Section A-Research paper
ISSN 2063-5346

DETECTION OF ALZHEIMER DISEASE (AD2) AUTOIMMUNE DISORDER USING MACHINE LEARNING

Dr. P. Pritto Paul[1],B Joseph Joel[2] ,Dr.M.USha[3], D.Saral Jeeva Jothi[4]

Associate Professor[1.3], Department of CSE,Velammal Engineering College, ,Chennai, India [1,3]
PG Student[2], Department of CSE, Velammal Engineering College, ,Chennai,TamilNadu,India[2]
Associate Professor[4], Department of CSE,Velammal Engineering College, ,Chennai, India [4]

ABSTRACT

Alzheimer's disease (AD) is a neurodegenerative disease that causes irreversible damage to several brain regions, including the hippocampus causing impairment in cognition, function, and behaviour. Early diagnosis of the disease will reduce the suffering of the patients and their family members.This condition affects around 45 million individuals worldwide.Machine learning refers to an probabilistic technique that allow system to acquire knowledge from huge amount of data.In particular, the performance of different ML algorithms has been evaluated against their detection accuracy. The detection techniqueuses K-Means groups the unlabeled datasets into different clusters and Decision Treeis a tree- structured classifier ,where internal nodes represent the features of a datasets where branches represent the decision rules and each leaf node represents the outcome.Alzheimer disease is inherited in an auto- somal dominant pattern , which means one copy of an altered gene in each cell is sufficient to cause the disorder. In most cases, an affected person inherits the altered gene from one affected parent.To overcome this problem,this paper introduces an detection technique using machine learning Predictive genetic testing by collecting the genetic samples using k mean algorithm the datasetare stored Into different clustersand using decision tree algorithm the genetic samples are structured into relevant classiferWith the help of algorithms it is trained to detect the genetic pattern which has high pior risk of Alzheimer'sDisorder.

KEYWORDS:Alzheimer'sDisorder, genetic pattern,machine learning,detection.

I. INTRODUCTION

Machine learning is programming computers to optimize a performance criterion using example data or past experience . We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data.

The development of algorithms and statistical models that enable computers to improve their performance in tasks through experience. These algorithms and models are designed to learn from data and make predictions or decisions without explicit instructions.

Alzheimer's disease is the most common type of dementia[1] Dementia isn't a single, specific disease instead, it's a term used to describe a combination of symptoms that includes the progressive loss of different cognitive functions, such as memory, orientation, language, and attention. As a result, Alzheimer's disease and other types of dementia have a significant impact on a person's everyday life, and in the lives of their loved ones and caregivers.Similarly to other types of dementia, Alzheimer's disease leads to progressive impairment and different symptoms,include:Memoryloss,Mood ,Agitation ,Anxiety,Disorientation,Depression,Apathy,Aggression,Hallucinations[1].

Indeed, certain genetic variants can increase your risk of developing Alzheimer's disease, but this increased risk doesn't necessarily have to translate into having the disease. There are many different preventative measures that you can

adopt in order to prevent Alzheimer's disease — in fact, practically anyone can benefit from these healthy strategies, whether they have a heightened risk of Alzheimer's disease or not.

One gene plays a particularly important role in determining whether you develop Alzheimer's or not: the APOE gene[2].

The APOE gene plays different roles in our brain function; among other things, it produces a protein called apolipoprotein E,[11] which helps eliminate deposits of a substance called amyloid-beta peptide that can accumulate and form amyloid plaques.[9][11] These amyloid plaques hinder brain function and eventually lead to Alzheimer's disease.

This gene also aids in the maintenance of a normal brain function synaptic plasticity, which is the process through which our brain cells establish synapses to communicate with each other. For these reasons, the APOE gene and its alleles play a very important role in the development of Alzheimer's disease.3

APOE2:[3][4] this is the least common allele of this gene, and it enables the APOE gene to synthesize larger amounts of apolipoprotein E[11], thus protecting your brain against Alzheimer's.

APOE3:[3][4] this is the "neutral" allele of the APOE gene since it doesn't seem to affect the production of apolipoprotein E. As a result, this variant neither decreases nor increases your susceptibility to Alzheimer's. However, research has found that APOE3 in combination with specific variants in other genes[11] could lead to an increased risk of Alzheimer's disease for some people.

APOE4:[3][4] finally, the APOE4 allele hinders the production of apolipoprotein E[11], which could put you at an increased risk of developing the disease.

Alzheimer disease is inherited in an autosomal dominant pattern , which means one copy of an altered gene in each cell is sufficient to cause the disorder. In most cases, an affected person inherits the altered gene from one affected parent[5].Machine learning, an AI branch use technique that allow system to acquire knowledge from huge amount of data.In particular, the performance of different ML algorithms has been evaluated against their detection accuracy.TheClustering model is firstly pre-trained on large text datasets to get the pre-trained language model, and then,based on such a model,[2] an AD classification model is performed on small training sets. using machine learning deductive genetic testingBy collecting the genetic samples using k mean algorithm the datasetare stored Into different clusters[4][6]and using decision tree algorithm the genetic samples are structured into relevant classiferWith the help of algorithms it is trained to detect the genetic pattern which has high pior risk of Alzheimer'sDisorder.

II.RELATED WORKS

A. Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models

Alzheimer's disease (AD) is the leading cause of dementia in older adults. There is currently a lot of interest in applying machine learning to find out metabolic diseases like Alzheimer's and Diabetes that affect a large population of people around the world. Their incidence rates are increasing at an alarming rate every year. In Alzheimer's disease, the brain is affected by neurodegenerative changes. As our aging population increases, more and more individuals, their families, and healthcare will experience diseases that affect memory and functioning. These effects will be profound on the social, financial, and economic fronts. In its early stages, Alzheimer's disease is hard to predict. A treatment given at an early stage of AD is more effective, and it causes fewer minor damage than a treatment done at a later stage. Several techniques such as Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting, and Voting classifiers have been employed to identify the best parameters for Alzheimer's disease prediction. Predictions of Alzheimer's disease are based on Open Access Series of Imaging Studies (OASIS) data, and performance is measured with parameters like Precision, Recall, Accuracy, and F1-score for ML models. The proposed classification scheme can be used by clinicians to make diagnoses of these diseases. It is highly beneficial to lower annual mortality rates of Alzheimer's disease in early diagnosis with these ML algorithms. The proposed work shows better results with the best validation average accuracy of 83% on the test data of AD. This test accuracy score is significantly higher in comparison with existing works

B. A Transfer Learning Method for Detecting Alzheimer's Disease Based on Speech and Natural Language Processing

Alzheimer's disease (AD) is a neurodegenerative disease that is difficult to be detected using convenient and reliable methods. The language change in patients with AD is an important signal of their cognitive status, which potentially helps in early diagnosis. In this study, we developed a transfer learning model based on speech and natural language processing (NLP) technology for the early diagnosis of AD. The lack of large datasets limits the use of complex neural network models without feature engineering, while transfer learning can effectively solve this problem. The transfer learning model is firstly pre-trained on large text datasets to get the pre-trained language model, and then, based on such a model, an AD classification model is performed on small training sets. Concretely, a distilled bidirectional encoder representation (distilBert) embedding, combined with a logistic regression classifier, is used to distinguish AD from normal controls. The model experiment was evaluated on Alzheimer's dementia recognition through spontaneous speech datasets in 2020, including the balanced 78 healthy controls (HC) and 78 patients with AD. The accuracy of the proposed model is 0.88, which is almost equivalent to the champion score in the challenge and a considerable improvement over the baseline of 75% established by organizers of the challenge. As a result, the transfer learning method in this study improves AD prediction, which does not only reduces the need for feature engineering but also addresses the lack of sufficiently large datasets.

C. Alzheimer Disease Detection Methods: Automatic Pipelines and Machine Learning Techniques

Alzheimer's Disease (AD) is becoming increasingly prevalent across the globe, and various diagnostic and detection methods have been developed in recent years. Several techniques are available, including Automatic Pipeline Methods and Machine Learning Methods that utilize Biomarker Methods, Fusion, and Registration for multimodality, to pre-process medical scans. The use of automated pipelines and machine learning systems has proven beneficial in accurately identifying AD and its stages, with a success rate of over 95% for single and binary class classifications. However, there are still challenges in multi-class classification, such as distinguishing between AD and MCI, as well as sub-stages of MCI. The research also emphasizes the significance of using multi-modality approaches for effective validation in detecting AD and its stages.

D.Alzheimers Disease Detection Using Different Machine Learning Algorithms

Alzheimer's disease is the most common form of dementia affecting the brain's parts. A broad term used to describe illnesses and conditions that causes a deterioration in memory, language, and other cognitive abilities severe enough to interface with daily life is "dementia". According to estimates, this disease affects 6.2 million Americans and 5 million people in India aged 65and older. In 2019, the most recent year for which data are available, official death certificates reported 121,499 deaths from AD, making Alzheimer's the "sixth leading cause of death in the country and the fifth leading cause of death for people 65 and older". In this paper, we suggest several machine Learning algorithms like Decision trees, SVM, Logistic regression, and Naive Bayes identify AD at an early stage. The Alzheimer\'s Disease Neuroimaging Initiative (ADNI) and the Open Access Series of Imaging Investigations (OASIS) provide data sets white used to detect the disease in its early stage. The datasets consist of longitudinal MRI data (age, gender, mini mental status, CDR) By taking into account many factors in each method, such as precision, F1 Score, Recall, and specificity are calculated. The results obtained 93.7% of maximum accuracy for the Decision Tree Algorithm.

E. Detection and analysis of Alzheimer's disease using various machine learning algorithms

Alzheimer's is a dynamic ailment that decimates the mind's memory and its general functioning. Unfortunately till now, no single test can diagnosis this disease. Cerebrum checks alone can't be considered as a key factor to decide if the individual is experiencing it or not. As of now, the physician is in a conclusion that an individual is suffering from Alzheimer's on premise of the reports of the relations in regards to the social proclivity and checking the past clinical record. Artificial intelligence along with Machine Learning calculations perhaps in a

situation to adjust this model. Big processing, in light of the fact that the data is taken through various sources with complex and creating circumstances that make certain to develop later on. Along these lines, in that, we'll take consequences of what extent level of patients get the illness as positive data and negative data. The proposed arrangement shows a big processing model, from the data mining perspective. Utilizing classifiers, this paper presents the work by preparing Alzheimer's rate and qualities are appearing as a disarray framework using different machine learning algorithms. The earlier research proved that the detection of Alzheimer's disease using Support Vector Machine classifier and obtained very less accuracy. In view of this there is need of increasing the accuracy. So, this paper presenting different algorithms to classify the data to improve the efficiency in detecting the mentioned disease and observed that the Support Vector Machine with linear kernel model gives better accuracy than other models.

F. Artificial intelligence and machine learning for Alzheimer's disease

As the world population ages, it is estimated that the population worldwide above the age of 65 years old will increase from 420 million in 2000 to almost 1 billion by 2030.1 Dementia, with Alzheimer's disease (AD) as the leading cause, is expected to rise in tandem. AD accounts for 60%–80% of all dementia cases,2 with an estimated 5–7 million new cases diagnosed each year.3 Despite intensive research, the diagnosis of AD is currently made through a combination of clinical assessment, neuroimaging and detection of biomarkers from positron emission tomography or cerebrospinal fluid examination,4 with patients facing issues including high costs, invasiveness of the procedures.5 Hence, alternative identification of AD without the use of costly or invasive tests remains a challenge that is difficult to surmount.

III.ARCHITECTURE DIAGRAM

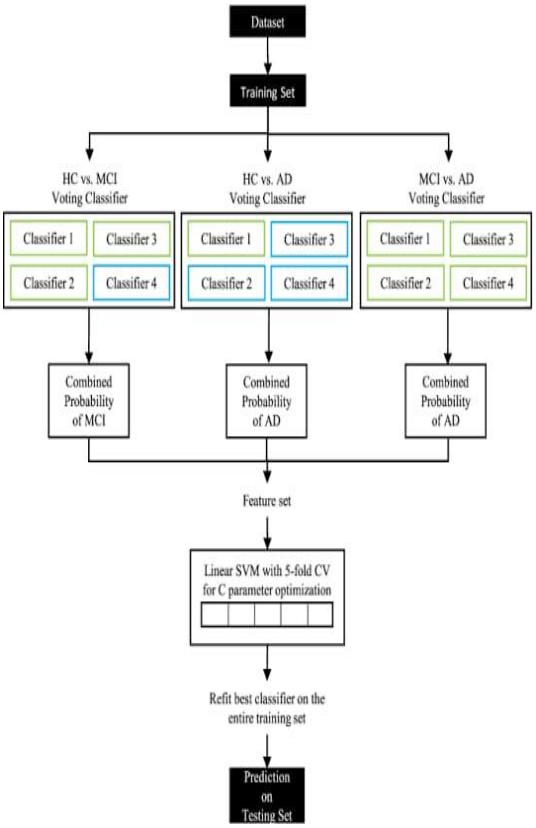

Fig.3.1 Architecture Flow diagram of classifier

Architecture diagram shows how[5][9]*Classes can be called as targets/labels or categories.The main goal of the Classification model is to identify the category of a given* dataset is converted into training set from this training set each and every gene samples are classified according to their genetic pattern into a clusters named[3][4] (HC vs MCI ,HC vs AD,MCI vs AD) from this clusters they are combined according to their probability by using genetic pattern through which we can obtain *classification deductive modeling is the task of approximating the mapping function[10] from input variables to discrete output variables* classifier for entire training set and prediction on testing set.

IV.PROPOSED MODEL

4.1 CLUSTERING MODEL:

Clustering model nests data together by common attributes. It works by grouping things or people

with shared characteristics or behaviors and plans strategies for each group at a larger scale[10]. An example is in determining credit risk for a loan applicant based on what other people in the same or a similar situation did in the past.Clustering or cluster analysis[2][4] is a machine learning technique, which groups the unlabelled dataset. It can be defined as A way of grouping the data points into different clusters, consisting of similar data points. The objects with the possible similarities remain in a group that has less or no similarities with another group.It does it by finding some similar patterns in the unlabelled dataset[3][6] such as shape, size, color, behavior, etc., and divides them as per the presence and absence of those similar patterns.After applying this clustering technique, each cluster or group is provided with a cluster-ID[9]. ML system can use this id to simplify the processing of large and complex datasets[5].The genetic samples are provided with an specific cluster -ID[9] and so that it's make it simple to use such type of data set without an large amount of complexity and process.

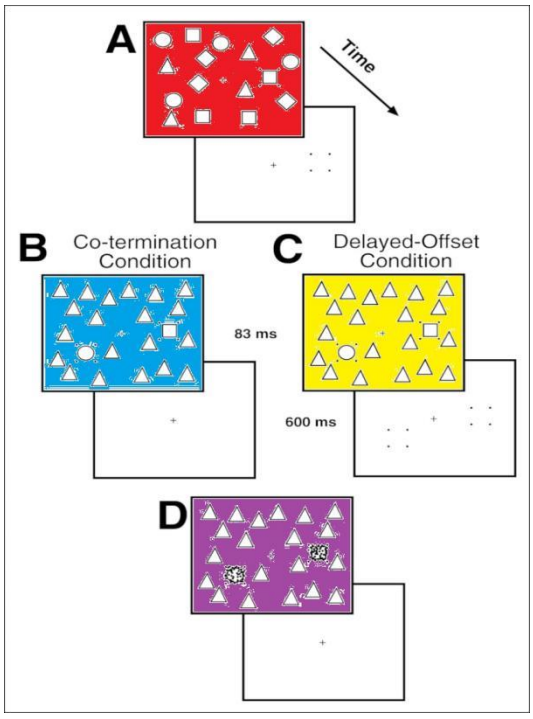

Fig.4.1.1 Flow diagram of clustering technique

4.2MODULES IMPLEMANTATION:

The Classification model is used to identify the category of new observations on the basis of training data[6]. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, Yes or No, 0 or 1[5][10], Spam or Not Spam,etc. Classes can be called as targets/labels or categories.The main goal of the Classification model is to identify the category of a given dataset, and these algorithms are mainly used to deduct the output for the categorical data.The model deducts or draws a conclusion to the input data given for training[4][6], it will predict the class or category for the data.Classification is a process of categorizing a given set of data into classes, The process starts with deducting the class of given data points[9]. The classes are often referred to as target, label or categories.The classification deductive modeling is the task of approximating the mapping function[10] from input variables to discrete output variables.[5][6] The main goal is to identify which class/category the new data will fall .Each genetic samples are categorised into corresponding genetic pattern so that high pior risk genetic pattern is obtained.

4.3 DECISION TREE ALGORITHM*:*

Decision tree learning employs a divide and conquer strategy by conducting a greedy search to identify the optimal split points[7] within a tree. This process of splitting is then repeated in a top-down, recursive manner until all, or the majority of records have been classified under specific class labels.Whether or not all data points are classified as homogenous[8] sets is largely dependent on the complexity of the decision tree.a decision tree starts with a root node, which does not have any incoming branches. The outgoing branches from the root node then feed into the internal nodes, also known as decision nodes. Based on the available features, both node types conduct evaluations to form homogenous[8] subsets, which are denoted by leaf nodes, or terminal nodes.[7][8] The leaf nodes represent all the possible outcomes within the dataset.Root node indicates the genetic pattern flow with respective genetic samples with relevant dataset.

V.CONCLUSION AND RESULTS

Alzheimer's disease is the most common type of dementia.1 Dementia isn't a single, specific disease instead, it's a term used to describe a combination of symptoms that includes the progressive loss of different cognitive functions, such as memory, orientation, language, and attention.Alzheimer's disease (AD) is a neurodegenerative disease that causes irreversible damage to several brain regions, including the hippocampus causing impairment in cognition, function, and behaviour. Early diagnosis of the disease will reduce the suffering of the patients and their family members.This condition affects around 45 million individuals worldwide.Clustering technique and classification model play an major role to detect genetic pattern along with the genetic samples which let us known about the high pior risk of Alzheimer's disorder.

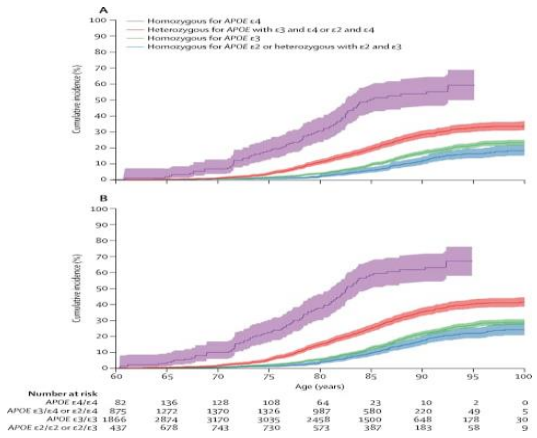

Fig.5.1 APOE E3 and E4 graphical representation

REFERENCES

1. Xianglian Meng,Yue Wu,Wenjie Liu,YingWang,Zhe Xu,Zhuqing Jiao.Research on Voxel-Based Features Detection and Analysis of Alzheimer's Disease Using Random Survey Support Vector MachineMarch 2022Frontiers in Neuroinformatics 16DOI:10.3389/fninf.2022.856295.

2. Arshad Hashmi;King Abdulaziz University,Omar BarukabDementia Classification Using Deep Reinforcement Learning for Early Diagnosis. January 2023Applied Sciences 13(3):1464DOI:10.3390/app13031464.

3. Jhansi KakaJawaharlal Nehru Technological University, Kakinada,K. Satya PrasadJawaharlal Nehru Technological University, KakinadaAlzheimer's Disease Detection using Correlation based Ensemble Feature Selection and Multi Support Vector MachineJuly 2022IJCDS Journal12(1):920DOI:10.12785/ijcds/120102.

4. Neha Rajawat,Bharat Singh Hada,Mayank Meghawat,Soniya LalwaniAdvanced Identification of Alzheimer's Disease from Brain MRI Images Using Convolution Neural NetworkFebruary 2022DOI:10.1007/978-981-16-6332-1_21.

5. Nikhil Mundey ,Shridevi KarandeDeep Learning Applications to Detect Alzheimer's—A ReviewNovember 2022DOI:10.1007/978-981-19-5292-0_33.

6. Batzaya Tuvshinjargal,Heejoung HwangVGG-C Transform Model with Batch Normalization to Predict Alzheimer's Disease through MRI DatasetAugust 2022Electronics

7. Front Comput Neurosci. 2023; 11(16):2601DOI:10.3390/electronics11162601.

8. Fabrizio StasollaUniversità Telematica "GiustinoFortunato"ofBenevento,Maria carla Di Gioia Università Telematica "Giustino Fortunato"Combining reinforcement learning and virtual reality in mild neurocognitive impairment: a new usability assessment on patients and caregiversMay

2023Frontiers in Aging Neuroscience 15:1-7DOI:10.3389/fnagi.2023.1189498.

9. Nair Bini Balakrishnan,P. S. SreejaJisha Jose PanackalAlzheimers Disease Diagnosis using Machine Learning: A ReviewApril 2023https://www.sciencedirect.com/science/article/pii/S1672022919301573.

10. Zhen Zhao, 1 Joon Huang Chuah,corresponding author 1 , * Khin Wee Lai,corresponding author 2 , * Chee-Onn Chow, 1 Munkhjargal Gochoo, 3 Samiappan Dhanalakshmi, 4 Na Wang, 5 Wei Bao,corresponding author 6 , * and Xiang Wu 7Front Comput Neurosci. 2023; 17: 1038636. Published online 2023 Feb 6. doi: 10.3389/fncom.2023.1038636.

11. Early-Stage Alzheimer's Disease Prediction Using Machine Learning ModelsC. Kavitha1*, Vinodhini Mani1, S. R. Srividhya1, Osamah Ibrahim Khalaf2 and Carlos Andrés Tavera Romero3Front. Public Health, 03 March 2022Sec. Digital Public HealthVolume 10 - 2022 | https://doi.org/10.3389/fpubh.2022.853294.