Alzheimer Disease Prediction using Machine Learning Techniques

Dr.A.Tamilarasi
Professor
Department of Computer Applications
Kongu Engineering College
Perundurai,Erode

Ms. K. Chitra
Assistant Professor(Sr.G)
Department of Computer Applications
Kongu Engineering College
Perundurai,Erode

Prakash. G, Priyasri. E
Sumanth .VM , PG Scholars
Department of Computer Applications
Kongu Engineering College
Perundurai,Erode
priyasrie.22mca@kongu.edu
prakashg.22mca@kongu.edu
sumanthvm.22mca@kongu.edu

Abstract— **This study explores machine learning approaches being used, namely Naive Bayes, Voting classifier, and Extra Trees Classifier, for Alzheimer's disease prediction. By leveraging these algorithms, the study aims to develop accurate models for Alzheimer's disease early recognition and detection. These techniques utilize different classification approaches, including probabilistic modeling, ensemble learning, and decision tree-based methods. The study evaluates the performance of these algorithms on relevant datasets and investigates their potential to improve Alzheimer's disease prediction accuracy. The results of this study make a meaningful contribution to the progression of Alzheimer's disease prediction through the application of machine learning techniques.**

Keywords-URL; Alzheimer Disease Classification, Machine Learning

I. INTRODUCTION

In today's digitally connected world, the brain is a vital organ that controls all actions of the human body. The discovery of numerous brain disorders has propelled machine learning into a potent instrument for crafting intelligent systems capable of decision-making and predictions grounded in data. In the specific context of Alzheimer's disease, forecasts can be generated by leveraging the Open Access Series of Imaging Studies (OASIS) dataset, and the model's effectiveness is gauged through metrics like Precision, Recall, and Accuracy.

Machine learning, a branch of Artificial Intelligence (AI), involves training and testing models to learn from natural phenomena and emulate human abilities. In this proposal, the focus is on utilizing biological parameters, including MR Delay, EDUC, sex, age, and more, as testing data. These parameters are employed to assess the precision of various machine learning methodologies, including the voting classification algorithm, naive Bayes model, and logistic regression. The objective is to compute the accuracy of these algorithms and identify the most efficient approach based on the analyses conducted in this study.

II. LITERATURE SURVEY

During the development of the system, several papers were referred to. On study [1] In their study titled "Alzheimer's Disease Detection Using Various Machine Learning Algorithms" (Harika S and Yamini T, 2022), the authors concentrated on identifying the most prevalent form of Alzheimer's disease, which is characterized by memory loss. They employed four distinct machine learning classification methods, including the Support Vector Classifier, Logistic Regression, Naive Bayes, and Decision Tree, with the aim of enhancing accuracy and overall performance in detecting the disease. The research's primary objective was to elevate the accuracy rate in Alzheimer's detection by implementing these machine learning classifier techniques.[2] "Considering PET images, Autoencoder-based Value Sorting for Classification of Memory Disorder" (Pham Minh Tuan and Trong-Le Phan, 2021), the authors propose a method that utilizes Autoencoder networks to rank the effectiveness of brain regions in distinguishing Normal PET imaging used to differentiate healthy (NC) from Alzheimer's disease (AD) brains. They employ a Support Vector Classifier (SVC) as the machine learning classification technique in their research. [3] "An Overview of ML Methods for dementia Detection" (Tanveer M and Richhariya B, 2020), the authors analyze 165 publications, published between 2005 and 2019, examine different extraction of features and machine learning methods. for the classification and identification of Alzheimer's disease in MR images.

[4] In their study titled "Machine Learning-Based Prediction of Alzheimer's Disease" (Malavika G and Rajathi N, 2020), the authors set out to forecast Alzheimer's disease by leveraging machine learning algorithms alongside psychological parameters like age, number of visits, MMSE, and education. They explored six prominent categories of machine learning techniques, including Logistic Regression, Decision Tree Classifier, K-Nearest Neighbor, Support Vector Machine, AdaBoost Classifier, and Random Forest Classifier, and their findings revealed accuracy results spanning from 66.9% to 86.8%. [5], In their study titled "Machine Learning-Based Prediction of Early-Stage Alzheimer's Syndrome" (Kavitha, Vinodhini Mani, and Srividhya R S, 2022), the authors tackle the rising prevalence of Alzheimer's disease and the crucial requirement for early prediction. Their research explores machine learning methodologies encompassing six primary categories, such as Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting, and Voting classifiers. The outcomes of their investigation demonstrate accuracy levels that span from 80.46% to 86.92%.[6]."An Innovative Machine Learning Approach for Alzheimer's Disease Detection" (Lin Liu and Shenghui Zhao, 2020), the authors propose a novel approach that uses spectrogram characteristics taken from conversations to determine whether someone has Alzheimer's disease. This method offers a non-invasive and cost-effective way to collect real-time and accurate data, enabling early-stage understanding of disease development and proactive measures.Linear SVC, Logistic Regression CV, Decision Tree, and Bagging are among the machine learning algorithms investigated, with a voice dataset comprised of Alzheimer's disease sufferers as well as healthy control participants [7]. "Identification of Alzheimer's Disease Using Modular Machine Learning from Retina Vasculature" (Jianqiao Tian and Glenn Smith, 2021), the authors emphasize the importance of routine screening and early diagnosis for Alzheimer's disease to maximize treatment benefits. They highlight the limitations of current diagnostic methods, such as invasive procedures and expensive neuroimaging techniques. The survey of machine learning techniques in their research focuses on Support Vector Machine (SVM) for Alzheimer's disease classification using retinal vasculature data.

[8]. "Deep Learning Used in MRI for Automatic Alzheimer's Disease Detection“(C Muhammed Raees and Vinu Thomas, 2021), the authors propose an early and automated deep learning-based system to predict Alzheimer's disease using MRI data, achieving high accuracy rates of 80-90%. [9], "Deep learning techniques in MRI for automatic Alzheimer's disease detection" (Mahda Nasrolahzadeh and Shahryar Rahnamayan, 2021),The authors highlight the usefulness of GP-based techniques in diagnosing Alzheimer's disease through impulsive speech analysis by creating a novel Computer-Aided Diagnosis (CADx) system for the condition and demonstrating its effectiveness compared to cutting-edge approaches.. [10] "A More Accurate Multi-Modal Dependent Machine Learning Approach for Alzheimer's Disease Prognosis" (Afreen Khan and Swaleha Zubair, 2022), the authors propose a five-stage machine learning pipeline using MRI brain images to diagnose Alzheimer's disease with higher accuracy. They utilize supervised learning-based generic framework and employ the Random Forest classifier, resulting in improved performance compared to other classifiers.

III. ALZHEIMER’S DISEASE

Alzheimer's illness, likewise suggested as Alzheimer's, is a neurodegenerative sickness that generally begins moderately, intuitively deteriorates. As it grounds for 60-70% of occurrence of dementia. Extensively concede untimely side effect troubles in recalling late occasions. sickness propels, complication can incorporate issues with language, confound, state of mind wags, inspiration loss, self-disregard, and behavioralists. An individual's circumstances declines, they frequently back out from family and society. Steadily, physical processes are lost, eventually precisely demise. Albeit the speed of movement can change, the average future following termination is three to nine years.

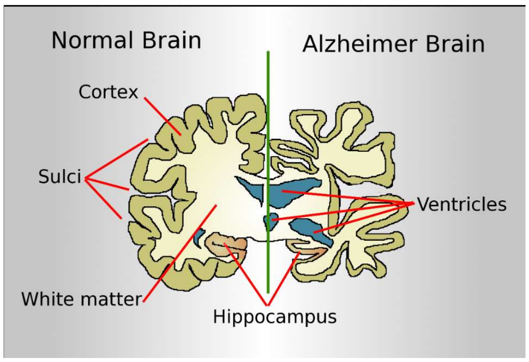

Figure 1. Alzheimer's disease

A mind cancer happens when unusual cells structure inside the cerebrum. There are two primary kinds of growths: destructive (threatening) cancers and harmless (non-malignant) cancers. Malignant growths can be partitioned into essential cancers, what begin inside the cerebrum, and optional cancers, ordinarily have span from cancers situated external the mind, admitted as cerebrum metastasis growths. A spacious compass of cancers might deliver complications that fluctuate fortuitous. These side effects might incorporate cerebral pains, seizures, issues with vision, spewing and mental changes. The migraine is traditionally more regrettable in the first part of the day and disappears with spewing. Different side effects might incorporate trouble strolling, talking or with sensations. As the sickness advances, obviousness might happen.

IV.METHODOLOGY

The process consists of five key steps. The first stage is to gather the dataset, which is immediately followed by data preprocessing, in which the original data set is turned into an organized and suitable format. Next, feature extraction techniques are applied to convert the dataset into numerical features that can be used by the classification algorithm. The dataset is then fed into the classification algorithm for analysis and prediction. Finally, the effectiveness of each categorization algorithm is assessed. To assess its effectiveness in handling the given dataset. The careful execution of each step in the process is crucial to ensure accurate and reliable classification results, ultimately contributing to better decision-making and insights from the dataset.

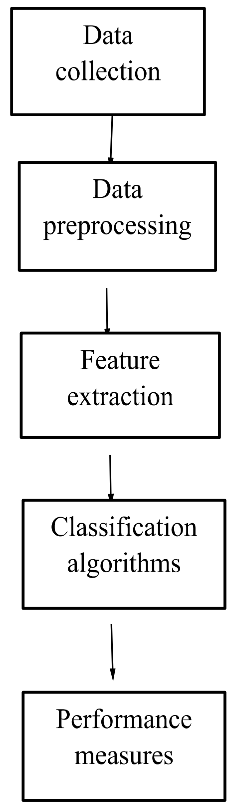

Figure 2. Methodology of experiment

A. DATA COLLECTION

The collection of data is a planned method of gathering knowledge on a specific subject. The dataset was compiled using Kaggle. There is 378 OASIS Dataset in the dataset. A project The Open Access Series of Imaging Studies (OASIS) intends to provide openly available MRI of the brain data sets to the the scientific world. Both non demented and demented data make up the dataset.

Subject ID	MRI ID	Group	Visit	MR Delay	M/F	Hand
OAS2_0001	OAS2_0001_MR1	Nondemented	1	0	M	R
OAS2_0001	OAS2_0001_MR2	Nondemented	2	457	M	R
OAS2_0002	OAS2_0002_MR1	Demented	1	0	M	R
OAS2_0002	OAS2_0002_MR2	Demented	2	560	M	R
OAS2_0002	OAS2_0002_MR3	Demented	3	1895	M	R
OAS2_0004	OAS2_0004_MR1	Nondemented	1	0	F	R
OAS2_0004	OAS2_0004_MR2	Nondemented	2	538	F	R
OAS2_0005	OAS2_0005_MR1	Nondemented	1	0	M	R
OAS2_0005	OAS2_0005_MR2	Nondemented	2	1010	M	R

Figure 3. Dataset

B. DATA PREPROCESSING

A raw dataset is turned into a dataset that is clean during the data preparation process and then it may be used by a machine learning classifier. Every time data is gathered from numerous sources, it is done in such a raw form that processing is difficult. Datasets usually have noise, data that is missing, and can even arrive in an incorrect format, rendering them insufficient for direct usage with machine learning approaches. A machine learning model is more accurate and effective when the data has been preprocessed.

- Data integration - merging multiple datasets to generate a unified and distinct body of information.
- Data reduction - technique used to reduce the volume of data by eliminating inaccurate or irrelevant information.
- Data Transformation - is the process of turning unstructured data into an organized form while retaining the original information.

C. FEATURE EXTRACTION

The collected characteristics are subsequently fed into machine learning models, such as Naive Bayes, Voting Classifier, and Extra Trees Classifier. These programme performance is assessed using common assessment measures such as Precision, Recall, and Accuracy. The results of this study provide insights into the effectiveness of feature extraction techniques applied to the OASIS dataset for Alzheimer's disease prediction. By capturing the relevant patterns and characteristics of the neuroimaging data, these extracted features improve the efficiency of machine learning algorithms. The findings contribute to advancing the field of Alzheimer's disease research by demonstrating the potential of feature extraction from OASIS and its impact on accurate prediction and diagnosis.

D. CLASSIFICATION ALGORITHMS

The dataset is classified using various machine learning algorithms, Naive Bayes, Voting classifier and Extra trees Classifier. The set of data is separated into two parts: training and testing sets for model evaluation and validation.

1. Naive Bayes

Popular supervised learning techniques include the Bayes' theorem-based Naive Bayes algorithm, which is particularly effective for classification problems. It is especially useful for classifying text because it makes heavy use of training data. This straightforward yet effective classifier produces machine learning models that make precise predictions quickly. As a probability-based classifier, it makes predictions by calculating the likelihood that a certain event will occur. Spam filtering, sentiment analysis, and article categorization are just a few of the areas where naive bayes is used, showing how versatile it is in dealing with classification problems.

APPLICATION ASSOCIATED WITH THE DATASET:

Step 1: Using data cleaning techniques, pre-process the OASIS dataset.
Step 2: Using Naive Bayes to fit the training set, which is made up of subject IDs.
Step 3: Using the trained Naive Bayes model to predict the test set's labels.
Step 4: Establishing a confusion matrix to evaluate the forecasts' accuracy.
Step 5: Visualizing the test set results for additional analysis and interpretation is step five.

2. Voting Classifiers

Voting classifiers are neural network approaches that predict a result (class) based on the category with the greatest likelihood to generate the result. They gain aptitude by studying a variety of scenarios. The results of each algorithm that was entered into the system for voting are simply combined together to estimate what would emerge out of class based on the most votes. The idea is to design a single model that learns from other models and predicts based on their aggregate votes or consensus for each probable output class, as opposed to building different, specialised models and evaluating their accuracy separately.

APPLICATION ASSOCIATED WITH THE DATASET:

Step 1: Preprocess the OASIS dataset in step one.
Step 2: Apply the Bootstrap methodology and choose samples from the practice data.
Step 3: Dividing the voting classifier nodes by picking the feature from a random sample of n features (N) that has the lowest Gini value.
Step 4: To establish a K group, repeat Steps 1 and 2 K times.
Step 5:Voting is used to choose the classification result (LABEL) by fusing naive bayes into a voting classifier.

3. Extra trees classifier

A number of decision trees are produced using the Extra Trees algorithm, which is similar to the Random Forest method. It uses a different approach, though. Each decision tree in Extra Trees is created by a random and non-repetitive sampling approach, ensuring that each tree is trained on a dataset made up completely of unique samples. A preset number of features from the entire collection of accessible features are further randomly selected for each unique tree. This novel strategy for diversity and randomness significantly improves the stability and effectiveness of the Extra Trees algorithm, making it a useful tool in a variety of machine learning applications.

APPLICATION ASSOCIATED WITH THE DATASET:
Step 1: Commence the tree at the OASIS root node, using the complete dataset as the starting point.
Step 2: Progress to step 2 by applying the Attribute Selection Measure (ASM).
Step 3: Subset attribute with potential values in step three.
Step 4: In step four, create more tree branches with the relevant properties.
Step 5: Using the produced subsets, iteratively build fresh Extra trees.

E. PERFORMANCE MEASURES

Evaluating model performance is crucial in building an effective machine learning model. Metrics of performance evaluate the model's effectiveness. The proportion of accurate predictions to all other predictions is known as accuracy.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \qquad (1)$$

Were,
True Positive (TP) - correctly affected people;
False Positive (FP) – not affected;
True Negative (TN) - correctly affected people.
False Negative (FN) - people misclassified as safe.
Confusion matrix - a performance evaluation table for classification algorithms.

TABLE II. ACCURACY OF ALGORITHMS

MODEL	ACCURACY SCORE
Naive Bayes	87.5
VT	80.373
ET	86.6

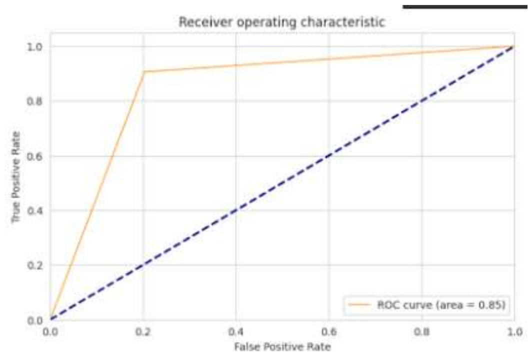

Figure 4. Accuracy chart

V. CONCLUSION AND FUTURE WORK

Alzheimer's syndrome is a degenerative neurologic condition affecting cognitive abilities, memory, and personality. It is the primary factor in a large majority of instances of dementia, making it the leading cause.The disease is characterised by the buildup of amyloid and tau tangle plaques in the brain, leading to the loss of brain cells and impaired brain function. The application of machine learning algorithms such as Extra Trees, Voting Algorithm, and Naive Bayes can provide valuable insights and predictive capabilities in the context of Alzheimer's disease. By comparing various features of the condition, such as diagnostic procedures, comparative algorithms have performed an important role in expanding our knowledge of Alzheimer's disease. treatment interventions, and disease progression. These algorithms enable researchers and clinicians to identify patterns, evaluate effectiveness, and make informed decisions in managing the disease. Extra trees algorithm shows the best performance with an accuracy of 87.5%.

Future work: The machine learning methods bring this aspect to reality, explore advanced techniques for feature selection and extraction to identify the most relevant biomarkers, genetic factors, or imaging features associated with Alzheimer's disease. This can help improve the accuracy and interpretability of the models built using these algorithms.

REFERENCES

[1] Harika S and Yamini T, et.al., [2022], "Alzheimer's Disease Detection Using Different Machine Learning Algorithms "in 2022 Decision Tree 93.7%, SVM 56.7%, Logistic Regression 57.1% ,Naïve Bayes 83.8%

[2] Pham Minh Tuan and Trong-Le Phan et al. (2021) developed an autoencoder-based feature ranking method for Alzheimer's disease classification using PET images.

[3] Tanveer M and Richhariya B et al. (2020) conducted a review of machine learning techniques for the diagnosis of Alzheimer's disease, focusing on feature extraction and classification using support vector machines, artificial neural networks, deep learning, and ensemble methods.

[4] Malavika G and Rajathi N et al. (2020) conducted a study on forecasting Alzheimer's disease using machine learning algorithms based on psychological parameters, achieving accuracies ranging from 66.9% to 86.8% with various classifiers.

[5] Kavitha, Vinodhini Mani and Srividhya R S et al. (2022) explored machine learning models for early-stage prediction of Alzheimer's disease, achieving accuracies ranging from 80.46% to 86.92% with decision tree, random forest, support vector machine, gradient boosting, and voting classifiers.

[6] Lin Liu and Shenghui Zhao et al. (2020) proposed a new machine learning method using spectrogram features extracted from speech data for identifying Alzheimer's disease, aiming to enable early-stage understanding and intervention using non-invasive speech data.

[7] Jianqiao Tian and Glenn Smith et al. (2021) utilized modular machine learning techniques for Alzheimer's disease classification based on retinal vasculature, aiming to enhance routine screening and early diagnosis without invasive procedures or expensive neuroimaging modalities.

[8] C Muhammed Raees and Vinu Thomas et al. (2021) developed an automated deep learning-based system using MRI data for the detection of Alzheimer's disease, achieving high accuracy and emphasizing the importance of early diagnosis for better clinical, social

[9] Mahda Nasrolahzadeh and Shahryar Rahnamayan et al. (2021) developed a novel computer-aided diagnosis system using genetic programming for the automated detection of Alzheimer's disease based on higher-order spectra features, achieving better performance compared to state-of-the-art approaches and emphasizing the practicality of the proposed method.

[10] Afreen Khan and Swaleha Zubair et al. (2022) proposed an improved multi-modal machine learning approach for the prognosis of Alzheimer's disease, utilizing an automated classification system and a five-stage ML pipeline on MRI brain images to achieve higher accuracy and improved performance.

[11] Z. Fan, F. Xu, X. Qi, C. Li, and L. Yao, "Classification of Alzheimer's disease based on brain MRI and machine learning," Neural Comput. Appl., vol. 32, no. 7, pp. 1927–1936, Apr. 2020.

[12] G. Battineni, N. Chintalapudi, F. Amenta, and E. Traini, "A comprehensive machine- learning model applied to magnetic resonance imaging (MRI) to predict Alzheimer's disease (AD) in older subjects," J. Clin. Med., vol. 9, no. 7, p. 2146, Jul. 2020.

[13] R. Binaco, N. Calzaretto, J. Epifano, S. McGuire, M. Umer, S. Emrani, V. Wasserman, D. J. Libon, and R. Polikar, "Machine learning analysis of digital clock drawing test performance for differential classification of mild cognitive impairment subtypes versus Alzheimer's disease," J. Int. Neuropsychological Soc., vol. 26, no. 7, pp. 690–700, Aug. 2020.

[14] E. Lella, A. Lombardi, N. Amoroso, D. Diacono, T. Maggipinto, A. Monaco, R. Bellotti, and S. Tangaro, "Machine learning and DWI brain communicability networks for Alzheimer's disease detection, " Appl. Sci., vol. 10, no. 3, p. 934, Jan. 2020.

[15] G. Uysal and M. Ozturk, "Hippocampal atrophy based Alzheimer's disease diagnosis via machine learning methods," J. Neurosci. Methods, vol. 337, May 2020, Art. no. 108669.

[16] F. Vecchio, F. Miraglia, F. Alü, M. Menna, E. Judica, M. Cotelli, and P. M. Rossini, "Classification of Alzheimer's disease with respect to physiological aging with innovative EEG biomarkers in a machine learning implementation," J. Alzheimer's Disease, vol. 75, no. 4, pp. 1253–1261, Jun. 2020.

[17] A. Khan and S. Zubair, ‘‘An improved multi-modal based machine learning approach for the prognosis of Alzheimer's disease,'' J. King Saud Univ.-Comput. Inf. Sci., pp. 1–19, Apr. 2020.

[18] L. Liu, S. Zhao, H. Chen, and A. Wang, ‘‘A new machine learning method for identifying Alzheimer's disease,'' Simul. Model. Pract. Theory, vol. 99, Feb. 2020, Art. no. 102023.

[19] J. Sheng, M. Shao, Q. Zhang, R. Zhou, L. Wang, and Y. Xin, ‘‘Alzheimer's disease, mild cognitive impairment, and normal aging distinguished by multi-modal parcellation and machine learning,'' Sci. Rep., vol. 10, no. 1, pp. 1–10, Dec. 2020.

[20] J. Ren, B. Zhang, D. Wei, and Z. Zhang, ‘‘Identification of methylated gene biomarkers in patients with Alzheimer's disease based on machine machine learning,'' J. Clin. Med., vol. 9, no. 9, p learning,'' BioMed Res. Int., vol. 2020, pp. 1–11, Mar. 2020.

[21] M. Karaglani, K. Gourlia, I. Tsamardinos, and E. Chatzaki, ‘‘Accurate blood-based diagnostic biosignatures for Alzheimer's disease via automated. 3016, Sep. 2020.

[22] H. Guo and Y. Zhang, ‘‘Resting state fMRI and improved deep learning algorithm for earlier detection of Alzheimer's disease,'' IEEE Access, vol. 8, pp. 115383–115392, Jul. 2020.

[23] C. Lian, M. Liu, J. Zhang, and D. Shen, ‘‘Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 4, pp. 880–893, Apr. 2020.

[24] S. El-Sappagh, T. Abuhmed, S. M. Riazul Islam, and K. S. Kwak, ‘‘Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data,'' Neurocomputing, vol. 412, pp. 197–215, Oct. 2020.