ISSN No : 2454-4221(Print)
ISSN No : 2454-423X(Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

Brain Stroke Severity Analysis Using Intelligent Methods

Sreenubabu dasari*, J.Neelima, T.Swapna, Y.Haritha, J.Gayatri.
Department Of Computer Science and Engineering.
Aditya Institue Of Technology and Management,Tekkali.

Abstract: The arteries inside and leading to the brain are affected by the condition known as stroke. A blood vessel that supplies the brain with nutrients and oxygen becomes blocked by a clot or bursts, resulting in a stroke. Stroke is the second most common cause of death worldwide, according to the WHO. Subarachnoid haemorrhage affects 3% of people worldwide, intracerebral haemorrhage 10%, and ischemic stroke 87% of the time. Since these strokes are avoidable 80% of the time, it's crucial to educate people about the warning symptoms of a stroke. The ability of the current studies to identify risk factors for different types of strokes is limited**.** Using the use of various machine learning techniques and the occurrence of hypertension, body mass index, average glucose level, smoking status, prior stroke, and age, this research work suggests an early prediction of stroke disorders. On a scale from 0 to 3, the severity of future stroke incidence is predicted using machine learning techniques including Logistic Regression, Random Forest, Decision Trees, Naive Bayes, SVM, MLP, etc. The study not only forecasts a person's future risk of having a stroke if they have never had one, but it also forecasts their future chance of developing a more serious variety of stroke if they have already experienced one.

Keywords: Machine Learning, Brain Stroke, AdaBoost, Gradient Boost, XG Boost, Severity, Risk.

1.Introduction

One of the common illnesses that today can shorten a person's lifespan is a brain stroke. In 2018, stroke was the cause of 1 in 6 fatalities from cardiovascular disease. Every 40 seconds, someone in the US suffers a stroke. A stroke claims the life of someone every 4 minutes. In the US, there are more than 795,000 stroke victims annually. Of these, about 610,000 are new or first strokes. A major factor in serious long-term impairment is stroke. More than half of stroke survivors aged 65 and older experience reduced mobility. In medical centres, the study of various data sets sometimes involves the use of data mining techniques and machine learning algorithms. The approaches and algorithms can be applied directly to a dataset to build models or make important deductions and inferences from it.

Stroke is defined as "rapidly developing clinical evidence of focal (or global) impairment of brain function, with symptoms lasting 24 hours or longer, or leading to death, with no evident cause other than of vascular origin" by the World Health Organization in 1970. A stroke happens when blood flow to various parts of the brain is disturbed, depriving those cells of nutrients and oxygen, which causes them to begin to degenerate.

Cite this article as: Sreenubabu dasari, J.Neelima, T.Swapna, Y.Haritha & J.Gayatri, "Brain Stroke Severity Analysis Using Intelligent Methods", International Journal of Research in Advanced Computer Science Engineering, (IJRACSE), Volume 8 Issue 11, April 2023, Page 54-60.

ISSN No : 2454-4221(Print)
ISSN No : 2454-423X(Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

Stroke is a medical emergency requiring quick care.

Early identification can help prevent additional health issues in the body and limit further damage to the damaged parts of the brain. There are primarily two types of strokes. An ischemic stroke occurs when the blood flow between blood tissues declines. A hemorrhagic stroke, on the other hand, happens when there is internal bleeding within the brain's tissues. It is possible to forecast the likelihood and severity of a stroke in a human using machine learning techniques. The ability of the current studies to foretell whether a stroke will occur or not is limited. Our research aims to predict the risk of stroke using a scoring system based on the following factors: Danger levels range from 0 (low risk) to 3 (severe risk). In contrast to the binary classification carried out by the majority of authors earlier, this classification uses many classes. To predict the risk of stroke, we have employed characteristics such as hypertension, body mass index level, average glucose level, smoking status, previous stroke severity (Nihss score), age, and gender.

2. Literature Survey

"Prediction of brain stroke severity by using machine learning'' by Vamsi Bandi, et al.(2020) have been used machine learning techniques to predict the severity of brain stroke.The study used a Random forest algorithm to predict the modified ranking scale (MRS) score, which is used to access functional outcome after stroke. The study used SPR model UI for predicting risk factors. The study involves Stroke_analysis (.CSV file) dataset to train the model.The result results showed that the random forest model had an accuracy of 93% in predicting the stroke severity **[1] .**

"Stroke risk prediction with machine learning techniques" by Elias Dritsas**.** (2022)
The study analyzed by Elias Dritsas used machine learning techniques to predict stroke risk. The study used features such as age, hypertension, bmi, avg_glucose_level, gender, smoking status to train the model. The results showed that the stacking model had an accuracy 92% **[2].**

"Analysing the performance of stroke prediction using ml classification algorithms by Gangavarapu sailasya**,** et al(2021) .The study by Gangavarapu sailasya used the Naïve bayes classification algorithm to predict the severity of stroke from stroke dataset. The study used features such as age, hypertension, bmi, avg_glucose_level, gender, smoking status to train the model. The result showed that the Naïve bayes classification model had an accuracy of 82% in predicting the stroke severity **[3].**

"Computer Methods and Programsin Biomedicine**"** by Jae–woo Lee, Hyun-sun Lim, Dong-wook Kim, Soon-aeShin, Jinkwon Kim, Bora Yoo, Kyung-hee Cho – The Purpose of this paper was Calculation of 10-year stroke prediction probability and classifying the user's individual probability of stroke into five categories **[4].**

"Detection of brain stroke using machine learning algorithm" by KD Mohan Sundaram, G. Haritha, A. Abhilash, K. Sona, E. Divyasri C. Bharat kumar _ The Purpose of paper was to predict the stroke by using machine learning techniques and flask. The study used Random forest, Decision tree, SVM, KNN techniques **[5].**

ISSN No : 2454-4221(Print)
ISSN No : 2454-423X(Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

"Predicting Stroke Outcome Using XGBoost Machine Learning Algorithm" by Shahbazi et al. (2021). The study by Shahbazi et al. used the XGBoost machine learning algorithm to predict the severity of stroke in 213 patients. The study used features such as age, sex, time from symptom onset to treatment, and imaging data to train the model. The results showed that the XGBoost model had an accuracy of 0.86 in predicting stroke severity **[6].**

"Prediction of Acute Ischemic Stroke Severity on Non contrast Computed Tomography" by Dhar et al. (2021) . Dhar et al. used machine learning techniques to predict the severity of acute ischemic stroke on noncontrast computed tomography (CT) images. The study used a convolutional neural network (CNN) to extract features from the CT images and a support vector machine (SVM) to predict the severity of the stroke. The results showed that the SVM model had an accuracy of 0.76 in predicting the severity of stroke **[7].**

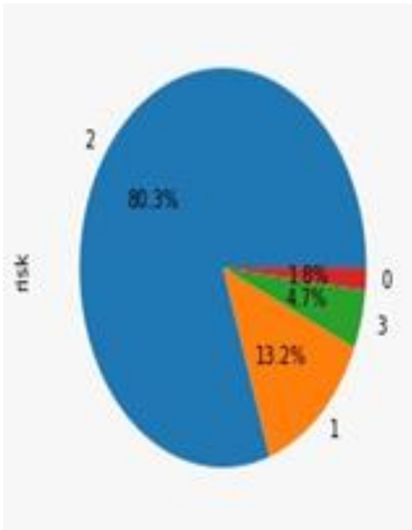
"Predicting Stroke Severity and Outcome Using Deep Learning and Random Forests" by Li et al. (2019) Li et al. used deep learning and random forests to predict stroke severity and outcome. The study used a CNN to extract features from MRI images and a random forest algorithm to predict the severity and outcome of stroke. The study used demographic, clinical, and imaging data to train the model. The results showed that the random forest model had an accuracy of 0.81 in predicting stroke severity and an accuracy of 0.79 in predicting stroke outcome **[8].**

"Machine Learning Approaches for Predicting Stroke Severity: A Systematic Review and Meta-Analysis" by Wang et al. (2018). Wang et al. conducted a systematic review and meta-analysis of machine learning approaches for predicting stroke severity. The study analysed 25 studies that used machine learning techniques to predict stroke severity. The results showed that machine learning models had an overall accuracy of 0.81 in predicting stroke severity **[9]**.

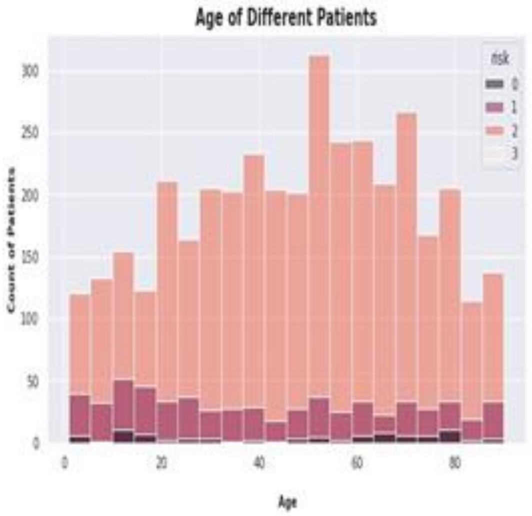
3. Methodology

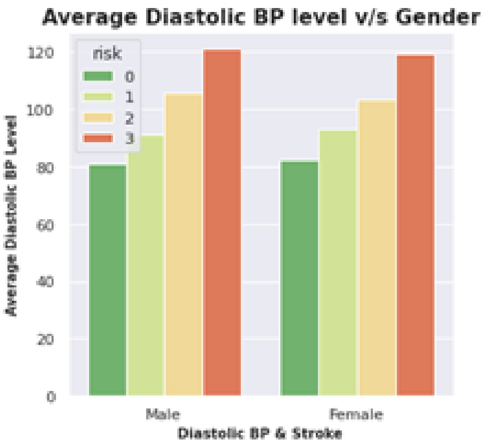
3.1. Dataset

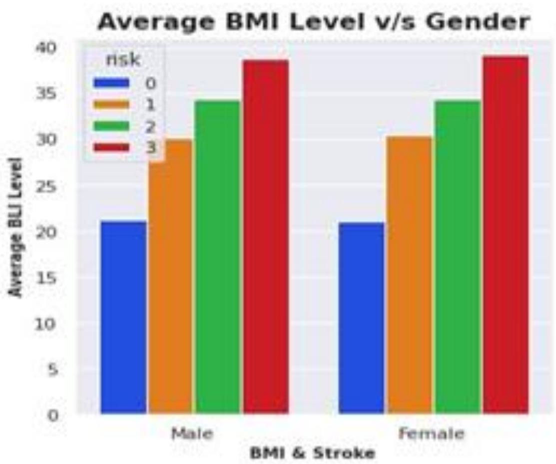
We obtain the data set from a website such as Kaggle, which includes text data. This data set contains different attributes such as pid, age, glucose level, bmi, gender, diastolic, systolic, smoking, tos, paralysis and risk. The size of the dataset is 4798 rows and 15 columns. Here, Risk is the target label in this instance, and its values range from 0 to 3. 0 represents Low risk, 1 represents Moderate risk, 2 represents Higher risk, and 3 represents Severe risk. This data set was then passed to the model for training.



ISSN No : 2454-4221(Print)
ISSN No : 2454-423X(Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com







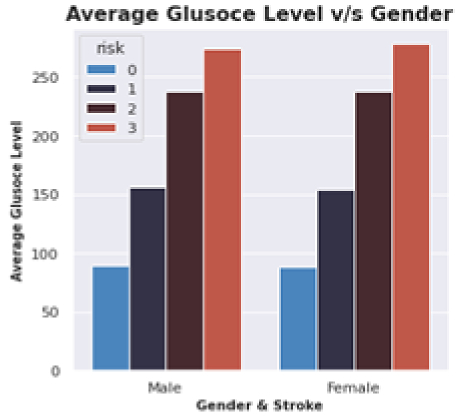


Fig.1. Figures for attributes in Brain Stroke Severity Dataset

4. Proposed Model

Figure 2 shows the steps in the proposed workflow which involves the pre-processing of data, training, and testing with specified models, evaluation of results and prediction of Brain Stroke Severity.

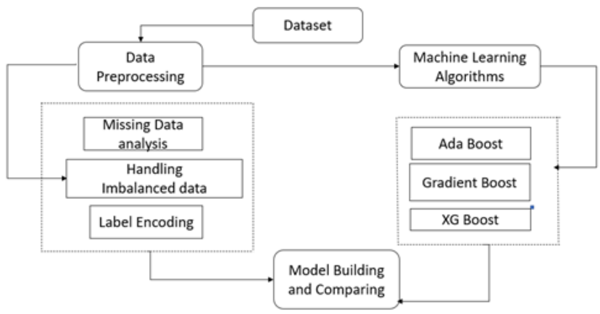


Fig 2. Steps Involved in Proposed Model.

4.1. Data Pre-Processing

Pre-processing involves converting raw data into a format that is meaningful and comprehensible. Real-world data frequently has errors and null values, which causes it to be inconsistent and incomplete. A good outcome is always produced by well pre-processed data. To deal with incomplete and inconsistent data, a variety of pre-processing techniques are utilised, including handling missing values,

ISSN No : 2454-4221(Print)
ISSN No : 2454-423X(Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

outlier detection, data discretization, data reduction (dimension and numerosity reduction), etc. Imputation has been used to solve the dataset's concerns with missing values.

4.2. Training and Testing Model

The entire dataset has been divided into two parts, with the ratio of each being 80:20: the training dataset and the testing dataset. the sets used for training, testing, and validation while classifying data.

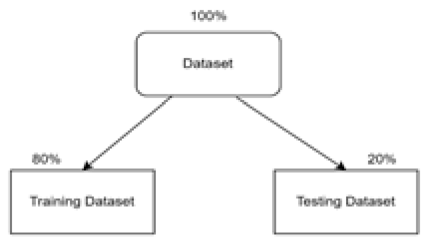


Fig 3. Final Training and Testing Data.

4.2.1. AdaBoost

AdaBoost (Adaptive Boosting) is a method for group learning that turns several poor learners into a strong learner. A weak learner is iteratively trained using the algorithm on a weighted subset of the training data, with the weights changed according to the accuracy of the previous weak learner. The AdaBoost method produces a weighted mixture of the results from the weak learners. Both continuous and categorical data can be handled using AdaBoost, which is effective and adaptable. For classification and regression problems, it is widely utilised in a variety of industries, including bioinformatics, computer vision, and speech recognition.

4.2.2. Gradient Boost

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. When a decision tree is the weak learner, the resulting algorithm is called gradient boosted trees, which usually outperforms random forest.

4.2.3. XG Boost

XG Boost (Extreme Gradient Boosting) is a gradient boosting algorithm that is designed for speed and performance. The algorithm works by iteratively adding weak learners to a model, where each learner tries to correct the errors of the previous learner. XG Boost uses a regularized objective function to prevent overfitting and can handle both categorical and continuous data. The algorithm also supports parallel processing, tree pruning, and early stopping to improve performance and reduce computation time. XG Boost is widely used in various fields, such as finance, healthcare, and computer vision, for classification, regression, and ranking tasks.

5. Results and Discussion

The result is measured in terms of Accuracy, F1- Score, Precision Score and Recall Score by using the confusion matrix and classification report. The result depends on how accurate the model is trained.

5.1 Performance Evaluation metrics

Measuring performance is key to check how well a classification model work to achieve a target. Performance evaluation metrics are used to evaluate the effectiveness and performance of the classification model on the test dataset. It

ISSN No : 2454-4221(Print)
ISSN No : 2454-423X(Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

is important to choose the correct metrics to evaluate the model performance such as confusion matrix, accuracy, precision, recall, etc. Following formulas are used to find the performance metrics.

$$\text{Precision Score} = \frac{TP}{(TP+FP)}$$

$$\text{Recall Score} = \frac{TP}{(TP+FP)}$$

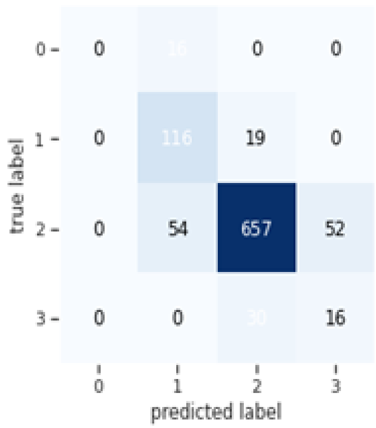
$$\text{F1- Score} = \frac{TP}{(TP+TN+FP+FN)}$$

Experimental results of various machine learning algorithms approach with all features selection have been shown for Brain Stroke Risk data. In this find the precision, F1-Score, Recall and accuracy of the predicted model. The Overall performance measures of all machine learning and classifiers with the mentioned dataset have been shown below in details:

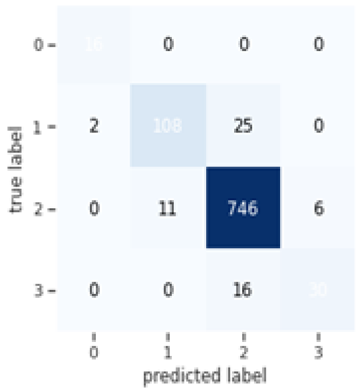
Table 1: Overall Results for Brain Stroke Severity prediction using Machine Learning Algorithms.

Classifier	Precision score	Rec all Score	F1 Score	Testing Accuracy
AdaBoost	0 | 62 | 93 | 23	0 | 85 | 86 | 34	0 | 72 | 89 | 28	82
GradientBoost	88 | 90 | 94 | 83	1 | 8 | 97 | 65	94 | 85 | 96 | 73	93
XGBoost	94 | 90 | 95 | 92	1 | 81 | 98 | 82	96 | 85 | 97 | 87	95

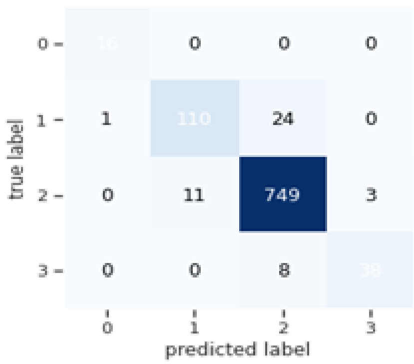
Evaluation of various machine learning models on Stroke Analysis dataset observed an accuracy in the range of (92% to 95 %) on the original dataset. AdaBoost has produced the least accuracy of 82.18%. GradientBoost Produces 93.7% and XGBoost have produced 95.10 % prediction accuracy on the original dataset. The confusion matrixes of all Machine Learning algorithms also describe the results of the prediction model.



(a)



(b)



(c)

ISSN No : 2454-4221(Print)
ISSN No : 2454-423X(Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

6. Conclusion

machine learning approaches were used to try and detect Brain Stroke Severity. To evaluate the performance of the models used to Predict Brain Stroke Severity on the stroke_analysis Dataset, a variety of performance evaluation indicators were employed. When XGBoost's classifier was put up against all other machine learning models, it exhibited high performance with 95% accuracy

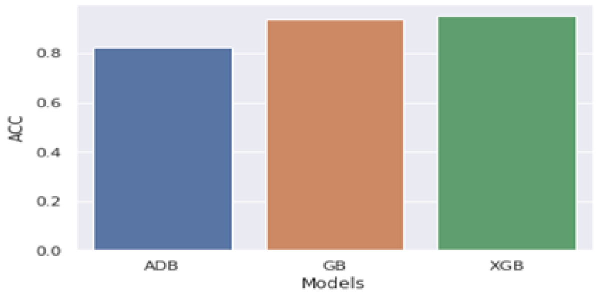


Fig 4: Testing accuracy comparison of Various Machine Learning Algorithms

7. References

1. Sailasya, G., & Kumari, G. L. A. (2021). Analyzing the performance of stroke prediction using ML classification algorithms. International Journal of Advanced Computer Science and Applications, 12(6).

2. Krishna, V., Kiran, J. S., Rao, P. P., Babu, G. C., & Babu, G. J. (2021, October). Early detection of brain stroke using machine learning techniques. In 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1489-1495). IEEE

3. Dritsas, E., & Trigka, M. (2022). Stroke risk prediction with machine learning techniques. Sensors, 22(13), 4670.

4. Bandi, V., Bhattacharyya, D., & Midhunchakkravarthy, D. (2020). Prediction of Brain Stroke Severity Using Machine Learning. Rev. d'Intelligence Artif., 34(6), 753-761.

5. Islam, R., Debnath, S., & Palash, T. I. (2021, December). Predictive Analysis for Risk of Stroke Using Machine Learning Techniques. In 2021 International Confer ence on Computer, Communication, Chemical, Materials and Electronic Engineeri ng (IC4ME2) (pp. 1-4). IEEE.

6. Uppal, S., Goel, S., Randhawa, B., Maheshwary, A.(2020). Autoimmune-associat ed vasculitis presenting asischemic stroke with hemorrhagic transformation: A Case report and literature review. Cureus, 12(9): e10403.

7. Monteiro, M., Fonseca, A.C., Freitas, A.T., Melo, T.P.,Francisco, A.P., Ferro, J., Oliveira, A.L. (2018). Using machine learning to improve the prediction of functional outcome in ischemic stroke patients. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 15(6): 1953-1959.

8. Sung, S.F., Lin, C.Y., Hu, Y.H. (2020). EMR-based phenotyping of ischemic stroke using supervised machine learning and text mining techniques. IEEE Journal of Biomedical and Health Informatics, 24(10).