

# Classification and Investigation of Alzheimer Disease Using Machine Learning Algorithms

Shweta Madiwalar<sup>1</sup>, Sujata Patil<sup>2</sup>, Shashidhar H<sup>3</sup>, and Parameshchhari B. D<sup>4</sup>

<sup>1</sup>Department of Electronics and Communications, KLE Dr. MSSCET, Belagavi, India

<sup>2</sup>Department of Electronics and Communications, KLE Dr. MSSCET, Belagavi, India

<sup>3</sup>Department of KLE Dr. MSSCET, Belagavi, India

<sup>4</sup>Department of Electronics and Communications, GSSSIETW, Mysuru, India

## ABSTRACT

Dementia is a globally identified problem. The occurrence of dementia increases abruptly with growing age. It is an irreversible brain disease which causes degeneration in the cognitive ability of a person affecting his thinking, memory and judgment. Throughout the world around 50 million people have dementia and around 10 million new cases are diagnosed every year. Hence addressing this issue has become need of the hour and early diagnosis of dementia is essential for the progress of more prevailing treatments. Early diagnosis of this disease is done using cognitive tests to determine the mental ability of a person. Some of the cognitive tests include CDR, MMSE, and Adden Brooke's cognitive examination. In present research work using machine learning techniques we have tried to detect the dementia in early stage. The data composed for investigation consists of the gender, age education, MMSE, CDR, ASF, Handedness, number of visits of the patient to the hospital who are clustered as demented or non-demented. We have used different machine learning algorithms like Random forest classifier, (SVM), Decision Tree Classifier, Extra Tree Classifier, Neighbors Classifier and Logistic Regression to analyze the data. The comparison study of each algorithm is done. The algorithm with highest accuracy will be used to further data analysis. In our proposed work we have used extra tree classifier is used for more examination of the facts.

**KEY WORDS:** DEMENTIA. ALZHEIMER'S DISEASE, DIAGNOSIS MACHINE LEARNING, CONFUSION MATRIX.

## INTRODUCTION

The 60%-70% case of dementia causes Alzheimer's. Continuous deterioration of memory and other significant mental functions leads to Alzheimer's disease (AD). The person with Alzheimer's sickness might get slight signs of misperception, partial relations and memorizing the

things and struggle in understanding visual pictures. The risk of getting Alzheimer diseases increase as the age increases (Memon M. H. 2019). As the brain functions are irreversible but the illness is progressive. The potential drug therapies need to be effective in early process. The demand of the timely prediction of diagnosis for testing therapies might delay the progression or limit the early diagnosis done through learning assessment steps. (Kishore. C. 2020). The other steps comprise psychiatric past, cerebral history and the behavioral fluctuations in the patient.

## MATERIAL AND METHODS

**Objectives:** The main objective of this research work is to build the methods and algorithms to classify the types of neurodegenerative diseases. The main research

## ARTICLE INFORMATION

Received 19th Oct 2020 Accepted after revision 30th Dec 2020  
Print ISSN: 0974-6455 Online ISSN: 2321-4007 CODEN: BBRBCA

Thomson Reuters ISI Web of Science Clarivate Analytics USA and Crossref Indexed Journal



NAAS Journal Score 2020 (4.31)  
A Society of Science and Nature Publication,  
Bhopal India 2020. All rights reserved.  
Online Contents Available at: <http://www.bbrc.in/>  
Doi: <http://dx.doi.org/10.21786/bbrc/13.13/3>

focus is on inspecting and refining features extraction of information of neurogeneration. These features are then given as the input to model to classify the dementia types .we have implemented six different types of machine learning algorithms to check the accuracy of the each model

**Prevailing Technique:** In the prevailing techniques (Ahmed, M. R. 2019) the SVM algorithms is used to classify the demented and non-demented subjects with respect to the data set. This algorithm used for bunch processing and it will isolate (Kishore, C. 2020) the individual who are suffering with Alzheimer's Disease, the detection process involved using this method is very slow . Time proficiency is the primary burden. The initial investment is too high to get the data set for Alzheimer Disease (AD).

**J):Machine Learning Approach:** The dataset set created here is taken from The OASIS dataset contains two categories of data That is cross sectional MRI data and longitudinal data. The data set created makes use of 150 subjects between the age group of 60 to 96. Scanning of all the subjects were carried out once and all the subjects were right handed .During the study 72 subjects were clustered as non-demented and 64 subjects were clustered as demented at the preliminary visits but the subsequently they all are characterized as demented at the later visit .The following parameters are used in the evaluation process.

Mini-Mental State Examination (MMSE) or Folstein test is used to measure the cognitive impairment. The test consists of 30 point questionnaire which are widely used in clinical research .Normal cognition is indicated if the score is 24 points (out of 30 points) in the screening test. If the score is (<27 points) then scan indicate severe moderate. Then the score is in the range of (10-18 points)it is mild. if the score is in the range of (18-23 points) then the scan indicates cognitive impairment ( Escudero J. 2013).

Clinical Dementia Rating (CDR): The six areas of the cognitive and functional routines applicable to disease which are related to dementias are characterized by using CDR which is five point scales (Marcus, D.S. 2007). Estimated Total intracranial volume- ETIV gives the information about the negated spaces present in the brain. The exact measurement procedure is given by bunker (Batmanghelich. 2013).

**Feature Selection Methods:** The Feature selection techniques in this research work used are Univariate Selection, Feature Importance and Correlation Matrix with Heatmap. Univariate selection method is used to extract the features which have the strongest correlation with output variable. The statistical test help us to select the required features. Feature importance property of the model helps us to get the importance of each feature in the dataset for each feature the feature importance give the score for the data (Prabu, S. 2019). Features will be more important if the score is high. To extract the top

features we use extra tree classifiers for the dataset. How best the features are related to target variable can be found using correlation matrix .The correlation can be positive or negative .To identify which features are most related to the target variables Heatmap are used. Using Seaborn library we can plot the Heatmap of correlated features.

**Proposed Algorithms:** Initially the work carried out with the existing Machine Learning Techniques with the available data set. Random Forest is a classification algorithm comprising of many decisions trees. It uses feature randomness and catching to develop the individual tree and then create an uncorrelated forest of tree whose prediction is more accurate as compared to individual tree (Prabu.S. 2019). Support Vector Machine is linear model for regression and classification problems. This algorithms focuses on finding the hyperplane in an N-dimensional space with N number of features which will clearly classifies the data points. The hyperplane is just line when the number of input features are two the hyperplane becomes the two dimensional plane when the input features are three. When the features exceed the number three it becomes difficult to imagine the hyper plane (Kumari. C.U.2020).

Decision Tree is a decision support tool is the one way to display an algorithm which contains conditional control statements. This algorithm uses the tree like model of decisions and likely consequences, including chance event outcomes and utility. K-Neighbor algorithm is a method of data cataloging which estimates depending on what group of data points are nearest to it and are in likely a data point is to be a member of one group or the other. Logistic regression is a classification algorithm when the output variable required is of categorical in nature using logistic sigmoid function . logistic regression transforms its output to return the probability value which can be mapped to two or more discrete class.

Extra Tree Classifier, Extremely Randomized Trees Classifier (Extra Trees Classifier) is a type of joint learning method which sums the grades of several de-correlated decision trees collected in a "forest" to output its classification result it is similar to random forest classifier. The only difference is the method of developing the decision trees in forest from the original training sample each decision tree in the extra trees forest is built.

## RESULTS AND DISCUSSION

In this section we are mainly exploring the relationship between dementia of the patient and features of MRI test (G. Uysal 2019). The main reason for conducting this experimental investigation is to find the relationship of the state of data through graph so that we can predict the correlations between the data extraction. To select the suitable technique understanding nature of data is important through the correlation becomes important .The minimum, maximum and average values of each feature of graph implementation are listed in (Table 1).

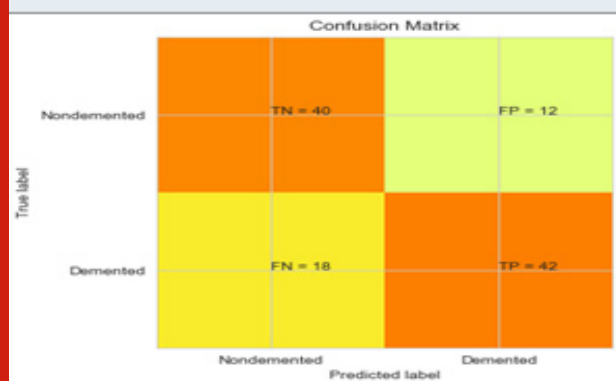
Table 1. Min, Max and Avg values of features

Features	Min	Max	Mean
Education	5	20	12.6
SES	1	4	1.34
MMSE	15	25	26
CDR	0	1	0.39
ETIV	1125	1170	1390
NWBV	0.52	0.79	0.66
ADF	0.73	1.363	1.3

In processing the data we would identify the 8 rows with mislaid SES (Socioeconomic Status). Column with this missing Socioeconomic Status issue is taken care by dropping the missing values or by replacing the missing values by the equivalent values by the process of imputation. The imputation will help us to improve the performance of the model. Imputation is completed by the median.

**Performance Measures:** In most of the neurodegenerative diseases it is significant to take the high positive rate so that all the subjects are detected as Alzheimer's as soon as possible but in the mean while we need to make false positive rate is small .Area under the curve is the best technique for performance measurement .During the process of calculating the accuracy of the models the confusion matrices are obtained as shown in (Figure 1(a-e)). correlation matrix gives the data of certain feature or data. (Escudero. J. I.2013).

Figure 1(a): Logistic regression



The maximum correlated features are considered then the appropriate algorithm is applied. For characterizing the data points hyperplanes are used .dataset arguments might lie on any sides of the hyper plane which can be further mapped as several classes. These hyperplanes mostly depend on the quantity of features which are existing in dataset. The range of hyper parameters are decided by the correlational values of the selected features. Then the data set is classified as training and test data set and feeded to the intended model .The technique is recurring for all the designated algorithms

and accuracy is practical calculated as shown in (Table 2).

Figure 1(b): K neighbor classification SVM Classifier

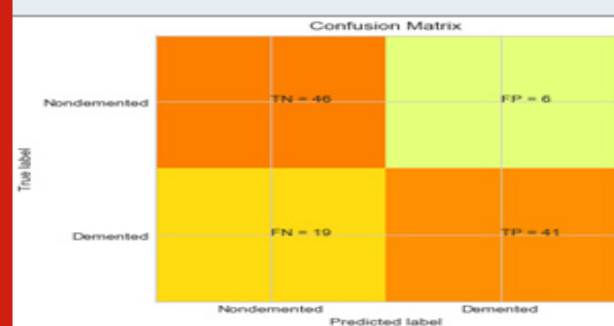


Figure 1(c): SVM

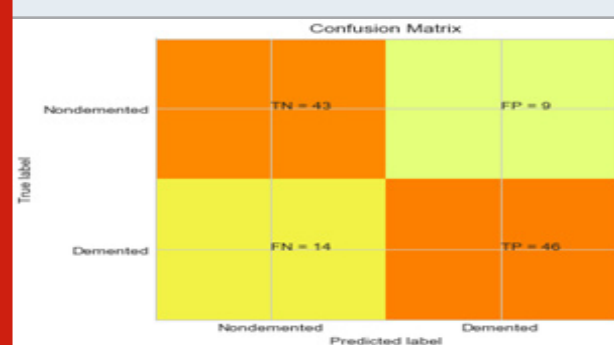


Figure 1(d): Decision Tree

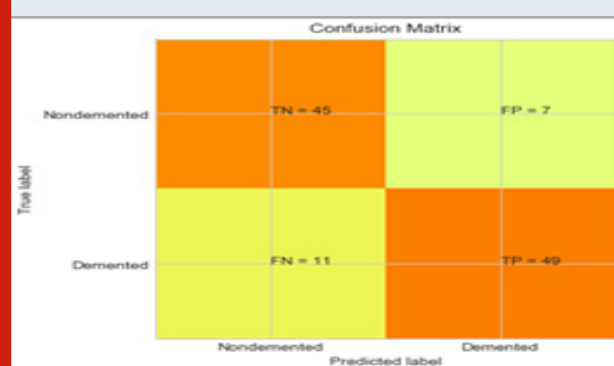
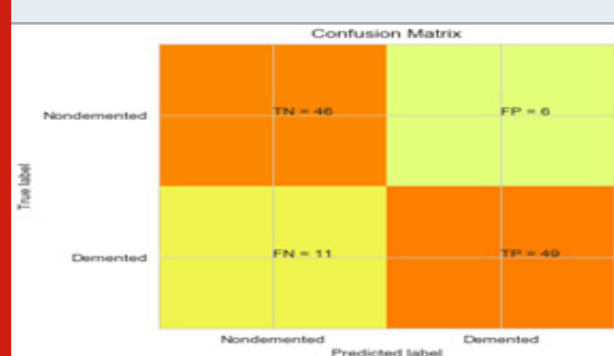


Figure 1 (e): Extra Tree Classifier



**Performance Evaluation Metrics:** For the evaluations of the models the performance evaluations metrics such as accuracy, precision, recall are described mathematically by equation (1), (2), (3) and (4) respectively. Whereas TP (true positive), TN (true negative), FP (false positive), FN (false negative).

$$ACCURACY = \frac{(TP+FP)}{(TP+TN+FP+FN)} \quad (1)$$

$$RECALL = \frac{TP}{(TP+FN)} \quad (2)$$

$$PRECISION = \frac{TP}{(TP+FP)} \quad (3)$$

$$F \text{ Score} = \frac{2*(RECALL*PRECISION)}{(RECALL+PRECISION)} \quad (4)$$

The algorithms used in research work help in faster detection of Alzheimer's. These predictions can benefit the patients for taking the early stage treatment. Complete curing of the disease is not found yet but the treatment have been given in the early stage so that the person with this disease could be able to fight against the disease in the early stage.

Table 2. Performance statistics (Precision, Recall, F1 score, Efficiency) comparison table for algorithms (PR-previous work, CW-current work)

Algorithms		Precision	Recall	F1 score	PR	CW
Logistic regression	0	0.69	0.77	0.73	0.90	0.95
	1	0.78	0.70	0.74		
	Avg/Total	0.74	0.73	0.73		
K-neighbour Classifier	0	0.61	0.81	0.69	0.644	0.669
	1	0.77	0.55	0.64		
	Avg/Total	0.69	0.67	0.67		
Support vector machine	0	0.80	0.88	0.79	0.58	0.77
	1	0.87	0.68	0.77		
	Avg/Total	0.80	0.78	0.78		
Decision tree classifier	0	0.75	0.83	0.79	0.83	0.90
	1	0.84	0.77	0.80		
	Avg/Total	0.80	0.79	0.79		
Random forest classifier	0	0.80	0.87	0.83	0.90	0.92
	1	0.88	0.82	0.84		
	Avg/Total	0.84	0.84	0.84		
Extra tree classifier	0	0.81	0.88	0.84	0.90	0.9314
	1	0.89	0.82	0.85		
	Avg/Total	0.85	0.85	0.85		

With the support of algorithms we can find out the stage in which patient is suffering so that early treatment can be initiated. The machine learning algorithms are compared with previous work which was (Kishore, C.2020) carried out. We found that SVM algorithm

produced better results for classification and extra tree classifier produced the better results are compared to the previous work as shown in the Table 2. In upcoming years if the treatment for Alzheimer's infection is detected then these algorithms benefit us to detect the

severity stage of the patients .depending on the severity levels we will be able to decide whether the treatment undertaken will be helpful to the patients or not .

## CONCLUSION

In our research work more concentration is given on early detection of dementia instead of diagnosis after the dementia has occurred in adults. As per the previous study lot of research is undergoing by using various techniques for the detection of dementia .There are many advantages of using Machine learning algorithms as they reduce human errors and provide accurate and efficient results. The time taken to solve a problem is reduced with very less or no human intervention. The algorithms we applied we found that Extra Tree classifier gave the most prominent, reliable and accurate results, with an accuracy of 93.14%. In Future since the database will have large amount of information the processing of the data and analysis of the data becomes the time consumption process it may affect the accuracy of the algorithm.

Thus only the significant features or dimension involved can be reduced using dimensionality reduction algorithm. This algorithm selects only the features which are prominent role in analysis and other features are just eliminated .This techniques help in achieving better performance with high accuracy. In future to access the mental abilities the patients are subjected to cognitive tests which provide the information about MMSE,CDR using Addenbrooke's Cognitive Examination form by collaborating with patient's family and relatives the tests are carried out.

## REFERENCE

- Ahmed, M. R. Y. Zhang, Z. Feng, B. Lo, O. T. Inan and H. Liao (2019). Neuroimaging and Machine Learning for Dementia Diagnosis: Recent Advancements and Future Prospects. *IEEE Reviews in Biomedical Engineering*. vol. 12. pp. 19-33. doi: 10.1109/RBME.2018.2886237.
- Ardekani B.A, Bachman. H (2018). Figarsky Callosum shape changes in early Alzheimer's disease: an MRI study using the OASIS brain database. *Brain Struct. Funct.* 219pp. 343-352.
- Asl, E. Hosseini, G. Gimelfarb El-Baz A. (2016). Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network. *arXiv preprint arXiv :1607.00556* . 2016.
- Batmanghelich, K.N. Ye, D.H.,PohlK.M., B. Tasker, C. Davatzikos (2011). Disease classification and prediction via semi-supervised dimensionality reduction. *Biomedical Imaging: From Nano to Macro. IEEE International Symposium on. IEEE* . 1086-1090.
- Bhat, P. Paul, M. (2016) Early diagnosis of Alzheimer's disease: a multi-class deep learning framework with modified k-sparse auto encoder classification. *IEEE International conference on Image and vision computing New Zealand (IVCNZ)*.
- Chang ,C.C. and Lin, C.J. (2011). LIBSVM library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*. vol. 2. pp. 27.
- Escudero, J. I. feachor, E.J Zajicek, ,P. Green, J. Shearer and S. Pearson, (2013). Alzheimer's Disease Neuroimaging Initiative Machine Learning-Based Method for Personalized and Cost-Effective Detection of Alzheimer's Disease. *IEEE Transactions on Biomedical Engineering*. vol.60. No. 1. pp. 164-168. Jan. doi: 10.1109/TBME.2012.2212278.
- G. Uysal and M. Ozturk, (2019)"Using Machine Learning Methods for Detecting Alzheimer's Disease through Hippocampal Volume Analysis," 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 2019, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895135.
- Khan A. and Usman M., (2015). Early diagnosis of Alzheimer's disease using machine learning techniques: A review paper. 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K). Lisbon. pp. 380-387. <https://doi.org/10.5220/0005615203800387>
- Kishore, C. Kumari Usha,(2020) Detection and analysis of Alzheimer's disease using various machine learning algorithms *Materials Today: Proceedings* <https://doi.org/10.1016/j.matpr.2020.07.645>
- Kostas. Pang, E.W and Rudzicz( F 2019). Machine learning for MEG during speech tasks. *Nature. Scientific Reports* 9. 1609 <https://doi.org/10.1038/s41598-019-38612-9>
- Kumari, C.U. Panigrahy A.K, Vignesh. N.A. (2020) Sleep Bruxism Disorder Detection and Feature Extraction Using Discrete Wavelet Transform, Springer, Cham, pp. 833-840.
- Kumari, C.U., Kora, P. K. , Swaraja, K. T., Padma, A.K. Panigrahy, N.A. Vignesh,( 2020) in: Feature Extraction and Detection of Obstructive Sleep Apnea from Raw EEG Signal. Springer. Singapore. pp. 425-433.
- Kumari,C.U. Prasad,S.J. G. Mounika, (2019). Leaf Disease Detection: Feature Extraction with K-means clustering and Classification with ANN. *IEEE*. pp. 1095-1098.
- Marcus, D.S. Wang, TH, Parker, J, Csernansky, JG, Morris, JC, Buckner, RL. (2007)Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience* ,1498-1507.
- Memon M. H. J. Li, A. U. Haq and M. Hunain Memon,(2019) "Early Stage Alzheimer's Disease Diagnosis Method," 2019 16th International Computer Conference on Wavelet Active Media Technology and

Information Processing, Chengdu, China, 2019, pp. 222-225, doi: 10.1109/ICCWAMTIP47768.2019.9067689.

Prabu, S., M. Lakshmanan, and V. Noor Mohammed. "A multimodal authentication for biometric recognition system using intelligent hybrid fusion techniques." *Journal of medical systems* 43, no. 8 (2019): 249.

Parameshachari, B. D., H. T. Panduranga, and Silvia liberata Ullo. "Analysis and Computation of Encryption

Technique to Enhance Security of Medical Images." In *IOP Conference Series: Materials Science and Engineering*, vol. 925, no. 1, p. 012028. IOP Publishing, 2020.

Prabu, S., V. Balamurugan, and K. Vengatesan. "Design of cognitive image filters for suppression of noise level in medical images." *Measurement* 141 (2019): 296-301.