International Journal for Modern Trends in Science and Technology, 9(05): 501-505, 2023

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0905085
Available online at: http://www.ijmtst.com/vol9issue05.html

Machine Learning Algoritham for the Detection of Brain Haemorrhage and Ischemic Stroke

S. Vishnuvardhan | R.V.Avinash | M.Sathish | U.Bhavan | Dr.D.Bight Anand

Department of CSE, Narayana Engineering College , Gudur , India.

To Cite this Article
S. Vishnuvardhan, R.V.Avinash, M.Sathish, U.Bhavan and Dr.D.Bight Anand. Machine Learning Algoritham for the Detection of Brain Haemorrhage and Ischemic Stroke. International Journal for Modern Trends in Science and Technology 2023, 9(05), pp. 501-505 https://doi.org/10.46501/IJMTST0905085

Article Info
Received: 16 April 2023; Accepted: 10 May 2023; Published: 18 May 2023.

ABSTRACT

Stroke is a medical disorder in which the blood arteries in the brain are ruptured, causing damage to the brain. When the supply of blood and other nutrients to the brain is interrupted, symptoms might develop. According to the World Health Organization (WHO), stroke is the greatest cause of death and disability globally. Early recognition of the various warning signs of a stroke can help reduce the severity of the stroke. Different machine learning (ML) models have been developed to predict the likelihood of a stroke occurring in the brain. -is research uses a range of physiological parameters and machine learning algorithms, such as Logistic Regression (LR), Decision Tree (DT) Classification, Random Forest (RF) Classification, and Voting Classifier, to train four different models for reliable prediction. Random Forest was the best performing algorithm for this task with an accuracy of approximately 96 percent. -e dataset used in the development of the method was the open-access Stroke Prediction dataset. -e accuracy percentage of the models used in this investigation is significantly higher than that of previous studies, indicating that the models used in this investigation are more reliable. Numerous model comparisons have established their robustness, and the scheme can be deduced from the study analysis

KEYWORDS: Brain stroke, machine learning, data analysis, prediction,

1. INTRODUCTION

Stroke occurs when the blood flow to various areas of the brain is disrupted or diminished, resulting in the cells in those areas of the brain not receiving the nutrients and oxygen they require and dying. A stroke is a medical emergency that requires urgent medical attention. Early detection and appropriate management are required to prevent further damage to the affected area of the brain and other complications in other parts of the body. -e World Health Organization (WHO) estimates that fifteen million people worldwide suffer from strokes each year, with one person dying every four to five minutes in the affected population. Stroke is the sixth leading cause of mortality in the United States according to the Centers for Disease Control and Prevention (CDC) [1]. Stroke is a noncommunicable disease that kills approximately 11% of the population. In the United States, approximately 795,000 people suffer from the disabling effects of strokes on a regular basis [2]. It is India's fourth leading cause of death. Strokes are classified as ischemic or hemorrhagic.In a chemical stroke, clots obstruct the drainage; in a hemorrhagic stroke, a weak blood vessel bursts and bleeds into the brain. Stroke may be avoided by leading a healthy and

balanced lifestyle that includes abstaining from unhealthy behaviors, such as smoking and drinking, keeping a healthy body mass index (BMI) and an average glucose level, and maintaining an excellent heart and kidney function.

Stroke prediction is essential and must be treated promptly to avoid irreversible damage or death. With the development of technology in the medical sector, it is now possible to anticipate the onset of a stroke by utilizing ML techniques. -e algorithms included in ML are beneficial as they allow for accurate prediction and proper analysis. -e majority ofprevious stroke-related research has focused on, among other things, the prediction of heart attacks. Brain stroke has been the subject of very few studies. -e main motivation of this paper is to demonstrate how ML may be used to forecast the onset of a brain stroke. -e most important aspect of the methods employed and the findings achieved is that among the four distinct classification algorithms tested, Random Forest fared the best, achieving a higher accuracy metric in comparison to the others. One downside of the model is that it is trained on textual data rather than real time brain images. -e implementation of four ML classification methods is shown in this paper.

Numerous academics have previously utilized machine learning to forecast strokes. Govindarajan et al. [3] used text mining and a machine learning classifier to classify stroke disorders in 507 individuals. -ey tested a variety of machine learning methods for training purposes, including Artificial Neural Network (ANN), and they found that the SGD algorithm provided the greatest value, 95 percent. Amini et al. [4, 5] performed research to predict a stroke occurrence. -ey classified 50 risk variables for stroke, diabetes, cardiovascular disease, smoking, hyperlipidemia, and alcohol consumption in 807 healthy and unhealthy individuals. -ey used two of the most accurate methods: the c4.5 decision tree algorithm (95 percent accuracy) and the K-nearest neighbor algorithm (94 percent accuracy). Cheng et al. [6] presented a study on estimating the prognosis of an ischemic stroke. In their study, they used 82 ischemic stroke patient data sets, two ANN models, and the accuracy values of 79 and 95 percent. Cheon et al. [7–9] conducted research to determine the predictability of a stroke patient death. -ey identified the stroke incidence using 15,099 individuals in their research.-ey detected strokes using a deep neural network method. -e authors utilized PCA to extract information from the medical records and predict strokes. -ey have 83 percent area under the curve (AUC). Singh et al. [10] conducted research using artificial intelligence to predict strokes. -ey employed a new technique for predicting stroke in their research using the cardiovascular health study (CHS) dataset. Additionally, they used the decision tree method to do a feature extraction followed by a principal component analysis. In this case, the model was built using a neural network classification method, and it achieved 97 percent accuracy.

medical experts when used to categorize strokes. -e majority of studies had an accuracy rate of around 90%, which was considered to be quite good. However, the novelty of our research is that we used several well-known machine learning methods to get the best result. Random forest (RF), decision tree (DT), voting classifier (VC), and logistic regression (LR) were the most successful algorithms, with 96, 94, 91, and 87 percent F1-scores, respectively. -e accuracy percent of the models used in this research is much greater than the accuracy percent of the models used in previous investigations, suggesting that the models used in this investigation are more trustworthy. -ey have been shown to be resilient in many model comparisons, and the scheme may be generated from the results of the study's analysis.

2. RELATED WORK

One person dies from a stroke every four to five minutes, according to World Health Organization (WHO) estimates of the fifteen million individuals who experience them globally each year. Stroke is the sixth most common cause of death in the United States, according to the Centers for Disease Control and Prevention (CDC) [6]. About 11% of people die from non-communicable diseases like stroke each year. Approximately 795,000 Americans experience the incapacitating symptoms of strokes regularly [7]. To categorize stroke conditions in 507 people, Govindarajan et al. [8] employed text mining and a machine learning classifier. They investigated several artificial neural networks (ANN)- based machine learning techniques for training purposes and discovered that the SGD algorithm delivered the highest value, 95%. Research on stroke prognosis was conducted by Amini et al. [9]. Fifty

risk factors for stroke, diabetes, cardiovascular disease, smoking, hyperlipidemia, and alcohol consumption were categorized by them in 807 healthy and unhealthy people. The c4.5 decision tree algorithm and the K-nearest neighbor algorithm, both of which have accuracy rates of 95%, were employed by the researchers (94 percent accuracy). A study on determining the prognosis of an ischemic stroke was given by Cheng et al. [10]. They used two ANN models, 82 ischemic stroke patient data sets, and accuracy values of 79 and 95 percent in their study. To ascertain whether a stroke patient's death may be predicted, Cheon et al. [11] conducted research. In their study, they calculated the stroke incidence using 15,099 people. They used a deep neural network technique to find strokes. To extract data from the medical records and forecast strokes, the authors used PCA. They have an area under the curve of 83 percent (AUC). Artificial intelligence was used in the study by Singh et al. [12] to forecast strokes.

They used the cardiovascular health study (CHS) dataset in their research and applied a novel method for predicting stroke. Additionally, they performed a feature extraction followed by a principal component analysis using the decision tree method. In this instance, a neural network classification method was used to build the model, and it had a 97 percent accuracy rate. To ascertain the efficacy of automated early ischemic stroke detection, Chin et al. [13] conducted research. Their research's main goal was to develop a Convolutional Neural Network technique for automating primary ischemic stroke (CNN). For the goal of developing and testing the CNN model, the author gathered 256 images.

To increase the gathered picture for their system's image preprocessing, used the data lengthening technique. CNN method had an accuracy rate of 90%. The research was done by Sung et al. [14] to create a stroke severity index they collected information on 3577 people who suffered an acute ischemic stroke. They built their predictive models using several data mining techniques, including linear regression. The k-nearest neighbor algorithm fared worse than their ability to anticipate (95% confidence interval). Machine learning was utilized by Monteiro et al. [15] to forecast the functional outcome of an ischemic stroke. They used a patient who passed away three months after admission to test this procedure. They achieved an AUC value of over 90. The research was undertaken to ascertain the risk of stroke by Kansadub et al. The authors of the study used Naive Bayes, decision trees, and neural networks to analyze the data and predict strokes. In their study, they evaluated the accuracy and AUC of their pointer. All of these algorithms were characterized by them as decision trees, with naive Bayes producing the most precise outcomes. To identify the classification of an ischemic stroke, Adam et al. [16] undertook a study. They used the decision tree method and the k-nearest neighbor method to categorize ischemic strokes. Medical professionals found the decision tree method to be more helpful in their study when categorizing strokes. 90% accuracy was regarded to be a good accuracy rate for the majority of investigations.

With a classification accuracy of 96%, Khan et al. [17] use random forest classification that exceeds the other investigated techniques. According to the study, the random forest method performs better than other methods when forecasting brain strokes using cross-validation measures. In this study, similar to the previous studies, different machine learning algorithms were trained and the best performing model is selected for the used data set.

In addition to the above approaches, some researchers have proposed hybrid systems that combine multiple sensors and techniques for finger tracking and virtual mouse control. For example, Lee et al. (2018) proposed a hybrid system that combines a depth camera and a gyroscope sensor to provide more accurate and robust finger tracking and virtual mouse control.

3. METHODOLOGY

Data Description: The dataset used in this research was obtained from the Kaggle data repository. The total number of participants was 4981 among which 2074 were male and 2907 were female. The dataset has 10 attributes as input for machine learning models and one target class. The attributes gender, age, hypertension, heart_disease, ever_married, work_type, residence_type, avg_glucose_level, bmi, and smking_status are the main attributes that are used as input for the machine learning model. The attribute stroke is used as the output variable. The number '0' denotes the absence of any stroke risk, while the number '1' denotes the possibility of stroke risk. The dataset used in this research is highly imbalanced as it has 248 rows with a value of '1' whereas 4733 rows have a value of '0' in the stroke column. To

attain better accuracy, the data pre-processing technique is used to balance the dataset. The details of the dataset are described in table 1.

Attribute Name	Type (Possible Values)	Description
Gender	String(Male, Female)	Describes the gender of the participant
Age	Floating point number (0.08 to 82)	Age of the participant
Hypertension	Numeric (0, 1)	Participants hypertension status
Heart_disease	Numeric (0, 1)	Participant's heart disease status
Ever_married	Nominal (Yes, No)	Tells whether the participant is ever married or not.
Work_type	String (private, self-employed, govt_job, children)	Describes the nature of the work of the participants
Residence_type	Nominal (Urban, Rural)	Tells the residence type of the participants
Avg_glucose_level	Floating point number (55.12 to 271.74)	Shows the average glucose level of the participants.
Bmi	Floating point number (14 to 48.9)	Gives the body mass index of the participant
Smoking_status	String (formerly smoked, never smoked, smokes, unknown)	Shows the smoking status of the participant
Stroke	Numeric (0, 1)	Response variable which describes the stroke status

Data Preprocessing: The raw data might contain noise and/or missing values affecting negatively in the final prediction. Hence, preprocessing of the data is necessary. This stage deals with anything preventing the model from operating more effectively. Preprocessing of the dataset includes feature selection, values reduction, and discretization. The dataset taken for the research has 11 attributes including the response variable. Firstly the dataset is checked for null values and if occurs it is filled. After adjusting null values, the string values are converted to nominal as WEKA cannot process string values.

The dataset used for this study is very unbalanced. The entire dataset contains 4981 rows, among which 248 rows in the stroke column have the value '1' whereas 4733 rows have the value '0'. If such uneven data is not managed, the predictions and outcomes are ineffective. The undersampling technique was used in this research to handle the imbalanced distribution of data between the stroke and non-stroke classes. By implementing undersampling the majority class is undersampled to match the minority class. More specifically, the majority of class 'stroke' with value '0' was undersampled for the class 'stroke' with value '1' to equally distribute the participants. After implementing undersampling, the dataset contains 248 rows with value '0' and 248 rows with value '1' resulting in 496 total rows. Figure 1 displays a graphical depiction of the response variable before and after applying undersampling in the dataset.

The next step after handling the imbalanced dataset and finishing data preprocessing is to develop a model. The dataset obtained after under-sampling is split into training and testing data. In this research, we have used a 10-fold cross-validation technique to get a better result. In the 10-fold cross-validation technique, 90% of the total training dataset is randomly selected as training data and the remaining 10% data as test data. After splitting the dataset, we used classification algorithms to train the model. The different machine learning algorithms used to train the model are described below.

4. Conclusion

Stroke causes a significant number of fatalities and is increasing every day. Several stroke risk factors are responsible for different types of strokes. The design of a machine learning model can aid in the early detection of stroke and minimize its severe effects. These performance indicators of different machine learning algorithms used in this research show the successful prediction of stroke. Five machine learning algorithms such as Naïve Bayes, AdaBoost, Decision Table, k-NN, and Random Forest were used to detect the stroke. All these machine learning models were trained using WEKA environment to achieve the performance of the models. The result obtained after training all five models shows that Decision Table performs better than other methods. The model can be helpful for the clinical prediction of stroke in a better way. In the future, we can extend the research by applying multiple classification techniques and designing a framework to display the predicted results.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

References

[1] Worldwide Stroke affect, WHO AvailableLink:http://www.emro.who.int/healthtopics/stroke-cerebrovascularaccident/index.html,Last Access:[6-10-2021].

[2] G. Vijayadeep and N. N. M. Rao, "A hybrid feature extraction based optimized random forest learning model for brain stroke prediction, "Turkish Journal of Computer and Mathematics Education (TURCOMAT),vol. 11, no. 3, pp. 1152-1165, 2020.

[3] SJ. Park, I. Hussain, S. Hong, D. Kim, H. Park, and HC. Benjamin, "Realtime Gait Monitoring System for Consumer Stroke Prediction Service,"IEEE International Conference on Consumer Electronics (ICCE), pp. 1-4,2020.

[4] T. Badriyah, N. Sakinah, I. Syarif, and D. R. Syarif, "Machine Learning Algorithm for Stroke Disease Classification," International Conference on Electrical, Communication, and Computer Engineering (ICECCE), pp. 1-5, 2020.

[5] T. Liu, W. Fan, and C. Wu, "A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset," Artificial intelligence in medicine, vol. 101, pp. 101723, 2019.

[6] G. Fang, W. Liu, and L. Wang, "A machine learning approach to select features important to stroke prognosis," Computational Biology and Chemistry, vol. 88, pp.107316, 2020.

[7] P. Govindarajan, RK. Soundarapandian, AH. Gandomi, R. Patan, P.Jayaraman, and R.Manikandan, "Classification of stroke disease using machine learning algorithms," Neural Computing and Applications, vol.32, no. 3, pp. 817-828, 2020.

[8] M. Emona, M. Keya, T. Meghla, M. Rahman M, A. Mamun, and M. Kaiser, "Performance Analysis of Machine Learning Approaches in Stroke Prediction," 4thInternational Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, pp. 1464-1469, 2020.

[9] T. Shoily, T. Islam, and S. Jannat, "Detection of stroke disease using machine learning algorithms," 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, pp. 1-6,2019. 40

[10] M. Amin, Y. Chiam and K. Varathan," Identification of significant features and datamining techniques in predicting heart disease," Telematicsand Informatics, vol.36, pp.82-93, 2019

[11] S. Cheon, J. Kim, and J. Lim, "The use of deep learning to predict stroke patient mortality," International Journal of Environmental Research and Public Health, vol. 16, no. 11, 2019.

[12] M. S. Singh and P. Choudhary, "Stroke prediction using artificial intelligence," in Proceedings of the 2017 8th Annual Industrial Automation And Electromechanical Engineering Conference (IEMECON), Aug. 2017, pp. 158–161.

[13] C.-L. Chin, B.-J. Lin, and G.-R. Wu et al., "An automated early ischemic stroke detection system using CNN deep learning algorithm," in Proceedings of the 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), Nov. 2017, pp. 368–372.

[14] S.-F. Sung, C.-Y. Hsieh, and Y.-H. Kao Yang et al., "Developing a stroke severity index based on administrative data was feasible using data mining techniques," Journal of Clinical Epidemiology, vol. 68, no. 11, pp. 1292–1300, 2015.

[15] M. Monteiro, A. C. Fonseca, and A. T. Freitas et al., "Using machine learning to improve the prediction of functional outcome in ischemic stroke patients," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 6, pp. 1953– 1959, 2018.

[16] S. Y. Adam, A. Yousif, and M. B. Bashir, "Classification of ischemic stroke using machine learning algorithms," International Journal of Computer Application, vol. 149, no. 10, pp. 26–31, 2016.

[17] T. Tazin, M. N. Alam, N. N. Dola, M. S. Bari, S. Bourouis, and M. Monirujjaman Khan, "Stroke Disease Detection and Prediction Using Robust Learning Approaches," Journal of Healthcare Engineering, vol. 2021, 2021, doi: 10.1155/2021/7633381.

[18] T. I. Shoily, T. Islam, S. Jannat, S. A. Tanna, T. M. Alif, and R. R. Ema, "Detection of stroke disease using machine learning algorithms," in In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE., 2019, pp. 1–6. 41

[19] D.-C. Feng et al., "Machine learning-based compressive strength prediction for concrete: an adaptive boosting approach."

[20] G. Wets, J. Vanthienen, and S. Piramuthu, "Extending a tabular knowledge-based framework with feature selection," Expert Systems with Applications, vol. 13, no. 2, pp. 109–119, 1997

[21] J. Vanthienen and E. Dries, "Illustration of a decision table tool for specifying and implementing knowledge based systems," International Journal on Artificial Intelligence Tools, vol. 3, no. 2, pp. 267–288, 1994.

[22] G. Sailasya and G. L. Aruna Kumari, "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms." [Online]. Available: www.ijacsa.thesai.org

[23] A. Pandey and A. Jain, "Comparative Analysis of KNN Algorithm using Various Normalization Techniques," International Journal of Computer Network and Information Security, vol. 9, no. 11, pp. 36–42, Nov. 2017, doi: 10.5815/ijcnis.2017.11.04.

[24] "WEKA Tool," Available Online: https://www.weka.io/. [Accessed: August 8, 2022]