International Scientific Journal of Engineering and Management
Volume: 03 Issue: 06 | June – 2024
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI: 10.55041/ISJEM02039

Intelligent Stroke Analysis: Utilizing Machine Learning Algorithms for Enhanced Clinical Decision Making

Geeta Bharti , Rajnandani Patil , Vedant Kulkarni , Pratik Kulkarni

Bachelors of Technology, Department of Computer Science and Engineering, SOET

DY PATIL UNIVERSITY, AMBI, PUNE.

e-mail: geetabhrti789@gmail.com, rajnandnipatil2@gmail.com, vedantk0501@gmail.com,

Abstract

Strokes happen when blood flow to the brain is cut off, causing serious damage. Predicting who might have a stroke beforehand is crucial for early treatment and better outcomes. This study explores a new way to predict stroke risk using a powerful type of artificial intelligence (AI) called XGBoost.

Imagine XGBoost as a super-sleuth that analyzes patient information like age, weight, medical history, and blood pressure. But before feeding this data to XGBoost, we clean and organize it for the best results. Unlike some AI, XGBoost is transparent, revealing the key factors that put each patient at higher risk of stroke.

Our research has two main goals:

1. Improved Stroke Prediction: By leveraging XGBoost's strengths, we aim to predict stroke risk even more accurately than previous methods.

2. Understanding Risk Factors: XGBoost's transparency allows us to see why someone might be at higher risk. This knowledge helps doctors create personalized plans to prevent strokes.

This research has the potential to revolutionize stroke prevention. By combining high accuracy with clear explanations, XGBoost can become a valuable tool for early intervention and customized prevention strategies.

Keywords: Stroke, AI, Prediction, Risk factors, Prevention

I. Introduction

Stroke, a leading cause of death and disability worldwide, remains a formidable foe in the fight for global health. This devastating neurological event, caused by a disruption in blood flow to the brain, can have catastrophic consequences. When a blood vessel is blocked by a clot (ischemic stroke) or bursts (hemorrhagic stroke), vital oxygen and nutrients are cut off, leading to brain cell death and lasting impairments.

The urgency of stroke treatment is paramount. Every minute that passes without intervention increases the risk of permanent brain damage and even death. Early and accurate diagnosis is crucial for maximizing patient outcomes and improving quality of life.

While modern medicine has made significant strides, stroke prevention isn't always straightforward. While maintaining a healthy lifestyle with regular exercise, balanced diet, and controlled blood sugar can help reduce risk factors, the complexities of stroke make it challenging to predict based solely on traditional methods.

This is where machine learning (ML) emerges as a powerful tool in the fight against stroke. ML's ability to analyze vast amounts of data and identify hidden patterns offers a promising avenue for early stroke detection. By leveraging various algorithms and patient data, researchers can potentially develop accurate prediction models that empower healthcare professionals to intervene before it's too late.

This study delves into the potential of ML algorithms for early stroke prediction, moving beyond traditional risk assessment approaches. We explore the effectiveness of a diverse set of classifiers, including established algorithms like Support Vector Machines (SVMs) and cutting-edge techniques like XGBoost. These algorithms will be trained on a dataset rich with physiological factors known to be associated with stroke risk, such as age, blood pressure, medical history, and potentially even bio-signals.

Our investigation has two primary objectives:

1. Enhanced Stroke Prediction Accuracy: By utilizing the strengths of various ML algorithms, we aim to achieve superior accuracy in predicting stroke risk compared to traditional methods. This could lead to earlier interventions and improved patient outcomes.

ISSN: 2583-6129
DOI: 10.55041/ISJEM02039

2. Interpretability and Explainability: While achieving high accuracy is crucial, understanding the factors most influential in stroke risk is equally important. By choosing algorithms with interpretability features, we hope to gain valuable insights into the underlying relationships between patient data and stroke risk. This knowledge can empower healthcare professionals to tailor preventive strategies for individual patients.

This research has the potential to significantly improve stroke care. By combining high prediction accuracy with interpretable insights, ML algorithms like XGBoost can become valuable tools for early intervention and personalized stroke prevention strategies. Ultimately, this could lead to a future where stroke's devastating impact is significantly reduced, saving lives and improving patient outcomes.

II. Literature Review

In [4], the authors focused on predicting stroke occurrences using the Cardiovascular Health Study (CHS) dataset through the application of five distinct machine learning techniques. They identified an optimal solution by combining the Decision Tree algorithm with C4.5, Principal Component Analysis (PCA), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs). This combination aimed to enhance the prediction accuracy. However, a notable limitation of this study was the smaller number of input parameters available in the CHS dataset, potentially restricting the model's predictive power and generalizability.

In [5], stroke prediction was approached from a novel angle by analyzing social media posts. The authors employed the DRFS (Deep Recurrent Feature Selection) method to identify various stroke-related symptoms mentioned in the posts. This innovative use of Natural Language Processing (NLP) techniques to extract relevant text from social media data introduced a unique dimension to stroke prediction. Nevertheless, the model's overall execution time increased due to the computational complexity of NLP processes, which was identified as a significant drawback affecting its practicality for real-time applications.

In [6] explored stroke prediction through an enhanced version of the random forest algorithm. This method was designed to evaluate the risk levels associated with strokes more effectively than traditional algorithms. The authors reported superior performance of their method in comparison to existing algorithms. However, the study's scope was limited to a specific subset of stroke types, indicating that its applicability to other or newer types of strokes might be constrained. This limitation suggests the need for further research to broaden the model's utility across a wider range of stroke categories.

In [7], the authors trained a stroke prediction model using Decision Tree, Random Forest, and Multi-layer Perceptron (MLP). The performance metrics showed close accuracy scores for these methods: Decision Tree achieved 74.31%, Random Forest 74.53%, and MLP 75.02%. The paper concluded that MLP provided a slightly better accuracy, but it emphasized that relying solely on accuracy as a performance metric might not yield the most favorable results, suggesting the need for additional evaluation criteria.

In [8] focused on heart stroke prediction using various machine learning techniques, including Decision Tree, Naïve Bayes, and SVM. The performance comparison revealed a maximum accuracy of 60%, indicating a relatively lower effectiveness of the applied algorithms in this context. This outcome suggests the necessity of exploring more advanced or alternative methods to enhance predictive accuracy.

In [9], the authors employed different data mining classification techniques on a dataset from the Ministry of National Guards Health Affairs Hospitals in Saudi Arabia to predict stroke. They used C4.5, Jrip, and Multi-layer Perceptron (MLP), achieving an accuracy of approximately 95%. Despite this high accuracy, the study noted that the training and prediction times were substantial due to the complexity of the combined algorithms used, highlighting a trade-off between accuracy and computational efficiency.

In [10] compared the performance of Naïve Bayes, Decision Tree, and Neural Networks for stroke prediction. The Decision Tree algorithm emerged with the highest accuracy at around 75%. However, the paper pointed out that the model's performance, as indicated by the confusion matrix, might not translate well to real-world scenarios, suggesting limitations in its practical application.

In [11], the researchers utilized the Cardiovascular Health Study (CHS) dataset to propose a novel automatic feature selection algorithm that identifies robust features using their proposed conservative mean. This method was combined with the Support Vector Machine (SVM) algorithm to enhance efficiency. However, the generation of numerous vectors reduced the model's overall performance, indicating a potential drawback of the proposed approach.

In [12] explored the prediction of thromboembolic stroke disease using Artificial Neural Networks (ANNs) with the Back-propagation algorithm. This method achieved an accuracy of around 89%. However, the complexity of Neural Networks, particularly with an increasing number of neurons, requires significant training time and higher processing power, which could limit its practical utility in real-time applications.

Jeena et al. [13] investigated several risk factors associated with stroke by employing a regression-based methodology to understand the relationship between these factors and stroke likelihood. This study emphasized the importance of identifying significant predictors in stroke occurrence.

Adam et al. [14] compared the decision tree method and the k-nearest neighbor algorithm for stroke prediction. Their

ISSN: 2583-6129
DOI: 10.55041/ISJEM02039

research revealed that medical professionals found the decision tree method to be more practical for predicting stroke occurrences, underlining its utility in clinical settings.

Singh and Choudhary [15] utilized the Cardiovascular Health Study (CHS) dataset to predict stroke, demonstrating the dataset's application in stroke prediction models. Emon et al. [10] implemented multiple learning-based classification algorithms, including XGBoost, Random Forest, Naive Bayes, Logistic Regression, and Decision Tree, on a dataset from Kaggle. This comprehensive approach allowed them to compare the effectiveness of different algorithms in predicting stroke.

Kansadub et al. [16] explored stroke prediction using decision trees, neural networks, and Naive Bayes analysis. They assessed their models' precision and AUC, aiming to optimize prediction accuracy. Tazin et al. [12] proposed an early-stage stroke prediction model using Logistic Regression, Decision Tree Classification, Random Forest Classification, and a Voting Classifier, with Random Forest emerging as the top performer.

Chetan Sharma et al. [17] applied a supervised random forest algorithm on an openly accessible dataset to predict stroke occurrence in the near future. Their findings reinforced the utility of random forest in stroke prediction.

Hung et al. [18] conducted a comparative analysis of machine learning and deep learning models for constructing stroke prediction models from electronic medical claims databases. Their study provided insights into the strengths and weaknesses of different modeling approaches.

III. RESEARCH METHODOLOGY

Data pre-processing is a crucial step in model construction to eliminate undesirable noise and outliers, which can cause the model to deviate from its intended training objectives. This phase addresses all issues that may hinder the model's operational effectiveness. After collecting the relevant dataset, data cleansing and processing are necessary for model development. The dataset in question comprises twelve attributes. Initially, the column 'id' is ignored as it does not impact model creation. Subsequently, the dataset is checked for null values, which are filled using the most frequent value for the 'BMI' column.

String literals in the dataset are transformed into integer values through label encoding, making the data comprehensible for computational training, which typically operates on numerical data. The dataset contains five columns (gender, ever married, work type, residence type, smoking status) with string data types. During label encoding, all strings are encoded, converting the entire dataset into a numerical format.

The stroke prediction dataset is highly imbalanced, with 5110 rows: 249 indicating a likelihood of stroke and 4861 indicating its absence. Training a machine learning model with such imbalanced data can lead to high accuracy but poor performance on other metrics like recall and precision, resulting in incorrect and unreliable predictions. To address this imbalance, the Random Oversample (ROS) approach was employed, balancing the dataset by ensuring both classes have the same number of instances, as depicted in Fig. 3.

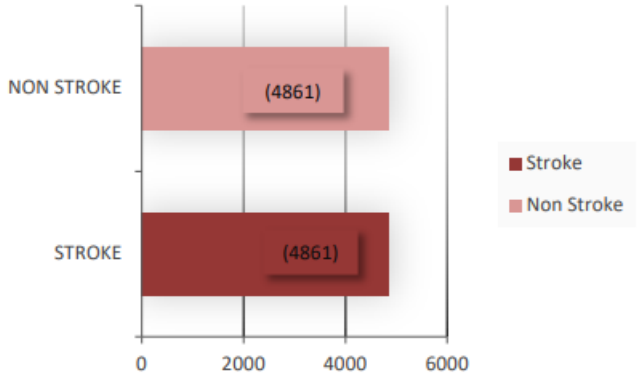

Fig. 3. Total count of stroke and non-stroke data after pre-processing.

Normalization of the features was performed using a MinMaxScaler, scaling the features between -1 and 1. Principal Component Analysis (PCA) was then applied, retaining 95% of the variance with the minimum number of principal components. Upon completing data preparation and addressing the imbalance, the model construction phase began. The dataset was split into training and testing subsets with an 80/20 split to enhance the model's accuracy and efficiency.

Various classification techniques were employed for model training, including deep neural networks (3-layer and 4-layer ANNs), Extreme Gradient Boosting (XGBoost), AdaBoost, Light Gradient Boosting Machine (LightGBM), Random Forest, Decision Tree, Logistic Regression, K Nearest Neighbors (KNN), SVM with a linear kernel, and Naive Bayes. The complete workflow is illustrated in Fig. 4.

International Scientific Journal of Engineering and Management
Volume: 03 Issue: 06 | June – 2024
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI: 10.55041/ISJEM02039

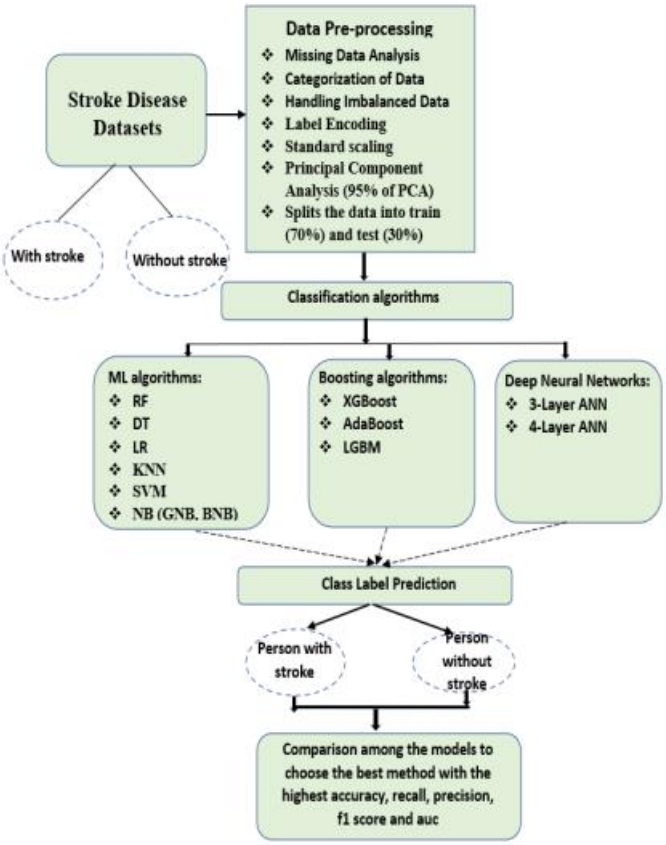

Fig. 4. The workflow of the proposed methodology.

This comprehensive pre-processing and modeling approach ensures that the data is adequately prepared for training, the class imbalance is rectified, and a variety of robust classification algorithms are utilized to develop a reliable and accurate stroke prediction model.

The proposed methodology for this work is outlined in several key stages: data collection and description, data pre-processing, and the implementation of various machine learning classifiers. Figure 1 illustrates the workflow implemented in our study, and the systematic procedure is detailed below:

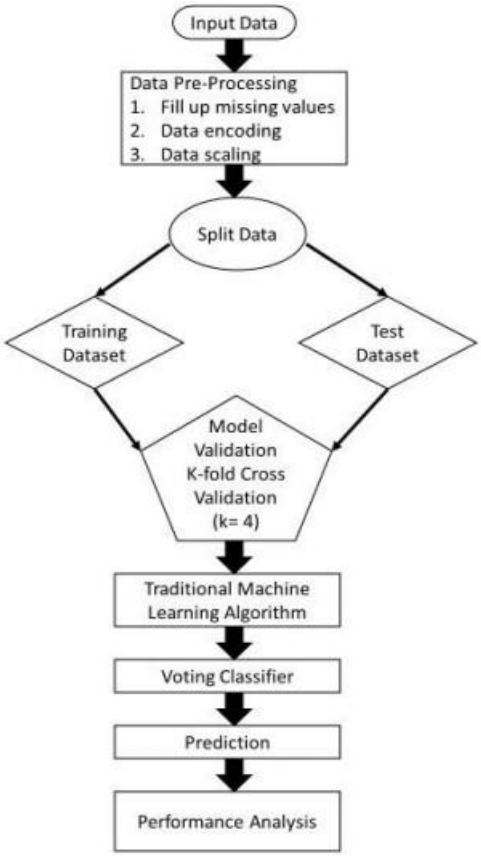

Fig. 1. Schematic working process diagram of our system

A. Collection of Data and Description of Data:

The stroke disease dataset was collected from various hospitals in Bangladesh and comprises data from 8600 subjects, 2500 of whom had experienced a stroke. Each record in the dataset includes 11 attributes, with some attributes being numerical and others categorical. The numerical attributes include age (the subject's age), hypertension (indicating whether the subject has hypertension), average glucose level (the subject's glucose consumption level), and BMI (the ratio of the subject's height to weight). The categorical attributes include smoking status (the subject's smoking condition), stroke (indicating whether the subject had a stroke in the past), gender (the subject's gender), work type (the subject's employment status), residence type (the subject's residence condition), and marital status (whether the subject is ever married). The 'stroke' attribute serves as the decision class, while the remaining attributes are the input class. This comprehensive dataset provides a solid foundation for further analysis and model development.

B. Data Pre-processing:

To make the dataset effective, we began by identifying and eliminating duplicate data. We then addressed null values,

ISSN: 2583-6129
DOI: 10.55041/ISJEM02039

filling missing numerical data with the mean value of the respective attribute and filling missing categorical data with the median value. Next, we used the One Hot Encoding method to convert all categorical values into numeric form, a necessary step for machine learning algorithms that cannot handle categorical data directly. After encoding, we normalized all numerical values to ensure consistency. The dataset was then split into two groups: a training dataset and a testing dataset. The training data was used to train the machine learning models, while the testing data was used to evaluate the model's performance. This comprehensive pre-processing ensured that the dataset was clean, balanced, and ready for effective model training and evaluation.

C. Machine Learning Classifier:

In our paper, we employed nine different machine learning classifiers: AdaBoost, Artificial Neural Networks (ANN), Decision Tree, K-Nearest Neighbors (KNN), Random Forest, Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), and XGBoost. These well-known classifiers allow for comparison with similar research works. We used 75% of the dataset to train our algorithms and reserved the remaining 25% to assess the trained models. For model validation, we utilized the k-fold cross-validation technique. In this method, the dataset is divided into k parts (folds). During the training process, k-1 folds are used to train the model, while one fold is used to test it. This process is repeated k times, with each fold serving as the test dataset once. This technique ensures that all samples in the dataset are used for both training and testing, thus reducing high variance. We evaluated the model's performance using confusion matrices, calculating metrics such as accuracy, recall, precision, F1 score, false-positive rate, and false-negative rate. By analyzing these values, we identified the best model for predicting stroke.

IV. CLASSIFICATION ALGORITHMS

A. Machine Learning Approaches

1.Decision Tree (DT): Classification with DT addresses both regression and classification issues using a supervised learning model and an output variable. The decision tree comprises decision nodes and leaf nodes, where data is split at decision nodes and combined at leaf nodes to generate the output. This method mimics human decision-making processes and assumes that the existence or absence of a feature depends on others, aiding in categorizing the target class .

2.Random Forest (RF): RF is an ensemble learning technique used for classification and regression problems. It involves distributed training of many decision trees, with the majority vote determining the final class. This method enhances model performance by using predictions from multiple trees, combining them to boost accuracy. RF demonstrated the highest accuracy using the stroke prediction dataset when configured with the entropy criterion.

3.Naive Bayes (NB)**:** NB is a supervised learning technique based on the naive theorem, which assumes that the presence of one feature is independent of others. It uses Bayes' theorem to estimate the probability of an event occurring given certain conditions. Variants include Gaussian NB and Bernoulli NB, both used in this study to handle different data types and distributions .

4.K-Nearest Neighbors (KNN)**:** KNN is a lazy learning algorithm where computations are deferred until classification. It uses the Euclidean distance metric to predict the target class based on the closest training data points. The parameter k, determining the number of neighbors considered, is set to 3 in this study to optimize classifier performance.

5.Support Vector Machine (SVM)**:** SVM is a supervised learning system that classifies data using labeled training data. It employs hyperplanes to separate data into different classes. The study used the Radial Basis Function (RBF) kernel with a setting of 1 to maximize classification accuracy by creating the largest possible margin while minimizing error .

6.Logistic Regression (LR)**:** LR is widely used in supervised learning for forecasting categorical dependent variables based on independent factors. Unlike linear regression for regression issues, LR addresses classification problems. The study utilized ridge regression with L2 regularization, employing solver='liblinear' and max iter=100 to handle multicollinear data effectively .

7.XGBoost**:** XGBoost is an optimized gradient boosting technique designed for high accuracy and efficiency. It enhances model performance by building strong learners from weak learners through gradient boosting and ensemble methods. Parameters such as base score=0.5, max bin=256, learning rate=0.9, max depth=25, and min child weight=1 were chosen to achieve high accuracy, making XGBoost one of the top performers in this study.

8.AdaBoost: AdaBoost, or Adaptive Boosting, is an ensemble method that redistributes weights to misclassified instances, giving them higher importance. This study combined 100 DT algorithms with the CART algorithm, iterating the boosting process to improve classification accuracy. Parameters included learning rate=0.9, n_estimators=20, and random_state=42 to enhance model performance.

International Scientific Journal of Engineering and Management
Volume: 03 Issue: 06 | June – 2024
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI: 10.55041/ISJEM02039

TABLE I: DESCRIPTION OF THE PARAMETERS USED FOR THE XGBOOST

Name of the parameter	Default Value	Description of the parameters
learning_rate	0.9	Reduce the weights with each step.
n_estimators	100	Number of trees to fit
objective	binary	logistic regression for binary classification
booster	gbtree	Select the model for each iteration
nthread	max	Input the system core number
min_child_weight	1	Minimum sum of weights
max_depth	25	Maximum depth of a tree
gamma	0	The minimum loss reduction needed for splitting
reg_lambda	1	L2 regularization term on weights
reg_alpha	0	L1 regularization term on weights

9.Light Gradient-Boosting Machine (LightGBM)**:** LightGBM is a gradient boosting system that utilizes tree-based learning techniques. It is designed to be highly efficient in training, with lower memory usage and improved accuracy compared to traditional gradient boosting methods. LightGBM aims to be distributed and efficient, making it suitable for handling large datasets and complex models efficiently. Its superior training efficiency and memory utilization make it a valuable addition to the ensemble of machine learning classifiers in this study.

B. Deep Learning Approaches

Deep learning encompasses artificial neural networks used in machine learning, including convolutional neural networks (CNNs), deep belief networks (DBNs), recurrent neural networks (RNNs), and deep neural networks (DNNs). These architectures are inspired by the human brain and are adept at recognizing complex patterns across various domains such as speech recognition, natural language processing, computer vision, and more. DNNs consist of an input layer, multiple hidden layers, and an output layer, with backpropagation being a common method for training these networks to minimize the error between desired and actual outputs. In our study, two ANN models were implemented: three-layer and four-layer ANNs, both employing sigmoid activation functions. Fig. 5 illustrates the flowchart depicting the proposed ANN techniques, with data collection and preprocessing procedures aligning with those used for machine learning approaches. This integration of deep learning techniques expands the scope of analysis and enhances the model's ability to recognize intricate patterns within the stroke prediction dataset.

VI. PERFORMANCE METRICS AND RESULTS

A. Performance Metrics

In evaluating classifier performance, five key statistical measures were utilized: accuracy, precision, recall (or sensitivity), F1 score, and Area Under the Curve (AUC). Accuracy denotes the proportion of correctly classified instances, precision assesses the accuracy of positive predictions, recall quantifies the ability to identify all positive instances, F1 score balances precision and recall, and AUC measures the classifier's overall discriminatory ability across different threshold settings, providing a comprehensive evaluation of stroke prediction efficacy.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall/Sensitivity} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1 Score} = \frac{2*(Recall*Precision)}{(Recall+Precision)} \quad (4)$$

where,

TP = true positive

FN = false negative

FP = false positive

TN =true negative

In this work, AUC curves were used to determine how well the probabilities from the positive classes and the negative classes could be separated. The degree of True Positive and False Positive rate is represented by the AUC curve, which also shows the model's overall performance.

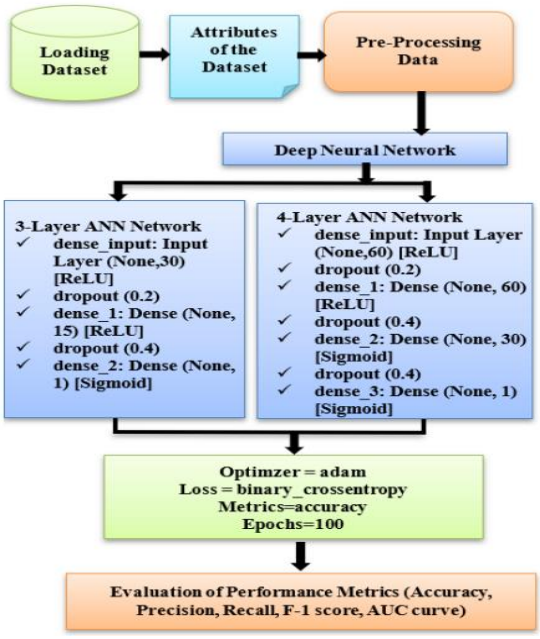

Fig. 5. The flowchart of the two proposed ANN techniques

International Scientific Journal of Engineering and Management
Volume: 03 Issue: 06 | June – 2024
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI: 10.55041/ISJEM02039

B. Results

Random Forest outperforms other classifiers in terms of accuracy (0.99) which is calculated using equation (1). It shows the highest accuracy among the machine learning algorithms, whereas 3-layer ANN demonstrated promising results among deep learning techniques. The comparison among the machine learning approaches is shown in Table III and a chart is shown (in Fig. 6) to represent the superiority of RF method over the other ML algorithms using the performance metrics which were calculated using equation (1-4). The area under roc curve for Random Forest method is given in Fig. 7. The performance metrics for the ANN approaches are described in Table IV. The Area under roc curve for the 4-layer ANN is shown in Figure 8. The comparison between the Random Forest and 3-layer ANN method is depicted in the bar chart of Fig. 9. From the comparison, it is clear that RF algorithm outperforms all the Boosting algorithms and deep neural approaches in all aspects.

TABLE III: COMPARISON AMONG THE MACHINE LEARNING APPROACHES

Algorithm	Accuracy	Precision	Recall	F1-score	AUC
LR	0.71	0.69	0.73	0.71	0.79
DT	0.98	1.00	0.95	0.97	0.98
RF	0.99	1.00	0.98	0.99	1.00
KNN	0.96	1.00	0.92	0.96	0.98
SVM	0.82	0.84	0.77	0.81	-
GaussianNB	0.70	0.68	0.73	0.70	0.78
BernoulliNB	0.67	0.69	0.59	0.63	0.71
XGBoost	0.97	1.00	0.93	0.97	0.98
AdaBoost	0.78	0.75	0.82	0.78	0.76
LGBM	0.95	1.00	0.90	0.95	0.96

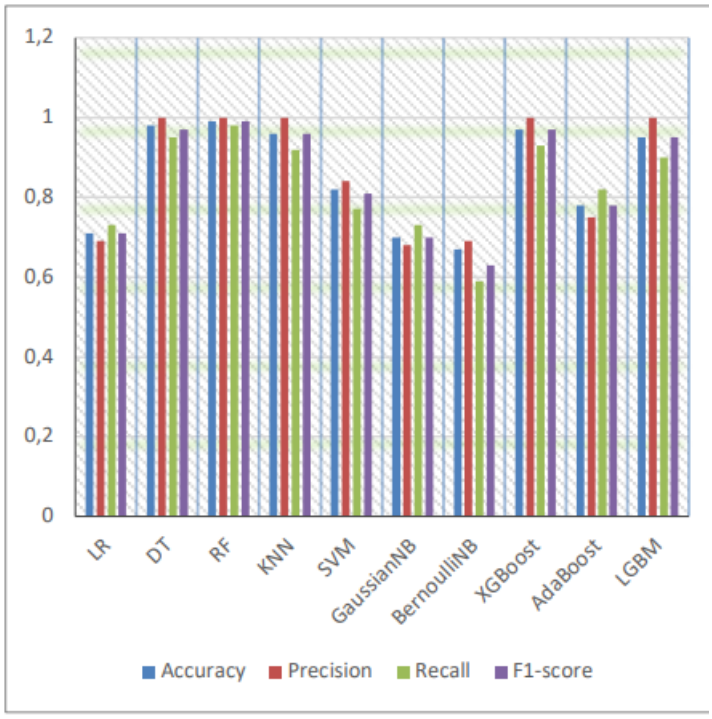

Fig. 6. A comparison chart of evaluation metrics of machine learning algorithms.

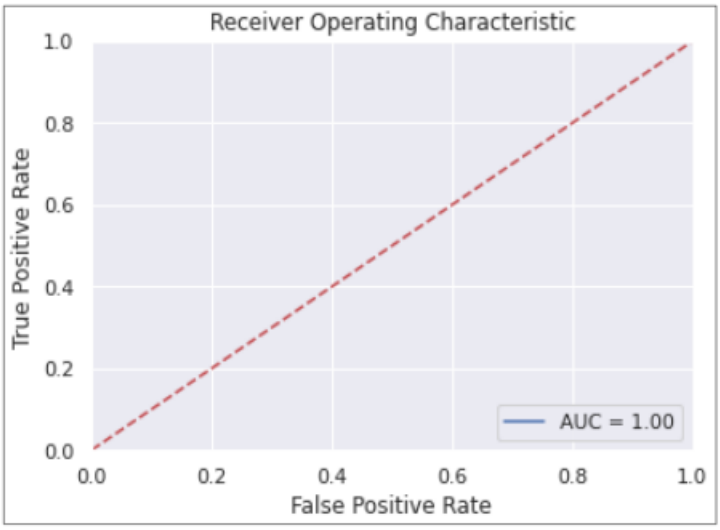

Fig. 7. The area under roc curve for Random Forest method

Algorithm	Accuracy	Precision	Recall	F1-score	AUC
4 -layer ANN	0.9239	0.8867	0.992	0.9364	0.97
3-layer ANN	0.8401	0.7709	0.974	0.8606	0.91

TABLE IV: COMPARISON BETWEEN THE DEEP LEARNING APPROACHES

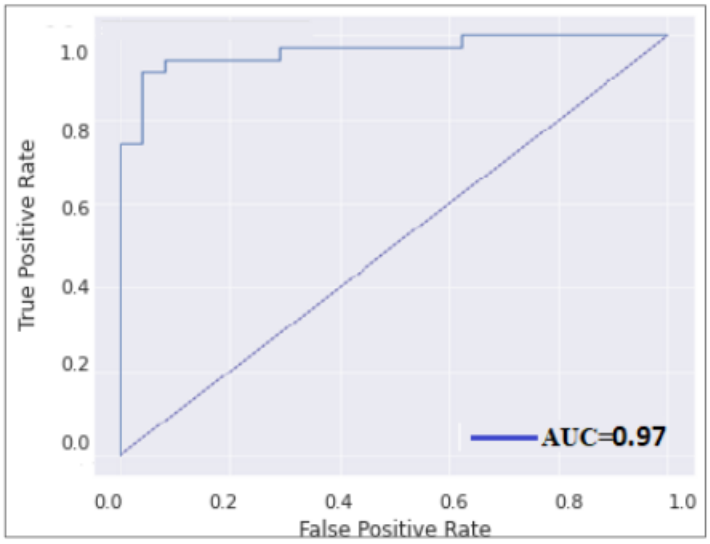

Fig. 8. The area under roc curve for 4-layer ANN method.

ISSN: 2583-6129
DOI: 10.55041/ISJEM02039

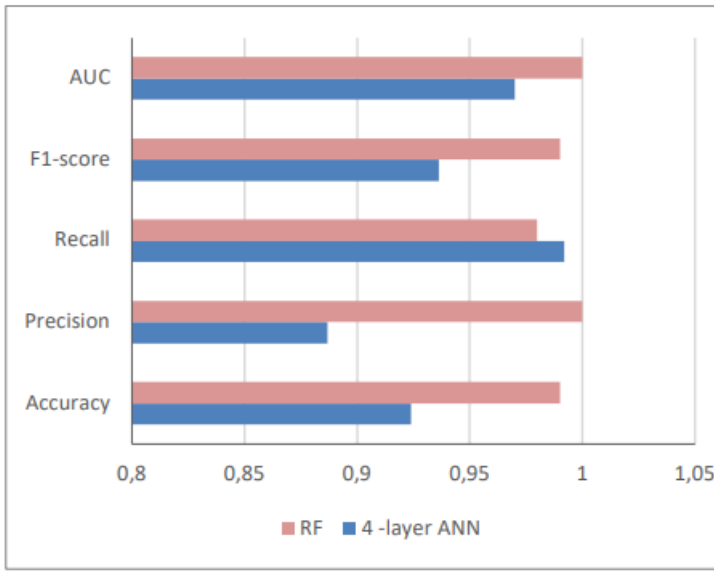

Fig. 9. A comparison chart of evaluation metrics of machine RF and 4-layer ANN algorithms

VII. Comparison with Existing Work

The proposed method for recognizing stroke patients demonstrates notable novelty when compared with previous studies in this field. Table V provides a comparative analysis, clearly showing that the proposed approach achieves higher performance metrics across the Random Forest, XGBoost, and 4-layer ANN models. These improved values underscore the originality and effectiveness of the current work.

VIII. CONCLUSION

Stroke is a potentially fatal medical condition that requires immediate treatment to prevent serious long-term consequences. Developing machine learning (ML) and deep learning models can aid in the early diagnosis of stroke and help mitigate its severe effects. This study evaluates the effectiveness of various ML and Boosting algorithms in predicting stroke based on different biological factors. The Random Forest classifier outperforms the other techniques investigated, achieving a classification accuracy of 99% and an AUC of 1. The findings indicate that the Random Forest method is superior to other methods in forecasting brain strokes using cross-validation measures.

References

[1] Pikula A, Howard BV, Seshadri S. Stroke and Diabetes. In: Cowie CC, Casagrande SS, Menke A, et al., editors. Diabetes in America. (3rd ed.). Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases (US), 2018, ch.19.

[2] Gary H, Gibbons L. National Heart, Lung and Blood Institute. 2022 [updated 2022 March 24]. Available from: https://www.nhlbi.nih.gov/health/stroke.

[3] Jeena RS, Kumar S. Stroke prediction using SVM, International Conference on Control. Instrumentation, Communication and Computational Technologies (ICCICCT), 2016: 600–602.

[4] Hanifa SM, Raja SK. Stroke risk prediction through non-linear support vector classification models. Int. J. Adv. Res. Comput. Sci., 2010; 1(3).

[5] Chantamit-o P, Madhu G. Prediction of Stroke Using Deep Learning Model. International Conference on Neural Information Processing, 2017: 774-781.

[6] Khosla A, Cao Y, Lin CCY, Chiu HK, Hu J, Lee H. An integrated machine learning approach to stroke prediction, in: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 2010: 183–192.

[7] Hung CY, Lin CH, Lan TH, Peng GS, Lee CC. Development of an intelligent decision support system for ischemic stroke risk assessment in a population-based electronic health record database. PLOS ONE, 2019;14(3):e0213007. https://doi.org/10.1371/journal.pone.0213007.

[8] Adam SY, Yousif A, Bashir MB. Classification of ischemic stroke using machine learning algorithms. International Journal of Computer Application, 2016;149(10):26–31.

[9] Singh MS, Choudhary P. Stroke prediction using artificial intelligence. 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), 2017:158–161.

[10] Emon MU, Keya MS, Meghla TI, Rahman MA, Mamun SA, Kaiser MS. Performance Analysis of Machine Learning Approaches in Stroke Prediction, International Conference on Enumerative Combinatorics and Applications, Nov. 2021.

[11] Kansadub T, Ammaboosadee S, Kiattisin S, Jalayondeja C. Stroke risk prediction model based on demographic data, in Proceedings of the 2015 8th Biomedical Engineering International Conference (BMEiCON), Pattaya, Thailand, November 2015: 1-3.

[12] Tazin T, Alam MN, Dola NN, Bari MS, Bourouis S, Khan M. Stroke Disease Detection and Prediction Using Robust Learning Approaches. Journal of Healthcare Engineering, 2021:1-12. doi: 10.1155/2021/7633381.

[13] Sharma C, Sharma S, Kumar M, Sodhi A. Early Stroke Prediction Using Machine Learning. International Conference on Decision Aid Sciences and Applications; Mar. 2022.

[14] Teoh D. Towards stroke prediction using electronic health records. BMC Medical Informatics and Decision Making, 2018; Dec.(1): 1–11. doi: 10.1186/s12911-018-0702-y.

ISSN: 2583-6129
DOI: 10.55041/ISJEM02039

[15] Hung CY, Lin CH, Lan TH, Peng GS, Lee CC. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017: 3110–3113.

[16] Fang G, Huang Z, Wang Z. Predicting Ischemic Stroke Outcome Using Deep Learning Approaches. Front Genet. 2022 Jan 24;12:827522. doi: 10.3389/fgene.2021.827522.

[17] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 1991 May-June; 21(3): 660-674. doi: 10.1109/21.97458.

[18] Navada A, Ansari AN, Patil S, Sonkamble BA. Overview of use of decision tree algorithms in machine learning. IEEE Control and System Graduate Research Colloquium, ICSGRC, 2011: 37–42.

[19] Rahman MM, Rana MR, Alam NAA, Khan MSI. A web-based heart disease prediction system using machine learning algorithms; 2022 June; 12. 64-80.

[20] Dhillon S, Bansal C, Sidhu B. Machine Learning Based Approach Using XGboost for Heart Stroke Prediction. in International Conference on Emerging Technologies: AI, IoT, and CPS for Science & Technology Applications, September 06–07, 2021.

[21] Akash K, Shashank HN, Srikanth S, Thejas AM. Prediction of Stroke Using Machine Learning. June 2020.

[22] Aiello S, Cliff C, Roark H, Rehak L, Stetsenko P, Bartz A. Machine Learning with Python and H2O. (5th Ed.). H2O.ai Inc. Nov. 2017.

[23] Sailasya G and Kumari G. L. A. Analyzing the performance of stroke prediction using ML classification algorithms. International Journal of Advanced Computer Science And Applications. 2021; 12(6): 539–545.

[24] Gurjar R, Sahana K, Sathish BS. Stroke Risk Prediction Using Machine Learning Algorithms. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2022: 20-25. doi: 10.32628/CSEIT2283121.

[25] Tavares J-A. Stroke prediction through Data Science and Machine Learning Algorithms. 2021; doi: 10.13140/RG.2.2.33027.43040.

[26] Martín J.R, Ayala J.L, Roselló G.R and Camarasaltas J. M. Comparison of Different Machine Learning Approaches to Model Stroke Subtype Classification and Risk Prediction. Spring Simulation Conference (SpringSim), pp. 1-10, 2019.