Final Paper

by alphaq foryou

Submission date: 08-Apr-2023 11:24PM (UTC-0400)
Submission ID: 2059310150
File name: Journal_paper.docx (960.95K)
Word count: 2607
Character count: 15431

DEMENTIA PREDICTION USING SUPERVISIED MACHINE LEARNING

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

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

***Abstract*— Dementia is a medical condition characterized by the rupture of blood vessels in the brain, resulting in brain damage. Symptoms may arise when the supply of blood and other essential nutrients 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 Gaussian NB algorithm outperforms other algorithms, achieving an accuracy of 92% 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 blood supply to certain parts of the brain is compromised or reduced, which prevents the cells in those parts of the brain from getting the nutrients and oxygen they need, which causes them to die. A medical emergency such as dementia necessitates immediate medical care. Timely detection and appropriate treatment are essential to mitigate further damage to the affected brain region and prevent potential complications elsewhere in the body. The World Health Organization (WHO) reports that there are currently fifteen million people worldwide living with dementia, and every four to five minutes, one of them dies. According to the Centers for Disease Control and Prevention, dementia is the sixth most common cause of death in the US (CDC). Dementia, a noncommunicable disease, accounts for 11% of all deaths and afflicts around 795,000 Americans with debilitating symptoms [2]. In India, it is the fourth leading cause of death.

Dementia can be classified into two main types: ischemic and hemorrhagic. Hemorrhagic dementia occurs when a weakened blood vessel ruptures and bleeds into the brain, whereas ischemic dementia is caused by clots that obstruct proper drainage. By having a healthy, balanced lifestyle. Dementia may be avoided if a person avoids harmful habits such as smoking and drinking, keeps a normal body mass index (BMI), ordinary blood sugar levels, and excellent heart and kidney function. It is critical to predict dementia, and it needs to be treated right away to prevent irreparable harm or death The suggested method is to create a machine learning algorithm for brain dementia classification. 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

A methodology to forecast Alzheimer's disease patients using their brain MRI images was put out by Nripendra Narayan Das et al. [1]. The standard MRI data from Alzheimer's disease Neuroimaging Project will be used to test his hypothesis (ADNI). He obtained Mri Scanned images from two types of Alzheimer's patients: mildly mentally impaired and non-impaired. Using Grey Level Co-Occurrence Matrix (GLCM) and Random Forest mapping, the shape and dimensions of other major brain regions, including the temporal lobe, occipital lobe, frontal lobe, and insular, are ascertained. Several machine learning algorithms were employed to predict the presence of illness. The accuracy is 96.4% when applying a random forest classification system, 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. With the use of neuroimaging technology and prediction at a very early stage using the data accumulated for dementia patients, this effort focuses on developing an evolving framework to identify dementia disease effectively.

Carlton Chu et al.[3] employed a Bayesian learning approach called the Gaussian Process (GP) model, which is closely related to the traditional variance component estimation. By maximizing "marginal likelihood" or "evidence of the data," the The Bayesian approach allows the model to autonomously ascertain the hyperparameters of the model.. Without further cross-validation, the relevance of various grey matter (GM) areas is automatically weighted in the training procedures using marginal likelihood maximization. Two groups of individuals with Huntington's disease and healthy controls were subjected to the algorithms. Automatic feature selection increased classification accuracy for the first dataset from 70% to 73% and for the second dataset from 58% to 69%.

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 Predicting the onset of Alzheimer's disease through cognitive impairment was a significant objective, and ultimately, the Alzheimer's Disease Neuroimaging Project provided compelling evidence of improved outcomes in this area and prediction was achieved by combining polygenic and neuroimaging data.

According to Hongming Li et al. [5], Multimodal biochemical, imaging, and neuropsychological markers have shown encouraging results for separating dementia disease patients from elderly people with normal cognitive function. It is still challenging to foresee who and when among those with mild cognitive impairment would acquire dementia, though The researchers observed that pattern classifiers trained on longitudinal data outperformed those trained on cross-sectional data. To leverage this finding, they designed deep learning model utilizes recurrent neural networks to effectively capture both informative representations and temporal changes in cognitive function. assessments over extended periods for individual subjects. By integrating baseline hippocampal MRI with cognitive assessments, a prognostic system was established.

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 Support Vector Machine (SVM), Artificial Neural Network (ANN), 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.

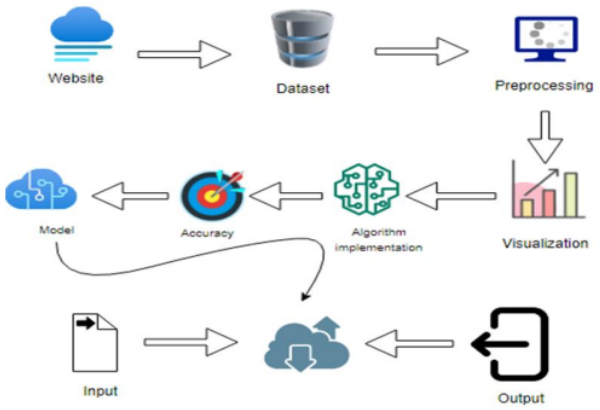


Fig 1. The System Architecture

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.

The provided dataset must be renamed, columns must be deleted, and other data cleaning and preparation steps must be completed in order various procedures and techniques will be employed to cleanse the data, allowing for the analysis of uni-variate, bi-variate, and multi-variate processes as appropriate for the dataset.. To boost the worth of datasets for analysis and decision-making, data cleaning aims to find and remove mistakes and anomalies.

Dataset:
With a total of 372 rows and 13 columns, the Dementia Prediction Dataset contains various attributes, including "Male/Female", "Hand", "Age", "EDUC", "SES", "MMSE", "CDR", "MR Delay", "Visit", "nWBV", "eTIV", and "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

Data visualizations can be used to convey and demonstrate important relationships using plots and graphs that are more visceral and attractive to stakeholders than measures of association or significance. Unless data is graphically represented, such as via charts and graphs, it might not always make sense. 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.

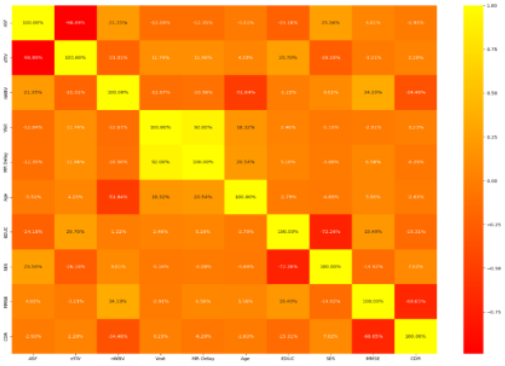


Fig 3. Data visualization

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 (CSS), Hyper Text Markup Language (HTML).

2. Framework Used:

Flask API based on Python for web-applications.

3. Runtime Environment:

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, obtained the highest accuracy when utilizing the dataset.

The output results generated was:

- The Decision tree classifier yielded the lowest accuracy. i.e. (86.91%)
- The XGB classifier produced the highest accuracy (94.36%)

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

Evaluation parameters	Accuracy Score	Recall Score	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 stroke prediction, was employed, The training sample was used to verify the new data provided by the user.

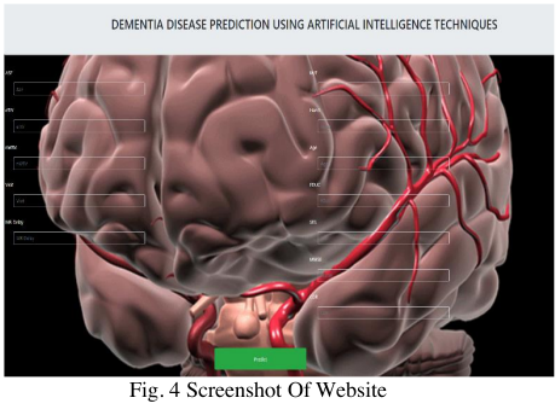


Fig. 4 Screenshot Of Website

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. Background knowledge from other datasets may help boost the accuracy of Dementia prediction models.

2. Our aim is to gather our institutional dataset to use as a benchmark for future machine learning techniques in predicting dementia.

3. As part of our upcoming work, we plan to externally validate our proposed method.

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.

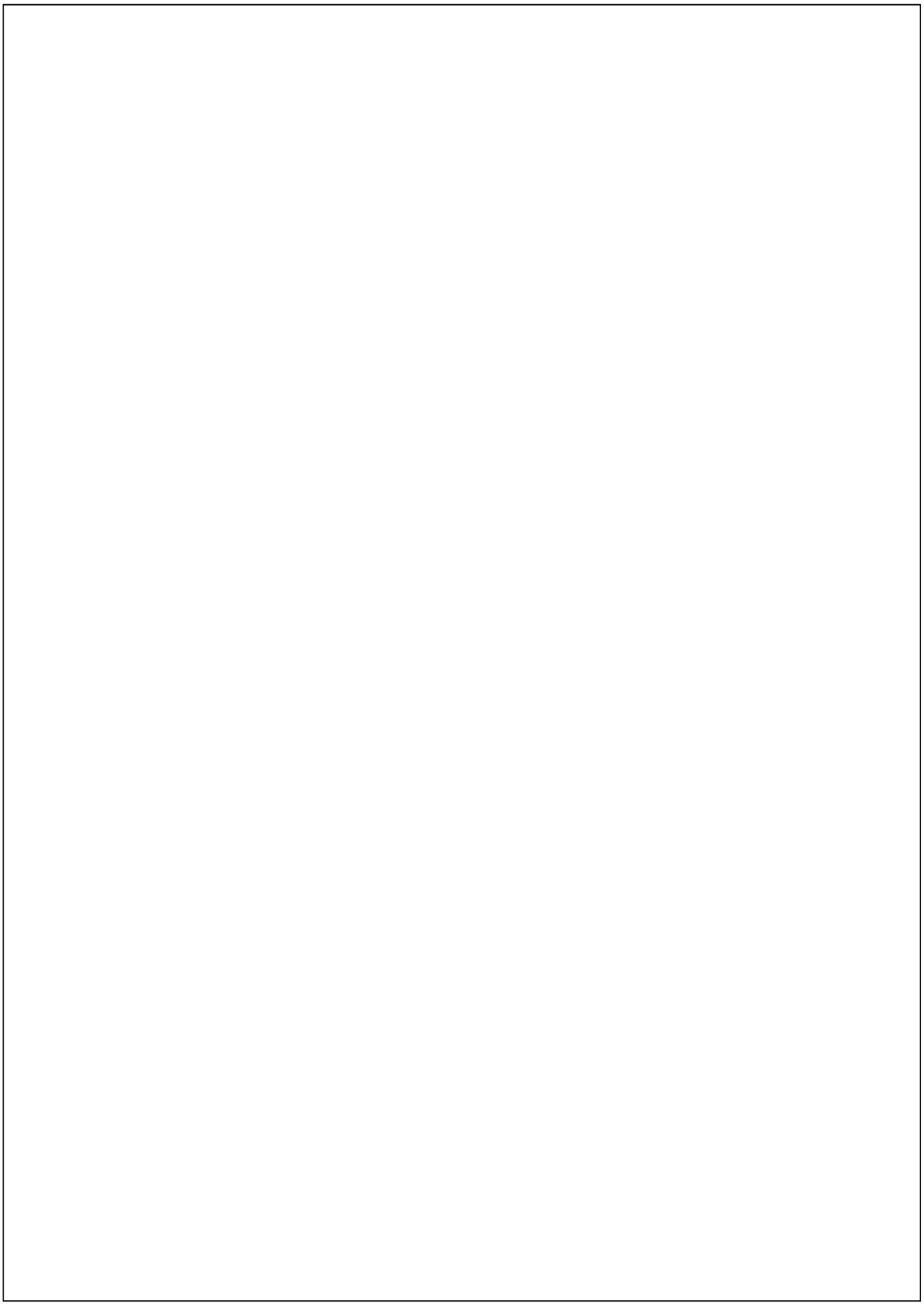
[5] Hongming Li , Yong Fan ; Alzheimer's Disease Neuroimaging Initiative, "EARLY PREDICTION OF ALZHEIMER'S DISEASE DEMENTIA BASED ON BASELINE HIPPOCAMPAL MRI AND 1-YEAR FOLLOW-UP COGNITIVE MEASURES USING DEEP RECURRENT NEURAL NETWORKS", Proc IEEE Int Symp Biomed Imaging, 2019,Apr;2019:368-371,doi:10.1109/ISBI.2019.8759397.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



Final Paper

ORIGINALITY REPORT

15%	6%	10%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	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	3%
2	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	2%
3	www.ijstr.org Internet Source	2%
4	www.jetir.org Internet Source	1%
5	D. Deepa, M. Sundar Raj, S. Gowthami, K. Hemalatha, C. Poongodi, P. Thangavel.	1%

"Identification and Analysis of Alzheimer's Disease using DenseNet Architecture with Minimum Path Length Between Input and Output Layers", 2022 Smart Technologies, Communication and Robotics (STCR), 2022
Publication

6	Submitted to Vel Tech University Student Paper	1 %
7	Carlton Chu, Peter Bandettini, John Ashburner, Andre Marquand, Stefan Kloeppel. "Classification of Neurodegenerative Diseases Using Gaussian Process Classification with Automatic Feature Determination", 2010 First Workshop on Brain Decoding: Pattern Recognition Challenges in Neuroimaging, 2010 Publication	1 %
8	Submitted to Indiana University Student Paper	1 %
9	shropsys.com Internet Source	1 %
10	Rashmi Kuksal, Musheer Vaqur, Ashish Bhatt, Harish Chander, Kapil Joshi. "Stroke Disease Detection and Prediction using Extreme Gradient Boosting", 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), 2023 Publication	1 %

11	Submitted to Southern New Hampshire University - Continuing Education Student Paper	1 %
12	ijasre.net Internet Source	<1 %
13	www.ijritcc.org Internet Source	<1 %
14	www.coursehero.com Internet Source	<1 %
15	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		