fff

by DEVANATHAN ARUL

Submission date: 14-Apr-2023 03:51AM (UTC-0500)
Submission ID: 2063779199
File name: Journal_paper.docx (230.11K)
Word count: 2569
Character count: 15223

DEMENTIA PREDICTION USING SUPERVISED MACHINE LEARNING

THYAGARAJAN C
Assistant Professor
Computer Science Engineering
Panimalar Engineering College
Chennai, India
thyaguwinner@gmail.com

PUGHAZENDI N
Professor
Computer Science Engineering
Panimalar Engineering College
Chennai, India
pughazendi@gmail.com

DEVANATHAN A
UG Scholar
Computer Science Engineering
Panimalar Engineering College
Chennai, India
devanathan.arul@gmail.com

DHARNISELVAN H
UG Scholar
Computer Science Engineering
Panimalar Engineering
*College*Chennai, India
dharniselvan002@gmail.com

ARAVIND S
UG Scholar
Computer Science Engineering
Panimalar Engineering College
Chennai, India
aravindan0078@gmail.com

***Abstract*— Dementia is a medical disorder that causes brain damage by rupturing blood vessels in the brain. Complications may arise when the flow of blood and other vital nutrients going to the brain is disrupted. Dementia is a major neurological disorder affecting millions of people worldwide. Early diagnosis of dementia is crucial for better management and treatment of the disease. Here, we explore the use of supervised machine learning (SMLT) algorithms to predict dementia from demographic, clinical, and cognitive data. The dataset used in this study includes information from a large sample of individuals diagnosed with dementia and a control group of healthy individuals. We compare the performance of four popular machine learning algorithms which are Decision tree classifier, Gradient boosting classifier, XGB classifier, Gaussian NB in predicting dementia. Our results show that the XGB Classifier algorithm outperforms other algorithms, achieving an accuracy of 94.36% in classifying individuals with dementia and healthy controls. After that we created a live website which predicts the disease is demented, non-dimented or converted. For that gaussian NB accuracy was accessed as a model file in deployment model.**

Keywords—Dementia, SMLT, Decision tree classifier, Gradient boosting classifier, XGB classifier, Gaussian NB, Demented, Non-demented, Converted.

I. INTRODUCTION

Dementia develops when the supply of blood to specific sections of the brain is interrupted or reduced, blocking the cells in specific sections of the brain from receiving the vitamins, minerals, and oxygen they require and ultimately leading them to die. A medical emergency, such as dementia, requires prompt medical attention. Early detection and treatment are critical to preventing further brain damage and potential problems in other sections of the body. The World Health Organisation (WHO) estimates that there are fifteen million dementia patients globally, with one dying every four to five minutes. Dementia is the sixth most significant cause of mortality in the United States. Dementia is a noncommunicable disease that kills 11% of adults. On average, approximately 795,000 Americans experience the incapacitating symptoms of dementia.

There are two types of dementia: vascular dementia and hemorrhagic dementia. A deficient vessels of the blood bursts and infiltrates the brain in hemorrhagic dementia, whereas clots slow down circulation in chemical dementia. By maintaining a healthy, well-balanced lifestyle. Dementia can be avoided if a person avoids dangerous habits like smoking and drinking, has a normal BMI, appropriate blood sugar levels, and excellent heart and kidney function. It is vital to predict dementia and treat it as soon as possible to avoid irreversible harm or death. The proposed method is to develop a machine learning algorithm for classifying brain dementia. The process begins with data collection, where previous data concerning brain dementia is gathered. In the healthcare domain, data mining is a common method for processing massive amounts of data. The timely detection of dementia can significantly improve patient outcomes. Deep learning has become a valuable tool in healthcare as it reduces manual effort and minimizes errors when using an optimized model. The dataset is carefully analyzed, and proper variable identification is performed, including both regression analysis and independent variable discovery. The most suitable machine learning techniques are applied to the dataset to uncover underlying patterns. After testing various algorithms, the best performing one is selected to predict the desired outcome.

II . RELATED WORKS

Nripendra Narayan Das et al. [1] published a strategy for forecasting Alzheimer's syndrome using brain magnetic resonance imaging scans. The Alzheimer's Disease Neuroimaging Project's standard MRI data will be utilised to test his theory (ADNI). He got Mri Scanned images from two sorts of Alzheimer's patients: those who were moderately mentally affected and those who were not. The form and dimensions of other key brain regions, such as the temporal lobes, occipital lobes, frontal lobes, and insular, are determined using Grey Level Co-Occurrence Matrix (GLCM) and Random Forest mapping. To anticipate the presence of disease, several machine learning methods were used. When using a random forest classification system, the accuracy is 96%, with a region of 82.1% for ROC-AUC.

Deepa, D et al .[2] This study used Magnetic Resonance Imaging (MRI) inputs to identify dementia disease, which led to speedier disease prediction and the contribution of the disease's evolution. By employing this technique, which involves screening dementia disease data and inducing machine learning classifiers, it is feasible to identify and predict the specific dementia of humans. This effort focuses on establishing an evolving framework to successfully identify dementia disease utilising neuroimaging technology and prediction at an early stage using data accumulated for dementia patients.

Carlton Chu et al.[3] used the Gaussian Process (GP) model, a Bayesian learning is similar as conventional variance component estimation. The Bayesian approach let the machine learning model to identify its hyperparameters on its own by maximising "marginal likelihood" or "data evidence."The relevance of various grey matter (GM) areas is automatically weighted in the training procedures using marginal likelihood maximisation without further cross-validation. The algorithms were tested on two groups of people with Huntington's disease and healthy controls. Automatic feature selection improved classification accuracy from 70% to 73% in the first dataset and from 58% to 69% in the second.

A multimodal imaging genetic method for predicting the progression of moderate cognitive impairment patients to Alzheimer's disease was presented by Roman Filipovych et al. [4]. He applied multimodal pattern recognition techniques to identify neuro-imaging and multigenic classification algorithm between healthy individuals and Alzheimer's disease patients. Then, in a linear fashion, he created a composite imaging-genetic score for patients with moderate cognitive impairment to predict the development of Alzheimer's disease. Eventually, the Alzheimer's Disease Neuroimaging Project demonstrated that better Alzheimer's disease prediction was achieved by combining polygenic and neuroimaging data.

Multimodal biochemical, image processing, and neurological markers have shown encouraging results in differentiating dementia patients from senior persons with normal cognitive function, according to Hongming Li et al. [5]. Although it is difficult to predict who and when among persons with mild cognitive impairment will develop dementia, the researchers discovered that pattern classifiers trained on longitudinal data outperformed those trained on cross-sectional data. To capitalise on this discovery, they developed a DL model based on recurring neural networks to capture both the informational presentation and periodic changes in cognitive evaluations for seperate subjects over the time. These cognitive tests were integrated with baseline hippocampus MRI to create a prognosis method.

AMOL K. KADAM ed al [6] The initiative aims to identify among stroke victims their acquaintance with the risk of having a stroke and the factors that influence it. The study has used ML techniques like Random Forest Classification, Decision Tree Classification, Logistic Regression, KNN, and SVM to accurately predict outcomes based on a variety of parameters.

Dharvi Soni et al.[7] The paper's objective is to describe a number of deep learning and machine learning techniques, as well as their advantages, and to choose a future research path for stage-by-stage Alzheimer stage prediction. By screening data on Alzheimer's disease and inducing Machine Learning classifiers, it is possible to use this technology to diagnose and predict the specific dementia in adults.

Govindarajan [8] managed the data assembled from Sugam Multispecialty Hospital. More than 500 patient records and a variety of intriguing class names for two important types of strokes were included in the collection. These methods were employed along with SVM classifier, Artificial Neural Network Classifier, Logistic Regression, Decision Tree, Bagging, and Boosting. Out of the aforementioned Machine Learning Algorithms, they had the highest accuracy (95%) with the ANN Algorithm.

III. Proposed Methodology

Building a model using machine learning for the categorization of brain dementia is the suggested approach. The process starts with data collecting, where previous information about brain dementia is gathered. Data analysis is a popular method for processing vast amounts of information in the health care sector. If detected early enough, a brain dementia can save lives.

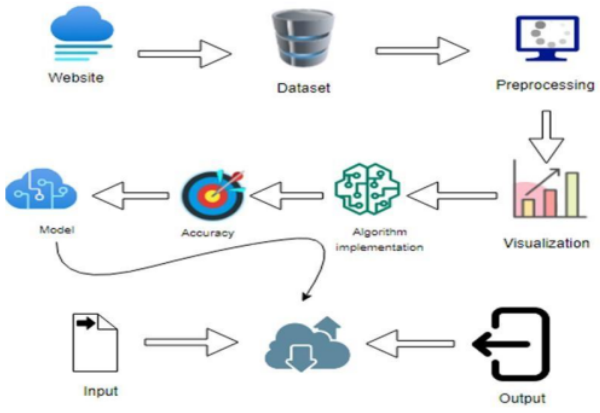

Fig 1. The System Architecture

A better model results in fewer errors, which saves lives, and machine learning is now used and used mostly in the healthcare industry. The dataset is subjected to a comprehensive data analysis, which includes the identification of the dependent and independent variables. On the dataset from which the data pattern is discovered, the appropriate machine learning algorithms are used. A better algorithm is utilized to anticipate the outcome after experimenting with several techniques.

The modules are:

A. Data Pre-processing

Loading the supplied dataset and importing library packages to investigate variable's identity by data type and shape and to assess duplicated and values that are missing. while fine-tuning models and using strategies to make the most of validate and test data while analyzing your models, you can use validation datasets, which are samples of data excluded from learning your model and used to gauge model competence.

In order to analyse the uni-variate, bi-variate, and multi-variate processes, the provided dataset must be renamed, columns must be eliminated, and other data cleaning and preparation stages must be accomplished. On understanding dataset various methods and techniques is been used to clean the data. Data must be cleaned before using a dataset inorder to increase their value for analysis and decision-making.

Dataset:

The Dementia Prediction Dataset consists of 372 rows of data in 13 columns, including qualities such as “Male/Female”, “Hand”, “Age” , “EDUC” , “SES” ,“MMSE” , “CDR” , “MR Delay” , “Visit” , “nWBV” , “eTIV” , “ASF” [9].

```
In [2]: df=p.read_csv('A.csv')
        df.head()
```

Out[2]:

	ASF	eTIV	nWBV	Visit	MR Delay	M/F	Hand	Age	EDUC	SES	MMSE	CDR	Group
0	0.883	1987	0.696	1	0	M	R	87	14	2.0	27.0	0.0	Nondemented
1	0.876	2004	0.681	2	457	M	R	88	14	2.0	30.0	0.0	Nondemented
2	1.046	1678	0.736	1	0	M	R	75	12	NaN	23.0	0.5	Demented
3	1.010	1738	0.713	2	560	M	R	76	12	NaN	28.0	0.5	Demented
4	1.034	1698	0.701	3	1895	M	R	80	12	NaN	22.0	0.5	Demented

```
In [3]: df.tail()
```

Out[3]:

	ASF	eTIV	nWBV	Visit	MR Delay	M/F	Hand	Age	EDUC	SES	MMSE	CDR	Group
368	1.037	1693	0.694	2	842	M	R	82	16	1.0	28.0	0.5	Demented
369	1.040	1688	0.675	3	2297	M	R	86	16	1.0	26.0	0.5	Demented
370	1.331	1319	0.801	1	0	F	R	61	13	2.0	30.0	0.0	Nondemented
371	1.323	1327	0.796	2	763	F	R	63	13	2.0	30.0	0.0	Nondemented
372	1.317	1333	0.801	3	1698	F	R	65	13	2.0	30.0	0.0	Nondemented

Fig 2. Dataset

B. Data visualization

Plots and graphs, which are more tactile and appealing to stakeholders than measurements of correlation or relevance, can be used to explain and highlight crucial linkages in data. Data may not always make sense unless it is graphically represented, such as through charts and graphs. Fast visualization of samples of data and other objects is valuable in both statistical applications and applied machine learning. It will demonstrate how to use different plot types to examine your own information as well as the wide variety of plot types you'll encounter while illustrating information using Python.

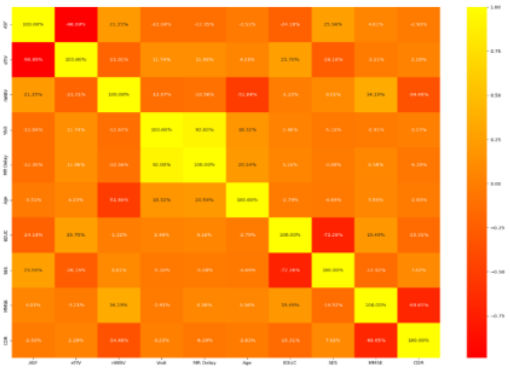

Fig 3. Data visualization

C. Training ML Models

To ensure the system produces the best results and is as accurate as possible, 4 distinct machine learning algorithms have been used in its implementation.

- Decision tree
- Gradient Boosting Classifier
- XGB classifier
- Gaussian NB

The libraries which were used in training these models are

1) Pandas
2) Numpy
3) Matplotlib
4) SKlearn

D. Deployment

After training all the model a live website is created which predicts the possibilities of dementia. For this the machine learning model which gave high accuracy is used by accessing its as a model file in our python program. The tools used in this deployment model are

1. Front End:

Bootstrap, Cascading Style Sheets), Hyper Text Markup Language.

2. Frameworks:

Flask-Python application programming interface for building web applications.

3. Environment for Runtime:

Jupyter notebook

E. Obtaining the results of the test

The last stage will be to use our web application to deliver exact and accurate results to the user, allowing them to proceed as needed in accordance with the results.

The system will finally produce the desired result after the approach, modules, algorithms, and codes have been implemented. The homepage will assist users in entering the information needed for the Dementia prediction system, and the GUI portion is designed to be user-friendly for regular people. The system has been developed utilizing four distinct ML algorithms, as specified in the Implementation, to get the high accuracy with the pre-processed dataset

The outputs generated was:

- The Decision tree classifier has the lowest accuracy (86.91%).
- The XGB classifier has the highest accuracy (94.36%).

The results of the evaluation parameters versus the models are shown in the table below. This was found using the optimal dataset split, which was 70% for training and 30% for testing.

Evaluation parameters	Model Accuracy	Model Recall Value	Model Precision Score	F1-Score
Decision Tree	86.91%	87%	86%	86%
Gradient Boosting	92.95%	92%	93%	92%
XGB Classifier	94.36%	94%	95%	93%
Gaussian NB	91.67%	93%	92%	92%

Table. 1 Accuracy Table

As shown in Figure 4, the suggested system helped us analyse the optimum manner to receive user inputs for the GUI implementation portion of our project. The XGB Classification, which has been trained with the datasets using the information supplied to the GUI for risk of Dementia prediction, was employed, and the training samples was used to compare the fresh data supplied by the user.

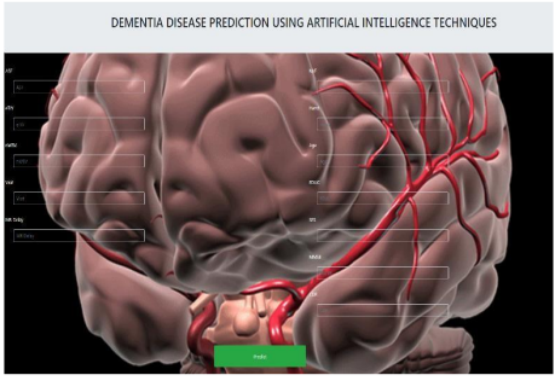

Fig. 4 Screenshot Of Website

IV. Conclusion

After conducting a literature review, we learned about the various benefits and drawbacks of various research papers and proposed a system that predicts Dementia in a cost-effective and efficient manner by requiring few inputs from the user and predicting accurate results using trained Machine Learning algorithms. Thus, the Dementia Prediction was constructed employing four Machine Learning algorithms with the greatest accuracy of 94.36. Consequently, the system is intended to offer a straightforward yet effective Design of user interfaces with a sympathetic viewpoint towards its clients and patients. The system has room for growth in the future, which can lead to greater results and improved customer service. This will assist the user in saving precious time and empower them to take appropriate action based on the results provided..

The implemented system's future scope could include:

1. We aim to gather our organizational dataset for future comparisons of these machine learning algorithms for dementia prediction.

2. A dementia prediction model must have a user-friendly interface in order for healthcare providers, patients, and carers to use it effectively.

3. Integrating the dementia prediction model with digital medical records could enable healthcare practitioners to assess individuals for dementia risk during routine check-ups and give tailored interventions as needed.

V. REFERENCES

[1] Nripendra Narayan Das; Neharika Srivastav; Sourabh Singh Verma, “Magnetic Resonance Imaging based Feature Extraction and Selection Methods for Alzheimer Disease Prediction” 2021, DOI: 10.1109/ICTAI53825.2021.9673337.

[2] D. Deepa; M. Sundar Raj; S. Gowthami; K.Hemalatha; C. Poongodi; P. Thangavel, “Identification and Analysis of Alzheimer’s Disease using DenseNet Architecture with Minimum Path Length Between Input and Output Layers” 2022, DOI: 10.1109/STCR55312.2022.10009552.

[3] Carlton Chu; Peter Bandettini; John Ashburner; Andre Marquand; Stefan Kloeppel, “Classification of Neurodegenerative Diseases Using Gaussian Process Classification with Automatic Feature Determination”, 2010, doi: 10.1109/WB"D.2010.11.

[4] Roman Filipovych; Bilwaj Gaonkar; Christos Davatzikos, “A Composite Multivariate Polygenic and Neuroimaging Score for Prediction of Conversion to Alzheimer's Disease”, 2012, DOI: 10.1109/PRNI.2012.9.

[5] Hongming Li , Yong Fan ; Alzheimer’s Disease Neuroimaging Initiative, “EARLY PREDICTION OF ALZHEIMER'S DISEASE DEMENTIA BASED ON BASELINE HIPPOCAMPAL MRI AND 1-YEAR FOLLOW-UP COGNITIVE MEASURES USING DEEP RECURRENT NEURAL NETWORKS”, Proc IEEE Int Symp Biomed Imaging, 2019 ,Apr;2019:368-371 ,doi:10.1109/ISBI.2019.8759397. Epub 2019 Jul 11.

[6] Priyanka Agarwal ; Mudit Khandelwal ; Nishtha ; Dr. Amol K. Kadam, “Brain Stroke Prediction using Machine Learning Approach”, 2022, Published In: Iconic Research And Engineering Journals , Paper ID: 1703646.

[7] Dharvi Soni; Safvan Vahora , “A Review On Alzheimer Disease Prediction Methodology” 2022, doi: 10.1109/ICAIS53314.2022.9743137.

[8] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, stroke disease classification using machine learning algorithms,” Neural Computing and Applications in 2019.

[9] Dementia Prediction w/ Tree-based Models. Kaggle.com, 2018 ,https://www.kaggle.com/code/ruslankl/dementia-prediction-w-tree-based-models/result

fff

ORIGINALITY REPORT

5%	3%	4%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijstr.org Internet Source	2%
2	Tahia Tazin, Md Nur Alam, Nahian Nakiba Dola, Mohammad Sajibul Bari, Sami Bourouis, Mohammad Monirujjaman Khan. "Stroke Disease Detection and Prediction Using Robust Learning Approaches", Journal of Healthcare Engineering, 2021 Publication	1%
3	shropsys.com Internet Source	1%
4	Nripendra Narayan Das, Neharika Srivastav, Sourabh Singh Verma. "Magnetic Resonance Imaging based Feature Extraction and Selection Methods for Alzheimer Disease Prediction", 2021 International Conference on Technological Advancements and Innovations (ICTAI), 2021 Publication	<1%
5	stacks.cdc.gov Internet Source	<1%

6	download.atlantis-press.com Internet Source	<1%
7	Filipovych, Roman, Bilwaj Gaonkar, and Christos Davatzikos. "A Composite Multivariate Polygenic and Neuroimaging Score for Prediction of Conversion to Alzheimer's Disease", 2012 Second International Workshop on Pattern Recognition in NeuroImaging, 2012. Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		