

Predicting Brain Stroke Using Supervised machine learning

Mr.M.THIRUNAVUKKARASU*,PADHARTHI VYSHNAVI**,K.GNANA THANUJA***

Assistant Professor*,BE Graduate**,BE Graduate***

*(Department of CSE,SCSVMV University,kanchipuram,Enathur)

***(Department of CSE,SCSVMV University,kanchipuram,Enathur)

****(Department of CSE,SCSVMV University,Kanchipuram,Enathur)

Date of Submission: 10-04-2023

Date of acceptance: 25-04-2023

Abstract—In many nations, stroke is the main cause of endomorph and death. The goal of this research was to figure out how to make things better. I used the stroke disease data set from Kaggle. Patients can be benefit from data that has been pre-processed. Ischemic stroke and stroke haemorrhagee are two forms of stroke, Individuals are divided into two categories using machine learning methods. Machine learning techniques were employed seven times in this investigation. Logistic Regression, Support Vector Machine (SVM), Random Forest, Cat Boost,Multi-layer Perceptron(MLP), Naive Bayes, K-Nearest Neighbors Because of this, our findings, Cat Boost makes the best accuracy, along with precision and recall values, and the f1-Score.

Keywords—Accuracy, Data preprocessing, Machine Learning, Prediction,Stroke

I. INTRODUCTION

When a blood vessel bleed or blockage lowers or stops the flow of blood to the brain, a stroke ensues. This causes the brain to receive less oxygen and nutrients, which damages brain cells begin to deteriorate. A cerebrovascular condition is stroke. Hence, it affects the blood vessels that supply oxygen to the brain. Insufficient oxygenation of the brain could lead to damage. Medical help is needed immediately. Many strokes are curable; however, some might be fatal or result in disability. A blocked or constricted artery causes an ischemic stroke. Restoring a sufficient blood supply to the brain is typically the main goal of treatment. Drugs that dissolve clots and stop others from forming are taken as part of the initial treatment. Patients with stroke are divided into ischemic and haemorrhagic groups.

A blood Vessel occlusion occurs most frequently the ischemic stroke's cause. While hemorrhagic stroke results from bleeding within the brain tissue. [1]

Now-a-days stroke is becoming the second largest death worldwide due to its causes Patients with stroke are divided into ischemic and hemorrhagic groups.

A blood vessel occlusion occurs most frequently the ischemic stroke's cause. While hemorrhagic stroke results from bleeding within the brain tissue.[2] The layer that pools and an entire connected layer, two convolutional layers, are used in the CNN architecture. Down-sampling, the main goal of this pooling layer is to lay off the quantity of data and parameters to reduce the possibility of overfitting. According to the results of classifying a dataset of 15,750 medical images using the MLP method, the greatest accuracy was 88.55 percent. 6,910 people were distributed among the begin class, and 8.84 percent were among the negative class. Blood leaking into the brain can result in hemorrhagic strokes.

Stopping the bleeding and relieving the pressure on the brain are the two main objectives of treatment. The initial step in treatment is typically to start taking drugs to lower brain pressure, regulate blood pressure generally, stop seizures, and stop any sudden blood vessel constriction.

I.HEADINGS

Literature survey is represented in section II . Problem statement illustrated in section III . Proposed system is described in section IV .Architecture is represented in section V as well as Algorithms and Implementation. Section VI contains Results And Outputs. Section VII consists of Conclusion And References.

II. LITERATURE SURVEY

“S.Keerthana[6].,” Proposed that Prediction of the stroke with the help of MRI scans and Histogram Based method.

“C.L.Chin et al[8].,”Proposed that, after cancer and heart disease, cerebrovascular diseases are the third leading cause of death worldwide. One of the most prevalent cerebral vascular disorders is brain haemorrhage.Brain heamorrhage is caused by the eruption of brain thruway leading to bleeding and can have a fatal impact on brain function and its performance.

“BadriyahTessy”[9] proposed that we can predict the stroke with the help of CT scan by improving image quality with the help of machine learning.

“M Sheetal Singh [11]” proposed that stroke prediction is done with cardiovascular health dataset using neural network classification network .

“Harshitha KV[12]” Proposed that We can predict the stroke with the help of user friendly application interface.

“Chetan Sharma[13]” proposed that the prediction of stroke is done with the help of datamining and determines the reduce of stroke.

“Gagana[14]” proposed that the Identification of stroke id done by using Brain CT images with the help of typical methods using Matlab.

“EdigaJyothsna[15]” Proposed that Deep learning technique is a time-consuming process for finding stroke and doesn’t provide accurate results.

III. PROBLEM STATEMENT

Now-a-days stroke is becoming the second largest death world wide due to its causes, In this project we areusing machine learning algorithms. In this project we are using machine learning algorithms to predict the strokes with the help of patients data. classification of stroke whether the person is having stroke or not with details or symptoms by using those algorithms and predicting which algorithm gives the highest accuracy in those algorithms

IV. PROPOSED SYSTEM

Here in the proposed system, For finding stroke disease classification, machine learning techniques are K-Nearest Neighbors(KNN),Support Vector Machine (SVM), & other machine learning algorithms like Random Forest, Cat Boost, Logistic Regression, Naive Bayes, and Multi-layer Perceptron.

Cat Boost has got best accuracy using in now a days that can improve the accuracy to finding out the stroke disease.

V.ARCHITECTURE

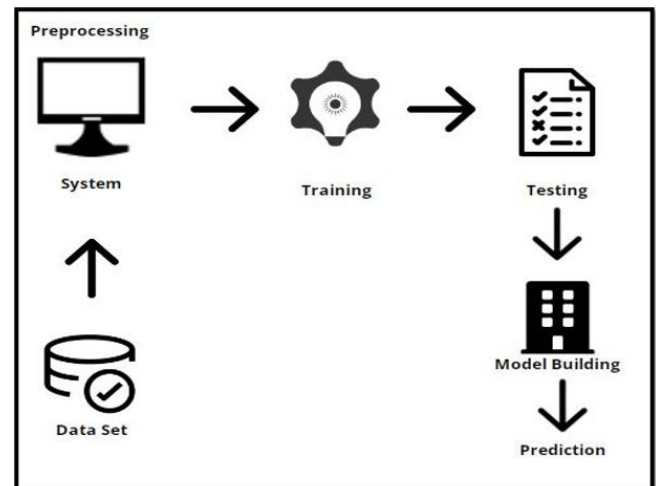


Fig 1 – System Architecture

First the user uploads the dataset and also can view the dataset and then it under goes testing and training of data and then model building and predict the stroke type and visualize in graphs

System:

Train data set:

System trains the data in dataset

pre-processing:

Pre-processing will be done using PCA module

model performance:

To evaluate a classification uses three main metrics model are accuracy, precision, and recall.

predictions:

Using the machine leaning algorithms, we can prognosticate the result.

IMPLEMENTATION

First, we took the hemorrhage diagnosis of CSV file.

- Load Dataset and view the dataset
- Here split the data in to train data and test data.
- Now, apply algorithms to the trained data.
- We got the best accuracy score for Cat Boost.
- Later, the total work is done with flask framework
- User can view the home, about , Upload Dataset, View Dataset, Splitting, ModelPerformance, Prediction And Graph

ALGORITHM

Random Forest:

A well-known example of machine learning programme that makes utilization of supervised learning method is called Random Forest. In order to handle regression and classification issues, this algorithm is used. This strategy is founded on the idea of ensemble learning, which can be used to combine different classifiers to categorise a challenging issue and enhance the model's performance.

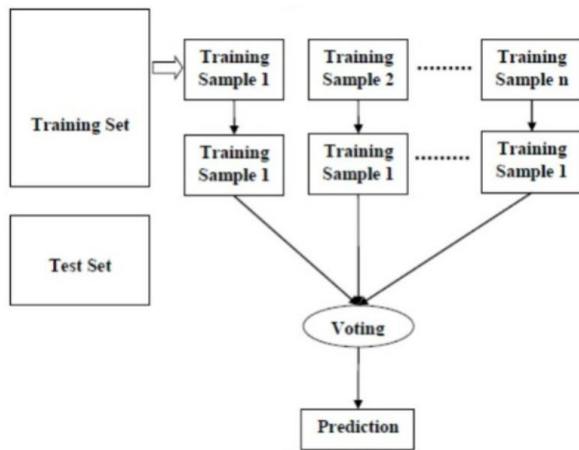


Fig 2 Random Forest

CatBooster Algorithm:

A high-performing open-source tool for gradient boosting on decision trees is the Cat Booster algorithm. This method boosts the gradient of decision trees. The Yandex and researchers created the Cat Booster Algorithm, which is utilised for many different applications, including Yandex and other companies like CERN, Cloudflare, and Careem taxi offer search, recommendation systems, personal assistants, self-driving cars, weather forecasting, and many other services. Because the source is open source, anyone can use this application programme. This new algorithm type, which has only been around for a little over a year, is already posing a danger to XG Boost and Light GBM. This algorithm gets the benchmark's best results, which is fantastic. Even though categorical features are important to the dataset, this improvement is substantial and unquestionable.

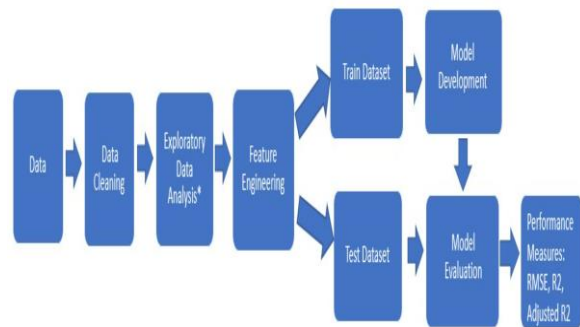


Fig 3: CatBooster

Multi Layer Perception:

There are multiple primary layers in the multilayer perceptron (combinations of neurons). If we use the three-layer network as a basic example, The output layer will now be referred to as the retired layer, while the first layer will continue to be the input layer, the middle layer will be the last. The input layer can receive our raw data, and the output layer can provide us with the output. To make the model more complex in accordance with our task, we are free to raise the number of retired layers as much as we like.

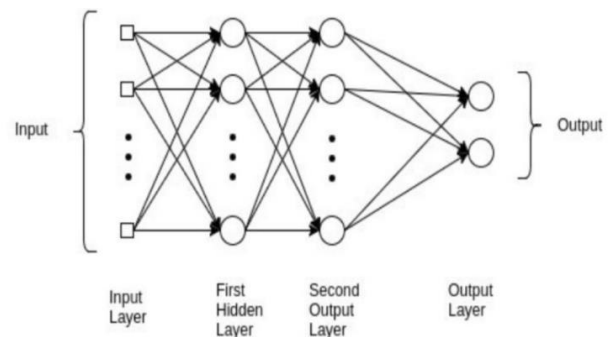


Fig 4:MLP

Logistic Regression:

Logistic regression falls under the domain of supervised learning. One of the most popular machine learning algorithms is supervised learning. This machine learning approach predicts the category dependent variable from a predefined set of independent variables. An outcome of a categorical dependent variable is predicted by this kind of algorithm. However, The outcome must be either a discrete value or a categorical value. Instead of exact values between 0 and 1, it provides probabilistic values that fall between those numbers. There are two possible outcomes: True or False, Yes (1) or No (0), etc.

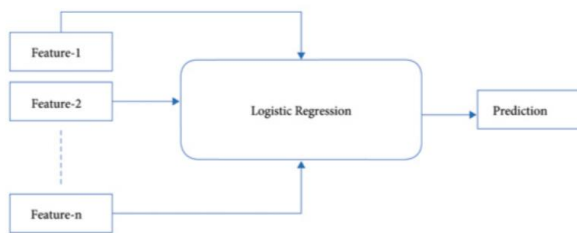


Fig 5: Logistic Regression

KNN:

KNearest Neighbor is one of the machine learning algorithms that is simple to grasp and that uses the supervised learning approach. This kind of algorithm places the new instance in a category that can match the current categories on the assumption that the data from existing cases and new cases can differ. It keeps all the current data and uses it to categorise a new data point based on similarity. This indicates that this algorithm can rapidly categorise newly generated data into appropriate groups. Although the K-NN approach can be applied to classification and regression problems, classification problems are where it is most frequently used. Because this approach is non-parametric, it does not make any assumptions about the underlying data.

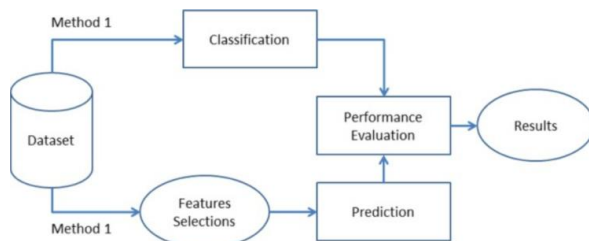


Fig 6: KNN

SVM:

One of the most well-known supervised learning methods, Support Vector Machine or SVM, is used for Regression and Classification issues. However, this algorithm main application that treats Categorization issues in machine learning. With the help of this algorithm, we can quickly kept new data points in the appropriate section then hereafter by creating the decision boundary or best line that can be divided into classes in n-dimensional space.

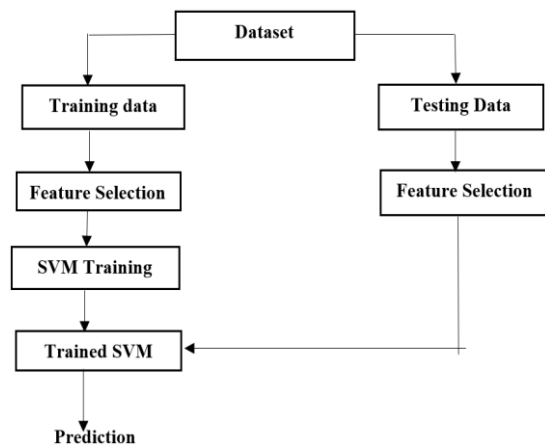


Fig 7: SVM

Naïve Bayes:

One supervised machine learning method, naive Bayes algorithm, make use of the Bayes theorem to clarify the classification issues. The primary implementation of this method is classification of text with a high-dimensional training set. One of the simplest and best Classification algorithms for creating quick machine learning models that may make immediate predictions in the naive bayes classifier. As a probabilistic classifier, these makes predictions based on the possibility then that an item exists. Spam detection, crime analysis, and false predictions are some instances of Naive Bayes algorithms are constantly used.

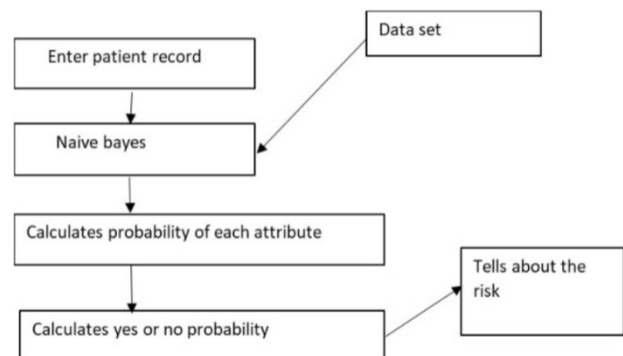


Fig 8:Naïve Bayes

VI.RESULTS AND OUTPUTS

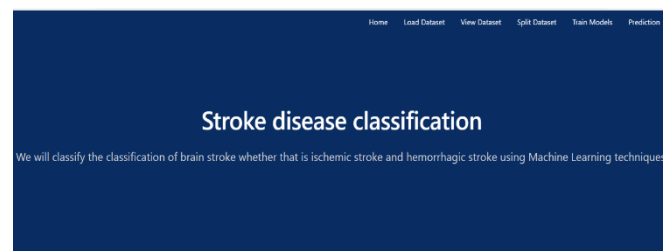


Fig 9: Home Page

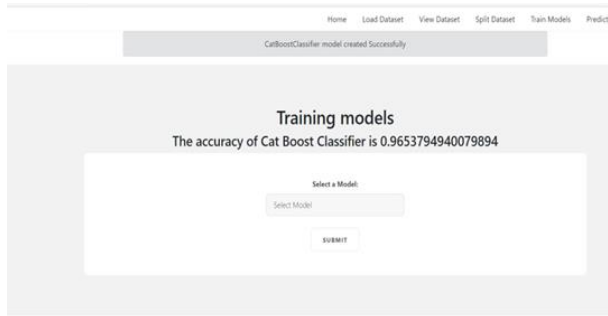


Fig 10: Training Models

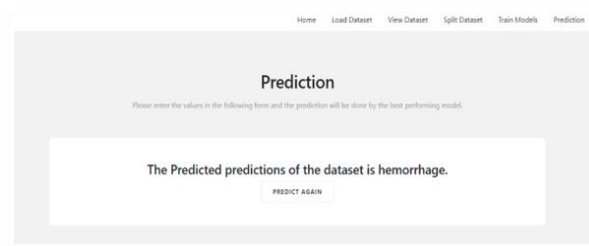


Fig 11: Predicting the type of stroke

Algorithm	Accuracy	Precision	Recall	F1 Score
Logistic regression	0.78	0.61	0.72	0.60
Naïve bayes	0.81	0.67	0.67	0.64
Random forest	0.96	0.85	0.94	0.88
KNN	0.94	0.70	0.97	0.78
MLP	0.925	0.04	0.50	0.07
SVM	0.72	0.56	0.90	0.58
Cat booster	0.97	0.86	0.96	0.90

Table 1: Algorithms comparison

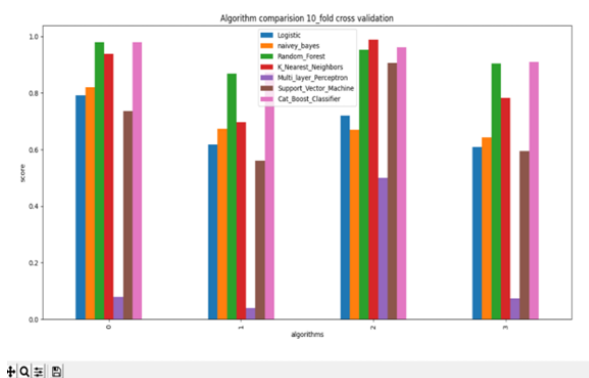


Fig 12: Algorithm Comparison 10_fold cross validation

VII. CONCLUSION

In this, we classify the stroke data using machine learning algorithms. And then the comparability of these eight algorithms are used to do classification, namely: Logistic Regression, K-Nearest Neighbors, Naive Bayes, Decision Tree, Random Forest, Multi-layer Perceptron, Deep Learning and Support Vector Machine and Cat Boost. According to our experiment, the Cat Boost method was used as a classification algorithm in comparison to other evaluated classification algorithms. Cat boost approach offers the best level of accuracy. The accuracy of the classification algorithm with the default optimisation parameter value has not been tested, nevertheless. The categorization model has the potential to be improved today. To increase the accuracy of the machine learning algorithm, parameter adjustment is required. More patients than patients with brain haemorrhages experienced an ischemic stroke.

REFERENCES

- [1]. Jenna R.S, Sukesh Kumar, (2019) 'Stroke Prediction Using SVM', International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT).
- [2]. N. Venketasubramanian, B. W. Yoon, J. Pandian, and J. C. Navarro, (2018) 'Stroke Epidemiology in South, East, and South-East Asia': A Review, vol. 20, no. 1.
- [3]. V. L. Feigin et al., (2018) 'Update on the global burden of ischemic and hemorrhagic stroke in 1990-2013': The GBD 2013 study, vol. 45, no. 3.
- [4]. S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, (2018) 'A gentle introduction to bilateral filtering and its applications'.
- [5]. J. T. Marbun, Seniman, and U. Andayani, (2018) 'Classification of stroke disease using convolutional neural network', vol. 978, no. 1.
- [6]. S. Keerthana (2016) 'Prediction of brain stroke using deep learning' IJRET vol-05, Issue-10.
- [7]. C. Y. Hung, W. C. Chen, P. T. Lai, C. H. Lin, and C. C. Lee, (2017) 'Comparing deep neural network and other machine learning algorithms for stroke prediction in a large population based electronic medical claims database'.
- [8]. C. L. Chin et al., (2019) 'An automated early ischemic stroke detection system using CNN deep learning algorithm', vol. 2018-Janua, no. iCAST.
- [9]. Badriyah Tessy, et al., (2019) 'Stroke Diagnosis using Deep Learning', International Journal of Advances in Intelligent Informatics, vol 5, n. 3, p.256272.

- [10]. Maria Sultana Keya et.al(2020) 'Performance Analysis of machine learning approaches in stroke prediction' ICEA pp. 1464-1469.
- [11]. M. Sheetal singh et.al (2017) 'Stroke prediction using Artificial intelligence' IEMECOM pp.158-161.
- [12]. Harshitha K V et.al (2021) 'Stroke Prediction Using Machine Learning' IJIREM ISSN:2350-0577,Vol8,Issue-4.
- [13]. Chetan Sharma (2022) 'Early stroke prediction using Machine Learning' Research gate, pp.890-894.
- [14]. Gagana (2021) 'Stroke Type Prediction using Machine Learning and Artificial Neural Networks' IRJET,vol-08,pp-06.
- [15]. EdigaJyothsna (2022) 'Prediction Of Brain Stroke Using Machine Learning' IJRES,Vol-10,Issue-5.