Automated Alzheimer's Disease Diagnosis

Ms. M. Sindhu
Assistant Professor,
Department of Computer Applications
Kongu Engineering College
Perundurai,India.

Nisanth G
Department of Computer Applications
Kongu Engineering College
Perundurai,India

Mrs.A.Deepa
Assistant Professor,
Department of Mathematics
Kongu Arts and Science College,Erode

Preteka A T
Department of Computer Applications
Kongu Engineering College
Perundurai, India

Varshini T
Department of Computer Applications
Kongu Engineering College
Perundurai, India

Abstract - Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline and memory impairment, primarily affecting older adults. It is distinguished by the accumulation of beta-amyloid plaques and tau protein tangles in the brain, leading to neuronal dysfunction and eventual cell death. This proposed work is to develop a computer-aided diagnosis system for Alzheimer's disease to assist healthcare professionals in accurate and timely diagnosis. This system will use machine learning algorithm to analyze clinical data, identifying potential indicators of Alzheimer's disease. The project involves data collection, algorithm development, and system testing. This work will collaborate with medical experts to gather relevant data and ensure the system's accuracy and reliability.A goal is to create a user-friendly system that provides comprehensive reports, enabling healthcare professionals to make informed decisions regarding patient care. Through the integration of automated algorithms, the system seeks to identify and quantify specific patterns within the datasets that are indicative of Alzheimer's pathology. This system will analyze whether diseases are present or not,byusing algorithms such as K Nearest Neighbor, Neural Network, Gradient Boosting Machine, Decision Tree, Random Forest, CatBoost. The aim of using these algorithms on this kind of data is to improve diagnosis's accuracy and make it more dependable and effective. Our objective is to enhance the accuracy of diagnosis and guarantee that the models are workable and efficient for practical implementation in healthcare environments by concentrating on data.

Keywords — Random Forest-KNN-Decision Tree-Gradient Boosting-Neural Networks-Catboost

1. INTRODUCTION

A prompt and precise diagnosis is essential for the efficient management of Alzheimer's disease, one of the 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 utilizing 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 cognitive decline, machine learning can have a significant impact on Alzheimer's care.

The study highlights the transformative potential of machine learning in healthcare, emphasizing the importance of integrating this technology into clinical practice to achieve better health outcomes. By leveraging advanced analytical capabilities, machine learning can play a pivotal role in the early detection and management of Alzheimer's disease, ultimately improving the quality of life for patients and easing the burden on caregivers and healthcare systems.

2. LITERATURE REVIEW

In older persons, Alzheimer's disease (AD) is the most common 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, such as diabetes and Alzheimer's. Every year, their incidence rates rise at a startling rate. Neurodegenerative alterations occur in the brain when an individual has Alzheimer disease. 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, including Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting, and Voting classifiers. The Open Access Series of Imaging Studies (OASIS) data is the basis for Alzheimer's disease predictions, and metrics such as Precision, Recall, Accuracy, and F1-score for machine learning 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 the best validation average accuracy of 83% on the test data of AD, the suggested work demonstrates superior outcomes. This test's accuracy score is noticeably greater than that of previous

studies.[1]

The most common neurodegenerative disease is Alzheimer's disease. Initially innocuous, the symptoms 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 using machine learning algorithms.[2]

Alzheimer's is a degenerative dementia that begins with a slight loss of memory and eventually results in the complete loss of mental and physical abilities. For the benefit of the patient, 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: OASIS (Open Access Series of Imaging Studies) is one public project that makes neuroimaging datasets freely accessible for scientific inquiry. This study proposes and compares a novel approach for MRI-based Alzheimer's diagnosis that is based on deep learning and image processing techniques to earlier research in the field. Findings: Our approach obtains a balance accuracy (BAC) of 0.88 for the creation of the disease stage (healthy tissue, very mild, and severe stage) and up to 0.93 for image-based automated diagnosis of the illness. Conclusions: Using the OASIS collection, the results produced outperformed the state-of-the-art suggestions. This shows that methods based on deep learning are useful in developing a strong solution for MRI data-driven Alzheimer's-assisted diagnosis.[3]

Timely 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 early phases, and explore future directions for this field of study, this review summarizes the most recent studies that use deep learning and machine learning techniques. A 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.[4]

Alzheimer's disease (AD) is a neurological 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 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 deep polynomial network, ensemble learning for regression using the support vector regression approach, and feature selection based on correntropy regularized 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 predict scores at the subsequent time point, 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.[5]

Relevance The pathophysiology 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 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.[6]

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 using these biomarkers and tested genetic variants associated with the biomarkers that may more accurately reflect the risk of biochemically defined Alzheimer's disease instead of the risk of dementia in order to define Alzheimer's disease using biomarkers rather than clinical assessment. The combination of all biomarkers, APOE, and polygenic risk score attained an area under the receiver operating characteristic curve (AUC) = 0.81 for the prediction accuracy of clinical diagnosis of Alzheimer's disease; the most significant contributors were ε4, Aβ40 or Aβ42, GFAP, and NfL[7].

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 strongest 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 standardized uptake value ratio (SUVR) and 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 sample size in a simulated clinical trial were then investigated using a power analysis. Conclusions and Pertinence Plasma p-tau217 with tau PET may work best for enrichment 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.[8]

Lately, Alzheimer's disease has emerged as a big worry. Approximately 45 million individuals are afflicted with this illness. Alzheimer's is a degenerative 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 sickness caused people to lose their capacity to read, think, and many other skills. By forecasting 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.[9]

Background: The health of the elderly is at risk due to Alzheimer's Disease (AD), a neurological condition that progresses over time. It is believed that mild cognitive impairment (MCI) represents the 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 investigate the use of radiomics analysis in the diagnosis and classification 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, there is still a long way to go until these computer-aided diagnostic techniques are used in clinical settings, even if radiomics analysis is already used for AD and MCI diagnosis and categorization.[10]

Alzheimer's disease permanently damages brain cells related to cognition and memory. Given that it results in death, it has a lethal outcome. Early 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 one of the most expensive diseases to treat, hence many researchers are focusing on developing an automated algorithm with great accuracy. Early detection and prediction 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 resolve problems in a range of fields, 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.[11]

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.[12]

Predicting the long-term course of Alzheimer's disease (AD), a chronic neurological illness, is undoubtedly crucial. When describing the cortical atrophy that is closely associated with clinical symptoms in AD and its prodromal stages, 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 3D brain sMRI and multi-information at the baseline time-point, the mi-GAN may provide individual 3D brain MRI images of superior quality. On the Alzheimer's Disease Neuroimaging

Initiative (ADNI), experiments are conducted. With a structural similarity index (SSIM) of 0.943 between the generated and real fourth-year MRI images, our mi-GAN demonstrates state-of-the-art performance. When mi-GAN and focused loss are used instead of conditional GAN and cross entropy loss, the pMCI vs. sMCI accuracy improves by 6.04%.[13]

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 set. We employed 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 modulated gray matter (GM) maps. Cross-validation was used in the Alzheimer's Disease Neuroimaging Initiative (ADNI; 334 AD, 520 CN) to optimize and assess classifiers. 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.[14]

One of 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 diagnosis and prognosis of AD using longitudinal or time 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 formulate a secondary problem of estimating those missing values and address it methodically by accounting for the multivariate and temporal linkages present in time series data. In particular, we suggest a deep recurrent network to jointly address four issues: (i) phenotypic measurements forecasting; (ii) trajectory estimation of a cognitive score; (iv) clinical status prediction of a subject based on longitudinal imaging 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 Alzheimers 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.[15]

The high prevalence of Alzheimer's disease (AD) and the high cost of traditional diagnostic methods make research into the automatic detection of AD crucial. Since AD substantially impacts the semantics and sound quality of spoken words, machine learning and natural language processing offer promising techniques for reliably detecting AD. Recently, there has been a proliferation of models for AD classification; however, these vary in terms 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.[16]

Alzheimer's disease (AD) is a gradual neurological illness that often affects middle-aged and older adults, 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 investigate the multi-layer autonomously recurrent brain network regression. The clinical score is predicted by the jointly suggested deep learning network, which also examines the correlation between the clinical score and magnetic resonance imaging. The anticipated clinical score values enable physicians to treat patients' illnesses promptly and with an early diagnosis.[17]

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 cognition (R2 = 24%) and memory (R2 = 25%) decline rates in sporadic 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 independently verified machine-learning approach may significantly lower the sample size required in AD clinical trials by predicting cognitive deterioration in sporadic prodromal AD. In order to predict rates of cognitive decline, we applied support vector regression to AD biomarkers obtained from structural magnetic resonance imaging (MRI), amyloid-PET, fluorodeoxyglucose positron-emission 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 utilizing model-based risk enrichment, the sample size required to identify treatment effects was calculated.[18]

The most prevalent type of dementia, Alzheimer's disease (AD), can result in a neurological 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 the 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 of 2480 AD, 2633 normal, and 1512 moderate cases, to develop Convolutional Neural Networks (CNNs) for the early 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.[19]

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 looked into whether an exhaustive study of survival could forecast the development of Alzheimer's disease in a comparable manner. 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.[20]

3. METHODOLOGY

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.

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.

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.

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.

Neural Networks:

- Neural networks consist of layers of neurons, with each layer transforming the input data before passing it to the next layer. The layers include an input layer, hidden layers, and an 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.

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

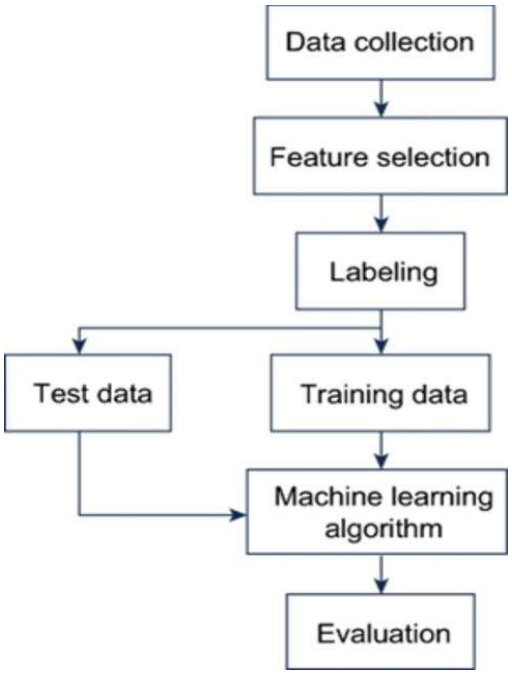

Fig1.Work Flow

Explanation for work flow:

Data Collection:

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

Feature Selection:

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

Lableing:

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

Data Split (test, train):

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

ML Algorithm:

- Choose a machine learning model based on the task. Train the model using training data. Learn patterns and relationships in the data.

Evaluation:

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

5. RESULTS

Table1.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

Diagrammatic representation of outputs

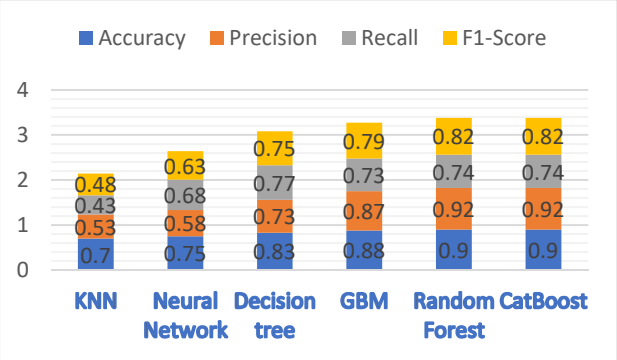

Fig2. Performance Metrics

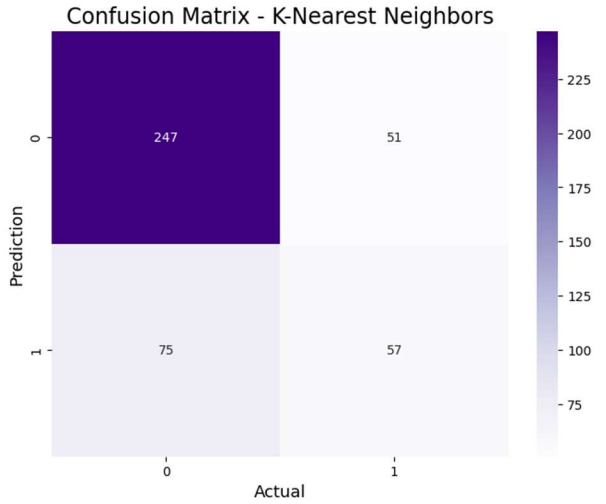

Fig3. Confusion matrix of KNN

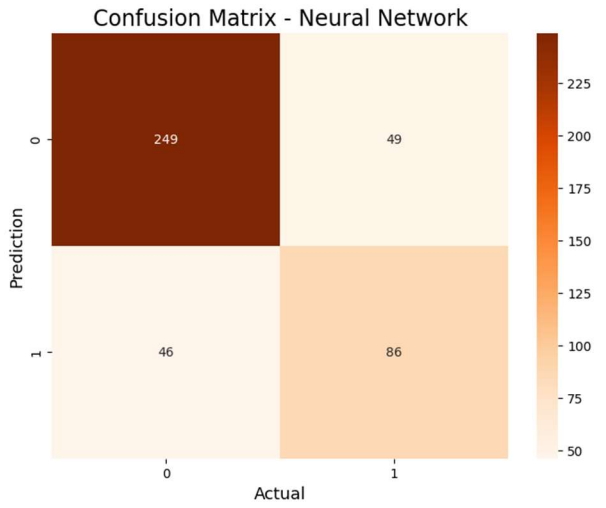

Fig4. Confusion matrix of NN

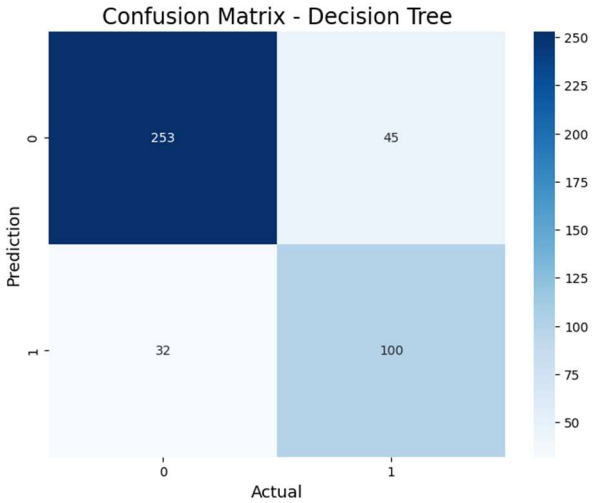

Fig5. Confusion matrix of DT

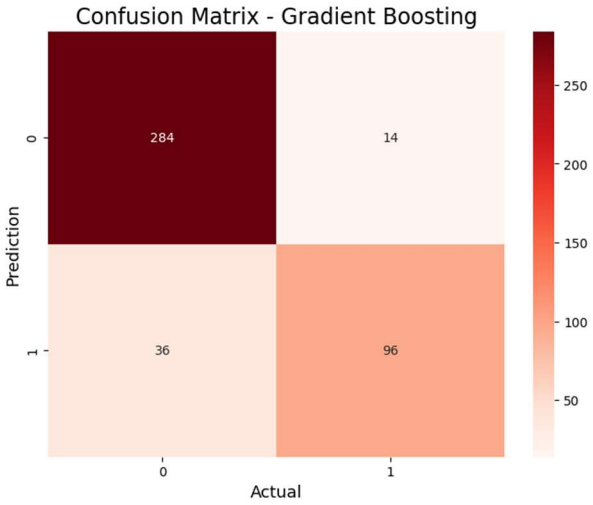

Fig6. Confusion matrix of GBM

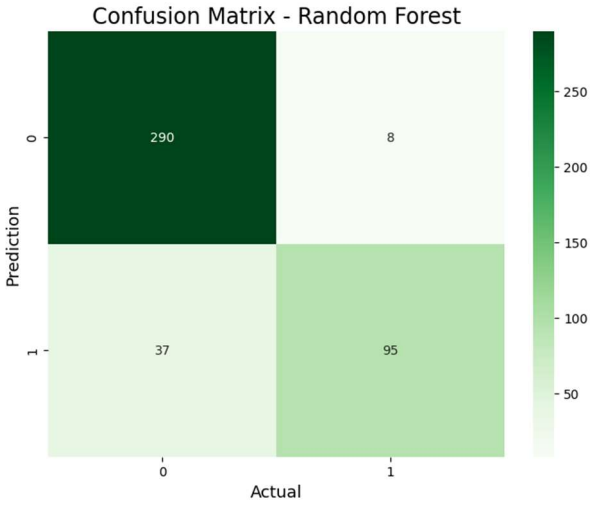

Fig7. Confusion matrix of KNN

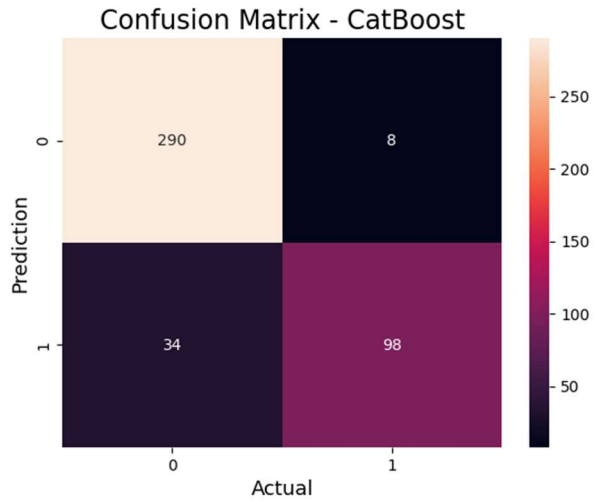

Fig8. Confusion matrix of CatBoost

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.

7. FUTURE SCOPES

In future work, hybridized algorithms may be explored to improve accuracy and robustness. For example, **ensemble methods** that combine multiple classifiers can be used to leverage the strengths of different models. Additionally, advanced techniques like **deep learning** may be investigated for more complex data.

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," *Front. Public Health*, vol. 10, p. 853294, Mar. 2022, doi: 10.3389/fpubh.2022.853294.

[2] J. Neelaveni and M. S. G. Devasana, "Alzheimer Disease Prediction using Machine Learning Algorithms," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India: IEEE, Mar. 2020, pp. 101–104. doi: 10.1109/ICACCS48705.2020.9074248.

[3] C. L. Saratxaga *et al.*, "MRI Deep Learning-Based Solution for Alzheimer's Disease Prediction," *JPM*, vol. 11, no. 9, p. 902, Sep. 2021, doi: 10.3390/jpm11090902.

[4] F. Saeed, "Applications of ML and DL Algorithms in The Prediction, Diagnosis, and Prognosis of Alzheimer's Disease," *AJBSR*, vol. 22, no. 6, pp. 779–786, Jun. 2024, doi: 10.34297/AJBSR.2024.22.003014.

[5] B. Lei *et al.*, "Deep and joint learning of longitudinal data for Alzheimer's disease prediction," *Pattern Recognition*, vol. 102, p. 107247, Jun. 2020, doi: 10.1016/j.patcog.2020.107247.

[6] N. Mattsson-Carlgren *et al.*, "Prediction of Longitudinal Cognitive Decline in Preclinical Alzheimer Disease Using Plasma Biomarkers," *JAMA Neurol*, vol. 80, no. 4, p. 360, Apr. 2023, doi: 10.1001/jamaneurol.2022.5272.

[7] J. Stevenson-Hoare *et al.*, "Plasma biomarkers and genetics in the diagnosis and prediction of Alzheimer's disease," *Brain*, vol. 146, no. 2, pp. 690–699, Feb. 2023, doi: 10.1093/brain/awac128.

[8] A. Leuzy *et al.*, "Biomarker-Based Prediction of Longitudinal Tau Positron Emission Tomography in Alzheimer Disease," *JAMA Neurol*, vol. 79, no. 2, p. 149, Feb. 2022, doi: 10.1001/jamaneurol.2021.4654.

[9] M. Bari Antor *et al.*, "A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–12, Jul. 2021, doi: 10.1155/2021/9917919.

[10] Q. Feng and Z. Ding, "MRI Radiomics Classification and Prediction in Alzheimer's Disease and Mild Cognitive Impairment: A Review," *CAR*, vol. 17, no. 3, pp. 297–309, May 2020, doi: 10.2174/1567205017666200303105016.

[11] K. N. Rao, B. R. Gandhi, M. V. Rao, S. Javvadi, S. S. Vellela, and S. Khader Basha, "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, Jun.

2023, pp. 85–90. doi: 10.1109/ICSCSS57650.2023.10169550.

[12] V. Sathiyamoorthi, A. K. Ilavarasi, K. Murugeswari, S. Thouheed Ahmed, B. Aruna Devi, and M. Kalipindi, "A deep convolutional neural network based computer aided diagnosis system for the prediction of Alzheimer's disease in MRI images," *Measurement*, vol. 171, p. 108838, Feb. 2021, doi: 10.1016/j.measurement.2020.108838.

[13] Y. Zhao, B. Ma, P. Jiang, D. Zeng, X. Wang, and S. Li, "Prediction of Alzheimer's Disease Progression with Multi-Information Generative Adversarial Network," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 3, pp. 711–719, Mar. 2021, doi: 10.1109/JBHI.2020.3006925.

[14] E. E. Bron *et al.*, "Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease," *NeuroImage: Clinical*, vol. 31, p. 102712, 2021, doi: 10.1016/j.nicl.2021.102712.

[15] W. Jung, E. Jun, and H.-I. Suk, "Deep recurrent model for individualized prediction of Alzheimer's disease progression," *NeuroImage*, vol. 237, p. 118143, Aug. 2021, doi: 10.1016/j.neuroimage.2021.118143.

[16] A. Balagopalan, B. Eyre, J. Robin, F. Rudzicz, and J. Novikova, "Comparing Pre-trained and Feature-Based Models for Prediction of Alzheimer's Disease Based on Speech," *Front. Aging Neurosci.*, vol. 13, p. 635945, Apr. 2021, doi: 10.3389/fnagi.2021.635945.

[17] B. Lei *et al.*, "Predicting clinical scores for Alzheimer's disease based on joint and deep learning," *Expert Systems with Applications*, vol. 187, p. 115966, Jan. 2022, doi: 10.1016/j.eswa.2021.115966.

[18] N. Franzmeier *et al.*, "Predicting sporadic Alzheimer's disease progression via inherited Alzheimer's disease-informed machine-learning," *Alzheimer's & Dementia*, vol. 16, no. 3, pp. 501–511, Mar. 2020, doi: 10.1002/alz.12032.

[19] A. W. Salehi, P. Baglat, B. B. Sharma, G. Gupta, and A. Upadhya, "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, Sep. 2020, pp. 156–161. doi: 10.1109/ICOSEC49089.2020.9215402.

[20] T. Nakagawa *et al.*, "Prediction of conversion to Alzheimer's disease using deep survival analysis of MRI images," *Brain Communications*, vol. 2, no. 1, p. fcaa057, Jan. 2020, doi: 10.1093/braincomms/fcaa057.