2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC) | 979-8-3503-0009-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICESC57686.2023.10193052

Early Prediction of Stroke using Machine Learning

Dr. G. Revathy
Assistant Professor III,
Department of CSE,
SRC, SASTRA DEEMED University,
Thanjavur.
revathyjayabaskar@gmail.com

Dr.U.Sesadri
Associate Professor,
Department of CSE,
Vardhaman College of Engineering,
Hyderabad
sesadri1601@vardhaman.org

Dr.Shaji.Theodore
Faculty of IT - Networking, Dept of IT,
University of Technology and Applied Sciences
Al-MUSANNA,
Sultanate of Oman,
theodore7733@hotmail.com

Mrs.J.Justina Princy Thilagavathy
Assistant Professor,
Department of CS,
Cherran's Arts Science College,
Kangeyam,
vinothilakviju@gmail.com

Mrs S. Senthilvadivu
Associate Professor,
Department of IT,
Arulmigu Meenakshi Amman College of Engineering,
Thiruvannamalai,
senthumukund@gmail.com

V.Senthil Murugan
Department of Networking and communications,
SRM Institute of Science and Technology,
Kattankulathur- 603203,
Tamilnadu, India.
senthilkmvs@gmail.com

Abstract.: **A stroke is the outcome of an abrupt cessation of blood flow to a region of the brain. Liable on the fragment of the brain that has been hurt, disability is caused by a loss of blood flow because brain cells gradually perish. In order to forecast stroke and maintain a healthy lifestyle, early symptom detection might be very beneficial. In order to afford a strong substance for the long-term peril prediction of stroke incidence, a quantity of replicas are developed and evaluated using machine learning (ML) in this study. The main contribution of this study is an ensemble, random forest, SVM, and XgBoost method that performs well and is validated by a variety of system of measurement, such as precision, recall, F-measure, and accuracy. According to the results of the experiment, random forest, XgBoost, SVM, and random forest classification have an accuracy of 96%, outperforming the other methods. Last but not least, it is recommended to take a number of preventative steps to lessen the risk of having a stroke, such as quitting smoking and abstaining from alcohol.**

Keywords—Machine learning, Stoke, Support Vector Machine, Random Forest, XgBoost, Linear Regression, Naïve Bayes, Decision Tree, Feature Extraction.

I. INTRODUCTION

The primary purpose of this procedure is to apply machine learning techniques to create a stroke prediction model [8-10]. Machine learning algorithms are useful in creating correct classifications and giving accurate assessments. The majority of previous stroke research has concentrated on heart stroke prediction; relatively little has been done on brain stroke[1-5]. The categorization of the occurrence of a brain stroke using machine learning is the foundation of this work. The most essential components of the techniques utilized, and Random Forest beat the other five classification approaches, earning a higher accuracy measure[6].

II. RELATED WORK

A retrospective investigation on a prospective database of acute ischemic stroke was undertaken in 2014 by Hamed Asadi et.al [3]. They have developed a well-known machine learning algorithm that can forecast whether endovascular treatment will be chosen over medical treatment for an acute stroke.

In order to accurately anticipate the occurrence of strokes, a new dataset is formed to examine these variables. The accuracy of the confusion matrix analysis used to determine the outcome was 95%.[4] To do this, they developed a ML methodology that assisted in a assessment of freshly fed data with survey data and using this assessment as a basis, the account was fashioned[15].

According to Tasfia et al comparison of various classification models.[11-13]. The proposed model will assist patients in determining whether or not they might experience a stroke trained with four different models.

In order to predict stroke, Joonet al. took into account three machine learning models, which reduces the need for simpler models.

The knowledge provided by Jaehak et.al says real-time variables to classify stroke severity into four categories. This information aids in predicting the potential timing of a stroke and associated handicap, allowing for the administration of additional drugs and the essential safety measures. Random Forest has a high accuracy of 88.9 percent, while Naive Bias has an accuracy rate of 85.4%[7][14].

III. PROPOSED WORK

A) PROBLEM STATEMENT

The main target of the projected model is to envisage the probability of stoke with the provided dataset.

The characteristics of those who are more prone than others to have a stroke are the first thing we're looking at the dataset, which was obtained from a freely available source, is subjected to a variety of classification techniques in order to predict the impending occurrence of a stroke. It has proven possible to use the random forest, xgb, and logistic regression methods to attain an accuracy of 96%.

This study's dataset is available (Kaggle)[5].

Importing the dataset:

Number of rows :5110
Number of columns: 12
Based on the parameters in the dataset Stroke is been predicted with the ML models.

B) METHODOLOGY

After reviewing the literature, seven distinct algorithms are chosen for the prediction. These seven algorithms—Decision Tree, Logistic Regression, Naive Bayes, Xgboost Random Forest, Support Vector Machine, and K Nearest Neighbour—are compared to one another.

Data Pre-processing:

Checking the NULL values
Dropping the NULL value and its column
Converting float values into int values
splitting dataset into 80:20

Training:80%
Test:20%
Feature extraction is done by removing unnecessary

FIGURE 1. Architecture Diagram

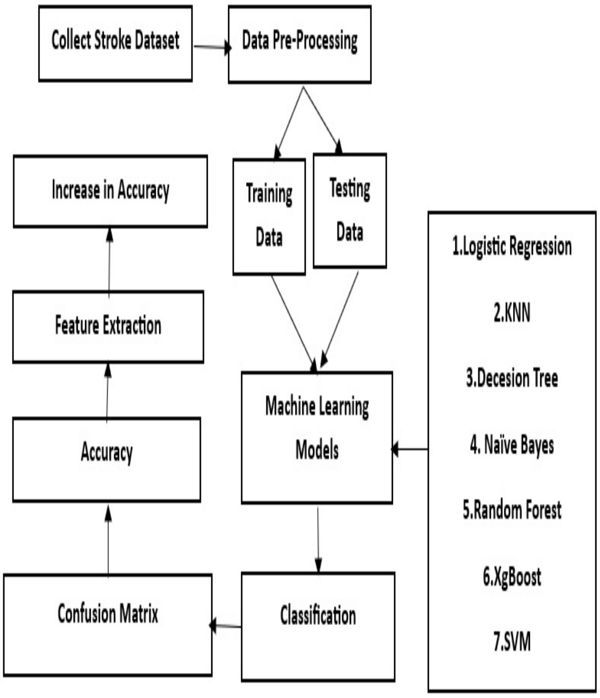

Figure 1 explains the workflow od the proposed methodology.

IV. RESULTS AND DISCUSSION

Machine learning models are automated procedures in which the computer does all of the data interpretation and analysis. Table 2 shows the feature significance of each parameter.

TABLE 2: FEATURE IMPORTANCE TABLE

Values	Importance
ever_married	0.017121
Heart_disease	0.23951
Hypertension	0.28203
Gender	0.32253
Residence_type	0.340346
Work_type	0.47666
Smoking_status	0.624044
Age	0.221072
BMI	0.24666
Avg_Glucose_level	0.286680

DND	LR	KNN	DT	NB	RF	XgB	SVM
Precision	0.96	0.96	0.96	0.95	0.96	0..96	0.96
recall	1.00	0.99	0.95	0.86	1.00	1.00	1.00
f1 score	0.98	0.97	0.95	0.90	0..98	0.98	0.98

TABLE 3: PERFORMANCE EVALAUTION WITHOUT FEATURE SELECTION

DND	LR	KNN	DT	NB	RF	XgB	SVM
Precision	0.00	0.11	0.02	0.02	0.00	0.00	0.00
recall	0.00	0.02	0.02	0.07	0.00	0.00	0.00
f1 score	0.00	0.04	0.02	0.03	0.00	0.00	0.00
Accuracy	96	95	97	82	96	96	96

TABLE 4: PERFORMANCE EVALAUTION WITH FEATURE SELECTION

Table 3 and Table 4 gives the result with and without feature selection. Based on the results we conclude that Random Forest performs better than all other compared ML models in early prediction of Stoke.

V. CONCLUSION

Many evaluations and classification models, including SVM, random forest, XgBoost, and logistic regression, demonstrated adequate accuracy in detecting stroke-prone individuals. In these difficult times, it is critical to be aware of and recognize the hazards of brain stroke. The model predicts the chance of a brain stroke based on extremely trivial daily elements that are known to all parameters. As a result, the project is both timely and vital for society. The future concept will be prepared for execution on a web platform in order to reach as many people as possible. Someone who is at danger of having a stroke can be spared if they receive an early warning.

REFERENCES

1. "Effective Analysis and Predictive Model of Stroke Disease using Classification Methods" - A. SUDHA, P. GAYATHRI, N. JAISANKAR.
2. Singh, M.S., Choudhary, P., THONGAM, K.: A comparative analysis for various stroke prediction techniques. In: Springer, Singapore (2020).
3. Potdar, Mrs Veena, Mrs Lavanya Santhosh, and Yashu Raj Gowda CY. "A Survey on Stroke Disease Classification and Prediction using Machine Learning Algorithms.".
4. Mohith, S., et al. "Development and assessment of large stroke piezo-hydraulic actuator for micro positioning applications." *Precision Engineering* 67 (2021): 324-338.
5. Shoily, Tasfia Ismail, et al. "Detection of stroke disease using machine learning algorithms." *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2019.
6. Heo, JoonNyung, et al. "Machine learning–based model for prediction of outcomes in acute stroke." *Stroke* 50.5 (2019): 1263-1265.
7. Yu, Jaehak, et al. "An elderly health monitoring system using machine learning and in-depth analysis techniques on the NIH stroke scale." *Mathematics* 8.7 (2020): 1115.
8. Dr G Revathy er.al, "MACHINE LEARNING ALGORITHMS FOR PREDICTION OF DISEASES", International Journal of Mechanical Engineering, volume 7, issue 1, January 2022.
9. Kansadub, T., Thammaboosadee, S., Kiattisin, S., Jalayondeja, C.: Stroke risk prediction model based on demographic data. In: 2015 8th Biomedical Engineering International Conference (BMEiCON), November 2015, pp. 1–3 (2015).
10. Freire, V.A., de Arruda, L.V.R.: Identification of residential load patterns based on neural networks and PCA. In: 2016 12th IEEE International Conference on Industry Applications (INDUSCON), November 2016, pp. 1–6 (2016).
11. P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman and R. Manikandan, "Classification of stroke disease using machine learning algorithms", Neural Computing and Applications, pp. 1-12.
12. R. Jeena and S. Kumar, "Stroke prediction using svm", 2016 International Conference on Control Instrumentation Communication and Computational Technologies (ICCICCT), pp. 600-602, 2016.
13. P. A. Sandercock, M. Niewada and A. Członkowska, "The international stroke trial database", Trials, vol. 13, no. 1, pp. 1-1, 2012.
14. S. Y. Adam, A. Yousif and M. B. Bashir, "Classification of ischemic stroke using machine learning algorithms", Int J Comput Appl, vol. 149, no. 10, pp. 26-31, 2016.
15. G. Kaur and A. Chhabra, "Improved j48 classification algorithm for the prediction of diabetes", International Journal of Computer Applications, vol. 98, no. 22, 2014.