International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

# DETECTION OF STROKE DISEASE USING ML ALGORITHM

**[1] Y. SANDHYA, [2] SAMEENA, [3] T. RAKSHITHA, [4] R. RAVITEJA, [5] C. SESHU VARDHAN**

[2,3,4,5] U.G. Scholor, Department of DS, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

[1]Assistant Professor, Department of DS, Sri Indu College Of Engineering & Technology, Ibrahimpatnam, Hyderabad.

## ABSTRACT

Using two machine learning methods, namely Logistic Regression and Novel Adaboost, the study aims to improve the identification of brain stroke using CT scans. A Brain Stroke Prediction Model Using Adaboost and Logistic Regression. There are 30 pictures of healthy human brains and 30 pictures of injured brains in the collection, which has a size of 64.4 MB. Assuming 500 records are employed for training, the test dataset may be recovered from the actual dataset, and the remaining 20% can be used for testing. enhance the precision to the current study, a comparison was made between logistic regression, which also used a sample size of 10 sets, and a new Adaboost method, both of which employed a sample size of 10 sets. Twenty sets were used for the comparison, with ten repetitions each set. The G power test, with $\alpha$=0.05 and $\beta$=0.2 as the parameters, yields an around 80%. Logistic Regression has an accuracy rating of 82.21% while the Novel Adaboost Classifier Algorithm has a value of 88.84%. Logistic Regression and Novel Adaboost were the subjects of this study. Novel AdaBoost outperforms Logistic Regression in terms of accuracy.Applying these two procedures properly is critical for diagnosing brain stroke, as shown by the findings. The study's dataset determines the observed degree of accuracy; other datasets may provide different results.

**Keywords:** Brain Stroke, Classifier, Health, Machine Learning, Logistic Regression, Novel Adaboost Classifier Algorithm, Prediction.

## INTRODUCTION

The brain is a crucial part of the human body. Humans cannot think or perform even the most basic tasks without a brain. The brain must communicate with the body in order for it to sit, stand, or perform work. A healthy lifestyle promotes brain activity. However, brain disorders like strokes are becoming more common today. Aside from this, health refers to a condition of

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

comprehensive wellbeing rather than just the absence of disease or infirmity on a physical, mental, or social level [1]. A human cannot survive without oxygen, just as they cannot survive without a functional nervous system. A brain hemorrhage occurs when a blood artery bursts and releases blood supply to the brain, a brain stroke may result. The breach or obstruction prevents oxygen and blood from reaching the brain tissue. The brain's tissues and cells swiftly deteriorate and start to die without oxygen, which results in a range of symptoms. Brain cells usually do not regenerate once they die, which can result in severe harm and impairment of the body, mind, and cognition. As soon as it is practical, the brain needs to have proper blood flow and oxygenation restored. Brain stroke is caused by stress, exhaustion, overthinking, and a lack of meditation [2].These numbers and the damaging effects of stress on human health underscore the need for a system that can recognise diseases linked to stress so that it can be treated with specific interventions or, in some cases, medication [3]. Psychological and physiological experts utilize stress analysis based on questionnaires to estimate an individual's level of stress. Because it only relies on individual responses from hesitant questionnaire responders, this strategy is highly speculative and inaccurate [4]. Since the time it takes for symptoms to appear is proportionate to the chance of survival and the efficacy of therapy, detecting strokes early is of the utmost importance. The likelihood of acquiring diabetes, ulcers, ulcerative colitis, asthma, headaches, skin disorders, health issues, epilepsy, and impotence is increased by stress [5]. Each of these illnesses as well as a large number by mental circumstances like stress in the environment.This study utilized Google Colaband LabView to study about the Brain stoke statistical factors like mean and standard deviation to distinguish the brain cells.

Eleven publications in Scopus and thirty-five in IEEE Xplore discuss this research. A structured data algorithm serves as the foundation for the Innovative Adaboost Classification Algorithm [6]. Machine learning was used to watch and identify stress. On the basis of the examination and evaluation of application model data, techniques for categorizing items were created. In order to detect brain strokes, machine learning techniques were used to construct this first Health Report [7]. As more and better algorithms for automatically identifying brain strokes were created [8]. Using these language characteristics, it can differentiate with astounding accuracy, according to a variety of machine learning classification algorithms. The work that used machine learning to detect brain strokes has generally garnered the most citations [9].

After conducting a review of existing literature, a research gap has been identified. This study aims to improve classification accuracy for predicting BrainStroke by comparing the performance of Adaboost Classifier with Logistic Regression using the current system. The proposed approach is expected to enhance the forecasting ability.

**MATERIALS AND METHODS**

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

The research was place in Chennai, India, at the Saveetha Institute of Medical and Technical Sciences, specifically at the Machine Learning Lab of the Saveetha School of Engineering.. The results of the process are compared between the two groups. The test dataset may be retrieved from the real dataset and used to identify the training dataset as long as there are 500 entries available. Around 80% of the records will be utilized for training, while the other 20% will be tested. 10 samples from each group are taken into consideration for this research. Two algorithms Novel Adaboost Classifier Algorithm and Logistic Regression use software for technical analysis in their implementation. By using GPower, a sample size of 10 was established for every group. 80%, threshold 0.05%, and CI 95% [10].

There was a column with 210 values in the dataset I was working with. I used a regression strategy to handle this problem and discovered some interesting results. I found a similar approach in a high-rated notebook, but it lacked performance evaluation, which I considered a disadvantage. Following an examination of the R-squared value, I noticed that the pipeline-based data assignment gave extremely bad results, with a value of less than 0.3, suggesting that it should not be regarded as a regression strategy (kaggle 2022).

The Windows 11 operating system, 16GB , together with an Intel Core i5 central processing unit, comprised the gear. setup for machine learning testing. The system was configured in 64-bit mode. While work is being done on the dataset and the code is running to verify correctness in the output process, Python was chosen as the programming language for developing the function.

**Novel Adaboost Classifier Algorithm**

Adaptive Boosting, also known as Novel Adaboost Classifier Algorithm, is a Boosting approach used to machine learning as an ensemble technique. The term "adaptive boosting" comes from the process of redistributing weights to each instance, with more weights assigned to instances that were incorrectly classified.

**Procedure for Adaboost Classifier Algorithm**

**Input**: Dataset K.
**Output**: A testing dataset class
1: Adaboost also called Adaptive boosting of Ensemble model. Adaboost is also modeled in decision tree format.

a. Each Decision tree in Adaboost is called Stump
b. Each node builds its own model
c. Assign weight value for each dataset.

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

d. Each value of the model is assigned equal weight.

2: The next step is that data will be divided into two divisions, one is the testing dataset and other one is the training dataset.

a. Patients will be divided based on height, weight, age and each patient problem category.

3: Initialize the weight and calculate the sum of each weight.

a. By calculating each stump's Gini index and choosing the one with the lowest value, you may determine which one categorizes the new collection of samples the best.

4: From different groups voting will be considered and by the voting output final prediction will be considered.

**Logistic Regression**

Logistic regression is often used when it is necessary to partition the data into many groups. The two choices are logistic regression with binary and multiple classes. The binary class has only two members, as the name suggests: yes or no, true or false, zero or one, etc. For multi-class classification, the number of classes utilized to group data is more than two.

**Procedure for Logistic Regression**

**Input**: Dataset K.
**Output**: A testing dataset class

1: Start Data Preprocessing

2: importing libraries

3: importing datasets

4: Extracting Independent and dependent Variable

5: Visualizing the training set result

It's important to keep in mind that there are software and hardware parameters under test conditions. The specs of this laptop include a 64-bit Windows operating system, 8 GB of RAM, a hard drive, and an x86-based processor—specifically, an Intel Core i7 10th generation CPU. At this time, the software is compatible with Windows 10 and is developed as a collaborative project using Python. Once the program is done, the accuracy value will be presented. Laptop connected to WiFi is the method.The current study makes use of the Google Collaborative tool to run python programs. Then it's possible to store the software for future uses. To get the final results of the graphs and mean values, save the file and retain the accuracy values in the SPSS program after obtaining 10 iterations of accuracy.

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

Kaggle is used to collect the dataset. Dataset pertains to patient reports for brain strokes. The collection comprises 1000 brain X-ray scans for each human brain. Age, weight, height, and a person's state of health are all included in the dataset.

**Statistical Analysis**

For this investigation, the IBM tool utilized is SPSS Software. Using or comparing machine learning algorithms for the purpose of predicting brain strokes is what this prediction is all about. [11]. The Independent t-test is measured with statistically significant values and the dependent variables are the dataset attributes like the patients gender, age and patients problem.

**RESULT**

Predicting a brain stroke using two ML methods is the focus of this study. The Novel Adaboost Classifier Algorithm and Logistic Regression are two machine learning algorithms. Table 1 displays the accuracy value together with the number of iterations of the Novel Adaboost Classifier Algorithm and Logistic Regression.
The Table. 1 is about the accuracy values obtained from Group 1 called Novel Adaboost Classifier Algorithm Classifier Algorithm accuracy values for iterations are 88.84,89.81,89.8,88.87,88.82,88.81,88.81,87.9,87.9,87.89,87.85. The Logistic Regression accuracy values for iterations are 82.211,82.210,82.28,82.25,82.20,82.20,82.17,82.16,82.13, 82.10.

Table 2. Adaboost and Logistic Regression Group Statistics. After analyzing ten samples, we may calculate their means, standard deviations, and standard error means. Adaboost outperforms Logistic Regression in terms of mean sustainability.
Table 3's independent sample test value is 0.01 which is under statistical significance. Fig.1 States the bar graph comparison between Novel Adaboost Classifier Algorithm Classifier Algorithm and Logistic Regression.

**DISCUSSION**

It seems that Adaboost is superior than Logistic in the provided research, since the independent sample test value obtained is Indicating statistical significance, the value is 0.01 ($p<0.05$). Regression analysis. An investigation of Adaboost's sustainability reveals that it is 88.84% sustainable. is 82.21.%.

In the previous work Novel Adaboost Classifier Algorithm has an accuracy value of 85% whereas for the proposed work the accuracy values for Novel Adaboost Classifier Algorithm is 88.84% and when compared with the Logistic Regression the accuracy values for existing work

is 80% and 82.21% for proposed work. Acquired brain injury (ABI) is most often caused by stroke and TBI. (ABI). The success of quantitative CT analysis in ABI depends on the availability of a completely automated artifact elimination technique [12]. Consequently, the goal of this research was to evaluate the suggested method using CT scans of patients with traumatic brain injuries and cerebrovascular accidents. There were a total of 522 CT images from stroke patients and 240 from TBI patients used to evaluate the efficacy of the automated artifact reduction algorithm. Using pseudo-object pixels, the fracture-sealing technique can help with both ROI delineation and automated artifact removal.

The limitations of this research is the considerable time required to train convolutional neural networks, particularly when dealing with enormous datasets. Expanding the system to accommodate more items in less time during data set training is the future goal of this project.

## CONCLUSION

This study compares Adaboost with Logistic Regression. While Logistic Regression has a sustainability value of 82.21%, Adaboost has a score of 88.84%. Analyzed results show that Adaboost outperforms Logistic Regression (82.21%).

## REFERENCES

[1] V. JalajaJayalakshmi, et.al., "Analysis and Prediction of Stroke using Machine Learning Algorithms," in *2021 I (ICAECA)*, Oct. 2021, pp. 1–5.
[2] B. Akter,et.al., "A Machine Learning Approach to Detect the Brain Stroke Disease," in *2022 4th (ICSSIT)*, Jan. 2022, pp. 897–901.
[3] O. Karadima, R. C. Conceição, and P. Kosmas, "Differentiation of brain stroke type by using microwave-based machine learning classification," in *(ICEAA)*, Aug. 2021, pp. 185–185.
[4] R. Dacosta-Aguayo *et al.*, "Predicting Cognitive Recovery of Stroke Patients from the Structural MRI Connectome Using a Naïve Bayesian Tree Classifier," *2017 IEEE 30th (CBMS)*. 2017. doi: 10.1109/cbms.2017.106.
[5] S. Gupta and S. Raheja, "Stroke Prediction using Machine Learning Methods," in *2022 12th (Confluence)*, Jan. 2022, pp. 553–558.
[6] W. Yang, "RETRACTED ARTICLE: Analysis of sports image detection technology based on machine learning," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1. 2019. doi: 10.1186/s13640-019-0409-8.
[7] S. Prasad and J. Chanussot, *Hyperspectral Image Analysis: Advances in Machine Learning and Signal Processing*. Springer Nature, 2020.
[8] C. N. Sujatha and V. Padmavathi, "Machine Learning Application for Detecting Leaf Diseases with Image Processing Schemes," *Machine Learning Methods for Signal, Image and Speech Processing*. pp. 47–76, 2022. doi: 10.1201/9781003338789-3.
[9] A. Saygılı, "A new approach for computer-aided detection of coronavirus (COVID-19)

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

from CT and X-ray images using machine learning methods," *Appl. Soft Comput.*, vol. 105, p. 107323, Jul. 2021.

[10] A. B. M. Sultan, S. Mehmood, and H. Zahid, "Man in the Middle Attack Detection for MQTT based IoT devices using different Machine Learning Algorithms," in *2022 2nd International Conference on Artificial Intelligence (ICAI)*, Mar. 2022, pp. 118–121.

[11] 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 *2019 10th (ICCCNT)*, Jul. 2019, pp. 1–6.

[12] D.-H. Park *et al.*, "Automated artefact elimination in computed tomography: A preliminary report for traumatic brain injury and stroke," in *The 3rd International Winter Conference on Brain-Computer Interface*, Jan. 2015, pp. 1–3.

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

## TABLES AND FIGURES

**Table 1.** A Study on the Logistic Regression and New Adaboost Classifier Algorithm Accuracy

| Iterations | Novel Adaboost Classifier Accuracy(%) | Logistic Regression Accuracy(%) |
|---|---|---|
| 1 | 88.84 | 82.211 |
| 2 | 89.81 | 82.210 |
| 3 | 89.8 | 82.28 |
| 4 | 88.87 | 82.25 |
| 5 | 88.82 | 82.20 |
| 6 | 88.81 | 82.20 |
| 7 | 88.81 | 82.17 |
| 8 | 87.9 | 82.16 |
| 9 | 87.89 | 82.13 |
| 10 | 87.85 | 82.10 |

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

**Table 2.** Analysis of Adaboost and Logistic Regression Groups. Central tendency, dispersion, and error The results are averaged from ten samples. Adaboost outperforms Logistic Regression in terms of mean sustainability.

| | **Group** | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
|---|---|---|---|---|---|
| **sustainability** | Adaboost | 10 | 88.8410 | .79246 | .025060 |
| | Logistic Regression | 10 | 82.2100 | .07257 | .02295 |

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

**Table 3.** With a p-value of 0.01 ($p<0.05$) in the Independent Sample T-test, Adaboost is not statistically superior than Logistic Regression.

| | | Levene's test for equality of variances | | T-test for equality means with 95% confidence interval | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | f | Sig. | t | Df | Sig. (2-tailed) | Mean difference | Std.Error difference | Lower | Upper |
| sustainability | qual variances assumed | 11.932 | .003 | 26.351 | 18 | .000 | 6.63100 | .25165 | 6.10231 | 7.15969 |
| | Equal Variances not assumed | | | 26.351 | 9.151 | .000 | 6.63100 | .25165 | 6.10231 | 7.15969 |

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 Mar 2025

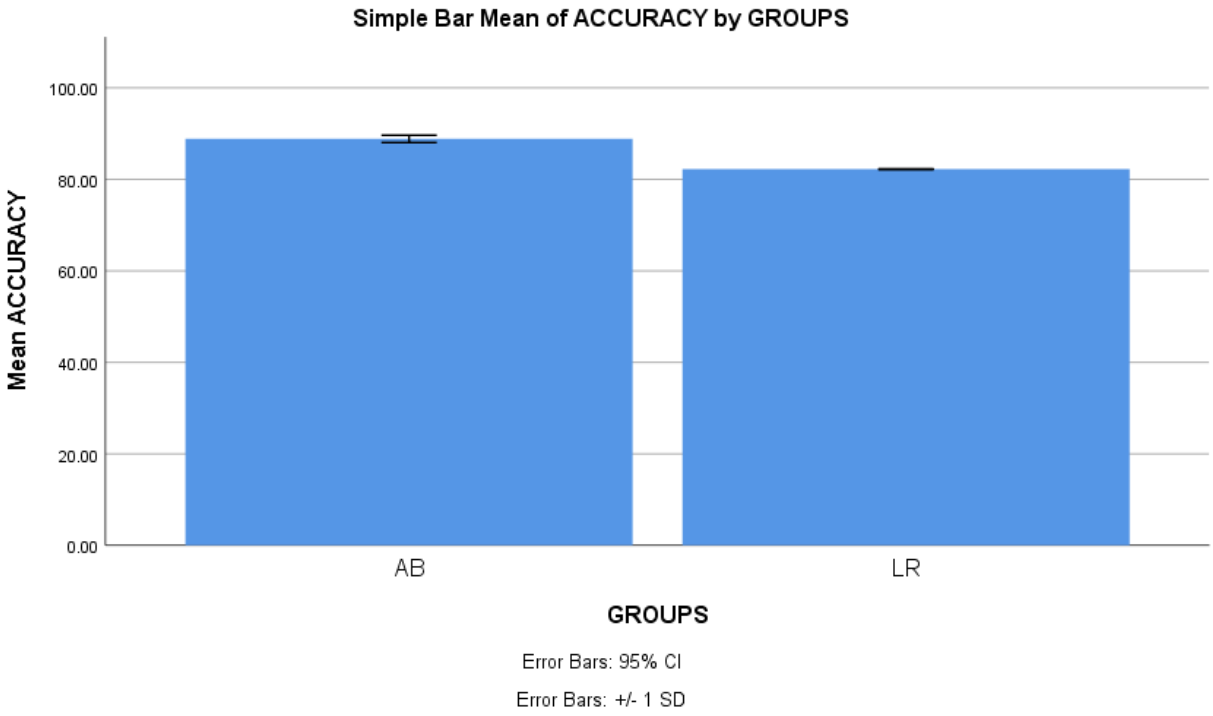


**Figure 1.** Evaluate Adaboost vs. Logistic Regression. Foretelling in light of average longevity. Adaboost outperforms Logistic Regression in terms of mean sustainability. Prediction: Compared to Logistic Regression, Adaboost's standard deviation is somewhat better. Adaboost vs. logistic regression on the X-axis Axis for prediction and Y-score: average detection sustainability with standard deviation +/-1.