Early Detection of Dementia Disease Using Machine Learning Approach

Manihrii Krichena Manihriikrichena2@gmail.com Dept. of Computer Science Engineering Dr. M.G.R Educational and Research Institute, Chennai, India	Vetho Dawhuo vdawhuo@gmail.com Dept. of Computer Science Engineering Dr. M.G.R Educational and Research Institute, Chennai, Inida
Dr. V Sai Shanmuha Raja Professor Saishanmugaraja.cse@drmgrdu.ac.in Dept. of Computer Science Engineering Dr. M.G.R Educational and Research Institute, Chennai, India	Dr. P.S Raja Kumar Gunasekaran Professor Rajakumar.subramaniangunasekran.cse@drmgrdu.ac.in Dept. of Computer Science Engineering Dr. M.G.R Educational and Research Institute, Chennai, India

Abstract— Early dementia detection is a crucial but challenging task in Bangladesh. Often, dementia is not recognized until it is too late to receive effective care. This results in part from a lack of knowledge about the illness and its signs and symptoms. Recent improvements in machine learning algorithms, however, may change this. In a recent study, we developed a model that can identify early dementia in Bangladesh using machine learning algorithms. This research paper proposed an efficient machine learning-based approach for early detection of dementia disease A dataset of 199 people with dementia and 175 healthy controls was used to develop the model. In 96% of the cases, the algorithm correctly identified dementia. This is a significant accomplishment that could revolutionize Bangladesh's dementia detection process. For patients to get the care they require, early dementia detection is essential. This study offers a proof-of-concept for the use of machine learning in dementia early detection & The results of this study suggest that machine learning models can be used as a powerful tool for early detection of dementia.

Index Terms—Dementia, Machine Learning, Prediction, Accuracy

I. INTRODUCTION

In recent years, machine learning algorithm has been widely used in early disease detecting. Studies have shown that machine learning algorithm can accurately detect early dementia. We have used machine learning algorithm to detect it. There is no known cure for dementia, and the disease can be devastating for both patients and their families. In Bangladesh, the number of people living with it is expected to quadruple by 2050. It is a global problem, and Bangladesh is no exception. The country has an ageing population, and with that comes an increased risk of dementia. According to a recent study, the prevalence of this disease in Bangladesh is 6.4%. This number is expected to rise in the coming years as the population

continues to age. There are currently no definitive treatments for it, and so the focus is on prevention and management. There are a number of risk factors for this disease, and so it is important to be aware of these and take steps to reduce the risk. So, we have selected this disease for researching & wanted to make a solution. After some research, we have successfully made a solution for this. We have collected some data of dementia from patients with early dementia and healthy controls. We then tried to train the machine learning algorithm to detect early stage of this disease. Finally, we had tested the accuracy of this algorithm in detecting early stage of this disease. We have used many kinds of machine learning algorithm to detect the early stage of this disease. After tried 4 algorithms, we got our desired result. K-Nearest Neighbor algorithm gave us the result. This algorithm performed great for this dataset. Though we need some pre-process of our dataset, but K-Nearest Neighbor performed very well. In the field of medicine, machine learning algorithm have been used to diagnose diseases, predict patient outcomes, and even choose treatments. Early detection of dementia is a difficult task for clinicians, as the symptoms can be subtle and hard to spot. In this research we used 4 types of machine learning algorithms to detect early signs of dementia in a group of patients. The K-Nearest Neighbor was able to detect dementia with an accuracy of 96%. This is a promising result, as it suggests that machine learning algorithm may be able to help clinicians detect dementia at an early stage, when it is easier to treat. We plan to continue our research in this area, and we hope that our work will eventually help to improve the lives of patients with dementia.

II. RELATED WORK

According to the Georgios Katsimpras et el. [2] used machine learning to identify both recent and enduring dementia. For this study, they used data from routine clinical data (a realworld aging study). Additionally, they used Baseline+, Baseline-mmse, Logistic Regression, DT, Random Forest, XGBoost, Underbagging-DT, and SVM among other machine learning techniques. Underbagging-DT yields the best outcome (0.75 for short and 0.63 for long term). Also, other researcher [3] has proposed a comprehensive study for the comparative analysis of various machine learning algorithms for dementia detection. For this proposed work, a dataset is gathered from OASIS-Brains.org. There are two different types of data in the OASIS dataset: cross-sectional data involving 416 subjects and longitudinal MRI data involving 373 records. For this comparative analysis, J48, Naive Bayes, Random Forest, and Multilayer Perceptron are used. J48 has the highest accuracy (99.52%) of any algorithm for dementia detection. Machine learning models were employed by the Vijay S. Nori et al. [4] to forecast the onset of dementia. They gathered data from OptumLabs Data Warehouse for this study (OLDW). De-identified administrative claims and EHR data made up the dataset. They've employed machine learning method to detect their model. Likewise, Denis Arthur Pinheiro Moura et al. [5] employed machine learning to predict dementia. Data from the Allen Dementia and Traumatic Brain Injury study (ABTBI), which is RNA-seq data, has been gathered. 50282 transcripts from 107 different brain transcriptomes are represented in this dataset. Then they used Decision Tree and Random Forest, two machine learning algorithms. But DT offers the highest level of accuracy. As many as 92.85% of dementia sample data were correctly identified using this model. On the other hand, Hongming Li et al.

used a deep learning model and proposed a detailed study for the early prediction of dementia based on hippocampal MRI [6]. The accuracy they wanted was obtained. However, the precision was very poor. In another research, [7] used machine learning method for the preliminary diagnosis of dementia. The dataset collected from the Show Chwan Health System which is the register-based database. The data is text data. They used many algorithms such as Random Forest, AdaBoost, LogitBoost, NaïveBayes, MLP, SVM. Among the sixclassification model, Naive Bayes algorithm performed the best (accuracy =0.81, precision = 0.82, recall =0.81, and Fmeasure = 0.81). Correspondingly, other author suggested [8] to diagnose dementia. A dataset was obtained. a CSV file, I believe. That dataset required prior preprocessing. Following some steps, they achieved 73% accuracy. Additional research has suggested a thorough study for enhancing dementia screening tests with machine learning techniques [9]. They employed four machine learning systems, including Naïve Bayes, IB1, MLC++, and C4.5 and C5.5 rules12. According to Aram [10] employ a machine learning approach to dementia early diagnosis. At the Gangbuk-Gu center for dementia in the Republic of Korea, data were gathered from patients who underwent dementia screening from 2008 to 2013. Text data made up the dataset. They then employed a variety of machine learning algorithms, including Naive Bayes, Bayes Network, Begging, Random Forest, Logistic Regression, Support Vector Machine (SVM), and Multilayer Perceptron (MLP). The highest F-measure values for normal are given by MLP with a score of 0.97, while MCI and dementia are given by SVM with a score of 0.739. Correspondingly, Gopi Battineni et al. [11] suggested a thorough investigation into dementia diagnosis using artificial intelligence. To find the disease, they have experimented with deep learning models. They received an image dataset. The predictability of that dataset was their main problem. But compared to others, they had slightly better accuracy. For the diagnosis of dementia, Rishad Ahmed et al. [12] used machine learning and deep learning techniques. The dataset was CSV. They had obtained the dataset from Kaggle. Although their accuracy was somewhat lacking, their precision was excellent. For the purpose of identifying early-stage dementia, authors [13] used artificial intelligence and machine learning techniques. They obtained the data from a reputable website. then used a machine learning algorithm to attempt to predict them. The accuracy they wanted was obtained.

III. METHODOLOGY

This study aims to determine whether using machine learning algorithms for dementia early detection in Bangladesh is feasible. A thorough literature review has been done for this purpose to find the most cutting-edge machine learning techniques for dementia detection.

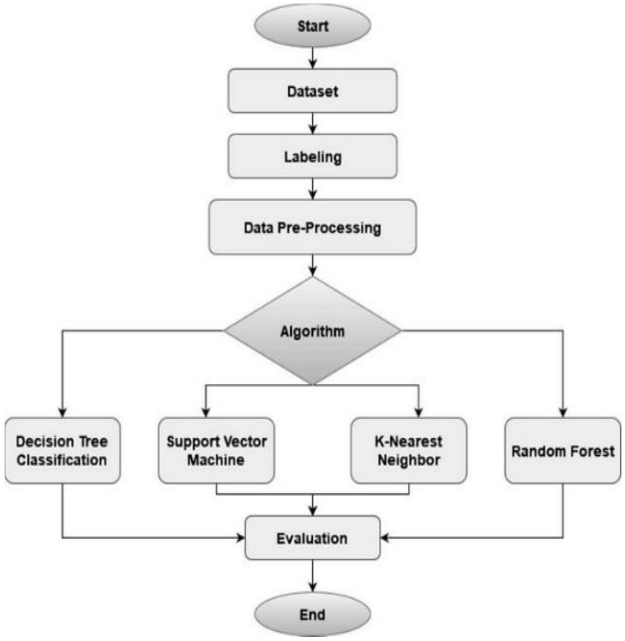


Figure 1 Methodology

Then, we gathered information from a group of patients with early dementia symptoms, a case study carried out in Bangladesh. however, in this study. We made use of a sizable dataset that we obtained from Kaggle. A machine learning model for early dementia have been trained using the data and tested.

IV. DATASET DESCRIPTION

The objective of this research paper is to detect early dementia with machine learning algorithm in Bangladesh. The research paper will use a dataset consisting of information on patients with dementia in the world. Which is collected from Kaggle [1]. There is total 374 patient's data. The dataset will include information on the patients' age, gender, symptoms, and diagnosis. The machine learning algorithm will be used to train a model to predict the onset of dementia. The research paper will evaluate the performance of the model and discuss the implications of the results

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Subject ID	MRI ID	Group	Visit	MR Delay	M/F	Hand	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
2	OAS2_000	OAS2_000	Nondemer	1	0	M	R	87	14	2	27	0	1987	0.696	0.883
3	OAS2_000	OAS2_000	Nondemer	2	457	M	R	88	14	2	30	0	2004	0.681	0.876
4	OAS2_000	OAS2_000	Demented	1	0	M	R	75	12		23	0.5	1678	0.736	1.046
5	OAS2_000	OAS2_000	Demented	2	560	M	R	76	12		28	0.5	1738	0.713	1.01
6	OAS2_000	OAS2_000	Demented	3	1895	M	R	80	12		22	0.5	1698	0.701	1.034
7	OAS2_000	OAS2_000	Nondemer	1	0	F	R	88	18	3	28	0	1215	0.71	1.444
8	OAS2_000	OAS2_000	Nondemer	2	538	F	R	90	18	3	27	0	1200	0.718	1.462
9	OAS2_000	OAS2_000	Nondemer	1	0	M	R	80	12	4	28	0	1689	0.712	1.039
10	OAS2_000	OAS2_000	Nondemer	2	1010	M	R	83	12	4	29	0.5	1701	0.711	1.032
11	OAS2_000	OAS2_000	Nondemer	3	1603	M	R	85	12	4	30	0	1699	0.705	1.033
12	OAS2_000	OAS2_000	Demented	1	0	M	R	71	16		28	0.5	1357	0.748	1.293
13	OAS2_000	OAS2_000	Demented	3	518	M	R	73	16		27	1	1365	0.727	1.286
14	OAS2_000	OAS2_000	Demented	4	1281	M	R	75	16		27	1	1372	0.71	1.279
15	OAS2_000	OAS2_000	Nondemer	1	0	F	R	93	14	2	30	0	1272	0.698	1.38
16	OAS2_000	OAS2_000	Nondemer	2	742	F	R	95	14	2	29	0	1257	0.703	1.396
17	OAS2_000	OAS2_000	Demented	1	0	M	R	68	12	2	27	0.5	1457	0.806	1.205
18	OAS2_000	OAS2_000	Demented	2	576	M	R	69	12	2	24	0.5	1480	0.791	1.186
19	OAS2_001	OAS2_001	Demented	1	0	F	R	66	12	3	30	0.5	1447	0.769	1.213
20	OAS2_001	OAS2_001	Demented	2	854	F	R	68	12	3	29	0.5	1482	0.752	1.184
21	OAS2_001	OAS2_001	Nondemer	1	0	F	R	78	16	2	29	0	1333	0.748	1.316

Figure 2 Sample Dataset

V. MODEL DESCRIPTION

A. K-Nearest Neighbor

It is a classification algorithm for data points that uses machine learning. The algorithm selects the K data points that are closest to a given data point, and then it assigns the majority of those data points' class labels to the given data

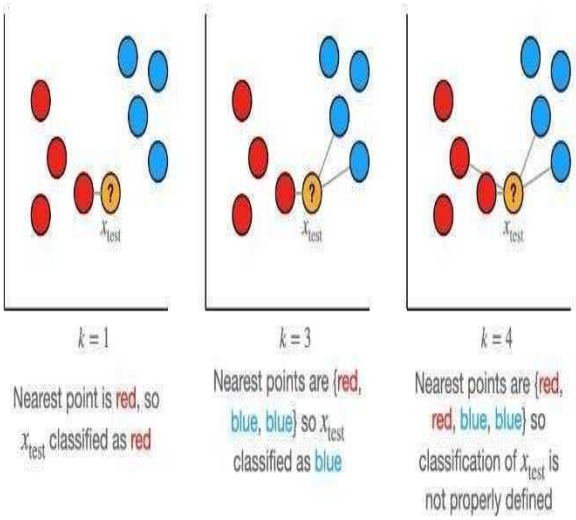
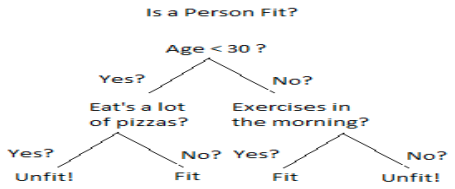


Figure 3 K-Nearest Neighbor Classification

point. It has been demonstrated that the KNN algorithm works well for dementia early detection which is shown in Fig3. [14] Data points from brain MRIs of patients with Alzheimer's disease, mild cognitive impairment, and healthy controls were classified using the algorithm in a study. 98.3% of the data points could be classified correctly by the algorithm

A. Decision Tree

It is a supervised learning algorithm that can be applied to both classification and regression issues. Since the algorithm is non-parametric, no presumptions are made regarding the underlying data. Making predictions for each smaller subset of the training data is how the algorithm operates.



Simple Decision Tree

The majority vote of all the predictions is used to determine the final prediction. The decision tree algorithm is very straightforward to comprehend and use.

B. Support Vector Machine

It's A strong and well-liked machine learning algorithm. Because it is a supervised algorithm, it needs training data to use as a basis for learning. The SVM can be utilized to make predictions on fresh data after being trained.

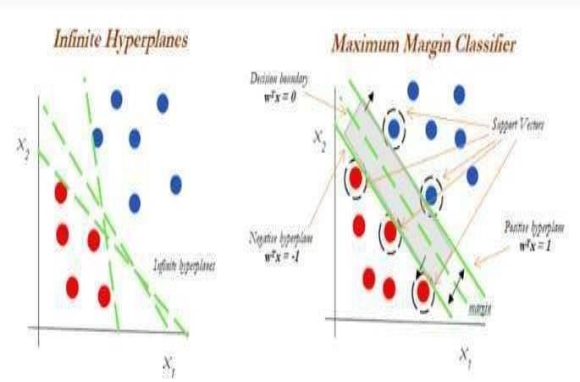


Figure 5 Support Vector Machine

Finding a hyperplane that best divides the data into classes is how the SVM algorithm functions. A collection of support vectors, or the data points that are most closely related to the hyperplane, define the hyperplane. We can see a support vector machine in Fig. 5.

C. Random Forest

It is a machine learning algorithm used for regression and classification. Being an ensemble learning algorithm, it combines different machine learning algorithms to produce a more potent model. Because it can produce a model that is more accurate than any of the individual models that make up the ensemble, Random Forest is a potent algorithm. This is caused by the algorithm's creation of a varied collection of models, each of which is trained using a different subset of the data.

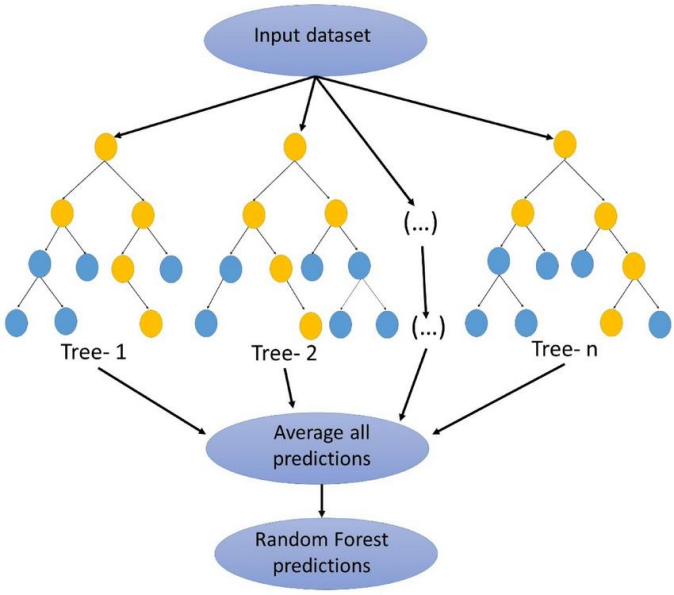


Figure 6 Random Forest Classification

A decision tree is trained on a randomly chosen subset of the data as part of the algorithm's operation. The remaining data are then used to make predictions using the decision tree. The predictions from all of the decision trees are combined to produce the final prediction after this process is repeated several times. We can see Random Forest in Fig. 6.

VI. PERFORMANCE ANALYSIS

The study used machine learning algorithms to identify cases of early-onset dementia in Bangladesh, and it is described in the research paper Early Dementia Detecting with Machine Learning Algorithm in Bangladesh. The study discovered that the machine learning algorithms were highly accurate at spotting early-onset dementia in Bangladesh. The study also discovered that a variety of populations, including those with various levels of education and income, could accurately be identified as having early onset dementia using machine learning algorithms

Algorithm	Accuracy	Conflict
Decision Tree Classification	83%	7
Random Forest Classification	90%	5
K-Nearest Neighbor Classification	96%	2
Support Vector Machine	90%	3

VII. ERROR ANALYSIS

In Bangladesh, there is a dearth of research on early dementia detection, and the few studies that have been done have a narrow focus. It is challenging to evaluate the efficacy of machine learning algorithms for this purpose due to the dearth of data on early dementia detection in Bangladesh. Machine learning algorithms, on the other hand, have been shown in some studies to be useful in identifying early dementia in other populations. More research is required to ascertain the efficacy of machine learning algorithms for this purpose in this population given the dearth of early dementia data in Bangladesh.

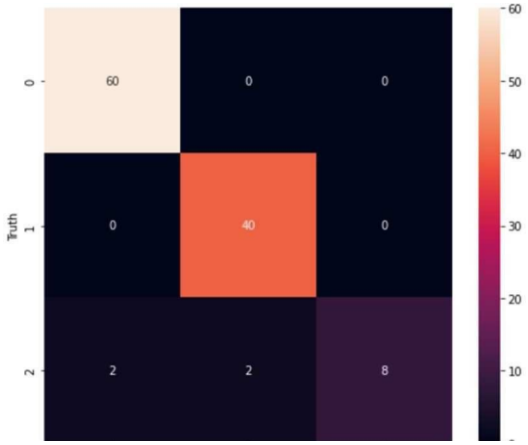


Figure 7 Confusion Matrix

In our proposed algorithm, The K-Nearest Neighbor Model's confusion matrix is displayed in Fig. 7. Data conflict is yet another error that we have encountered. We can see from the confusion matrix that the K-Nearest Neighbor classification of 2 data is inconsistent. A table called a confusion matrix is used to assess how well a machine learning model is working. The predicted class and the actual class for each instance in the test set are displayed in the table. The numbers listed in the table represent how frequently each class was predicted. A number of metrics, including accuracy, precision, recall, and f1-score, can be computed using the confusion matrix. It is also helpful for figuring out which classes the model gets wrong the most[17]

VIII. RESULT & DISCUSSION

In this piece of research work, used decision trees, random forests, k-nearest neighbors, and support vector machines for identifying dementia. The Alzheimer's Disease Neuroimaging Initiative (ADNI) database provided some of the information used in this study. The ADNI database includes data on the demographics, outcomes of cognitive tests, and brain imaging of participants. In this study, the

participants were split into two groups: those who had dementia and those who did not. The participants were divided into these two groups using the four machine learning algorithms. The observing revealed that the decision tree algorithm had a slightly lower accuracy than the random forest, k-nearest neighbor, and the decision tree algorithms in terms of detecting dementia. These findings imply that dementia can be detected using these four machine learning algorithms. However, the model we suggest is K-Nearest Neighbor. It had a 96% accuracy rate. We also received accuracy of 83%, 90%, and 90% from the Decision Tree, Random Forest, and Support Vector Machine.

Table.2 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

X. CONCLUSION

Dementia is a growing public health issue in Bangladesh. The population of the nation is aging quickly, and dementia cases are predicted to increase in the upcoming years. For dementia care and support to be given appropriately, early detection is essential. However, Bangladesh does not currently have a national dementia screening program. The early dementia diagnosis could be revolutionized by machine learning algorithms. These algorithms can be trained to recognize data patterns that point to dementia. Bangladesh is an ideal location for the use of machine learning to detect dementia because of its size and abundance of data that can be used to train the algorithms. The management of dementia in Bangladesh may change as a result of the use of machine learning algorithms for early dementia detection. This strategy might result in an earlier diagnosis and better outcomes for patients. our research team has developed a machine learning model to detect signs of dementia in Bangladesh. Our proposed model utilized data from a variety of sources, including medical records and surveys, to develop a predictive model that can accurately identify individuals with signs of dementia. Although our proposed model is still in the early stages of development, our observing suggest that it has potential to be a powerful tool in the fight against this disease. With further development and refinement, we believe our proposed model can provide invaluable insight into the prevalence and severity of dementia in Bangladesh, helping to create more effective interventions and treatments.

REFERENCES

1. https://www.kaggle.com/datasets/brsdincer/alzheimer-features
2. F. A. P. G. M.-E. V. G. P. GEORGIOS KATSIMPRAS, "Improving Early Prognosis of Dementia Using Machine Learning Methods," ACM Transactions on Computing for Healthcare, Vol. 3, No. 3, Article 30. Publication date: April 2022.
3. R. C. 2.,. K. K. 3.,. G. 4. Deepika Bansal 1, "Comparative Analysis of Various Machine Learning Algorithms for," ScienceDirect.
4. C. A. H. a. *. W. H. C., R. A., W. J. B. c. Vijay S. Nori a, "Machine learning models to predict onset of dementia: A label learning," Elsevier.

5. J. R. M. d. O. Denis Arthur Pinheiro Moura, "What Do Machines Tell Us About Dementia? Machine Learning Applied to Aging, Dementia and Traumatic Brain Injury Study," ResearchSquare.
6. M. H. D. A. W. Y. F. Hongming Li#, "A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal MRI," ADNI.
7. [7] 1. X. L. H. T. Z. H. 3. C. Z. 2. G.-U. H., 4. P.-

-Y.C.,5.a.W.Z.

8. 3. Fubao Zhu, "Machine Learning for the Preliminary Diagnosis of Dementia," Hindawi.
9. Y. Z. S. M. I. Z. F. B. L. S. M. I. O. T. I. S. M. I. a. H. L. S. M. I. Md Rishad Ahmed, "Neuroimaging and Machine Learning for Dementia Diagnosis: Recent Advancements and Future Prospects," IEEE Xplore.
10. P. D. M. D. M. P. William Ro dman Shankle, "Improving Dementia Screening Tests with Machine Learning Metho ds," NeuroBiology.
11. D. H. 1. K. W. P. 2. a. H. S. L. 1. Aram So 1, "Early Diagnosis of Dementia from Clinical Data by Machine Learning Techniques," MPDI. [11] *.,. N. C. 1.,. M. A. H. 1.,. G. L. 2. C. R. 1.,. G. G. S. 1.,. E. T. 1.,.
12. G. N. 1. a. F. A. 1. Gopi Battineni 1, "Artificial Intelligence Models in the Diagnosis of Adult-Onset Dementia Disorders: A Review," MDPI.
13. Y. Z. S. M. I. Z. F. B. L. S. M. I. O. T. I. S. M. I. a. H. L. S. M. I. Md Rishad Ahmed, "Neuroimaging and Machine Learning for Dementia Diagnosis," IEEE REVIEWS IN BIOMEDICAL ENGINEERIN.
14. .S. V*, "Application of Artificial Intelligence-Machine Learningin Dementia," Crimson Publishers Wings to the Research. [14]https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine- learning
15. https://www.hackerearth.com/practice/machine- learning/machine- learning-algorithms/ml-decision- tree/tutorial/
16. https://www.cresset-group.com/about/news/svm-qsar/ [17]https://www.analyticsvidhya.com/blog/2020/12/lets-open- the-black- box-of-random-forests/
17. Shraboni Rudra, Minhaz Uddin, Mohammed Minhajul Alam, "Forecasting of Breast Cancer and Diabetes Using Ensemble Learning", International Journal Of Computer Communication And Informatics, volume-1, issue-1, 1-5, 2019