IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

Brain Stroke Prediction using ML Techniques

DV Swetha Ramana1, Anshu Kumar Tiwari[2], Emmanuvel Daniel[3], L Vishal[4], Abhishek V[5]
Assistant Professor, Dept of CSE[1]
Students, Dept of CSE[2,3,4,5]
Rao Bahadur Y Mahabaleswarappa Engineering College, Ballari, , Karnataka, India

Abstract: *A Stroke is a health condition that causes damage by tearing the blood vessels in the brain. It can also occur when there is a halt in the blood flow and other nutrients to the brain. According to the World Health Organization (WHO), stroke is the leading cause of death and disability globally. Most of the work has been carried out on the prediction of heart stroke but very few works show the risk of a brain stroke. With this thought, various machine learning models are built to predict the possibility of stroke in the brain. This paper has taken various physiological factors and used machine learning algorithms like Logistic Regression, Decision Tree Classification, Random Forest Classification, K-Nearest Neighbors, Support Vector Machine and Naïve Bayes Classification to train five different models for accurate prediction. The algorithm that best performed this task is Logistic Regression that gave an accuracy of approximately 94.7%*

Keywords: Brain Stroke

I. INTRODUCTION

Stroke is the fifth-leading cause of death in the worldwide. Stroke is a non-communicable infection that is liable for around 11% of total deaths. It is the fourth significant reason for death in India. With the advancement of technology in the medical field, predicting the occurrence of a stroke can be made using Machine Learning. The algorithms present in Machine Learning are constructive in making an accurate prediction and give correct analysis. This paper is based on predicting the occurrence of a brain stroke using Machine Learning. The key components of the approaches used and results obtained are that among the five different classification algorithms used Naïve Bayes has best performed obtaining a higher accuracy metric. The limitation with this model is that it is being trained on textual data and

II. LITERATURE SURVEY

[1]. 'Stroke Prediction Using SVM' In this paper we were using Support Vector Machine for stroke prediction. This research work investigates the various physiological parameters that are used as risk factors for the prediction of stroke .Data was collected from International Stroke Trial database and was successfully trained and tested using Support Vector Machine (SVM). The objective of this work is to develop a machine learning based approach to predict the possibility of stroke in people having the symptoms or risk factors of stroke. In this work, we have implemented SVM with different kernel functions and found that linear kernel gave an accuracy of 90 %. The results were evaluated on a spectrum of patients of different age groups.

[2]. 'Stroke Prediction using Artificial Intelligence' The stroke deprives person's brain of oxygen and nutrients, which can cause brain cells to die. Numerous works have been carried out for predicting various diseases by comparing the performance of predictive data mining technologies. In this work, we compare different methods with our approach for stroke prediction on the Cardiovascular Health Study (CHS) dataset. Here, decision tree algorithm is used for feature selection process, principle component analysis algorithm is used for reducing the dimension and adopted back propagation neural network classification algorithm, to construct a classification model. The proposed method use Decision Tree algorithm for feature selection method, PCA for dimension reduction and ANN for the classification. The experimental results show that the proposed method has higher performance than other related well-known methods.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

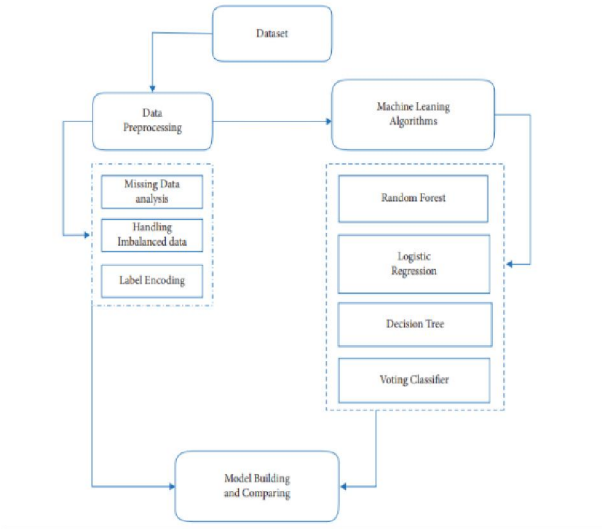
2.1 Problem Statement

Stroke is the second leading cause of death worldwide and remains an important health burden both for the individuals and for the national healthcare systems. Potentially modifiable risk factors for stroke include hypertension, cardiac disease, diabetes, and dysregulation of glucose metabolism, atrial fibrillation, and lifestyle factors.

Therefore, the goal of our project is to apply principles of machine learning over large existing data sets to effectively predict the stroke based on potentially modifiable risk factors. Then it intended to develop the application to provide a personalized warning on the basis of each user's level of stroke risk and a lifestyle correction message about the stroke risk factors.

III. PROPOSED SYSTEM

The proposed strategy focuses on a novel machine learning procedures for Stroke prediction, thus overcoming the existing problem. Different machine learning methods may not perform equally on the same feature set. Therefore, optimal feature sets for each machine learning methods were defined systematically. By utilizing multiple machine learning algorithms the model can be used in order to increase the performance and accuracy. jupyter was used to develop the proposed system. The proposed model describes the approach taken to develop the proposed solution which entails pre-processing, feature reduction and the final classification. The collected dataset is given as the input which is preprocessed to remove unwanted rows and columns to produce modified dataset. The data modified is analyzed and is compared to people with stroke and without stroke and the dataset is split into train and test data. The test data is used to get the predictions to produce accuracy report.



IV. METHODOLOGY

- Dataset will be collected and data will be cleaned and inconsistency will be removed.
- The most suitable algorithm will be selected from five machine learning algorithms and will be applied.
- The machine will be trained from the cleaned data as 80% data will use for training and 20% for testing.
- The pycharm or VS code software will be used
- After model building will be done. The user interface will be created in html.
- The machine learning algorithms will be pickle from the jupyter.
- The data of patient will be entered on application.
- The predicted output from given data will be shown by interface after machine learning algorithms will be applied.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

V. EVALUATION OF MODELS

Random Forest (RF)

The classification report for the RF model. In this case, the total F1-score obtained is 94.6 %. Prior to fine-tuning, the model had an accuracy of 92 percent. the random forest model's prediction. predicted outcome and the model's calculated performance are shown.

Decision Tree

The classification report for the decision tree classification is shown in figure, final score in this case is 90.5%. Also, the precision and recall are shown in Figure A fine-tuned decision tree model has also been implemented. However, after fine-tuning, the accuracy did not improve.

Logistic Regression

The classification report for the decision tree classification is shown in figure, final F1-score in this case is 94.7%.

KNN

The classification report for the decision tree classification is shown in figure, final F1-score in this case is 94.5% percent.

SVM:

The classification report for the decision tree classification is shown in figure, final F1-score in this case is 94.7%.

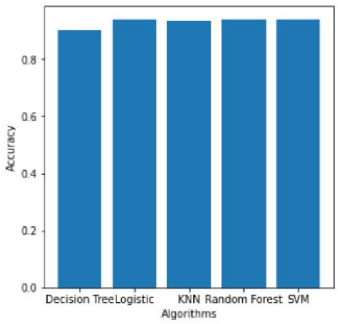


Fig: Accuracy of algorithms

5.1 Hardware Requirement

- System-Pentium 1V 2.4 GHz.
- Hard Disk -500 GB
- Ram -4 GB
- Any system with above configuration or higher level

5.2 Data Availability

The data utilized to support this research findings are accessible online at https://www.kaggle.com/fedesoriano/ stroke-prediction-dataset.

VI. CONCLUSION

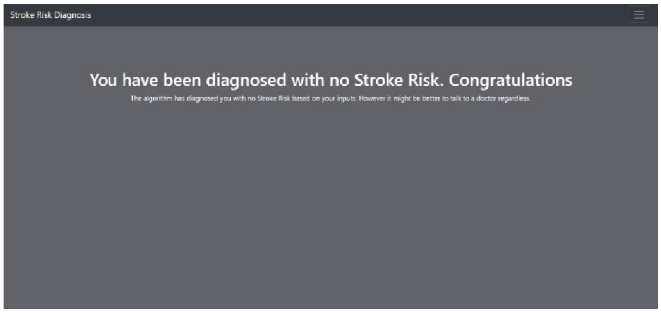
Stroke is a life-threatening medical illness that should be treated as soon as possible to avoid further complications. development of an ML model could aid in the early detection of stroke and the subsequent mitigation of its severe consequences. effectiveness of several ML algorithms in properly predicting stroke based on a number of physiological variables is investigated in this study. Random forest classification outperforms the other methods tested with a classification accuracy of 94.7%. According to the research, the random forest method outperforms other processes

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

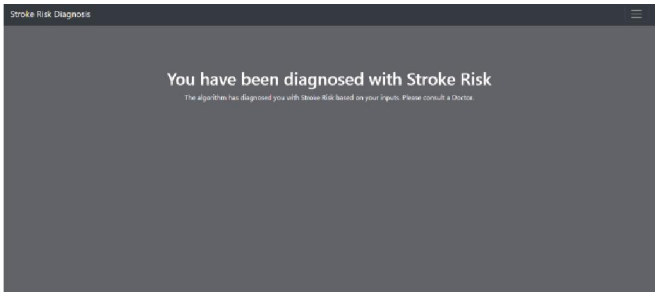
when cross-validation metrics are used in brain stroke forecasting. future scope of this study is that using a larger dataset and machine learning models, such as AdaBoost, SVM, and Bagging, the framework models may be enhanced. In exchange for just providing some basic information, the machine learning architecture may help the general public in determining the likelihood of a stroke occurring in an adult patient. In an ideal world, it would help patients obtain early treatment for strokes and rebuild their lives after the event.

VII. SNAPSHOTS

Snapshot 2: Home page



Snapshot 3: No Stroke page



Snapshot 4: Stroke page

REFERENCES

[1]. Singh, M.S., Choudhary, P., Thongam, K.: A comparative analysis for various stroke prediction techniques. In: Springer, Singapore (2020)

[2]. Pradeepa, S., Manjula, K. R., Vimal, S., Khan, M. S., Chilamkurti, N., &Luhach, A. K.: DRFS: Detecting Risk Factor of Stroke Disease from Social Media Using Machine Learning Techniques. In Springer (2020).

[3]. Dataset named „Stroke Prediction Dataset" from Kaggle: https://www.kaggle.com/fedesoriano/strokeprediction-dataset

[4]. Wang, S., Li, Y., Tian, J., Peng, X., Yi, L., Du, C., Feng, C., Liang, X. (2020). A randomized controlled trial of brain and heart health manager-led mHealth secondary stroke prevention. Cardiovascular Diagnosis and Therapy, 10 1192-1199. https://doi.org/10.21037/cdt-20-423

[5]. Wilkinson, D.A., Daou, B.J., Nadel, J.L., Chaudhary, N., Gemmete, J.J., Thompson, B.G., Pandey, A.S. (2019). Abdominal aortic aneurysm is associated with subarachnoid hemorrhage. Journal of Neuro Interventional Surgery. https://doi.org/10.1136/neurintsurg2020-016757

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 10, May 2023

[6]. JEENA R S, Dr. Sukesh Kumar , " Stroke Prediction Using SVM" , 2016 International Conferecnce on Control Instrumentation, Communication and Computational Technologies (ICCICCT),978-1-5240-076/$31.00@2016IEEE, https://ieeexplore.ieee.org/document/8079581

[7]. M. S. Singh and P. Choudhary, "Stroke prediction using artificial intelligence," 2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), 2017, pp. 158- 161,doi: 10.1109/IEMECON.2017.8079581. https://ieeexplore.ieee.org/document/8079581