ISSN 2063-5346

ALZHEİMER'S DİSEASE CLASSİFİCATİON USİNG RANDOM FOREST ALGORİTHM WİTH OPTİMAL FEATURE EXTRACTİON

***Narmada Kari[1], Sanjay Kumar Singh[2] , Roshan M. Bodile [3]**

Article History: Received: 10.05.2023 Revised: 29.05.2023 Accepted: 09.06.2023

Abstract

Alzheimer's disease (AD) is the most frequent cause of dementia and accounts for 60% to 80% of all dementia instances. It is a neurodegenerative form of dementia that initially manifests as moderate cognitive impairment (MCI) before progressing to more severe symptoms over time. It disrupts brain cells, causes a decline in memory and cognitive abilities, and makes it difficult to complete even the simplest tasks. In the early stages of Alzheimer's disease, there are medical treatments that can be used to reverse the effects of the disease. Therefore, this paper proposed the random forest classifier to classify longitudinal and cross-sectional MRI data. Furthermore, accuracy and sensitivity are used as quantitative evaluation parameters. The obtained results show that the random forest outperforms in accuracy and sensitivity compared to logistic regression and decision trees.

Keywords: Random forest, Alzheimer, Logistic regression, Decision tree classifier.

[1] Research Scholor, Dept of Electronics & Communication Engineering,Amity University Rajasthan,Jaipur,India.

[1*] narmadakari@gmail.com

[2] Associate Professor, Dept of Electronics & Communication Engineering, Amity University Rajasthan,Jaipur,India.

[2] sksingh.eee@gmail.com

[3] Assistant Professor, Dept of Electronics and Communication Engineering B R Ambedkar NIT, Jalandhar, India.

[3] roshanmbodile110@gmail.com

* Corresponding Author

DOI:10.48047/ecb/2023.12.9.92

1 Introduction

Alzheimer's disease (AD) is the most frequent cause of dementia and accounts for 60% to 80% of all dementia instances [1]. It is a neurodegenerative form of dementia that initially manifests as moderate cognitive impairment (MCI) before progressing to more severe symptoms over time. It disrupts brain cells, causes a decline in memory and cognitive abilities, and makes it difficult to complete even the simplest of tasks [2]. Therefore, Alzheimer's disease is a degenerative neurological brain illness that has multiple aspects. People who have MCI are more prone to develop Alzheimer's disease than other people [3]. People don't notice the effects of Alzheimer's disease until years after the disease has already caused changes in the brain. This is because AD starts at least two decades before the symptoms appear. The Alzheimer's Disease International (ADI) reported more than 50 million people across the globe are suffering from dementia and the number will rise further by the year 2050, this proportion would have increased to 152 million individuals, which indicates that one person develops dementia every three seconds.

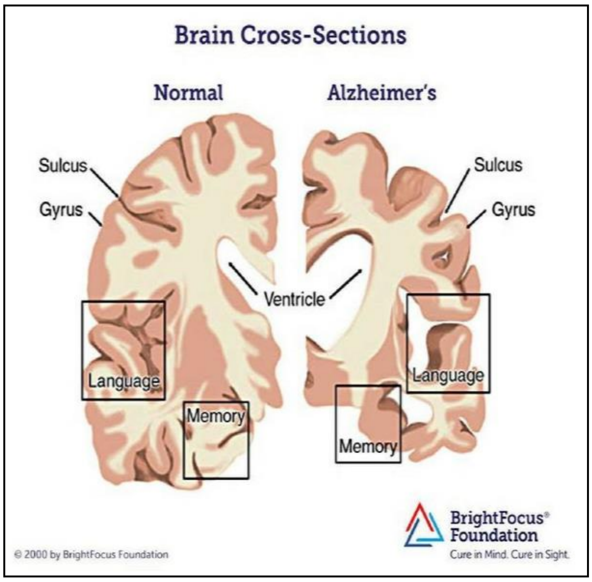

Fig. 1 Comparison of Alzheimer's disease brain with a healthy brain (Source: Bright Focus Foundation) [18].

The symptom of AD is short-term memory loss which makes it difficult for a patient to perform their day-to-day tasks. This problem is more prevalent in senior patients. Although Alzheimer's disease is not normal because of age, the risk factor for developing it grows as one gets older. In 2015, Alzheimer's disease affected around 29.5 million people all over the world. The majority of persons don't show symptoms until after the age of 65, although between 4% and 5% of instances show symptoms far earlier than these age groups. AD is one of the diseases that cost wealthy countries the most money to treat. More than 4 million people in India are afflicted with Alzheimer's disease or another type of dementia. It is projected that by the year 2050, the number of people suffering from AD will increase by a factor of three [4]. Figure 1, shows the comparison of Alzheimer's disease brain with a healthy brain [5]. It is possible to observe that the brain of a person with Alzheimer's disease is not only noticeably smaller than the brain of a healthy person, but it is also badly affected by neurological disorder and dysfunction. In addition, several frequent Alzheimer's disease symptoms are illustrated in Figure 2. Memory loss, changes in behaviour, difficulty with ordinary tasks, and bewilderment in familiar situations are the most frequent sorts of symptoms that people experience when they have dementia.

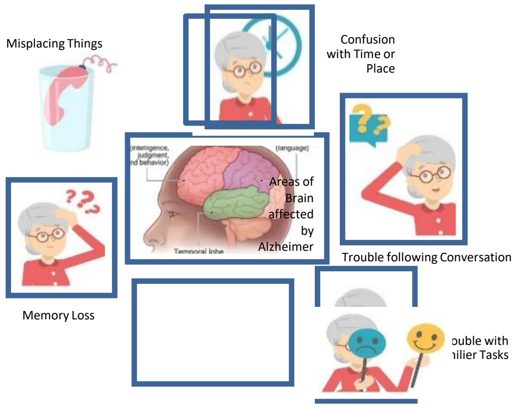
Misplacing Things
Confusion with Time or Place
(intelligence, judgment, nd behavior)
(language)
Areas of Brain affected by Alzheimer
Temporal lobe
Trouble following Conversation
Memory Loss
ouble with nilier Tasks

Fig. 2 Alzheimer's disease symptoms

In the early stages of Alzheimer's disease, there are medical treatments that can be used to reverse the effects of the disease. However, because Alzheimer's disease is characterized by a progression that cannot be reversed once it has begun, early diagnosis of the condition is of the utmost clinical, societal, and economic significance. The advances that have been made in imaging and computational technology have allowed medical science to diagnose the disease in its early stage and begin the therapeutic procedures that are necessary to treat it. But despite this, there is a need for computational approaches and algorithms that are both accurate and fast in order to facilitate the straightforward and timely identification of neurological conditions. Alternating sclerosis causes alterations in the brain's structure as well as its function. It produces morphological and anatomical patterns that are constantly developing through dynamic processes. Examining these alterations on both a qualitative and quantitative level and contrasting them with the typical activities and characteristics of the human brain is a helpful way to identify the new changes brought on by Alzheimer's disease.

In the present scenario, computer-aided diagnostics [6] makes use of sophisticated computer programs and various algorithms for pattern recognition and image processing. The purpose of such computer-aided diagnostics is to locate Features of Interest (FOI) or Regions of Interest (ROI) in the MR images that are to be analyzed. Proficiently developed computer programs yield much better accuracy than human neurologists and can be of great assistance to them in understanding the physiological changes occurring in the brain cells. It is expected that the developed programs will spot the vital features alongside the false-negative rate control. Because of this, a considerable amount of research is currently being conducted all over the world with the goal of classifying and detecting the various phases of neurodegenerative disorders including AD [7].

When it comes to the application of machine learning for the diagnosis of Alzheimer's disease, many models have been utilised rather frequently. For example, Alickovic and Subasi performed a various comparative study to assess the performance of supervised machine learning models to predict Alzheimer's disease [8]. The study was aimed to assess the usefulness of supervised machine learning models in AD predictions. The investigation concentrated mostly on the topics such as artificial neural network, k-nearest neighbours, logistic regression, support vector machine, random forest, and naive bayes. The authors used the ADNI data repository in the carrying out of their research [9]. The random forest classifier has produced the highest performance, while the k-nearest neighbours classifier has delivered the highest performance [10]. Classified the five stages of AD using five machine learning models viz. naive bayes, decision tree, k-nearest neighbour, rule induction, and generalised linear model Along with one deep learning model. The authors of the study carried out their research utilising the ADNI data repository [9]. The generalized linear model classifier, and deep learning (78.32% accuracy) both have produced the best performances so far.

Song et al. [11] proposed a method for the categorization and recognition of AD using the Gaussian Mixer Model (GMM) that relied on the cortical thickness in MR images [12]. Dimensionality reduction and the extraction of necessary features are both accomplished through the usage of the GMM algorithm. After that, a GMM model that operates within the range of the Bayesian framework is utilized for the purpose of AD categorization and detection [13]. Moreover, the authors compared the Bayesian framework with other classic classifiers such as Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM), to prove the accuracy and efficiency of the Bayesian framework. In addition to this, it was mentioned that the number of components that can be included in the distribution of each class cannot exceed 2. Using functional MR images [14], suggested a machine learning practice for the categorization and diagnosis of Alzheimer's disease [15]. The authors processed functional MRI images in initial stage with the help of statistical parametric mapping toolbox in order to obtain individual statistical maps of voxels [16] [17]. After that, active filters were used to pick the voxels that were active.

Random forest is a versatile and user-friendly machine learning method that consistently offers excellent results without the need for hyperparameter customization. Because of its flexibility and ease of implementation, it is also one of the most used algorithms. Therefore, this paper proposed the random forest classifier to classify longitudinal and cross-sectional MRI data. Furthermore, accuracy and sensitivity are used as quantitative evaluation parameters.

Remaining paper is divided into three sections. Section 2 gives the random forest for classification. Section 3 provides results on three different classifiers, and last sections provides conclusion.

2 Proposed Method

Random forest is a versatile and user-friendly machine learning method that consistently delivers excellent results without the need for hyperparameter customization. Because of its flexibility and ease of implementation, it is also one of the most used algorithms [8].

A random forest's hyperparameters are similar to those of a bagging classifier or a decision tree. However, using the random forest classifier class is more straightforward than combining a decision tree with a bagging classifier. The regressor in the random forest algorithm may also be used to handle regression jobs. The trees in a random forest model are grown with added unpredictability. When dividing a node, it doesn't look for the most relevant feature but rather the most significant component from a random selection of characteristics. More variety means a more robust model in most cases.

Consequently, the procedure for splitting a node in a random forest only considers a random subset of the characteristics. In addition to not looking for the optimal thresholds, applying arbitrary thresholds for each feature might produce more unpredictable trees [8].

RF is more stable in the company of outliers and is more resilient to overfitting since it adheres to a strict set of criteria for tree building, tree combination, post-processing, and self-testing. It also observed the same stability for extremely large (higher dimension) parameter space compared to other computer programs capable of learning new information. Considering the significance of variables as an implied RF-based feature selection using a completely at-random-subspace technology, the Gini impurity criterion index measures its quality. As a statistical indicator, the Gini index capacity for predicting outcomes in regression or classification based on the theory of reducing contaminants is non-parametric; it does not require the data to be categorized in a particular distribution.

It is optimal to maximize the increase in the Gini index when deciding how to divide a binary node. In other words, if the Gini coefficient is low, it indicates that a certain predictor trait is crucial in dividing the data into two groups. To this end, the Gini index may be utilized to prioritize features in a classification task.

The operation of Random Forest is split into two stages: the first stage involves building the random forest by mixing N decision trees, and the second stage consists of making predictions for each tree created in the first stage and algorithm shown in Figure 3.

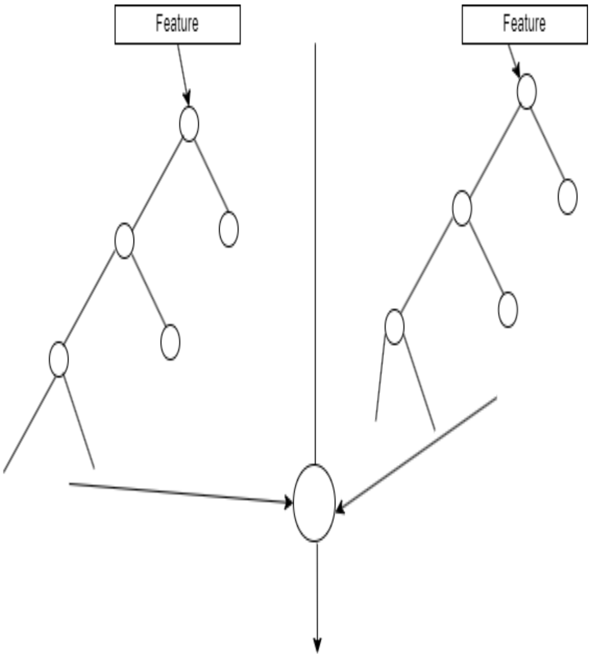

Fig. 3: Random Forest Classifier.

The steps for the Random Forest algorithm are as follows:

- ➢ Choose random data points.
- ➢ Develop the decision trees based on chosen points from data.
- ➢ Select the decision for development
- ➢ Repeat 1 and 2 steps.
- ➢ For the following new values, develop a new decision tree and allocate the new points to the selected category that wins the highest votes.

3 Result and Discussion

3.1 Data Used

Cross-sectional MRI Data

This dataset consists of 416 members ranging in age from younger to older (18 to 96). During a single scanning session, three or four separate T1-weighted MRI images are obtained for each person. All of the individuals are right-handed, and they are a mixed group of males and females. One hundred patients over 60 enrolled in the study were given a clinical diagnosis ranging from highly minor to severe Alzheimer's. In addition, a consistent data set that includes 20 healthy participants who underwent imaging during a second visit within the first three months after their original session is also included in the package [9].

Longitudinal MRI Data

Longitudinal data collection of 150 contributors varying in age from 60 to 96 completes this group. For a total of 373 imaging sessions, each individual had scanning at two or more visits, with each visit being spaced out by at least one year. During a single scanning session, three or four separate T1-weighted MRI images are obtained for each person. All of the individuals are right-handed, and they are a mixed group of males and females. Throughout the research, 72 of the contributors were deemed to be free of dementia. Sixty-four participants who participated in the study were classified as having dementia during their first visits, and this diagnosis was retained for all future scans; among these were 51 persons with minor to moderate Alzheimer's. Another 14 members were classified as having no signs of dementia during their first visit but were later classified as having dementia during a second examination [9]. Also scale of dementia is provided in Table 1.

Table 1: Estimating the CDR (scale of Dementia).

CDR Scale	Value
3	Severe cognitive impairment
2	Moderate cognitive impairment
1	MCI
0.5	Questionable dementia

Figures 4 and 5 show brain volume and density for both demented and nondemented group of people [19][20].

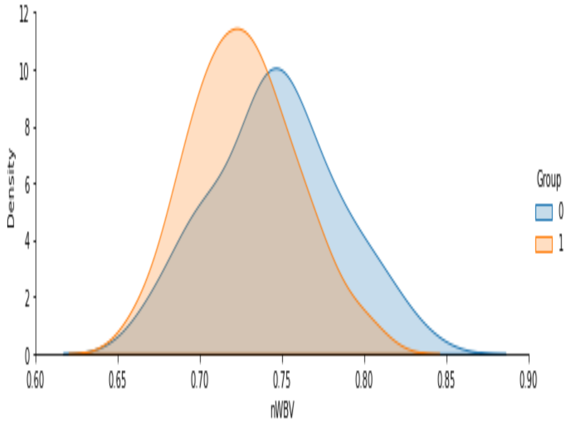

Fig. 4: Demented group Vs nondemented group brain volume.

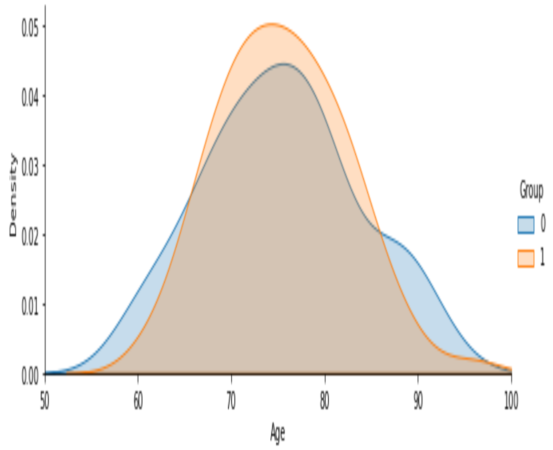

Fig. 5: Comparison of density in Demented group Vs nondemented group brain volume.

3.2 Performance Evaluation

As seen in Table 2, a confusion matrix may represent the quantitative evaluation of a classifier by utilizing valuable parameters. The number of instances for which the classifier made accurate predictions may be found on the diagonal. These values can be separated into true positives, also known as TP, which represents patients who have been accurately identified, and true negatives, also known as TN, which means controls that have been correctly recognized. The number of cases that were incorrectly stratified by the RF classifier can be broken down into two categories: false positives (FP), which are examples of controls that were wrongly classified as patients, and false negatives (FN), which are examples of patients that were incorrectly classified as controls [21].

The percentage of instances adequately identified by a classifier is used to measure accuracy and is provided by

Accuracy=(TP+TN)/(FN+TN+FP+TP)

If the class distribution of the dataset is not evenly distributed, this may not be the best performance statistic.

For instance, if the class is significantly larger, a classifier that assigns all examples to the class could get a high accuracy rating by labeling all instances as belonging. The rate of true positives (TP) is referred to as sensitivity. Definition of sensitivity includes the following:

Sensitivity=TP/(TP+FN)

The proportion of patients who are accurately recognized is what sensitivity assesses.

Table 2: Quantitative analysis for different classifier algorithms.

Sr. No.	Methods	Accuracy		Sensitivity	AUC
		Train	Test		
1	Logistic regression	0.829	0.746	0.703	0.746
2	Decision Tree Classifier	0.775	0.8	0.594	0.797
3	Random Forest	1	0.8	0.757	0.799

From Table 2, it is clear that the Random Forest provides better classification accuracy compared to others. Apart form the sensitivity and accuracy, the RF also shows better area under curve compared to Decision Tree and Logistic regression.

4 Conclusion

Alzheimer's disease (AD) is the most frequent cause of dementia and accounts for 60% to 80% of all dementia instances. It is a neurodegenerative form of dementia that initially manifests as moderate cognitive impairment (MCI) before progressing to more severe symptoms over time. It disrupts brain cells, causes a decline in memory and cognitive abilities, and makes it difficult to complete even the simplest tasks. In the early stages of Alzheimer's disease, there are medical treatments that can be used to reverse the effects of the disease. In this paper, three different classifier is presented for Alzheimer's disease. However, the random forest classifier provides better accuracy and sensitivity than Decision Tree and Logistic regression.

References

1. Kari, N., Singh, S.K., Velliangiri, S. (2022). Various Machine Learning Techniques to Diagnose Alzheimer’s Disease—A Systematic Review. In: Mahajan, V., Chowdhury, A., Padhy, N.P., Lezama, F. (eds) Sustainable Technology and Advanced Computing in Electrical Engineering . Lecture Notes in Electrical Engineering, vol 939. Springer, Singapore. https://doi.org/10.1007/978-981-19-4364-5_40
2. Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D Deep learning on medical images: a review, 2020;1–13.
3. Wen J, et al. Convolutional neural networks for classification of Alzheimer’s disease: overview and reproducible evaluation, Medical Image Analysis. 2020;63:101694
4. Physicians PC. 2020 Alzheimer’s disease facts and figures. Alzheimer’s Dement. 2020;16(3):391–460.
5. Jo T, Nho K, Saykin AJ. Deep learning in Alzheimer’s disease: diagnostic classification and prognostic prediction using neuroimaging data, Frontiers in Aging Neuroscience. 2019;11.
6. Kathleen Taylor. Dementia: A Very Short Introduction. Oxford University Press, 2020.
7. Padilla P, Lpez M, Grriz JM, Ramirez J, Salas-Gonzalez D, Alvarez I. NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer’s disease. IEEE Transactions on medical imaging. 2011 Sep 12;31(2):207-16.
8. Barik, S., Mohanty, S., Rout, D., Mohanty, S., Patra, A. K., and Mishra, A. K. (2020). Heart disease prediction using machine learning techniques. In

Advances in Electrical Control and Signal Systems, pages 879–888.

9. Alickovic, E. and Subasi, A. (2020). Automatic detection of alzheimer disease based on histogram and random forest. In Badnjevic, A., Škrbic, R., and Gurbeta Pokvi ´ c, L., editors, CMBEBIH. ´ Springer International Publishing
10. ADNI. Alzheimer’s disease neuroimaging initiative. (2017)
11. Shahbaz, M., Ali, S., Guergachi, A., Niazi, A., and Umer, A. (2019). Classification of alzheimer’s disease using machine learning techniques. In DATA, pages 296–303.
12. Song S, Lu H, Pan Z. Automated diagnosis of Alzheimer’s disease using Gaussian mixture model based on cortical thickness. In2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI) 2012 Oct 18 (pp. 880-883). IEEE.
13. Reynolds D. Gaussian mixture models. Encyclopedia of biometrics. 2015:827-32.
14. Bui DT, Hoang ND. A Bayesian framework based on a Gaussian mixture model and radial-basis-function Fisher discriminant analysis (BayGmmKda V1. 1) for spatial prediction of floods. Geoscientific Model Development. 2017;10(9):3391.
15. Armaanzas R, Iglesias M, Morales DA, Alonso-Nanclares L. Voxelbased diagnosis of Alzheimer’s disease using classifier ensembles. IEEE journal of biomedical and health informatics. 2016 Mar 4;21(3):778-84.
16. Zhu X, Sobhani F, Xu C, Pan L, Ghasebeh MA, Kamel IR. Quantitative volumetric functional MR imaging: an imaging biomarker of early treatment response in hypo-vascular liver metastasis patients after yttrium-90 transarterial radioembolization. Abdominal Radiology. 2016 Aug 1;41(8):1495-504.
17. Kurth F, Gaser C, Luders E. A 12-step user guide for analyzing voxelwise gray matter asymmetries in statistical parametric mapping (SPM). Nature protocols. 2015 Feb;10(2):293.
18. Altinkaya E, Polat K, Barakli B. Detection of Alzheimer ’ s disease and dementia states based on deep learning from MRI images: a comprehensive review. 2020;39–53.
19. BrightFocus Foundation, Clarksburg, Maryland, U.S. https://www.brightfocus.org/alzheimers?fbclid=IwAR0aqQjCkqMlVXBJq-Bu3E_Q0Dvn5Ybe9ibJvR6zearUGMvCPpHsLA8ujXA.
20. Yang Y, Li X, Wang P, Xia Y, Ye Q. Multi-Source transfer learning via ensemble approach for initial diagnosis of Alzheimer’s disease, IEEE J Transl Eng Heal Med. 2020;1–10.
21. Mendez M F 2012 Early-onset Alzheimer’s disease: non amnestic subtypes and type 2 AD Archives of Medical Research 43(8) pp 677–85
22. Asri, H., Mousannif, H., Al Moatassime, H., and Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science, 83:1064–1069.