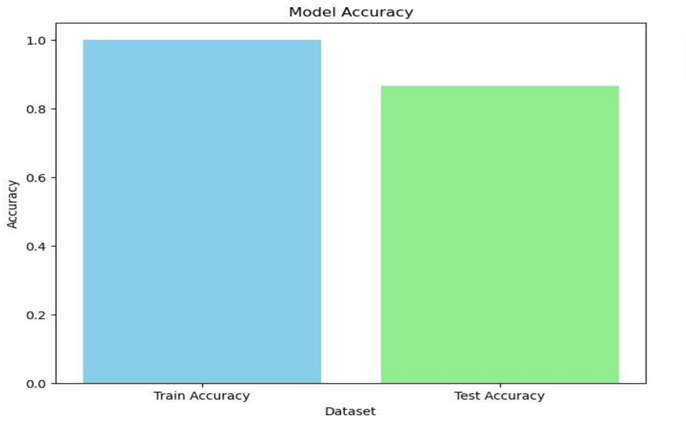

Fig.6: Model accuracy

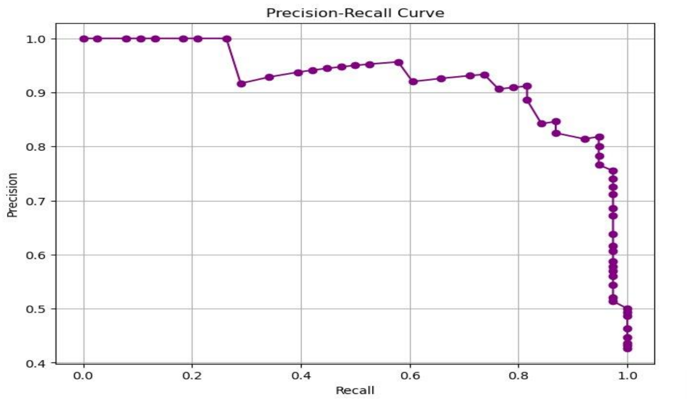

Fig.7: Precision-Recall

The Precision-Recall curve represented in Fig.7 how well the model is balancing precision (correct positive predictions) and recall (finding all actual positives). A higher curve means the model is performing better, especially when dealing with imbalanced data.

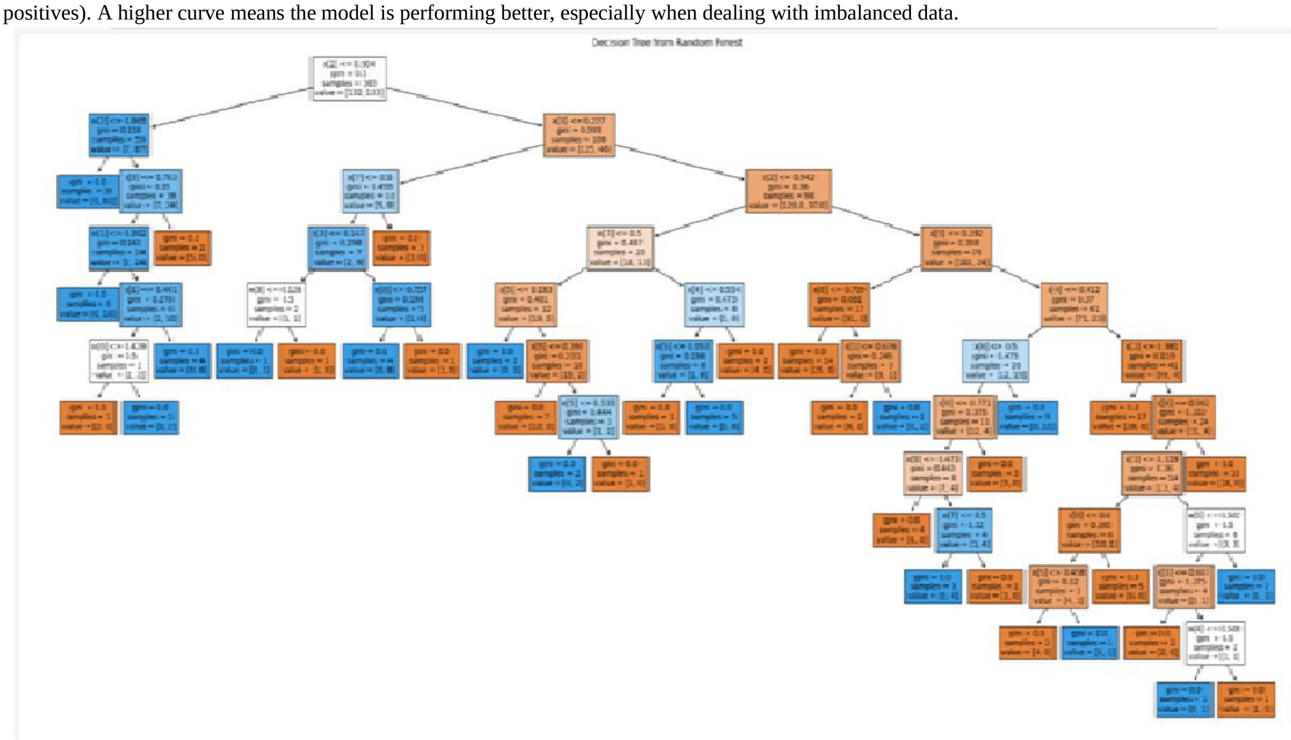

Fig.8: Decision Tree

Fig.8 shows Decision tree. This visualization helps in understanding how the model makes decisions by showing how each decision tree splits the data based on feature values. Each node represents a decision rule, and the leaves show the predicted class. By analyzing the tree structure, you can gain insights into how the model interprets different input features for classification.

Conclusion and Future Scope:

This work examined the use of the local outlier algorithm (LOF) and random forest in the prediction of Alzheimer's disease. A dataset that included different neuroimaging, clinical, and demographic characteristics is used. High accuracy, sensitivity, and specificity in differentiating between Alzheimer's patients and healthy controls were among the encouraging performance indicators displayed by the random forest model. Furthermore, by pointing out odd patterns or anomalies that would require more research, the LOF algorithm offered insightful information about possible outliers in the dataset. The comparison and assessment of contemporary research on the prognosis and prediction of Alzheimer's disease using machine learning techniques is the basis of this paper. The types of data being used and the effectiveness of machine learning techniques in predicting the early stages of Alzheimer's disease are two specific examples of recent advances in machine learning that have been made public. When compared to conventional statistical methods, it is clear that machine learning tends to increase prediction accuracy. But according to the models, the clinical diagnosis was not very accurate because there was no pathology confirmation, which raises doubts about the expected outcomes and the scope of further research. It can be incorporated into an appropriate application with a user interface so that everyone can utilize it with ease. Future research could focus on integrating machine learning models into CDSS platforms that provide actionable insights to clinicians, aid in risk assessment, patient stratification, and personalized treatment planning. It can be used in hospitals as a tool to assist patients suffering from Alzheimer's disease. Ensuring the ethical and appropriate use of AI in Alzheimer's detection is crucial, just like with any other medical application of machine learning. To guarantee the safe and efficient implementation of machine learning models in clinical settings, future research should address concerns pertaining to data privacy, openness, interpretability, and regulatory compliance.

REFERENCES :

1. Alzheimer’s Disease Facts and Figures. Alzheimer’s & Dementia; Alzheimer’s Association’s Publication & Wiley: Chicago, IL,USA, 2023.
2. Rajput, A. Does essential tremor increase the risk of dementia? No. Int. Rev. Neurobiol. 2022, 163, 233–253. [CrossRef] [PubMed]
3. Rasmussen, J.; Langerman, H. Alzheimer’sDisease-Why We Need Early Diagnosis. Degener. Neurol. Neuromuscul. Dis. 2019
4. F. Razavi, M. J. Tarokh, and M. Alborzi, “An intelligent alzheimer’s disease diagnosis method using unsupervised feature learning,” Journal of Big Data, vol. 6, no. 1, p. 32, 2019.
5. Bomasang-Layno, E.; Bronsther, R. “Diagnosis and Treatment of Alzheimer’s Disease” An Update. Del. J. Public Health 2021, 7, 74–85.
6. Y. Zhang, S. Wang, Y. Sui, M. Yang, B. Liu, H. Cheng, J. Sun, W. Jia,P. Phillips, and J. M. Gorriz, “Multivariate approach for alzheimer’s disease detection using stationary wavelet entropy and predator-prey particle swarm optimization,” Journal of Alzheimer’s Disease, vol. 65, no. 3, pp. 855–869, 2018.
7. R. Vidushi and A. K. Shrivastava, “Diagnosis of alzheimer disease using machine learning approaches,” International Journal of Advanced Science and Technology, vol. 29, no. 4, p. 7062–7073, 2019.
8. Alejandro Puente-castro, Alejandro pazos, Cristian R. Munteanu, Enrique Fernandez Blanco “Automatic Assessment of Alzheimer’s Disease Diagnosis based on Deep Learning Techniques.
9. Sitara Afzal, MuazzamMaqsood,Umair KhanandIfranMehmood.Alzheimer”Disease Detection Techniques and methods A Review”2021.
10. Gamal,M.Elattar,S. Selim” automatic Early Diagnosis of Alzheimer’s Disease using 3D Deep Ensemble Approach” 2022.
11. C.Chabib,L.J.Hadjileontiadis,A.A Shehhi “Deep Convolutional Curvelet Transform-based MRI Approach for Early Detection of Alzheimer’s Disease” 2023.
12. Association et al., “ 2020 alzheimer’s disease facts and figures, ”Alzheimer’s & Dementia, vol. 14, no. 3, pp. 367–429, 2020.
13. F. J. Martinez-Murcia, A. Ortiz, J.-M. Gorriz, J. Ramirez, and D. Castillo-Barnes, “Studying the manifold structure of alzheimer’s disease: a deep learning approach using convolutional autoencoders,” IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 1, pp.17–26, 2019.
14. D. Gopagoni, P. Lakshmi, and R. S. Ananthula, “An application of machine learning strategies to predict alzheimer’s illness progression in patients,” Technology, vol. 11, no. 6, pp. 1056–1063, 2020.