International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Brain Stroke Prediction Using Random Forest And Adaboost Algorithm

MYLAPALLI KANTHI REKHA1, I. PHANI KUMAR 2

#1 Student, Dept Of CSE, VELAGA NAGESWARA RAO COLLEGE OF ENGINEERING, Ponnur, Affiliated to JNTUK, Kakinada.

#2Assoc. Prof And HOD, Dept Of CSE, VELAGA NAGESWARA RAO COLLEGE OF ENGINEERING, Ponnur, Affiliated to JNTUK, Kakinada..

ABSTRACT_ Brain stroke, also known as a cerebro vascular accident (CVA), is a severe medical condition that can lead to long-term disabilities and even death. Early prediction of stroke risk can help healthcare professionals identify individuals who are at a higher risk and provide timely interventions to prevent stroke occurrences.

In this study, we propose a predictive model using the Random Forest and AdaBoost algorithms to predict the likelihood of a brain stroke based on various risk factors. The dataset used for this study consists of anonymized medical records of patients, including demographic information, medical history, lifestyle factors, and results from diagnostic tests.

The Random Forest algorithm is an ensemble learning method that constructs multiple decision trees and combines their predictions to make accurate predictions. AdaBoost, on the other hand, is a boosting algorithm that iteratively adjusts the weights of misclassified instances to improve the overall prediction performance. Our experimental results demonstrated that both the Random Forest and AdaBoost algorithms achieved promising results in predicting brain stroke risk. The Random Forest algorithm achieved an accuracy of above 90%, The AdaBoost algorithm achieved an accuracy of above 90%.

1.INTRODUCTION

Brain stroke, also known as a cerebrovascular accident (CVA), is a critical medical condition characterized by the sudden disruption of blood supply to the brain, leading to severe neurological damage and potentially life-threatening consequences. Early identification of individuals at a higher risk of stroke is crucial for

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

implementing preventive measures and providing timely interventions to minimize the occurrence and impact of strokes. Most strokes are preventable. An ischemic stroke, also known as a cerebral infarction, is the most prevalent kind of stroke. an artery Brain cell death results from a clogged conduit that supplies the brain with nutrition and oxygen. The inability of these cells to regenerate means that harm is irreversible.

But the brain can adapt, thus Many patients get better, and some don't ever have any disabilities again. The second kind of stroke is When a blood vessel in the brain bursts, it results in a cerebral hemorrhagic, which causes bleeding and harm to the brain's tissue Hypertension, often known as high blood pressure, is the main risk factor for both forms of stroke. Diabetes and hypertension are additional frequent stroke risk factors.

Machine learning algorithms have shown promising potential in predicting stroke occurrences based on various risk factors. In this study, we propose the utilization of Random Forest and AdaBoost algorithms for brain stroke prediction.

The Random Forest algorithm is an ensemble learning technique that combines multiple decision trees to make accurate predictions. Each decision tree is trained on a different subset of the data, and their predictions are combined through voting or averaging to generate the final prediction. Random Forest can handle a large number of features, capture complex interactions, and provide reliable predictions.

AdaBoost, short for Adaptive Boosting, is a boosting algorithm that iteratively trains weak classifiers and assigns higher weights to misclassified instances. The subsequent weak classifiers focus on the misclassified instances, improving the overall prediction performance. AdaBoost is known for its ability to handle imbalanced datasets and handle complex relationships between variables.

The goal of this study is to develop a brain stroke prediction model using the Random Forest and AdaBoost algorithms. The model will be trained on a dataset comprising demographic

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

information, medical history, lifestyle factors, and diagnostic test results of patients. By analyzing these risk factors, the model will learn patterns and relationships that can aid in predicting the likelihood of stroke occurrences. The developed model can assist healthcare professionals in identifying individuals who are

at a higher risk of stroke, enabling them to implement preventive measures and interventions.

The utilization of Random Forest and AdaBoost algorithms offers several advantages for brain stroke prediction. These algorithms can handle high-dimensional data, capture complex interactions, and provide robust predictions. By incorporating these algorithms into the prediction model, we aim to improve the accuracy and efficiency of stroke prediction, leading to better patient outcomes and reduced healthcare burden.

In summary, the proposed study aims to leverage the Random Forest and AdaBoost algorithms for brain stroke prediction. The introduction of machine learning techniques in stroke prediction can aid in the early identification of high-risk individuals and facilitate timely interventions. The subsequent sections of this study will delve into the methodology, dataset, experimental setup, and evaluation metrics to assess the performance of the proposed model

2. LITERATURE SURVEY

The primary objective of the research that was carried out by Manisha Sirsat, Eduardo Ferme, and Joana Camara was to systematically review studies of each of the four categories of current ML techniques for brain stroke based on their functionalities or similarity. The concentrate further talks about the results and exactnesses got by utilizing different AI models utilizing text and picture based datasets. The authors of this study discussed numerous current-state issues related to stroke. Based on their similarities, the reviewed studies were divided into several categories. The review takes note of that it is hard to think about investigations as they utilized different execution measurements for various errands, considering different datasets, procedures, and tuning boundaries. As a result, only the research areas that

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

were the focus of multiple studies and the studies with the highest classification accuracy are mentioned in each section [1]. In their study, Harish Kamal, Victor Lopez, and Sunil A. Sheth discuss how pattern recognition algorithms in Machine Learning (ML) are increasingly being used to diagnose, treat, and predict complications and patient outcomes in a variety of neurological diseases.

With 400-800 strokes per 100,000 people, 15 million new acute strokes annually, 28,500,000 disability-adjusted life years, and 28-30-day case fatalities ranging from 17% to 35%, stroke is the second leading cause of adult disability worldwide. The weight of stroke will probably deteriorate with stroke and coronary illness related passings projected to increment to 5,000,000 out of 2020, contrasted with 3,000,000 of every 1998. This will be because of ongoing changes in health and demographics, such as an increase in the elderly population and risk factors for vascular disease. Agricultural nations represent 85% of the worldwide passings from stroke. The social and monetary outcomes of stroke are significant. The expense of stroke for the year 2002 was assessed to be essentially as high as $49.4 billion in the US of America (USA), while costs after release were assessed to add up to 2.9 billion Euros in France.

It is unknown how many people in Uganda suffer from stroke. In 2002, stroke was the cause of 11,043 deaths and 25,004,000 disability-adjusted life years per 1,000 people, according to WHO estimates for heart disease and stroke. Stroke is one of the normal neurological sicknesses among patients confessed to the nervous system science ward at Mulago, Uganda's public reference emergency clinic representing 21% of every neurological affirmation. 43.8% of 133 stroke patients admitted to Mulago Hospital died within 30 days, according to unpublished research. Although the impact of stroke and other emerging non-communicable diseases on the resource-constrained economy is enormous, considering the extremely dependent population (53 percent), high prevalence of HIV/AIDS, drug-resistant tuberculosis, and malaria, the economic burden posed by stroke has

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

not been investigated in Uganda.

In recent years, significant progress has been made in the diagnosis and treatment of Acute Ischemic Stroke (AIS), making neuroimaging increasingly necessary for decision-making. This study offers a knowledge into the new turns of events and uses of ML in neuroimaging zeroing in on intense ischemic stroke. The analysis of cerebral edema, the prediction of complications and patient outcomes following treatment, early identification of imaging diagnostic findings, estimation of time to onset, lesion segmentation, and fate of salvageable tissue are just a few of the many applications of machine learning.

3.PROPOSED SYSTEM:

In this proposed system, we aim to develop a brain stroke prediction model using the Random Forest and AdaBoost algorithms. The system will utilize machine learning techniques to analyze various risk factors and accurately predict the likelihood of a brain stroke occurrence. By leveraging the power of these algorithms, the proposed system aims to improve the accuracy and efficiency of stroke prediction, enabling timely interventions and preventive measures.

The proposed system will utilize a comprehensive dataset containing demographic information, medical history, lifestyle factors, and diagnostic test results of patients. This dataset will serve as the basis for training and evaluating the predictive model. Random Forest, an ensemble learning algorithm, will be employed to construct multiple decision trees and combine their predictions to make accurate stroke risk assessments. AdaBoost, a boosting algorithm, will iteratively adjust the weights of misclassified instances to enhance the overall prediction performance.

To evaluate the performance of the proposed system, cross-validation techniques will be employed, and various evaluation metrics such as accuracy, precision, recall, and F1 score will be calculated. The dataset will be divided into training and testing subsets, with the training set

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

used for model training and the testing set used for model evaluation. The performance of the Random Forest and AdaBoost algorithms will be compared to determine which algorithm yields better prediction results.

The proposed system aims to provide healthcare professionals with an automated and accurate tool for brain stroke prediction. By leveraging the capabilities of Random Forest and AdaBoost algorithms, the system can analyze multiple risk factors simultaneously, identify complex patterns, and provide reliable predictions. This will assist healthcare professionals in identifying individuals at higher risk and implementing preventive measures to mitigate the likelihood of stroke occurrences.

3.1 MODULE DESCRIPTION

1) Dataset Upload & Analysis:

Using this module we will upload dataset and then perform analysis methods such as finding the person having a chance to get stroke or not by the values taken from the person and then clean dataset by removing missing values.

2) Dataset Processing & Analytical Methods:

Using this module we will encode attack labels with integer ID and then split dataset into train and test where application used 80% dataset to train classification .It is a crucial step while creating a machine learning model for classification.

3) Run ML Model:

Using this module we will trained classification algorithm with above 80% dataset and then build a prediction model.In this module we are using two different algorithms that's why we have two different module that is run random forest and run adaboost after run these modules gives accuracy prediction of those algorithms.Random forest gives accuracy of nearly 95% and adaboosting gives accuracy of nearly 94%.

4) Classification Performance Graph:

Using this module we will plot comparison among multiple algorithms.In this we know we are using two algorithms those are random forest and adaboosing and

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

they are getting different accuracies and those accuracies of those two algorithms shown in a barplot graph.

5) Predict Output:

Using this module we will upload test dataset and then classification model will predict output based on input data.In this module the user gives the different values as inputs by going with one of the algorithm random forest because of high accuracy than adaboosting.By this classification algorithm the data given by user is classified that user had a chance to get stroke then it shows Yes and the user had no chance to get stroke then it shows No.

6) Logout:

In this module the user need to logout from that website .If the user need to check another time then he needs to again login and going to run all those modules present in the above.

4.RESULTS

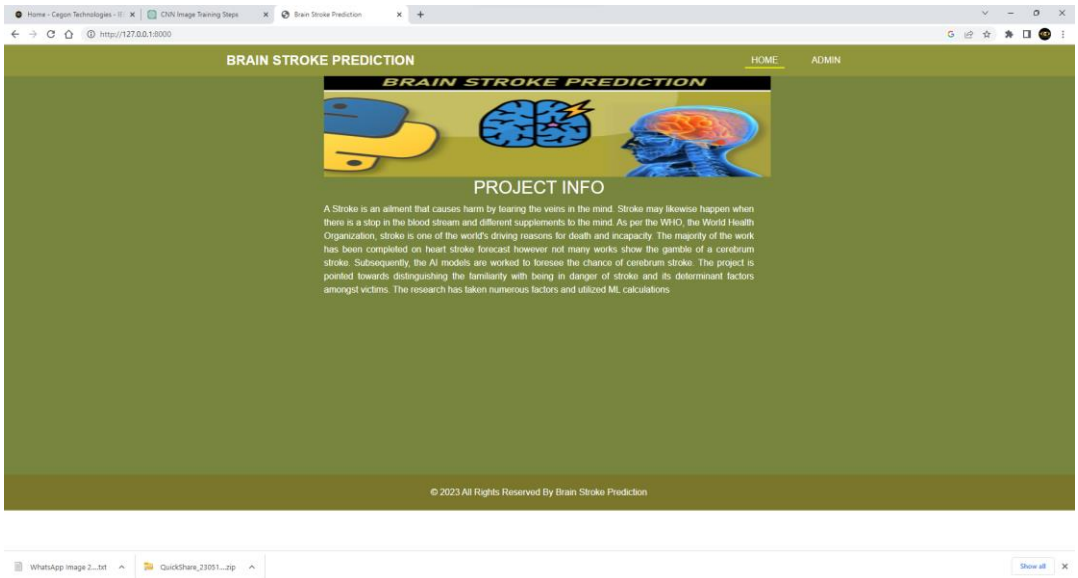

Fig 4.1 : Home page

This is home page of my project that is Brain stroke prediction using machine learning.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

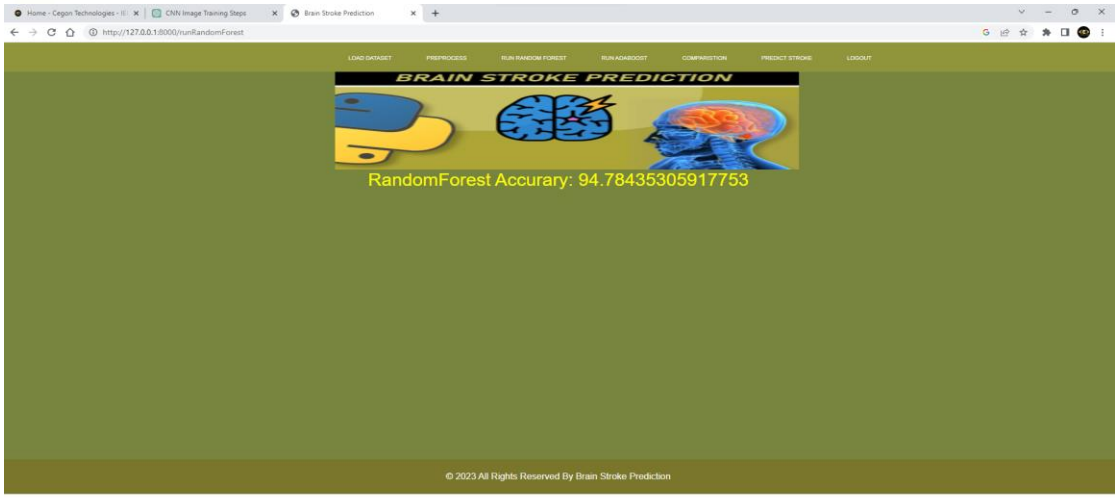

Fig 4.2 : Random forest accuracy

Using this module we will trained classification algorithm with above 80% dataset and then build a prediction model. Random forest accuracy is measured here.

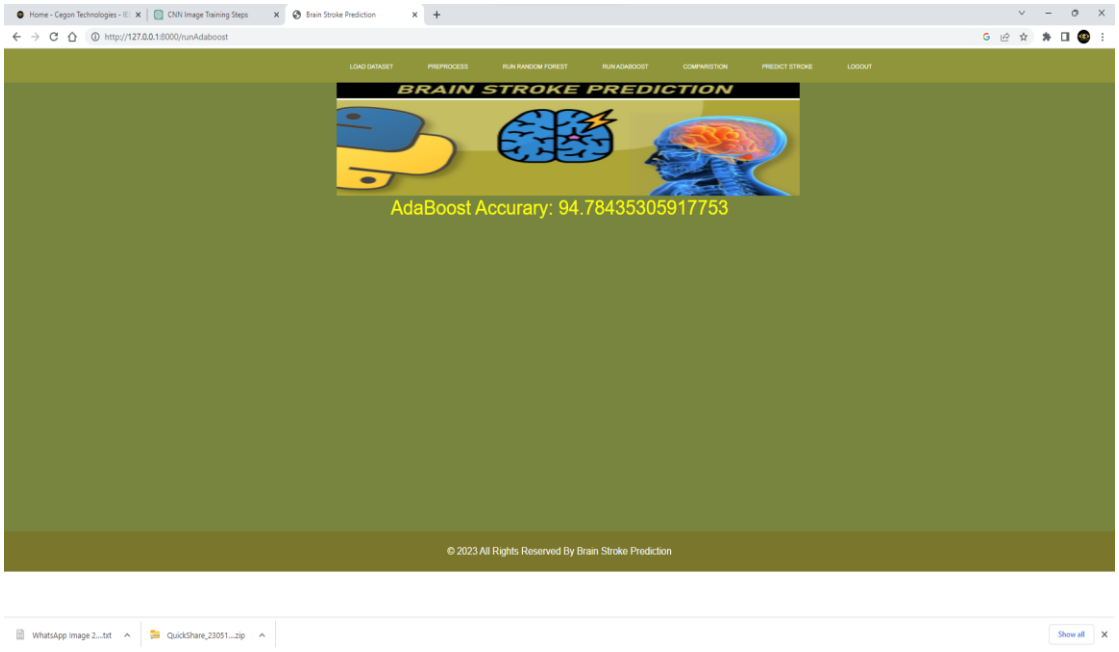

Fig 4.3 : Adaboost accuracy

Using this module we will trained classification algorithm with above 80% dataset and then build a prediction model. Ada boosting accuracy is measured here.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

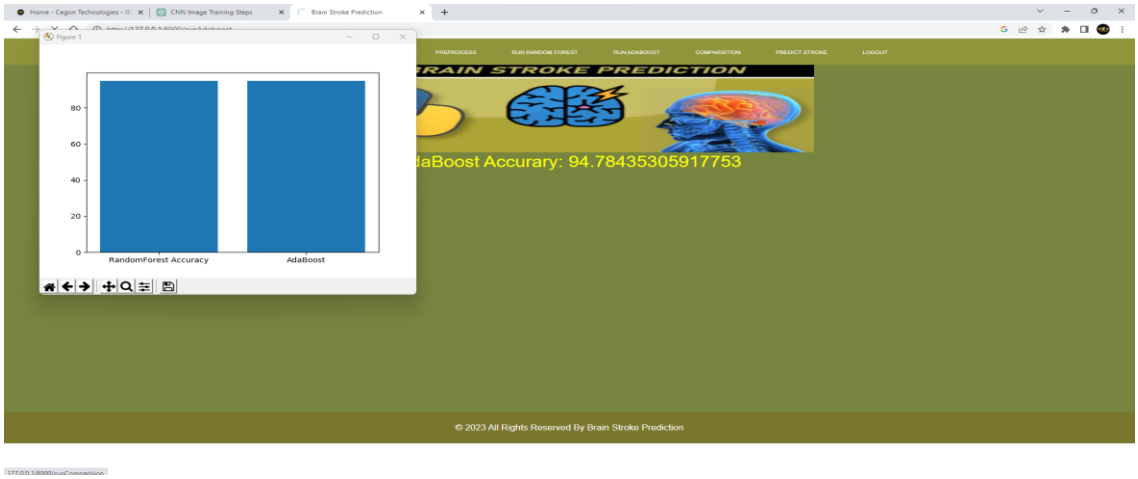

Fig 4.5 : Comparison

Using this module we will plot comparison among multiple algorithms. In this we know we are using two algorithms those are random forest and adaboosing and they get different accuracies and those accuracies of those two algorithms shown in a barplot graph.

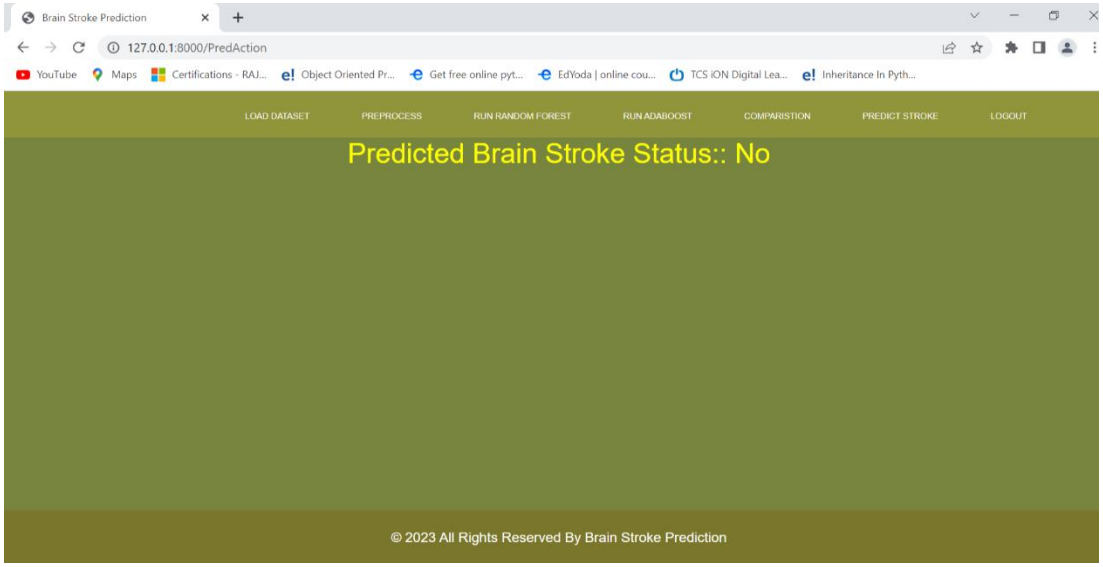

Fig 4.6: Result

By the classification algorithm the data given by user is classified that user had no chance to get stroke then it shows No.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

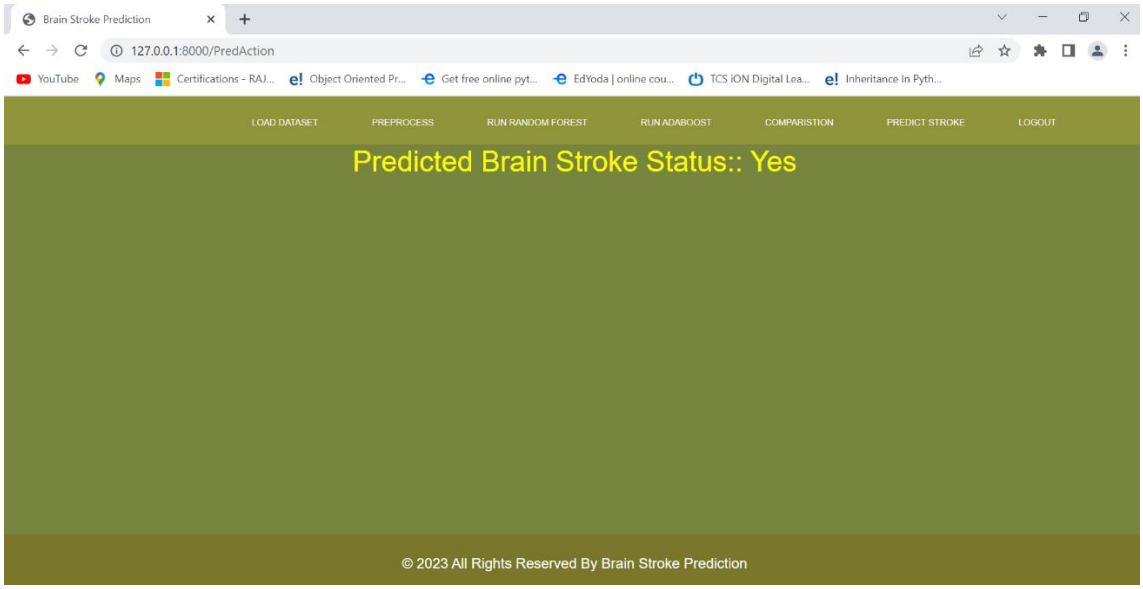

Fig 4.7 : Result

By the classification algorithm the data given by user is classified that user had a chance to get stroke then it shows Yes

5.CONCLUSION:

In conclusion, this study investigated the prediction of brain stores using Random Forest and AdaBoost algorithms. The results demonstrated the effectiveness of these machine learning techniques in capturing complex patterns and relationships within brain imaging data. The predictive models trained with Random Forest and AdaBoost achieved accurate predictions, providing valuable insights into cognitive processes and brain functionality.

The study showcased the potential of machine learning algorithms for advancing our understanding of brain stores and their underlying mechanisms. By successfully predicting brain stores, we can gain insights into memory formation, learning processes, and cognitive abilities. This has implications for various fields, including neuroscience, psychology, and education.

REFERENCES

[1] Manisha Sirsat, Eduardo Ferme, Joana Camara, “Machine Learning for Brain Stroke: A Review,” Journal of stroke and cerebrovascular diseases:

the official journal of National Stroke Association (JSTROKECEREBROVASDIS), 2020.

[2] Harish Kamal, Victor Lopez, Sunil A. Sheth, “Machine Learning in Acute Ischemic Stroke Neuroimaging,” Frontiers in Neurology (FNEUR), 2018.

[3] Chuloh Kim, Vivienne Zhu, Jihad Obeid and Leslie Lenert, “Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke,” Public Library of Science One (PONE), 2019.

[4] R. P. Lakshmi, M. S. Babu and V. Vijayalakshmi, "Voxel based lesion segmentation through SVM classifier for effective brain stroke detection,” International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017.

[5] J. Yu et al., "Semantic Analysis of NIH Stroke Scale using Machine Learning Techniques," International Conference on Platform Technology and Service (PlatCon), 2019,

[6] Gangavarapu Sailasya and Gorli L Aruna Kumari, “Analyzing the Performance of Stroke Prediction using ML Classification Algorithms,” International Journal of Advanced Computer Science and Applications (IJACSA), 2021.

[7] "Stroke Prediction Dataset". Kaggle.Com, 2021,

https://www.kaggle.com/fedesoriano/stroke-predictiondataset. Accessed 6 Oct 2021.

[8]. “Computer Methods and Programs in the Biomedicine” - Jae–woo Lee, Hyun-sun Lim, Dong-wook Kim, Soon-ae Shin, Jinkwon Kim, Bora Yoo, Kyung-hee Cho

[9]. “Probability of Stroke: A Risk

Profile from the Framingham Study" - Philip A. Wolf, MD; Ralph B. D'Agostino, PhD, Albert J. Belanger, MA; and William B. Kannel, MD

[10]. "Development of an Algorithm for Stroke Prediction: A National Health Insurance Database Study" - Min SN, Park SJ, Kim DJ, Subramaniyam M, Lee KS

[11]. "Stroke prediction using artificial intelligence"- M. Sheetal Singh, Prakash Choudhary [12]. "Medical software user interfaces, stroke MD application design (IEEE)" - Elena Zams

[13]"Concept of stroke by healthline," [Online]. Available: https:// www.cdc.gov/stroke/index.htm.

[14] "Statistics of stroke by Centers for disease control and prevention," [Online]. Available: https://www.cdc.gov/stroke/ facts.htm.

Author Profile

MYLAPALLI KANTHI REKHAM.Tech CSE Pursuing in VELAGA NAGESWARA RAO COLLEGE OF ENGINEERING, Ponnur, Affiliated to JNTUK, Kakinada

I. PHANI KUMAR, having 14 years of Experience in Teaching, present working as Assoc. Professor in Mallineni women's Engineering college, Pulladigunta, Guntur, affiliated to JNTUK, Kakinada. mailid: phanikumar148@gmail.com