Alzheimer Disease Prediction using Machine Learning Algorithms

J.Neelaveni
Computer Science and Engineering Department,
Sri Ramakrishna Engineering College-641022
Coimbatore, India
neelaveni.1852001@srec.ac.in

M.S.Geetha Devasana
Computer Science and Engineering Department,
Sri Ramakrishna Engineering College-641022
Coimbatore, India
msgeetha@srec.ac.in

***Abstract*---Alzheimer disease is the one amongst neurodegenerative disorders. Though the symptoms are benign initially, they become more severe over time. Alzheimer's disease is a prevalent sort of dementia. This disease is challenging one because there is no treatment for the disease. Diagnosis of the disease is done but that too at the later stage only. Thus if the disease is predicted earlier, the progression or the symptoms of the disease can be slow down. This paper uses machine learning algorithms to predict the Alzheimer disease using psychological parameters like age, number of visit, MMSE and education.**

Keywords- Alzheimer disease, mild cognitive impairment, machine learning algorithms, psychological parameters.

I. INTRODUCTION

Alzheimer disease is caused by both genetic and environmental factors, those affects the brain of a person over time. The genetic changes guarantee a person will develop this disease.

This disease breaks the brain tissue over time. It occurs to people over age 65. However people live with this disease for about 9 years and about 1 among 8 people of age 65 and over have this disease.

MMSE (Mini Mental State Examination) score is the main parameter used for prediction of the disease. This score reduces periodically if the person is affected. Those people having MCI have a serious risk of growing dementia. When the fundamental MCI results in a loss of memory, the situation expects to develop to dementia due to this kind of disease.

There is no treatment to cure Alzheimer's disease. In advanced stages of the disease, complications like dehydration, malnutrition or infection occurs which leads to death. The diagnosis at MCI stage will help the person to focus on healthy approach of life, and good planning to take care of memory loss.

II. LITERATURE SURVEY

Ronghui Ju et.al, suggested method of deep learning along with the brain network and clinical significant information like age, ApoE gene and gender of the subjects for earlier examination of Alzheimer’s [1]. Brain network was arranged, calculating functional connections in the brain region by employing the resting-state functional magnetic resonance imaging (R-fMRI) data. To produce a detailed discovery of the early AD, a deep network like autoencoder is used where functional connections of the networks are constructed and are susceptible to AD and MCI. The dataset is taken from the ADNI database. The classification model consists of the early diagnosis, initially preprocessing of raw R-fMRI is done [1]. Then, the time series data (90×130matrix) is obtained and that indicates blood-oxygen levels in each and every region of brain and changes a long period. Then, a brain network is built and transformed to a 90×90 time series data correlation matrix. The targeted autoencoder model is used which is a three layered model which gives intellectual growth of the nervous system then excerpts brain networks attributes completely [1]. When finite amount of data cases is taken, k-fold cross verification was implemented mainly to avoid the over fitting complication.

K.R.Kruthika et.al, proposed a method called multistage classifier by using machine learning algorithms like Support Vector Machine, Naive Bayes and K-nearest neighbor to classify between different subjects [2]. PSO (particle swarm optimization) which is a technique that best selects the features was enforced to obtain best features. Naturally image retrieving process requires two stages: the first stage involves generating features so that it reproduces the query image and then later step correlates those features with already gathered in database [2]. The PSO algorithm is used to select the finest biomarkers that show AD or MCI. The data is taken from Alzheimer's disease Neuroimaging Initiative (ADNI) database. The MRI scans are preprocessed first after taking from the database. The feature selection includes volumetric and thickness measurements. Then the optimum feature lists were obtained from PSO algorithm [2]. The Gaussian Naïve Bayes, K- Nearest Neighbor, Support vector machine was used to distinguish between the subjects. Here a 2 stage classifier was used where in the initial stage GNB classifier was used to classify the objects between AD, MCI and NC and in later stages SVM and KNN were used to analyze the object based on the performance of the initial one [2]. Control Based Image Retrieval was used for retrieving images from the database.

Ruoxuan Cuia et.al, proposed a model where longitudinal analysis is performed on consecutive MRI and is essential to design and compute the evolution of disease with time for the purpose of more precise diagnosis [3]. The actual process uses those features of morphological anomaly of the brain and the longitudinal difference in MRI and constructed classifier for distinguishing between the distinct groups. The MRI brain images of 6 time points that is for consecutive intervals in a gap of six months are taken as inputs from ADNI database [3]. Then feature learning is done with the 3D Convolutional Neural Network. The CNN is followed by a pooling layer and have many ways for pooling, like collecting mean value otherwise the maximal, or definite sequence of neuron in the section. But for studying the characteristics, the convolutional operation of $2\times2\times2$ is applied so that a linear combination is studied for pooling of neurons [3].

The fully connected layer has neurons that produce output of all neurons in a linear combination, which are taken from preceding layer and then is moved through nonlinearity. Finally for the last fully connected, a softmax layer is particularly used and then tuned finely for back-propagation to predict the class probability [3]. The result of each node varies from 0 to 1, and the total of nodes will always be 1. Finally the classification includes the deep network construction including the 3D CNN training and RNN model training. Then the results of fully connective layers are directly mapped using a softmax function [3]. The initial parameters that were trained by both 3 dimensional CNN and the RNN network are established and then only the uppermost fully connective layer parameters and the softmax layer that was used for prediction are adjusted so that the dimensional and longitude features were united for distinct identification.

Fan Zhang et.al, proposed a multi-modal model where medical images are used for training. It is done by two separate convolutional neural networks. Here an auxiliary diagnosis using a deep learning model is used [4]. The two separate independent CNN for extracting the characteristics from both the MR scans and also the PET scans is used. It is obtained by a sequence of forward procreation convolution and the downer sampling method. The outputs cohesion is calculated by correlation analysis for the two networks. The structure of CNN consists of sampling layer in the down, a convolution row or a fully connective row, pooling layer and finally the output row [4]. Then it computes the correlation using Pearson correlation coefficient method in between the prognosis of MRI scans and PET images. The main idea of the correlation search is for regulating the output of both neuroimaging examinations whether they were persistent or not. The purpose of identifing and classifing is finished by using the output layer called softmax logistic regression method [4]. The benefit of this process is, it merges clinical neuropsychological results with neuroimaging results.

III. IMPLEMENTATION OF SUPPORT VECTOR MACHINE

SVM is directed study model that classifies by separating the objects using a hyperplane. It can be used for both classification and regression. The hyperplanes are drawn with the help of the margins. The main goal is to maximize the distance between the hyperplane and the margin.

The margins are drawn with the help of support vectors that are belonging to the objects. The main advantage of SVM is that it can distinguish linear and non-linear objects. Fig.1 shows the steps in predicting the Alzheimer disease using machine learning algorithms.

```
classifier = svm (formula=age, visit, MMSE, EDUC
                 .,data = train, type = 'C-
                 classification', kernel = 'linear')
```

The required packages for SVM classifier in R are caret and e1071 packages. The formula consists of the fields that are considered for prediction. The basic type c-classification and linear kernel is chosen. They both mostly depend on the data used.

The psychological parameters are given as the input for the classifier. When the classifier is trained and given for testing, it predicts the output with an accuracy of 85%.

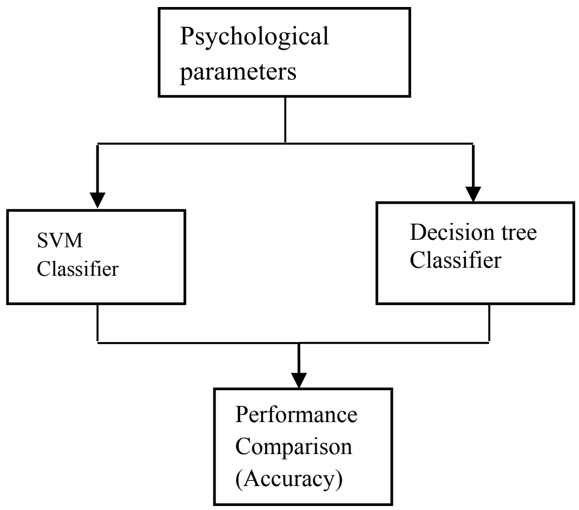

Fig 1 Block Diagram

IV. IMPLEMENTATION OF DECISION TREE

Decision tree is a supervised learning model that uses a set of rules to find a solution. It can solve any type and variety of problems. It can also be used for both classification and regression. A small change in the data also can give a great impact on the output.

For a continuous variable regression trees can be used and for categorical variable classification trees can be used. The decision tree consists of the following nodes:

- Root node: It is the starting point of the tree.
- Internal node: It represents the decision point of the problem that leads to the solution.
- Leaf node: They are the final or last nodes of the entire tree.

The algorithm for decision tree classifier is as follows:

```
model <- rpart (formula = age, visits, MMSE, EDUC~ .,
                data = alzhe,
                method = "class")
```

The formula consists of the fields that are considered for the prediction of the Alzheimer disease. The method class indicates the classification trees. The packages used here are party, rpart, and rpart.plot. The package ctree() can also be used to analyze the decision tree.

V. RESULTS

TABLE I COMPARISON OF PREDICTION ACCURACY OF MACHINE LEARNING ALGORITHMS

ALGORITHM	**ACCURACY**
Support Vector Machine	85%
Decision Tree	83%

The dataset has various parameters but only the significant parameters that greatly help in predicting the disease like MMSE score, age, number of visits, and education of the patients were used. When the machine learning algorithms like Support Vector Machine, Decision Tree were used, they predict the disease with different accuracies. Each algorithm is trained with 70% training dataset and tested with 30% test dataset.

The conduct of the algorithms is compared based on their accuracy. Then the dataset is partitioned according to that ratio and when the algorithms are compared the best one is selected and can be used for next stage of prediction.

VI. CONCLUSION

Machine learning approach to predict the Alzheimer disease using machine learning algorithms is successfully implemented and gives greater prediction accuracy results. The model predicts the disease in the patient and also distinguishes between the cognitive impairment.

The future work can be done by combining both brain MRI scans and the psychological parameters to predict the disease with higher accuracy using machine learning algorithms. When they are combined, the disease could be predicted with a higher accuracy in the earlier stage itself.

REFERENCES

[1] K.R.Kruthika, Rajeswari, H.D.Maheshappa, “Multistage classifier-based approach for Alzheimer’s Disease prediction and retrieval”, Informatics in Medicine Unlocked, 2019.

[2] Ronghui Ju , Chenhui Hu, Pan Zhou , and Quanzheng Li, “Early Diagnosis of Alzheimer’s Disease Based on Resting-State Brain Networks and Deep Learning”, IEEE/ACM transactions on computational biology and bioinformatics, vol. 16, no. 1, January/February 2019.

[3] Ruoxuan Cuia, Manhua Liu “RNN-based longitudinal analysis for diagnosis of Alzheimer’s disease”, Informatics in Medicine Unlocked, 2019.

[4] Fan Zhang , Zhenzhen Li , Boyan Zhang , Haishun Du , Binjie Wang , Xinhong Zhang, “Multi-modal deep learning model for auxiliary diagnosis of Alzheimer’s disease”, NeuroComputing, 2019.

[5] Chenjie Ge , Qixun Qu , Irene Yu-Hua Gu , Asgeir Store Jakola “Multi-stream multi-scale deep convolutional networks for Alzheimer’s disease detection using MR images”, NeuroComputing, 2019.

[6] Tesi, N., van der Lee, S.J., Hulsman, M., Jansen, I.E., Stringa, N., van Schoor, N. et al, “Centenarian controls increase variant effect sizes by an average twofold in an extreme case-extreme control analysis of Alzheimer's disease”, Eur J Hum Genet. 2019;27:244–253

[7] J. Shi, X. Zheng, Y. Li, Q. Zhang, S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer’s disease", IEEE J. Biomed. Health Inform., vol. 22, no. 1, pp. 173-183, Jan. 2018.

[8] M. Liu, J. Zhang, P.-T. Yap, D. Shen, "View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data", Med. Image Anal., 2017 vol. 36, pp. 123-134.

[9] Hansson O, Seibyl J, Stomrud E, Zetterberg H, Trojanowski JQ,Bittner T, “CSF biomarkers of Alzheimer’s disease concordwith amyloid-bPET and predict clinical progression: A study of fullyautomated immunoassays in BioFINDER and ADNI cohorts”. Alzheimers Dement 2018;14:1470–81.

[10] Van der Lee SJ, Teunissen CE, Pool R, Shipley MJ, Teumer A,Chouraki V, “Circulating metabolites and general cognitive abilityand dementia: Evidence from 11 cohort studies”, Alzheimer’s Dement2018;14:707–22

[11] Kauppi Karolina, Dale Anders M, “Combining Polygenic Hazard Score With Volumetric MRI and Cognitive Measures Improves Prediction of Progression from Mild Cognitive Impairment to Alzheimer’s Disease”, Frontiers in Neuroscience, 2018.

[12] Grassi M, Loewenstein DA, Caldirola D, Schruers K, Duara R, Perna G, “A clinically-translatable machine learning algorithm for the prediction of Alzheimer's disease conversion: further evidence of its accuracy via a transfer learning approach”, Int Psychogeriatr, 2018 14:1–9. doi: 10.1017/S1041610218001618.

[13] Nation, D.A., Sweeney, M.D., Montagne, A., Sagare, A.P., D’Orazio, L.M., Pachicano, M. et al, “Blood-brain barrier breakdown is an early biomarker of human cognitive dysfunction”, Nat Med. 2019;25:270–276.