Automated Alzheimer's Disease Diagnosis

K. Saraswathi, N. T. Renukadevi, Harita T, Neha Parveen N and Vibinikha E M
Department of Computer Technology-UG, Kongu Engineering College, Perundurai, Erode, India

Keywords: Random Forest, KNN, Decision Tree, Gradient Boosting, Neural Networks, CatBoost.

Abstract: Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline and the accumulation of beta-amyloid plaques and tau protein tangles in the brain, leading to neuronal dysfunction and eventual cell death. This research aims to develop a computer-aided diagnostic system utilizing machine learning algorithms to analyze clinical data and identify indicators of Alzheimer's disease. The project involves data collection, algorithm development, and rigorous testing to ensure accuracy and reliability. Algorithms such as K Nearest Neighbor, Neural Network, Gradient Boosting Machine, Decision Tree, Random Forest, and CatBoost are employed to enhance diagnostic precision. Hyperparameter tuning and an ensemble model with a Voting Classifier will further improve accuracy. The objective is to deliver a user-friendly system that provides comprehensive reports, enabling healthcare professionals to make informed decisions. This system strives to improve diagnostic accuracy, ensuring its practical utility in healthcare environments by leveraging advanced machine learning techniques.

1 INTRODUCTION

A prompt and precise diagnosis is essential for the efficient management of Alzheimer's disease is leading causes of mortality worldwide. However, the large amount of complex medical data required for an accurate diagnosis presents a major challenge. The goal of this research is to enhance the prediction of Alzheimer's disease by using machine learning algorithms. The intention is to assist medical professionals in making more rapid and accurate diagnoses by developing reliable models that undergo stringent accuracy and precision testing. Machine learning algorithms can analyze large amounts of complex data, identifying patterns that indicate Alzheimer's disease. This approach improves patient outcomes by facilitating early intervention and increasing diagnostic accuracy. By efficiently evaluating patient data and detecting early warning signs of intelligent decline, machine intelligence can have a high impact on Alzheimer's care.

The study climaxes the transformational potential of machine intelligence in healthcare, emphasizing the importance of integrating this technology into clinical practice to achieve better health outcomes. By leveraging state-of-the-art examining capacities, machine intelligence can play a pivotal act in the early discovery and administration of Alzheimer's disease, eventually reconstructing the status of growth for patients and alleviating the work on caregivers and healthcare methods.

2 LITERATURE REVIEW

In older persons, Alzheimer's disease (AD) is ultimate prevailing cause of dementia. Currently, there is a lot of interest in using machine learning to discover metabolic disorders that impact a huge number of people globally. Diseases that impair memory and functionality will affect an increasing number of people, their families, and the healthcare system as our aging population grows. There will be significant social, financial, and economic repercussions from these. Alzheimer's disease is unpredictable when it is first developing. When AD is treated early on, less mild damage is caused and the efficacy of the treatment is higher than when it is treated later. To determine the optimal parameters for Alzheimer's disease prediction, a number of methods have been used, containing Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting, and Voting classifiers. The Open Access Series of Imaging Studies (OASIS) data is the action for Alzheimer's disease prognoses, and versification to a degree Precision, Recall, Accuracy, and F1-score for

Saraswathi, K., Renukadevi, N. T., T, H., N, N. P. and E M, V.
Automated Alzheimer's Disease Diagnosis.
DOI: 10.5220/0013588000004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 2*, pages 113-122
ISBN: 978-989-758-763-4

machine intelligence models are used to assess model performance. Clinicians can diagnose these disorders using the proposed classification approach. When these ML algorithms are used for early diagnosis, the annual death rates from Alzheimer's disease can be significantly reduced. With high-quality confirmation average veracity of 83% on the test dossier of AD, the suggested work demonstrates superior outcomes. This test's accuracy score is noticeably greater than that of previous studies. (Kavitha, Mani, et al. 2022)

The most common neurodegenerative disease is Alzheimer's disease. Initially innocuous, the manifestations worsen with time. One type of dementia that is common is Alzheimer's disease. The difficulty with this illness is that there isn't a cure. The disease is diagnosed, but only in its final stages. Therefore, the disease's progression or symptoms may be slowed down if the illness is diagnosed early. In this study, psychological variables such age, visits, MMSE, and education are used to predict the likelihood of Alzheimer's disease utilizing machine learning algorithms. (Neelaveni, Devasana, et al. 2020)

Alzheimer's is a backward senility that starts accompanying a slight loss of memory and someday results in the complete deficit of insane and tangible skills. For the patient benefits, the diagnosis should be made as soon as possible to begin treatment and preventive measures. While assessments like the Mini-Mental State Tests Examination are typically utilized for preliminary detection, brain analysis through magnetic resonance imaging (MRI) provides the basis for diagnosis. Techniques like OASIS (Open Access Series of Imaging Studies) is individual public project that makes neuroimaging datasets freely accessible for scientific inquiry. This study proposes and compares a novel approach for MRI-located Alzheimer's disease that is established deep learning and image processing techniques to earlier research in the field. Findings: Our approach obtains a balance veracity (BAC) of 0.88 for the production of the ailment stage (healthful tissue, very gentle, and harsh stage) and until 0.93 for concept-located computerized diagnosis of the illness. Conclusions: Using the OASIS collection, the results produced outperformed the state-of-the-art suggestions. This brings that methods based on deep learning are useful in developing a strong solution for MRI data-driven Alzheimer's-assisted diagnosis. (Saratxaga, et al. 2021)

The time of prevention and treatment outcomes of Alzheimer's disease depend on a precise and early diagnosis. In order to evaluate the course of Alzheimer's disease (AD), pinpoint its previous phases, and exploring this for future fields of study, this review summarizes the most recent research studies that use the deep learning methods and machine learning techniques. Several number of modern AI techniques, including Support Vector Machines, Random Forest, Logistic Regression, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Transfer Learning, are covered in this review in relation to their use in AD diagnosis. Their efficacy is also investigated, along with their advantages and disadvantages. The talk includes an overview of the key conclusions and medical imaging preprocessing techniques from the earlier investigations. Finally, we discuss the limitations and opportunities going forward. As a result, we emphasize that further data are required and that advanced neuroimaging technologies will be created. (Saeed, 2024)

Alzheimer's disease (AD) is a nerve condition that progresses irreversibly. The patient's treatment approach must be adjusted in light of the disease's close AD monitoring. Clinical score prediction using neuroimaging data is ideal for AD monitoring because it can accurately reveal the disease status. The majority of the earlier research on this task concentrated on a single time point and ignored the correlation between clinical scores at various time periods and neuroimaging data, such as magnetic resonance imaging (MRI). In contrast to previous research, we suggest developing a framework for predicting clinical scores using longitudinal data collected at several time periods. The three components of the proposed system are as follows: feature encoding using a multi-layer or deep polynomial network, ensemble learning techniques for regression using the support vector regression approach, and feature selection utilizing correntropy regularized in joint learning. There are two scenarios created for score prediction. To be more precise, scenario 1 makes use of baseline data to forecast longitudinal scores, but scenario 2 makes use of all data from prior time points to forecast scores at the subsequent at time, potentially increasing the accuracy of score prediction. To address the incompleteness of the data, the missing clinical scores at several longitudinal time periods are imputed. (Lei, et al. 2020)

Relevance of the pathological process of Alzheimer's disease (AD) begins with a protracted phase of amyloid (Aβ) buildup that is symptomless. The length of this stage varies widely from person to person. The optimal way to forecast the start of clinical progression is still unknown, despite the significant relevance of this disease phase for clinical

trial designs. Goal to assess the efficacy of various plasma biomarker combinations in predicting cognitive deterioration in cognitively unimpaired (CU) people who test positive for Aβ.The main result was a series of longitudinal cognitive tests that assessed over a median of 6 years (range, 2–10) using the Mini-Mental State Examination (MMSE) and the modified Preclinical Alzheimer Cognitive Composite (mPACC). The development of AD dementia was the secondary result. Linear regression models were employed to estimate the rates of longitudinal cognitive change (determined independently) using baseline biomarkers. The models were calibrated for baseline cognition, apolipoprotein E ε4 allele status, years of schooling, sex, and age. The revised Akaike information criterion and model R2 coefficients were used to compare multivariable models. (Mattsson-Carlgren, et al. 2023)

Over the past few years, there has been a tremendous advancement in the discovery of plasma biomarkers for pathologies associated with Alzheimer's disease. Blood tests that are validated for neurodegeneration, astrocytic activation, and amyloid and tau pathology are now available. We evaluated the prediction of research-diagnosed disease status by using these biomarkers and examined genetic variants associated with the biomarkers that may provide a more accurately reflect the risk of biochemically defined Alzheimer's disease instead of the risk of dementia to define Alzheimer's disease using biomarkers rather than clinical evaluation. The combination of all biomarkers, APOE, and polygenic risk score attained an area of using the receiver operating characteristic curve (AUC) showed a prediction accuracy of 0.81 for clinical diagnosis of Alzheimer's disease; the most significant contributors were ε4, Aβ40 or Aβ42, GFAP, and NfL. (Stevenson-Hoare, et al. 2023)

Relevance Regarding which biomarkers are most useful in predicting longitudinal tau buildup at various clinical stages of Alzheimer disease (AD), there is currently little agreement. Goal In order to identify which biomarker combinations demonstrated the most significant relationships with longitudinal tau PET and best optimal clinical trial enrichment, as well as to characterize longitudinal [18F]RO948 tau positron emission tomography (PET) findings across the clinical continuum of AD. Principal Results and Measures Using a data-driven method that combines clustering and event-based modeling, baseline tau PET using standardized uptake value ratio (SUVR) and also annual percent change in tau PET SUVR across regions of interest were determined. In order to determine which combinations best predicted longitudinal tau PET, regression models were utilized to investigate relationships between certain biomarkers and longitudinal tau PET. The effects of using these combinations as an enrichment method on a number of participants in a simulated clinical trial were then investigated using a power analysis. Conclusions and Pertinence's Plasma p-tau217 combined with tau PET may work best for improvement in preclinical and prodromal AD in studies where tau PET is the endpoint. Nonetheless, tau PET was more significant in prodromal AD, but Plasma p-tau217 was more significant in preclinical AD. (Leuzy, et al. 2022)

Lately, Alzheimer's disease has emerged as a big worry. Approximately 45 million individuals are afflicted with this illness. Alzheimer's is a declining brain disease that mostly affects the elderly and has an unclear etiology and pathophysiology. Dementia is the primary cause of Alzheimer's disease, as it gradually affects brain cells. This disease caused people to lose their capability to read, think, and many other skills. By estimating the illness, a machine learning system can lessen this issue. The primary goal is to identify dementia in a range of people. The investigation and findings of the detection of dementia using several machine learning models are presented in this research. The method has been developed using the Open Access Series of Imaging Studies (OASIS) dataset. Many machine learning models have been applied and the dataset examined. For prediction, decision trees, random forests, logistic regression, and support vector machines have all been employed. The system has been used both with and without fine-tuning. After comparing the outcomes, it is discovered that the support vector machine outperforms the other models. Among a large number of patients, it had the highest accuracy in identifying dementia. The technique is easy to use and can identify individuals who may be suffering from dementia. (Bari et al. 2021)

Background: The health of the elderly is at risk due to Alzheimer's Disease (AD), a nerve condition that progresses over time. It is believed that mild cognitive impairment (MCI) is through to be prodromal stage of AD. As of right now, the diagnosis of AD or MCI is made following permanent changes to the structure of the brain. Thus, the creation of novel biomarkers is essential to the early diagnosis and management of this illness. Currently, a few studies have demonstrated that radiomics analysis can be a useful diagnostic and classification technique for AD and MCI. Goal: To look into how the use of radiomics analysis can be used diagnosis and categorize of AD patients, MCI patients, and Normal

Controls (NCs), a thorough evaluation of the literature was conducted. Results: In the end, thirty finished MRI radiomics investigations were chosen for inclusion. The acquisition of picture data, Region of Interest (ROI) segmentation, feature extraction, feature selection, and classification or prediction are typically steps in the radiomics analysis process. The majority of the radiomics techniques were devoted to texture analysis. The histogram, shape-based, texture-based, wavelet, Gray Level Co-Occurrence Matrix (GLCM), and Run-Length Matrix (RLM) are additional characteristics that were retrieved. In conclusion, even though randomics analysis is now employes for the diagnosis and classification of AD and MCI, there is still a long way to go before these computer-aided diagnosis approaches are applied in clinical settings. (Feng, and Ding, 2020)

Alzheimer's disease permanently damages brain cells related to cognition and memory. Given that it results in death, it has a lethal outcome. In previous identification of Alzheimer's disease is so crucial. Accurately diagnosing this illness in its early stages is essential for clinical research as well as patient care. Alzheimer's disease (AD) is among the most costly illnesses to cure, hence many researchers are focusing on developing an automated algorithm with great accuracy. Early detection and estimation of Alzheimer's disease may provide difficulties. An ML system that can predict the sickness can solve this problem. The potential of machine learning (ML) to address issues in a variety of domains, including the interpretation of medical imaging, has recently led to ML's significant rise in popularity. Current research uses machine learning algorithms and 3D magnetic resonance imaging (MRI) images to predict and classify Alzheimer's disease. Using 3D MRI technology, this study integrates the white and grey matter found in MRI images to produce 2D slices in the axial, sagittal, and coronal orientations. In order to forecast and categorize Alzheimer's disease, Multi-Layer Perceptron (MLP) and SVM algorithms are used for feature extraction after the most pertinent slices have been chosen. The precision, recall, accuracy, and F1-score are among the criteria the researchers use to evaluate the system's effectiveness. (Rao, Gandhi, et al. 2023)

This section discusses experimental results and presents an actual MRI image using the suggested methods. The trials are conducted using several grayscale MRI image standards that vary in size. As seen in Fig. 5(a), the MRI pictures are distorted by speckle noise, random noise, and salt and pepper noise generated by MRI scanning equipment. These three noise characteristics serve as the basis for the de-noising procedure. In summary is using a variety of algorithms, the Computer Aided Diagnosis (CAD) method is suggested as a means of identifying and categorizing Alzheimer disease on authentic MRI scans. An extremely expensive diagnostic tool for Alzheimer's is the picture of the disease, which is quite dangerous. The biomedical field has gained popularity recently as a result of computer-aided diagnosis (CAD), which uses digital image processing to diagnose clinical patients accurately and quickly. For people with Alzheimer's disease (AD), early and appropriate diagnosis and treatment planning lead to increased life expectancy and quality of life. Modern methods that consider multimodal analysis to be accurate and efficient have been demonstrated to be superior to manual analysis. Although numerous technologies have been developed to diagnose Alzheimer's disease, the diagnosis system is still very expensive and provides low-accuracy and inefficient disease detection because of the limitations of Magnetic Resonance Imaging (MRI) scanning machines. This study suggests a fresh approach for CAD procedure that predicts AD utilizing a variety of algorithms. (Sathiyamoorthi, Ilavarasi, et al. 2021)

Predicting the long-term course of Alzheimer's disease (AD), a chronic neurological illness, is undoubtedly crucial. When describing the cortical atrophy that is strongly linked with AD prodromal stages and clinical symptoms, structural magnetic resonance imaging, or sMRI, might be utilized. A large number of current techniques have concentrated on employing a set of morphological traits obtained from sMRI to predict the cognitive scores at future time-points. More extensive information can be obtained from the 3D sMRI than from the cognitive scores. Nevertheless, relatively few studies attempt to forecast a single brain MRI scan at a later period. In order to forecast the overall appearance of a person's brain over time, we present a disease progression prediction framework in this paper that includes a 3D multi-information generative adversarial network (mi-GAN). and a multi-class classification network tuned with a focal loss based on 3D DenseNet that determines the estimated brain's clinical stage. With respect to the individual of Multi-information and 3D brain sMRI at the baseline time point, the mi-GAN may provide individual 3D brain MRI images of superior quality. On the use Alzheimer's Disease Neuroimaging Initiative (ADNI), experiments are conducted. With a structural similarity index (SSIM) of 0.943 between the produced and real fourth-year MRI images, our mi-GAN demonstrates advanced performance. When mi-GAN and focused loss are

used in place of conditional GAN and cross entropy loss, the pMCI vs. sMCI accuracy improves by 6.04%. (Zhao, Ma, et al. 2021)

In order to predict the likelihood that someone with mild cognitive impairment (MCI) will develop Alzheimer's disease (AD), this study confirms the generalizability of the MRI-based classification of AD patients and controls (CN) to an external data collection. A deep convolutional neural network (CNN) and a traditional support vector machine (SVM) method based on structural MRI data that were either minimally or heavily pre-processed into maps of the modulated gray matter (GM). The Alzheimer's Disease Neuroimaging Initiative (ADNI; 334 AD, 520 CN)employed cross-validation. After that, trained classifiers were used in the independent Health-RI Parelsnoer Neurodegenerative Diseases Biobank data set as well as in ADNI MCI patients (231 converters, 628 non-converters) to predict conversion to AD. We enrolled 199 AD patients, 139 participants with subjective cognitive impairment, 48 MCI patients who converted to dementia, and 91 MCI patients who did not convert to dementia from this multi-center trial, which represented the population of a tertiary memory clinic. For AD classification, deep and conventional classifiers performed similarly well, with just a minor drop in performance when applied to the external cohort. We anticipate that this external validation study will help translate machine learning into clinical settings. (Bron, et al. 2021)

The main causes of dementia, Alzheimer's disease (AD) is characterized by a gradual course that takes years to complete with no known cure or medication. In this sense, attempts have been made to determine the likelihood of acquiring AD at an early age. More recent research has concentrated on the disease and prognosis of AD utilizing long or period series data in a manner of disease progression modeling, whereas many earlier works used cross-sectional analysis. In this study, we provide a unique computational framework that can predict, under the same problem settings, cognitive scores at various future time points, coupled with the trajectories of clinical status and phenotypic measures of MRI biomarkers. However, it typically encounters a large number of unexpected missing observations when handling time series data. Given such an adverse scenario, we plan a subordinate question of estimating those missing principles and address it accurately by accounting for the multivariate and temporal linkages present in period succession data. In particular, we plan a deep repeating network to jointly address four issues: (i) phenotypic calculations predicting; (ii) course guess of a cognitive score; (iv) dispassionate rank guess of a subject established longitudinal image biomarkers; and (iii) missing value imputation. Interestingly, our cautiously constructed loss function is used to train the learnable parameters of each module in our prediction models end-to-end using the morphological features and cognitive scores as input. We tested our approach using The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) challenge cohort, comparing it to rival approaches in the literature and measuring performance for a number of measures. Furthermore, ablation tests and thorough analysis were carried out to further verify the efficacy of our approach. (Jung, Jun, et al. 2021)

The extreme predominance of Alzheimer's disease (AD) and the extreme cost of usual demonstrative patterns create research into the mechanical discovery of AD critical. Since AD materially impacts the meaning and sound character of spoken conversation, machine intelligence and robotics offer hopeful methods for dependably detecting AD. Recently, skilled has happened a conception of models for AD categorization; still, these change in agreements of the types of models, datasets used, and training and testing paradigms. In this work, we analyze the efficiency of two prevalent methods to mechanical recognition of AD from speech on the same, appropriate dataset, in order to ascertain the benefits of using expertise in the field vs. had trained transfer models. In order to identify the best predictive model, it is important to assess its effectiveness on carefully crafted datasets using compatible same variables for training and self-sufficient test datasets. This approach supports the usefulness of productive machine learning and linguistically-focused machine learning methods that identify AD from speech. (Balagopalan Eyre, et al. 2021)

Alzheimer's disease (AD) is a step-by-step affecting animate nerve organs illness that frequently influences middle-old and older persons, gradually impairing their cognitive function. There is currently no treatment for AD. In addition, it takes too long to diagnose AD clinically today. In order to predict AD clinical scores, we have designed a combined and deep learning system in this research. To be more precise, features of brain regions linked to AD are screened and dimensions are reduced using a process of feature selection that combines group LASSO and correntropy. In order to investigate the temporal association between longitudinal data and the internal connections between various brain regions, we interrogate the multi-layer alone repeating brain network reversion. The clinical score is concluded

apiece jointly submitted deep learning network, that likewise examines the equivalence betwixt the clinical score and drawing resonance depict. The expected clinical score principles enable physicians to treat patients' illnesses promptly and with an early diagnosis. (Lei, et al. 2022)

A crucial but unmet clinical issue is creating multi-biomarker models that are cross-validated to predict the rate of cognitive deterioration in Alzheimer's disease (AD). Global understanding (R2 = 24%) and thought (R2 = 25%) decline rates in rare AD were predicted by a model integrating all diagnostic categories and tested in ADAD over a 4-year period. By utilizing model-based risk-enrichment, the sample size needed to identify simulated intervention effects was decreased by 50% to 75%. Our alone confirmed machine-knowledge approach concede possibility significantly lower the sample amount necessary in AD clinical troubles by forecasting cognitive degeneration in scattered prodromal AD. In order to think rates of intelligent decline, we applied support heading reversion to AD biomarkers obtained from fundamental attractive resonance depict (MRI), amyloid-PET, fluorodeoxyglucose positron-diffusion tomography (FDG-PET), and cerebrospinal fluid. Prediction models were checked in sporadic premature AD (n = 216), after being trained in autosomal-dominant AD (ADAD, n = 121). When promoting model-located risk enrichment, the sample content necessary to identify situation belongings was premeditated. (Franzmeier, et al. 2020)

The most prevailing type of senility, Alzheimer's disease (AD), can influence a affecting animate nerve organs condition that damages brain cells and impairs function, ultimately leading to gradual memory loss and difficulty carrying out daily tasks. We can identify AD patients based on whether they currently have the lethal disease or may not in future by using MRI (Magnetic Resonance Imaging) scan brain images to aid in the identification and prediction of this disease. The primary goal of all of our work is to create the greatest tools for detection and prediction that radiologists, physicians, and other caregivers can use to treat patients with this illness and save time and money. Deep Learning (DL) algorithms have shown great promise in recent years for the diagnosis of AD due to their ability to operate on enormous datasets. In this study, we have used MRI images from the ADNI 3 class, which has a total 2480 records, 2633 normal, 1512 moderate cases, to cultivate Convolutional Neural Networks (CNNs) for earlier diagnosis and classification of AD. When compared to numerous other relevant papers, the model performed well, with a noteworthy accuracy of 99%.Additionally, we contrasted the outcome with our earlier research, which used the OASIS dataset to apply machine learning algorithms. This revealed that methods that use deep learning can be a better choice than standard methods for machine learning when handling large amounts of data, such as medical data. (Salehi, Baglat, et al. 2020)

It is difficult to anticipate when healthy people or people with modest cognitive impairment will progress to the stage of active Alzheimer's disease. Recently, a deep learning-based survival analysis was created to make predictions about when an event would occur in a dataset that contains censored data. Here, we studied either an comprehensive study of addition survive forecast the happening of Alzheimer's disease in a matching style. We employed the white matter dimensions of various brain regions in patients who were cognitively normal and those who had mild cognitive impairment as predictive variables. The prediction results of our deep survival model, which is based on a Weibull distribution, the DeepHit model, and the conventional standard Cox proportional-hazard model were then compared. Our model produced the highest correlation index of 0.835, which was similar to the DeepHit model's and greater than the Cox model's. As far as we are aware, this is the sole research that discusses using brain-MRI data to apply a deep survival model. Our findings show that this kind of study could accurately forecast when a person would develop Alzheimer's disease. (Nakagawa, et al. 2020)

3 METHODOLOGY

3.1 Random Forest:

- Ensemble Learning: Random Forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.
- Bagging Technique: It uses the bagging method, where each model is trained on a random subset of the data. This helps in reducing the variance and avoids overfitting.

3.2 K-Nearest Neighbor:

- K-NN stores all the data and classifies

the new data point according to the similarity. Therefore, when new data appears, it can easily be classified into the well suite category by K-NN algorithm.

- At the training phase, KNN only stores the datasets, when it receives new data, it classifies according to the similarity of the new data.

3.3 Decision Tree:

- The decision tree is built by recursively dividing the training data into sub-data sets based on the attributes' values until a threshold is reached, such as a maximum depth or minimum number of samples to split a node.
- The aim is to find an attribute that gets the most information or reduces the amount of impurity after splitting the data.

3.4 Gradient Boosting:

- It builds an ensemble of models sequentially, where each model attempts to correct the errors of its predecessor.
- This method is particularly known for its effectiveness in improving the accuracy of predictions.

3.5 Neural Networks:

- Neural networks comprise layers of neurons, accompanying each layer tr ansforming the input
data before passing it to the next lay er. The layers contain an input layer, hidden layers and output layer.
- They use activation functions to introduce non-linearity, enabling the network to learn from complex patterns and Neural networks are trained using backpropagation.

3.6 CatBoost:

- It builds an ensemble of trees sequentially, each one correcting errors from the previous one.
- Automatically handles categorical features without the need for extensive preprocessing and Uses ordered boosting to reduce overfitting and improve accuracy.

4 WORK FLOW

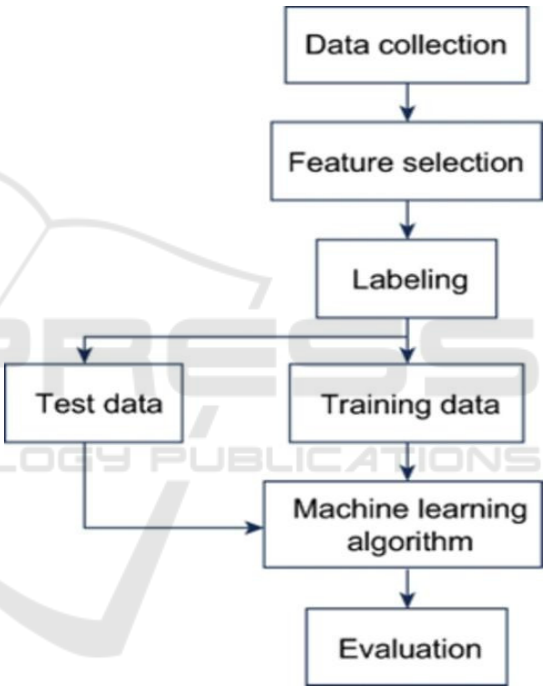

Figure 1: Work Flow

4.1 Explanation for work flow

4.1.1 Data Collection:

- Collect raw data from various sources. Ensure data is relevant to the problem. Organize it for processing.

4.1.2 Feature Selection:

- Identify key attributes or features. Eliminate irrelevant or redundant data. Prepare data for labeling and training.

4.1.3 Labeling:

- Assign labels to the dataset if it's supervised learning. Categorize data based on classes or outputs. Prepare it for training and testing.

4.1.4 Data Split (test, train):

- Split the data into training and test sets. Training data is used to model learning. Test data will validate model.

4.1.5 ML Algorithm:

- Choose a machine learning model and Use training data to train the model. Understand patterns and relationships in the data.

4.1.6 Evaluation:

- Apply the model to test set. Measure the performance using metrics (accuracy, precision, recall, etc.). Optimize the model based on the results if necessary.

5 RESULTS

Table 1: Output values.

Algorithm Used	Classification On Accuracy	Precision	Recall	F1-Score
KNN	0.73	0.54	0.44	0.50
Neural Network	0.81	0.71	0.70	0.70
Decision Tree	0.83	0.73	0.77	0.75
GBM	0.88	0.87	0.73	0.79
Random Forest	0.90	0.92	0.74	0.82
CatBoost	0.90	0.92	0.74	0.82

Ensemble Methods	Accuracy
Random forest with KNN	0.86
Random forest with CatBoost	0.89
Random forest with GBM	0.94

5.1 Diagrammatic representation of outputs

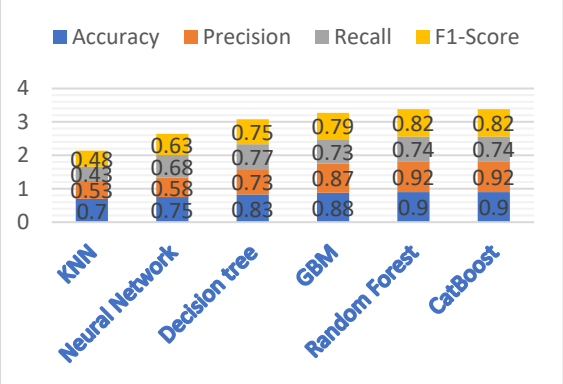

Figure 2: Performance Metrics.

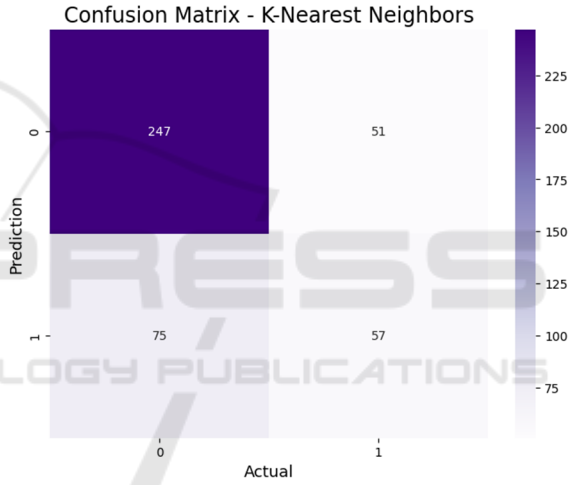

Figure 3: Confusion Matrix of KNN.

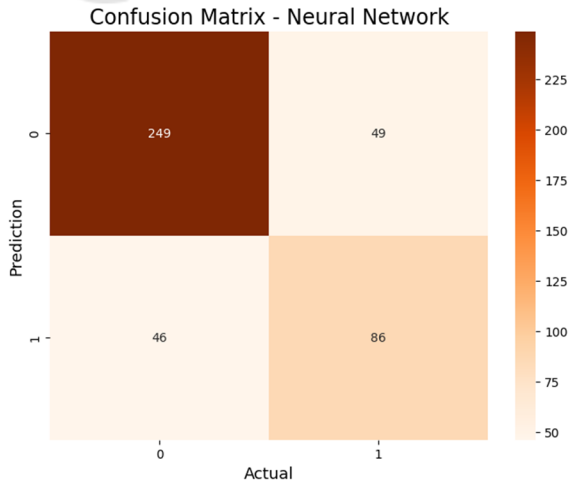

Figure 4: Confusion matrix of NN.

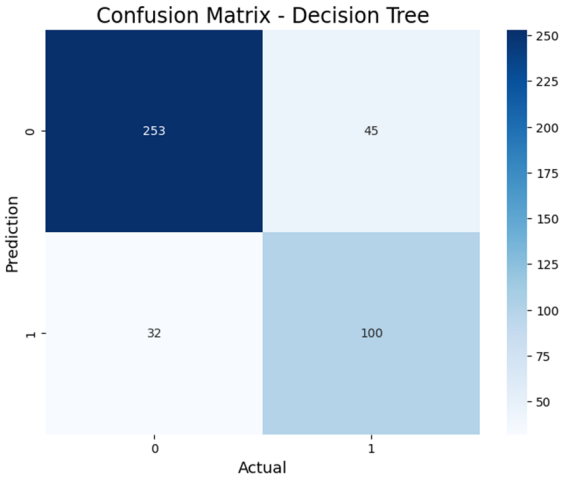

Figure 5: Confusion matrix of DT.

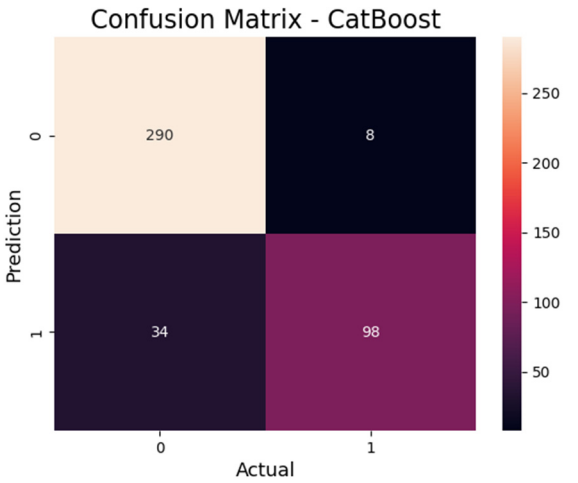

Figure 8: Confusion matrix of CatBoost.

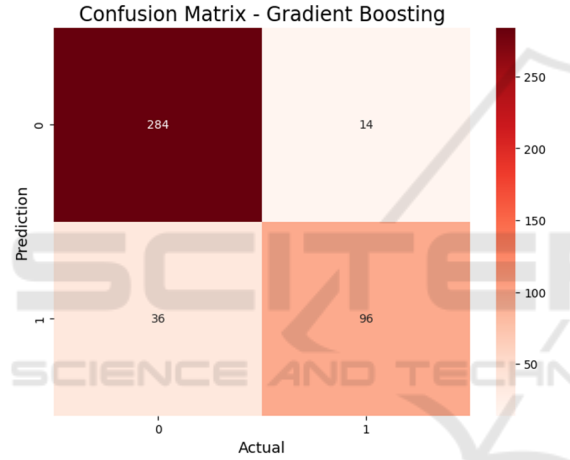

Figure 6: Confusion matrix of GBM.

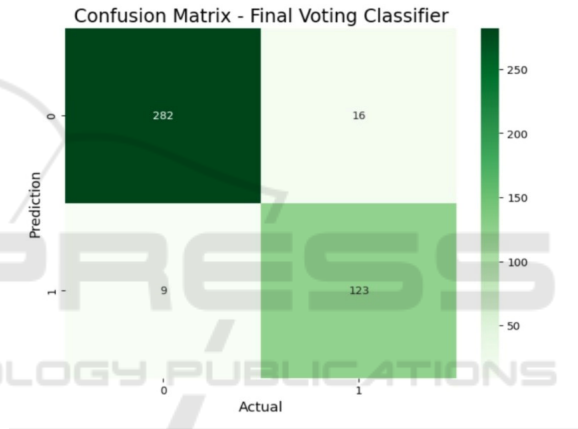

Figure 9: Confusion matrix of Ensemble Method.

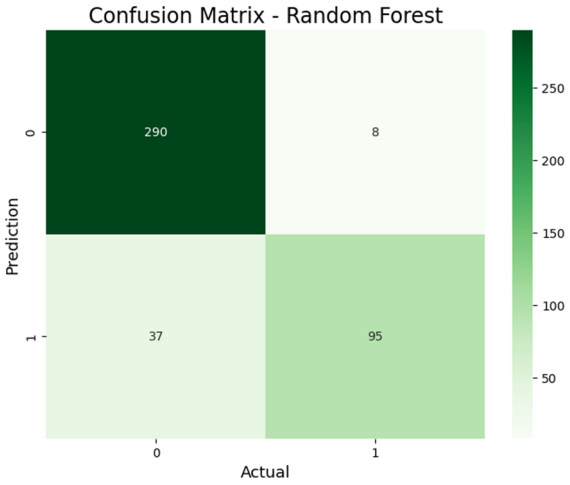

Figure 7: Confusion matrix of KNN.

6 CONCLUSIONS

In this proposed work, the prediction of the target variable was performed using classification techniques, including K-Nearest Neighbors (KNN), Neural Network (NN), Decision Tree (DT), Gradient Boosting, Random Forest and CatBoost. While comparing these algorithm results, Random Forest and CatBoost emerged as the best-performing algorithm with an accuracy of 90%, outperforming the other classification algorithms. The model's robustness to noise and ability to handle overfitting contributed to its superior performance. However, by applying hyperparameter tuning and creating an ensemble model with Voting Classifier (combining Random Forest and Gradient Boosting), the accuracy was significantly improved to 94%.

7 FUTURE SCOPES

In future work, exploring hybridized algorithms can help enhance both accuracy and robustness. Additionally, advanced techniques such as deep learning may be explored, especially when working with more complex data, offering improved feature extraction and predictive capabilities. These approaches hold excellent potential for further improving model performance and could contribute significantly to more accurate and reliable forecastings in Alzheimer's disease classification and additional healthcare uses.

REFERENCES

Kavitha, C., Mani, V., Srividhya, S. R., Khalaf, O. I., Tavera Romero, C. A., 2022. Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. Front. Public Health, 10, 853294.

Neelaveni, J., Devasana, M. S. G., 2020. Alzheimer Disease Prediction Using Machine Learning Algorithms. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India. IEEE, pp. 101–104.

Saratxaga, C. L., et al., 2021. MRI Deep Learning-Based Solution for Alzheimer's Disease Prediction. JPM, 11(9), 902.

Saeed, F., 2024. Applications of ML and DL Algorithms in the Prediction, Diagnosis, and Prognosis of Alzheimer's Disease. AJBSR, 22(6), 779–786.

Lei, B., et al., 2020. Deep and Joint Learning of Longitudinal Data for Alzheimer's Disease Prediction. Pattern Recognition, 102, 107247.

Mattsson-Carlgren, N., et al., 2023. Prediction of Longitudinal Cognitive Decline in Preclinical Alzheimer Disease Using Plasma Biomarkers. JAMA Neurol, 80(4), 360.

Stevenson-Hoare, J., et al., 2023. Plasma Biomarkers and Genetics in the Diagnosis and Prediction of Alzheimer's Disease. Brain, 146(2), 690–699.

Leuzy, A., et al., 2022. Biomarker-Based Prediction of Longitudinal Tau Positron Emission Tomography in Alzheimer Disease. JAMA Neurol, 79(2), 149.

Bari Antor, M., et al., 2021. A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease. Journal of Healthcare Engineering, 2021, pp. 1–12.

Feng, Q., Ding, Z., 2020. MRI Radiomics Classification and Prediction in Alzheimer's Disease and Mild Cognitive Impairment: A Review. CAR, 17(3), 297–309.

Rao, K. N., Gandhi, B. R., Rao, M. V., Javvadi, S., Vellela, S. S., Khader Basha, S., 2023. Prediction and Classification of Alzheimer's Disease using Machine Learning Techniques in 3D MR Images. In 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India. IEEE, pp. 85–90.

Sathiyamoorthi, V., Ilavarasi, A. K., Murugeswari, K., Thouheed Ahmed, S., Aruna Devi, B., Kalipindi, M., 2021. A Deep Convolutional Neural Network Based Computer Aided Diagnosis System for the Prediction of Alzheimer's Disease in MRI Images. Measurement, 171, 108838.

Zhao, Y., Ma, B., Jiang, P., Zeng, D., Wang, X., Li, S., 2021. Prediction of Alzheimer's Disease Progression with Multi-Information Generative Adversarial Network. IEEE J. Biomed. Health Inform., 25(3), 711–719.

Bron, E. E., et al., 2021. Cross-Cohort Generalizability of Deep and Conventional Machine Learning for MRI-Based Diagnosis and Prediction of Alzheimer's Disease. NeuroImage: Clinical, 31, 102712.

Jung, W., Jun, E., Suk, H.-I., 2021. Deep Recurrent Model for Individualized Prediction of Alzheimer's Disease Progression. NeuroImage, 237, 118143.

Balagopalan, A., Eyre, B., Robin, J., Rudzicz, F., Novikova, J., 2021. Comparing Pre-Trained and Feature-Based Models for Prediction of Alzheimer's Disease Based on Speech. Front. Aging Neurosci., 13, 635945.

Lei, B., et al., 2022. Predicting Clinical Scores for Alzheimer's Disease Based on Joint and Deep Learning. Expert Systems with Applications, 187, 115966.

Franzmeier, N., et al., 2020. Predicting Sporadic Alzheimer's Disease Progression via Inherited Alzheimer's Disease-Informed Machine-Learning. Alzheimer's & Dementia, 16(3), 501–511.

Salehi, A. W., Baglat, P., Sharma, B. B., Gupta, G., Upadhya, A., 2020. A CNN Model: Earlier Diagnosis and Classification of Alzheimer Disease Using MRI. In 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India. IEEE, pp. 156–161.

Nakagawa, T., et al., 2020. Prediction of Conversion to Alzheimer's Disease Using Deep Survival Analysis of MRI Images. Brain Communications, 2(1), fcaa057.