Towards The Early Detection of Alzheimer Disease Using Machine Learning Approach

MdShariarKabir[1], MossaddakHossain[2], Naznin Sultana[3]
[1, 2] Student, Daffodil International University, Dhaka, Bangladesh.
[3]Assistant Professor, Daffodil International University, Dhaka, Bangladesh.
[1] shariar15-2414@diu.edu.bd, [2] mossaddak15-2413@diu.edu.bd,
[3] naznin.cse@diu.edu.bd

Abstract

Alzheimer's is a part of dementia. The disease is a neurological condition that affects the patient's brain. It compresses the brain causing brain cells to become damaged. The patient loses memory and cannot remember anything. Although this disease can happen to anyone. But the disease is more prevalent among the elderly. It cannot be completely cured. According to the WHO, a re-port was published in Bangladesh in 2020 that a total of 14993 people or 2.09% died due to Alzheimer & Dementia. The death rate from those diseases are 13.89 people per 100,000 [1]. Our main focus in the paper is to use machine learning algorithm for detecting Alzheimer disease in its primitive stage. After the experiment we got the accuracy of our proposed model using K-Nearest Neighbor algorithm is 96% compared to other existing techniques.

Index terms Alzheimer, Dementia, Machine Learning, Prediction, Accuracy.

DOINumber:10.14704/nq.2022.20.8.NQ44612 **NeuroQuantology2022;20(8):5858-5865**

INTRODUCTION

Our brain controls almost everything we think, feel, say and do. It also preserves our memories. There are certain diseases that keep our brain from working normally. When someone has this kind of disease, then they suffer many kinds of problem like they can't remembering, thinking and speaking properly. They may say or act like strange. It may seem strange to others, and they may find it difficult to do things on a daily basis. Alzheimer's is one of the very dangerous diseases that cannot be totally cured. However, not everyone gets Alzheimer as they get older. It can be caused by many diseases. The effects of this disease can be different for each of the different parts of the brain, that is why it can affect the patients in different ways. Doctors and scientists are working tirelessly to learn more about Alzheimer. The risk of Alzheimer is higher in people over the age of 65, but it can also affect young people. Some people had a stroke, have Diabetes, High blood pressure, High cholesterol, Depression etc. They have more chance to have Alzheimer than others.

In this study, we have applied some machine learning algorithms on a dataset collected from Kaggle and tried to predict whether a person has this disease inside or not at an early stage. We got best accuracy

*e*ISSN1303-5150

by using K-Nearest Neighbor (KNN) algorithm and it was 96%. We Also tried with some other algorithms such as Decision Tree & SVM & Random Forest Classifications but got far less accuracy than KNN.

RELATED WORKS

Authors [2] have shown a comparative picture between four algorithms. These are SVM, LR, DT, and RF. Their system is able to predict 81% correct predictions. They have taken 70% of total data for training and 30% for the testing. In their dataset male-female ratio was 60% and 40%. And most of the patient's total education year was 12.5 years.

Authors [3] used longitudinal MRI data as a primary data source. This MRI data has been taken from 150 patients whose ages were between 60 to 96. 80% of the total data they selected for the testing and 20% of data has taken for training. They applied four algorithms and showed a comparative picture. These algorithms are the DT classifier with 80% accuracy, the RF classifier with 86.92% accuracy, SVM with 81.67% accuracy XGBoost with 85.92% accuracy, and Voting Classifier with 85.12% accuracy.

Authors [4] They analyzed three datasets using three classifier models. They got the highest accuracy on the DTI+GT dataset with the ANFIS classifier. The accuracy was 85.25% and 77.33% was the correct prediction rate.

Authors [5] applied four algorithms. These are SVM, KNN, LR, and GP classifiers. SVM performed well than the rest of the two algorithms. The accuracy was 95%. In their dataset 44.095% were male and 55.91% were female.

Authors [6] applied two models. SVM model performed better than the DT model. The accuracy was 85% for the SVM model and 83% for the DT model. They have chosen 70% of the total data for training and 30% for testing purposes.

Authors [7] got 96% accuracy using CNN model with neuroimaging data. They also applied CNN model and got 90% accuracy.

Authors [8] used 1,618 ADMI participants and their ages were between 55 to 91. They showed that MCI patients progress to AD at a rate of 10% to 15% per year.

Structural images and Functional images are used in Authors [9] dataset to perform a better result. The structural image provides information on the structure (size, shape, and volume) and on the other hand functional image analyzes the brain metabolism, chemical activities, cerebral blood flow, and blood oxygen level measurements. The applied four algorithms, LRC, SL, SVM, and ANN and achieved 92.31% accuracy.

In this work Authors [10] have shown a comparison picture between four algorithms got higher accuracy with SP. the accuracy was 80.50%. Respectively rest of the algorithms are SS (accuracy 75.16), PR (accuracy 77.68) and final one is AC (accuracy 78.05%).

In this study Authors [11] used 503 MRI images from the 808 patients. They performed 81% of total data for training and 9 % for internal cross-validation.

In this work, Authors [12] use Oasis-longitudinal MRI data. The MRI dataset was brain of the people of young, middle and older age who got Alzheimer. They've shown a picture nowadays of Alzheimer's affecting middle age also. They tried to find drug for this disease. For missing data, they used mean and mode method. The feature extracted by using EM algorithm & classified with Gaussian process algorithm.

METHODOLOGY

We have adapted for early detection of Alzheimer disease. Fig.1 shows our proposed methodology. Though there are numerous machine learning algorithms

available that can be applied for classification purpose but different algorithm shows different performance based on the size and type of dataset. Again, accuracy may not always be a good parameter to detect the best model. It depends on a number of other factors as well. In this research, we have used classic machine learning algorithms such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT) & Random Forest (RF) to evaluate the performance of the algorithm and found KNN as the most suited model for our research work.

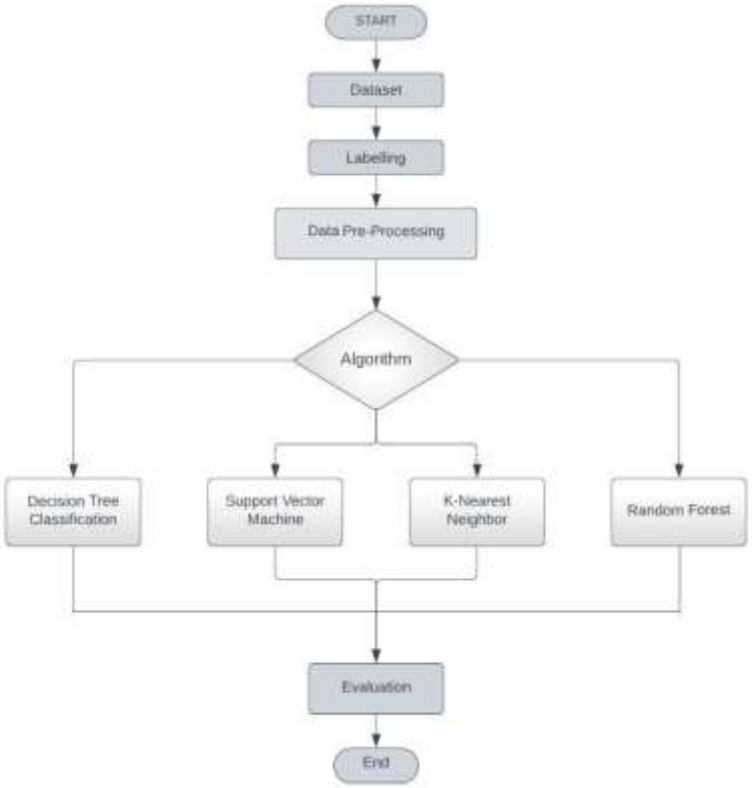

Figure 1: Block diagram of proposed methodology

Our work is divided into the following steps:

A. Dataset Preprocessing:

We encoded this dataset with the ordinal Encoder pre-processing technique. Again, we applied MinMaxScaler normalization technique on this dataset. We encoded total two attributes, these are Group & M/F. There were some null values in SES & MMSE column, we filled those null values.

B. Splitting For Training & Testing:

We split the dataset in two parts. For testing we have chosen 30% of total data & rest of the data for training.

DATASET DESCRIPTION

In our research, we have collected data from Kaggle [13]. It is a text data. The attributes of our dataset are Gender, Age, EDUC, SES, MMSE, CDR, eTIV, nWBV, ASF.

The full meaning of the attributes of the dataset are EDUC = Years of education, SES = Socioeconomic Status with the range of 1-5, MMSE = Mini Mental State Examination, CDR = Clinical Dementia Rating, eTIV =

*e*ISSN1303-5150

Estimated total intracranial volume, nWBV = Normalize Whole Brain Volume, ASF = Atlas Scaling Factor.

Our target attribute is Group. The group's attribute has two types of value one is affected and another affected. In our dataset 39% are affected, 10% are other. 57% of the patients are female and 43% are male. Fig.2 shows the sample of our dataset.

	Group	M/F	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
0	Nondemented	M	87	14	2.0	27.0	0.0	1987	0.696	0.883
1	Nondemented	M	88	14	2.0	30.0	0.0	2004	0.681	0.876
2	Demented	M	75	12	NaN	23.0	0.5	1678	0.736	1.046
3	Demented	M	76	12	NaN	28.0	0.5	1738	0.713	1.010
4	Demented	M	80	12	NaN	22.0	0.5	1698	0.701	1.034
5	Nondemented	F	88	18	3.0	28.0	0.0	1215	0.710	1.444
6	Nondemented	F	90	18	3.0	27.0	0.0	1200	0.718	1.462
7	Nondemented	M	80	12	4.0	28.0	0.0	1689	0.712	1.039
8	Nondemented	M	83	12	4.0	29.0	0.5	1701	0.711	1.032
9	Nondemented	M	85	12	4.0	30.0	0.0	1699	0.705	1.033

Figure 2: Dataset Sample

ALGORITHM DESCRIPTION

In this research, we have chosen 4 classifiers (K-Nearest Neighbor, Decision Tree, Random Forest & Support Vector Machine) since from literature review these classifiers perform better among other classic machine learning algorithms. And from our experimental analysis we got best accuracy from K-Nearest Neighbor classifier than the other 3 classifiers we have used. So, our suggested algorithm is K-Nearest Neighbor classification.

Our suggested algorithm is K-Nearest Neighbor Classification. It is a very simplest algorithm of machine learning & based on Supervised Learning. It is used for Regression and Classification. It can calculate distance between training points and test data. Then select the closest data from test data and predict.

We have also tried Decision Tree Classification. This algorithm is based on Super-vised Learning. It can solve regression and classification problems. In this algorithm, it's start with the roots of the tree to make predictions. Then it compares the value of the root attribute with the properties of the record. Then move on to the next node ac-cording to the branch related to that value based on the comparison.

Another algorithm is random forest classification. Random forest algorithm basically a group of Decision Trees. It operates as an ensemble. In this algorithm, at first it selects data point from training data set. Then build some Decision Tree model in the data point. Then we choose how many decision trees we want to make. Then it finds the predictions of all Decision Trees model. Then predict with the best accuracy from all Decision Trees model.

The last one is Support Vector Machine (SVM). It is a supervised learning model. It can work in high dimensional spaces. It can work with various kind of methods like classification, regression & detection.

RESULT AND DISCUSSION

In this paper, we worked with four different classifiers for predicting the Alzheimer disease at its early stage. In those four algorithms, we got the best accuracy by using K-Nearest Neighbor Classification. Our accuracy is 96% when we implemented the

model. Among all those models, the lowest accuracy we got from Decision Tree Classification, it gives us only 83% accuracy. Another two algorithms such as Random Forrest Classification & Support Vector Machine also generate a good accuracy of 90%. But our suggested model is K-Nearest Neighbor Classification since it generates highest accuracy among the other three algorithms. Table 1 shows the performance of our models.

Table-1 Performance

Model Name	**Accuracy**
Decision Tree Classification	83%
Random Forest Classification	90%
K-Nearest Neighbor Classification	96%
Support Vector Machine	90%

We evaluate the final result by analyzing the accuracy score.Fig.3 shows the Precision, Recall & F1-Score.

Table-2 Precision, Recall & F1-Score

	Precision	Recall	F1-score	Support
0.0	0.97	1.00	0.98	60
1.0	0.95	1.00	0.98	40
2.0	1.00	0.67	0.80	12
Accuracy			0.96	112

Fig.3 shows us the accuracy graph of K-Nearest Neighbor.

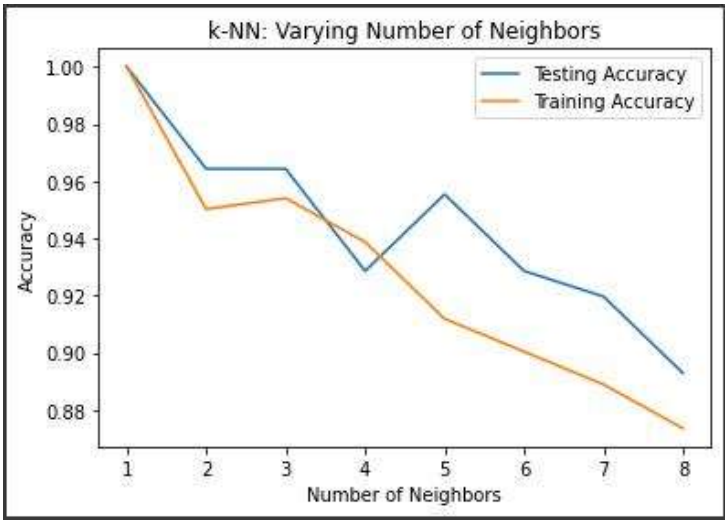

Figure-3: K-Nearest Neighbor Accuracy Graph

*e*ISSN1303-5150

We also compared our experimental results with some other works found in the literature for the detection of Alzheimer disease. Most of the authors used different machine learning algorithm in their model. Table 3 shows a comparison of our proposed model with some of the latest research work related to Alzheimer's Disease Detection in terms of accuracy and it has been found that our model achieved best accuracy so far.

Table-3 Performance Analysis

Author Name	Experiment on Different Algorithm	Best Model	Accuracy
Liu, et al., 2014 [14]	Binary Classification, SVM Classification	Binary Classification	87.76%
Kim, et al., 2022 [15]	AD classification model, AD progression monitoring model	AD classification model	84%
Minhas, Khanum, Riaz, A. Khan, & Alvi, 2016 [16]	SVM Classification, Decision Tree, Logistic Regression	SVM Classification	84%
Ye, M. Pohl, & Davatzikos [17]	LapSVM classification	LapSVM Classification	94.1%
Proposed model	Decision Tree Classifier, Random Forest, K-Nearest Neighbor, Support Vector Machine	K-Nearest Neighbor	96%

ERROR ANALYSIS

Fig.4 shows us the confusion matrix of our K-Nearest Neighbor. Here we got some error. There were some missing values in our dataset. Here conflicted 4 data with truth & predicted.

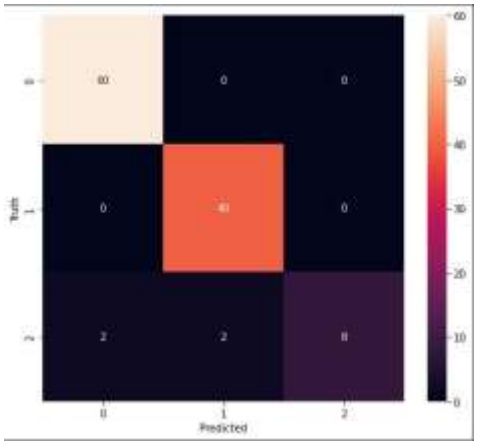

Figure-4: Confusion Matrix Of K-Nearest Neighbor Classification

Figure-5 shows us the graph of the missing value of dataset.

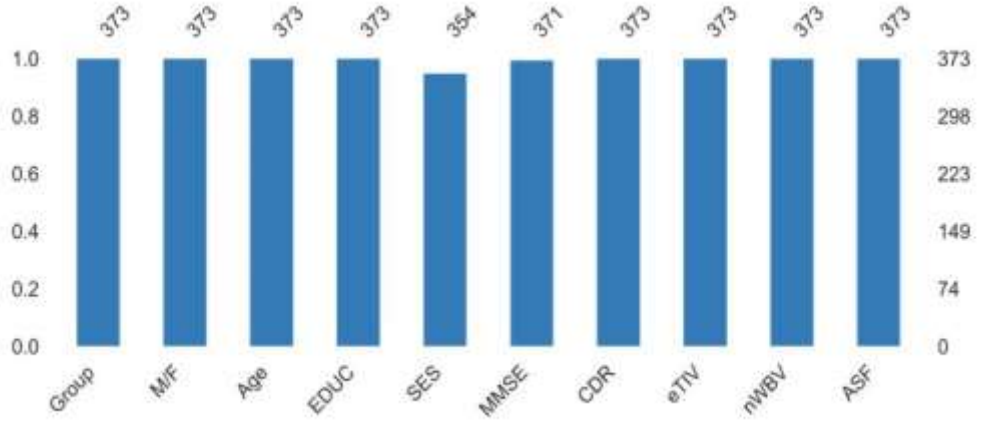

Figure-5: Missing Value

CONCLUSION

In our research work, we adopted K-Nearest Neighbor (KNN) as the main algorithm. We had total 9 attributes and a target attribute. We trained all the dataset through machine learning algorithm and got 96% accuracy by using K-Nearest Neighbor. There was total 374 data in the dataset. Our selected model gave us the best accuracy. In future it can be a perfect model where everyone can predict this disease. Since it is a disease that has not yet been cured so, if we follow the right rules, it is possible to stop this vein before it gets infected. Also, if someone already has this disease, it is possible to reduce the effect of the disease if we can detect at its earlier stage even if it cannot be cured. However, since we have used only some classic machine learning algorithms so deep learning algorithms like RNN, ANN or using some hybrid or ensemble approach would be our future work. We will also try to work with brain image dataset along with text dataset to evaluate the model performance in our future work.

REFERENCES

[1] https://www.worldlifeexpectancy.com /bangladesh-alzheimers-dementia

[2] Antor, M. B., Jamil, A. S., Mamtaz, M., Khan, M. M., Aljahdali, S., Kaur, M., . . .Masud, M. (2021). A Comparative Analysis of Machine Learning Algorithms to. Journal of Healthcare Engineering, 1-12.

[3] C. Kavitha, Mani, V., Srividhya, S., Khalaf, O. I., & Tavera Romero, C. A. (2022). Early-Stage Alzheimer's Disease. Frontiers in Public Health, 1-12.

[4] Castellazzi, G., Cuzzoni, M. G., Ramusino, M. C., Martinelli, D., Denaro, F., Ric-ciardi, A., . . . Wheeler-Kingshott, C. A. (2020). A Machine Learning Approach for the Differential Diagnosis of alzheimer and Vascula Dementia Fed MRI Selected Features. Frontiers in Neuroinformatics, 1-13.

[5] Grassia, M., Perna, G., Caldirola, D., Schruers, K., Duara, R., & A.

Loewenstein, D. (2018). A Clinically-Translatable Machine Learning. Journal of Alzheimer's Disease, 1555-1573.

[6] J.Neelaveni, &Devasana, M. (2020). Alzheimer Disease Prediction. IEEE, 101-104.

[7] Jo, T., Nho, K., & Saykin, A. (2019). Deep Learning in Alzheimer's. Frontiers in Ag-ing Neuroscience, 1-13.

[8] Lee, G., Nho, K., Kang, B., Sohn, K.-A., & Kim, D. (2018). Predicting Alzheimer's disease. SCIENCE REPORT, 1-12.

[9] Mirzaei, G., Adeli, A., & Adeli, H. (2016). Imaging and machine learning techniques for. 857–870.

[10] Alongi, P., Laudicella, R., Panasiti, F., Stefano, A., Comelli, A., Giaccone, P.,. Russo, G. (2022). Radiomics Analysis of Brain [18F] FDG PET/CT to Predict. MDPI, 1-13.

[11] Venugopalan, J., Tong, L., Hassanzadeh, H. R., & D. Wang, M. (2021). Multimodal deep learning models. Scientific Reports, 1-13.

[12] Sivakani. R, & Ansari, G. A. (2020). Machine Learning Framework for Implement-ing. IEEE, 0588-0591.

[13] https://www.kaggle.com/datasets/brsdincer/alzheimer-features

[14] Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., & Feng, D. (2014). Early diagnosis of Alzheimer's disease with deep learning. IEEE, 1015-1018.

[15] Kim, Y., Kim, J., Son, M., Lee, J., Yeo, I., Choi, K. Y., . . . Kim, Y. (2022). Plasma protein biomarker. Scientific Reports.

[16] Minhas, S., Khanum, A., Riaz, F., A. Khan, S., & Alvi, A. (2016). Predicting Pro-gression from Mild Cognitive Impairment to Alzheimer's Disease using Autoregres-sive Modelling of Longitudinal and Multimodal Biomarkers. IEEE.

[17] Ye, D. H., M. Pohl, K., & Davatzikos, C. (n.d.). Semi-Supervised Pattern Classifica-tion: IEEE, 1-4.

*e*ISSN1303-5150