JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Machine Learning Algorithms for Forecasting Brain Stroke Risk

Dr.TVS Divakar
Department of Electronics and Communication Engineering
GMR Institute of Technology
Rajam, Andhra Pradesh, India
divakar.tvs@gmrit.edu.in

Dr. P. Ravi Kumar
Department of Electronics and Communication Engineering
GMR Institute of Technology
Rajam, Andhra Pradesh, India
ravikumar.p@gmrit.edu.in

Duppatla Balaram
Department of Electronics and Communication Engineering
GMR Institute of Technology
Rajam, Andhra Pradesh, India
duppatlaram@gmail.com

Guntuku Rishitha
Department of Electronics and Communication Engineering
GMR Institute of Technology
Rajam, Andhra Pradesh, India
guntukurishitha@gmail.com

Gondu Sai Kumar
Department of Electronics and Communication Engineering
GMR Institute of Technology
Rajam, Andhra Pradesh, India
gondusaikumar97@gmail.com

Bepala Chandrika
Department of Electronics and Communication Engineering
GMR Institute of Technology
Rajam, Andhra Pradesh, India
chandrika216@gmail.com

***Abstract*—** Stroke, one of the most dangerous illnesses in the world, causes a large number of deaths, either directly or indirectly. Machine learning plays a crucial role in the prediction of strokes. It has the potential to employ multiple machine learning algorithms, like Support Vector Machine (SVM), Random Forest, Decision Tree, K-Nearest Neighbors, and Logistic Regression. Early recognition of the various warning signs of a stroke can help reduce the severity of the stroke. Early prediction of a stroke can save a few lives. The Health Care Stroke Dataset, which includes parameters like age, gender, hypertension, heart diseases, stroke, work type, BMI, smoking status, average glucose level, marriage status, residence type, and stroke, has been our working tool. The accuracy of each algorithm was assessed following data preprocessing, and subsequent comparisons were made to identify the algorithm with the highest accuracy in stroke prediction. Based on the comparison results, a Graphical User Interface individual is at risk of a stroke by analyzing the manual input provided by the user.

***Keywords*—** *Random Forest, Decision Tree, Support Vector Machine Machine Learning, GUI, Logistic Regression, KNN.*

INTRODUCTION

Stroke is the second-most common cause of mortality globally and one of the most serious illnesses affecting those over age 65. Similar to how a "heart attack" harms the heart, it harms the brain as well. A stroke can finally result in death, in addition to requiring expensive medical treatment and leaving the victim permanently disabled. A stroke claims a life every 4 minutes, although 80% of cases can be avoided if the condition is detected or anticipated in its early stages. A stroke is a cerebral haemorrhage or blood clot that can result in permanent impairment of one's ability to move, think, see, or speak. The majority of strokes are ischemic, embolic, or haemorrhagic stroke happens when there's an interruption or decrease in blood flow to particular brain regions. This results in a shortage of oxygen and nutrients for the cells in those locations, which eventually leads to cell death. Given that this is a medical emergency, getting help quickly is crucial**[1].** The prevention of stroke's long-term effects is mostly dependent on prompt intervention. Still, standard methods of estimating the risk of stroke are often inaccurate and time-consuming. Recent research has shown that machine learning algorithms can accurately predict the risk of stroke by examining a variety of clinical risk factors**[2].**In India, stroke ranks as the fourth-most common cause of death. Thanks to developments in clinical science, machine learning (ML) algorithms are now able to predict the likelihood of stroke. These algorithms are useful for producing exact analyses and forecasts**[3].** Machine learning is a subfield of artificial intelligence (AI) that is part of computer science. It lets systems self-improve and control datasets to produce output values within predetermined bounds. Algorithms in this technological field are defined precisely to represent the datasets that computers use to evaluate problems. Two methods for putting these algorithms into practice are supervised and unsupervised learning**[4].** The majority of research had acceptable accuracy rates of about 90%. But our work differs in that it uses multiple proven machine learning techniques to achieve the best possible outcomes. Interestingly, the algorithms that produced the highest F1 scores were Logistic Regression(LR), Decision Tree (DT), and Random Forest (RF)**[5].**

THEORY:

To predict brain strokes with machine learning, an organized technique is used. Begin by describing the problem and compiling a complete dataset of pertinent medical and lifestyle data. After preparing the data, which includes resolving missing values and normalizing numerical features, use feature selection to determine the most relevant predictors. Divide the dataset into training and testing sets, then select a suitable

machine learning model, such as Logistic Regression, Decision Trees, KNN, Random Forest, and Support vector machine. Generate the model using the training set, examine its results using the testing set, and subsequently modify the hyperparameters for optimization. Analyse feature contributions to ensure interpretability, then validate the model on separate datasets.

Depending on the nature of the data and the task's specific needs, many machine-learning techniques can be used to forecast brain strokes. Commonly used algorithms are:

1. K-Nearest Neighbors (KNN)

2. Logistic regression (LR)

3. Random Forest (RF)

4. Support Vector Machine (SVM)

5. Decision Tree (DT)

K-Nearest Neighbors (KNN):

Both classification and regression applications can benefit from the adaptable and easy-to-use K-Nearest Neighbors (KNN) machine learning technique. In the context of predicting brain strokes, KNN categorizes a new data point based on the majority class of its k-nearest neighbors in the feature space. During the training phase, the algorithm memorizes a dataset of labeled instances and assigns each data point to a given class, such as stroke or non-stroke, based on its attributes. During the prediction phase, the method computes the distances between the new data point and its neighbors, usually using metrics like Euclidean distance. The class assigned to the majority of these k-nearest neighbors is used to forecast the new data point.

Logistic Regression (LR):

When dealing with problems with binary classification with two alternatives leading to variables, statistical procedures like logistic regression are usually utilized. Contrary to its name, logistic regression is a classification algorithm rather than a regression tool. The logistic function is used by the algorithm to estimate the probability that a given input belongs to a specific class and award it a value between 0 and 1. Logistic regression can be used to predict brain strokes based on multiple input features. Logistic regression is beneficial in predicting brain strokes because of its interpretability, simplicity, and ability to handle linear connections between features and log odds. The model output is simply understandable as the chance of a stroke occurring, and the coefficients associated with each variable provide information about their impact on stroke risk.

Random Forest:

In terms of forecasting brain strokes, the Random Forest algorithm emerges as a potent and versatile tool. An effective technique for forecasting brain strokes is to collect a comprehensive dataset that includes critical factors such as age, blood pressure, cholesterol levels, smoking behaviors, and relevant medical history variables. To assure data quality, preprocessing activities are performed after data collection, including the management of missing values and outliers. The most influential variables are then identified using feature selection, with Random Forest automatically supplying feature importance scores.

Support Vector Machine:

A supervised machine learning method known as Support Vector Machine, or SVM, is useful for classification as well as regression. Support Vector Machine operates by determining the hyperplane that optimally divides distinct classes in the feature space. To forecast the occurrence of strokes, a dataset is compiled, often containing factors such as age, blood pressure, cholesterol levels, smoking behaviors, and other relevant medical history information. After a component of the dataset is pre-processed to guarantee the quality of the data by removing outliers along with missing values. The SVM technique converts the features into dimensions in a multidimensional space, and the algorithm seeks a hyperplane that best separates the data into different groups.

Decision Tree:

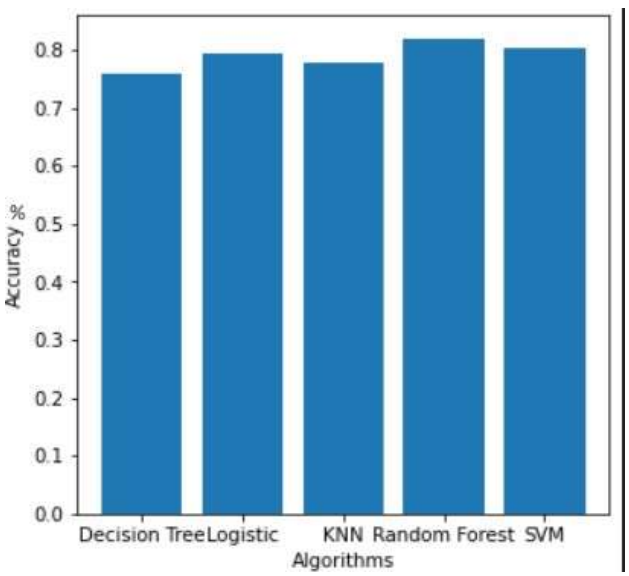
Regression and classification can both be accomplished with the help of the flexible and intuitive Decision Tree algorithm in machine learning. It creates a tree-like structure by recursively partitioning the dataset by feature, with each internal node representing a feature-based decision and each leaf node indicating the ultimate predicted outcome. The algorithm's goal is to build a tree that optimizes information gain or decreases impurity at each node, allowing for more accurate and interpretable predictions.

IV. RESULTS

A stroke is a medical emergency brought on by bleeding or blood clots that stop the blood supply to a portion of the brain. With a 5.5 million yearly mortality rate, it ranks as the second most common cause of death worldwide. Over 15 million individuals worldwide suffer a stroke each year, and one stroke-related death occurs every four minutes. Since a bad lifestyle is typically the cause of a stroke, it can be prevented in up to 80% of instances.

As a result, stroke prediction becomes essential and ought to be applied to stop stroke-related irreparable harm. Therefore, to determine if a person is experiencing a brain stroke or not, we have developed a brain stroke prediction model.

Fig1 Accuracies Graph



Accuracies Comparison:

Table 1

ALGORITHMS	ACCURACY(%)
Decision tree	75.769
Random Forest	81.923
Logistic Regression	79.230
KNN	77.692
SVM	80.384

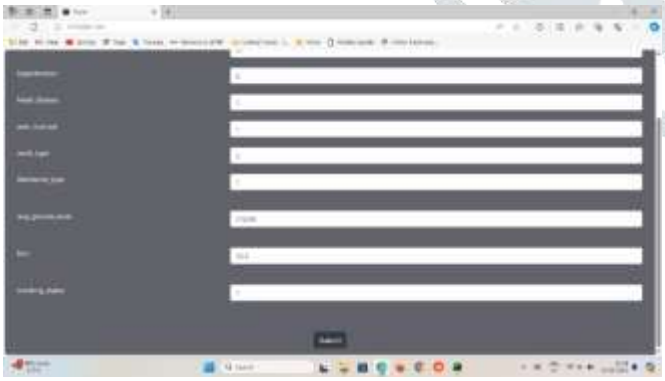
The healthcare dataset was pre-processed, and then it was split into two halves. 20 percent of the data is used for testing, while 80 percent is used for training. The accuracy of several machine methods, including Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors, Random Forest, and Decision Tree, has then been identified. The various machine-learning algorithms have then been placed on a graph.

Graphical User Interface(GUI):

The Graphical User Interface (GUI) will be utilized to manually enter the values, which will then be used to predict whether or not the individual in question is in danger of experiencing a stroke.

The following variables are collected for the diagnosis: gender, age, heart disease, hypertension, work type, residence type, average glucose level, body mass index, and smoking status.

Fig 2 Screenshot of GUI



Stroke-Risk Prediction:

The person who has been diagnosed with stroke risk or not is predicted by the machine learning algorithm based on the values of specified parameters.

The fig describes the person is diagnosed with no stroke risk which means the person doesn't have a risk of brain stroke. The fig describes the person is diagnosed with stroke risk which means the person is having a risk of brain stroke.

Fig 3 Screenshot of no stroke

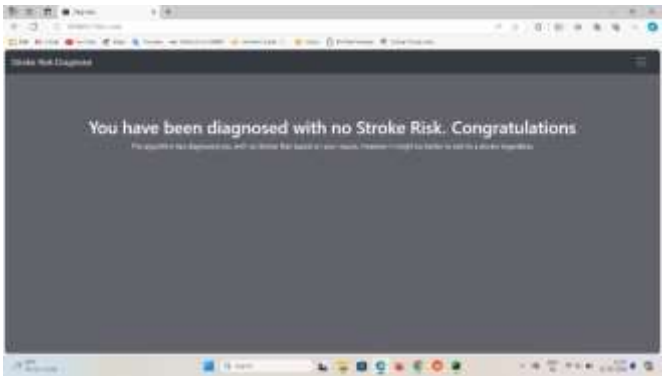


Fig 4 Screenshot of stroke

V. CONCLUSION

In conclusion, there is a chance to enhance healthcare by using machine learning algorithms to anticipate brain strokes. By analyzing vast amounts of data, these algorithms can detect patterns and risk factors associated with the illness, which enables the early detection and prevention of strokes. With its ability to provide personalized insights and treatments that may ultimately save lives and improve patient outcomes, this technology has the potential to fundamentally alter the way strokes are treated.

VI. REFERENCES

[1] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. A. Mamun and M. S. Kaiser, "Performance Analysis of Machine Learning Approaches in Stroke Prediction," *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore,India,2020,pp.14641469,doi:10.1109/ICECA49313.202 05.[2] K. Mridha, S. Ghimire, J. Shin, A. Aran, M. M. Uddin and M. F. Mridha, "Automated Stroke Prediction Using Machine Learning: An Explainable and Exploratory Study With a Web Application for Early Intervention," in *IEEE Access*, vol. 11, pp. 52288-52308, 2023, doi: 10.1109/ACCESS.2023.3278273.

[3] Priyanka Agarwal, Mudit Khandelwal, Nishtha, Dr. Amol K. Kadam "Brain Stroke Prediction using Machine Learning Approach" Iconic Research And Engineering Journals Volume 6 Issue 1 2022

[4] Bandi,V.Bhattacharyya, D.Midhunchakkravarthy, D. (2020). Prediction of brain stroke severity using machinelearning d'IntelligenceArtificielle,Vol.34,No.https://doi.org/10.18280/ria.340 609.

[5] 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, vol. 2021, Article ID 7633381, 12 pages, 2021. https://doi.org/10.1155/2021/7633381.

[6] Sirsat MS, Fermé E, Câmara J. Machine Learning for Brain Stroke:AReview.JStrokeCerebrovascDis.2020Oct;29(10):105162.doi :10.1016/j.jstrokecerebrovasdis.2020.105162.Epub2020Jul28.PMID: 32912543.

[7] Harshitha K V, Harshitha P, Gunjan Gupta, Vaishak P, Prajna KB (2021)**,** Stroke Prediction Using Machine Learning Algorithms, International Journal of Innovative Research in Engineering & Management (IJIREM), Vol-8, Issue-4, Page No-6-9], (ISSN 2347 - 5552). www.ijirem.org.