International Journal of Research and Development in Engineering Science
ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/
Regular Issue, Vol. 7, Issue.1 (January – February), 2025 , Pages: 46 - 52
Received: 25 July 2024 ; *Accepted:* 28 January 2025 ; *Published:* 23 February 2025
https://doi.org/10.63328/IJRDES-V7RI1P8
Original Paper : Inductive Research – Interpretative Research Paper

Enhancing Brain Stroke Prediction Using Machine Learning for Early Intervention

Bokka Suri Babu

Department of Computer Applications , Mohan Babu University, Tirupati, Andhra Pradesh, India;

* *Corresponding Author:* suribabubokka801@gmail.com

Abstract: Stroke is still a significant global health concern that requires sophisticated predictive methods for early detection and treatment. In this work, a novel machine learning (ML) framework for automated stroke prediction is presented, and its accuracy and generalization skills are evaluated against those of six well-known classifiers. In order to guarantee clear decision-making in clinical applications, SHAP and LIME approaches are also used to highlight model interpretability. By combining local and global analytical approaches, the suggested framework improves the standardization of intricate machine learning models. Notably, Random Forest routinely achieves higher predicting accuracy than other algorithms. An enhanced ensemble strategy that uses a voting mechanism to leverage numerous classifiers and incorporates CATBOOST and a Stacking Classifier is presented in order to further increase performance. This study offers a thorough and trustworthy approach to early stroke diagnosis and treatment, which will ultimately lessen the serious health and financial effects of this common illness.
Keywords: Data Integrity, Duplication detection, Data Validation, Anomaly Detection, Automated alerts.

1. Introduction

Stroke is one of the leading causes of death and permanent disability worldwide, making it a major health problem. Its rising incidence emphasizes how urgently sophisticated prediction and preventative techniques are needed. A disturbance in blood flow to the brain causes a stroke, which deprives brain cells of oxygen and can cause fatalities or serious neurological impairment. Conventional methods for predicting the risk of stroke frequently depend on manual evaluations, which can be laborious, prone to human error, and unable of deciphering intricate patterns in patient data. More effective, automated, and precise techniques to forecast stroke risk and facilitate early intervention are therefore in greater demand. With data-driven solutions for detecting high-risk patients based on a variety of clinical and demographic characteristics, machine learning (ML) and artificial intelligence (AI) have become game- changing tools in the healthcare industry. Large datasets can be processed by ML algorithms, which can also uncover hidden correlations between risk factors and produce accurate stroke risk forecasts that help with prompt decision-making. To assess a person's risk of stroke, these models take into account a number of factors, such as age, hypertension, cholesterol, diabetes, smoking, obesity, and prior medical history. Healthcare professionals can lower the incidence of strokes and increase patient survival rates by utilizing predictive analytics to conduct early interventions, lifestyle changes, and individualized treatment strategies.

Concerns regarding the interpretability and transparency of AI-driven models have also been raised by the growing use of ML in medical applications. ML models frequently operate as "black boxes," making it challenging for academics and medical professionals to comprehend how a prediction was made. Advanced interpretability methods like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) have been developed to overcome this difficulty. These techniques help doctors make well-informed treatment decisions by revealing the elements that most influence a patient's risk of stroke and boosting confidence in ML-based diagnostics. Explainable AI makes stroke prediction models more practical in real-world medical settings by ensuring that they not only produce accurate results but also provide concise, understandable explanations for their forecasts. The World Stroke Organization estimates that 13 million people have a stroke each year, which leads to 5.5 million fatalities. In addition to affecting individuals, this illness has a major negative impact on their families, healthcare systems, and economies around the globe. Long-term disabilities that impact a stroke survivor's quality of life and capacity to work or interact with others include

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 1,2025 , https://doi.org/10.63328/IJRDES-V7RI1P8

speech problems, paralysis, memory loss, and emotional discomfort. Stroke can happen to anyone, regardless of age, gender, or general health, despite the widespread belief that it only happens to the old or those with pre-existing illnesses. The need for improved knowledge, lifestyle changes, and early risk detection techniques is underscored by the rising prevalence of stroke in younger populations.

Using cutting-edge classification methods including Random Forest, CATBOOST, and Stacking Classifiers, this study suggests a novel machine learning (ML)-based stroke prediction framework that improves predictive accuracy and dependability. This strategy integrates several machine learning models to capitalize on their unique strengths by utilizing ensemble learning approaches, which produces a prediction system that is more resilient and broadly applicable. This work also highlights the need of both local and global analytical techniques in standardizing complicated machine learning models and guaranteeing their adaptability to various patient populations. Additionally, the study investigates how feature selection and data pretreatment methods might enhance model performance and reduce prediction bias. By combining big data analytics, AI- driven diagnostics, and explainable machine learning models, this research aims to transform stroke risk assessment and create a thorough, effective, and comprehensible stroke prediction system. This strategy seeks to lower the number of stroke-related fatalities, decrease long-term impairments, and lessen the financial strain on healthcare systems by promoting early detection, individualized treatment programs, and preventative care. To sum up, this study's results mark a substantial breakthrough in medical AI and a hopeful step toward a time when technology-driven healthcare solutions can improve clinical decision-making, improve patient outcomes, and ultimately save lives.

Types of Brain Strokes :

Ischemic Stroke: The most prevalent kind, which happens when a plaque or blood clot obstructs a blood vessel supplying the brain, resulting in a decreased oxygen supply and perhaps the loss of brain cells. It is responsible for almost 87% of all strokes and is frequently brought on by embolism or atherosclerosis.

Hemorrhagic Stroke: This condition is brought on by the rupture of a weak blood vessel in the brain, which puts pressure on the brain's tissues and causes bleeding. This category covers subarachnoid hemorrhage, which occurs in the area around the brain, and intracerebral hemorrhage, which occurs inside the brain and is frequently brought on by aneurysms or hypertension.

Transient Ischemic Attack: A "mini- stroke" is a brief interruption of blood supply to the brain that goes away in a day without causing long-term harm. Even though the symptoms go away, a transient ischemic attack (TIA) is a marker of a future stroke and needs to be treated right once to avoid worsening episodes.

Brain Stem Stroke: A brain stem stroke impairs vital processes like movement, breathing, and heartbeat. It may be ischemic or hemorrhagic in origin and can result in severe paralysis, vision issues, speech difficulties, and loss of coordination.

Cryptogenic Stroke: A stroke that, despite intensive medical investigation, has no known etiology. Advanced imaging and monitoring are necessary for an appropriate diagnosis because it is frequently considered to be connected to concealed embolisms, blood clotting abnormalities, or atrial fibrillation.

Cerebral Venous Sinus Thrombosis: An uncommon kind of stroke brought on by a blood clot in the venous sinuses of the brain that obstructs normal blood flow. Increased pressure and possible brain edema result from this, and it is frequently connected to coagulation issues, infections, or pregnancy.

2. Literature Survey

The primary goal of this research paper [1] is to use machine learning and deep learning techniques to forecast the risk of brain stroke early on. The work makes use of a stroke prediction dataset from Kaggle that is openly accessible and comprises 5110 records with 12 attributes. Extreme Gradient Boosting, AdaBoost, Light Gradient Boosting Machine, Random Forest, Decision Tree, Logistic Regression, K-Nearest Neighbours, Support Vector Machine, Naïve Bayes, and deep neural networks (3-layer and 4-layer ANN) were among the diverse categorization models used.

Model performance was assessed using accuracy, precision, recall, F1-score, and AUC following the application of data preparation techniques such as Principal Component Analysis (PCA) and oversampling to balance the dataset. The results demonstrated the efficacy of ML approaches in stroke prediction, with the 4- layer ANN achieving 92.39% accuracy and the Random Forest classifier achieving the maximum accuracy of 99%, beating other ML models. This study paper's primary goal is to present a thorough analysis of deep learning (DL) and artificial intelligence (AI) methods for brain stroke diagnosis, detection, and post- stroke rehabilitation [2]. The study examines a number of brain stroke datasets,

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 1,2025 , https://doi.org/10.63328/IJRDES-V7RI1P8

including those from ISLES, UCLA, Loma Linda University (LLU), and others, as well as data collected via CT, MRI, EEG, and DSA. The study examines a number of deep learning models, such as Generative Adversarial Networks (GAN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Deep Neural Networks (DNN). With CNN-based models being the most successful for stroke detection and AI-driven robotic systems supporting post-stroke rehabilitation, the study suggests that AI and DL techniques greatly enhance stroke diagnostic and rehabilitation procedures.

By examining several physiological characteristics, the author of this study[3] aims to employ machine learning techniques to predict brain stroke. The Kaggle Stroke Prediction dataset, which is used in this study, goes through preprocessing procedures such label encoding, addressing missing values, and data splitting for testing and training. To create prediction models, a number of machine learning techniques were used, such as Support Vector Machine (SVM), K- Nearest Neighbors (KNN), Random Forest, Decision Tree, and Logistic Regression. The most successful model for stroke prediction among these was Random Forest, which had the highest accuracy of 95.5%. Additionally, Tkinter was used to create an intuitive user interface that lets users enter data for a real-time stroke risk prediction.

In order to predict brain stroke, the author [4] can focus on developing a hybrid system that combines feature selection techniques with machine learning classifiers. The study uses the Stroke Prediction Dataset from and the Synthetic Minority Over Sampling Technique (SMOTE) to ensure class balance.

Kaggle. The machine learning methods that were evaluated included Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Adaptive Boosting (Adaboost), and Extreme Gradient Boosting (XGBoost). Three feature selection techniques are compared in the study: Feature Importance (FI), Pearson Correlation (PC), and Mutual Information (MI). With an accuracy of 97.17%, the results show that Random Forest (RF) and Feature Importance (FI) work well together. This high accuracy indicates that the suggested approach can accurately forecast brain stroke, facilitating prompt detection and treatment in medical settings.

Using an ensemble machine learning strategy based on a soft voting classifier, this study[5] aims to construct a brain stroke detection model. The Stroke Prediction Dataset, which comprises 4,981 records with 11 attributes, is taken from the UCI Machine Learning Repository and used in this study. In order to increase accuracy, the suggested model combines the predictions of three classifiers Random Forest, Extremely Randomized Trees, and Histogram-Based Gradient Boosting using a weighted average technique. Using the Synthetic Minority Oversampling Technique (SMOTE) to balance the dataset, addressing missing values, and label encoding are all examples of data preprocessing. The findings demonstrate that the ensemble model based on soft voting outperformed individual classifiers and offered a more dependable technique for stroke diagnosis, with an accuracy of 96.88%.

By examining numerous medical risk factors, this study[6] aims to optimize brain stroke prediction through machine learning approaches. The Stroke Prediction Dataset from Kaggle is used in the study, and data pretreatment methods like normalization, oversampling of the minority class, and outlier removal are used to improve model performance. Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), Decision Tree, Naïve Bayes, and Logistic Regression were among the several supervised learning algorithms that were assessed. The models' accuracy, precision, recall, and AUC-ROC scores were used to evaluate their performance after they were trained using k-fold cross-validation.

According to the results, SVM was the best model for predicting strokes, achieving the greatest accuracy of 94.6% and an AUC of 0.99. By examining numerous medical risk factors, the author[7] primarily concentrates on optimizing brain stroke prediction through machine learning approaches. The Stroke Prediction Dataset from Kaggle is used in the study, and data pretreatment methods like normalization, oversampling of the minority class, and outlier removal are used to improve model performance. Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), Decision Tree, Naïve Bayes, and Logistic Regression were among the several supervised learning algorithms that were assessed. The models' accuracy, precision, recall, and AUC-ROC scores were used to evaluate their performance after they were trained using k-fold cross-validation. According to the results, SVM was the best model for predicting strokes, achieving the greatest accuracy of 94.6% and an AUC of 0.99.

In order to help with patient selection for endovascular therapy, this research [8] focuses on creating a layered multimodal framework for autonomous collateral scoring in brain stroke. Computed Tomography Angiography (CTA) scans from 116 patients are used in the study; the dataset was authorized by SCTIMST, India. The suggested approach uses Automated Machine Learning (AutoML) for patient-level predictions, Convolutional Neural Networks (CNN) for picture selection, and a multimodal model to compare occluded and contralateral

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 1,2025 , https://doi.org/10.63328/IJRDES-V7RI1P8

brain regions. The framework's ability to automate collateral evaluation and enhance stroke patient triage is demonstrated by the results, which reveal that it achieves 91.17% accuracy for multi-class collateral scores and 94.118% accuracy for binary-class collateral scores.

3. Problem Statement

Stroke affects millions of people annually and places a significant strain on healthcare systems, making it one of the top causes of disability and death globally. Early detection and risk assessment are essential for successful medical intervention and better patient outcomes because of its unexpected nature and quick start. Traditional stroke detection methods, however, mostly rely on clinical evaluations and sophisticated imaging methods like CT scans and MRIs, which, although useful, are frequently expensive, time-consuming, and necessitate the knowledge of highly qualified radiologists and neurologists. Reliance on manual interpretation of medical data raises the possibility of interobserver variability and human error, which could result in inconsistent prognostic assessments, misdiagnosis, or treatment delays.

Stroke risk factors are also complex, impacted by lifestyle choices, medical history, genetic susceptibility, and underlying disorders like diabetes, hypertension, and cardiovascular diseases. It is becoming more and more challenging for medical personnel to proactively detect high-risk individuals before they exhibit crucial symptoms due to the lack of standardized, real-time, and automated predictive technologies. Current stroke prediction machine learning models have a number of drawbacks that make it difficult to integrate them into actual clinical settings, such as poor generalization across a range of patient populations, high sensitivity to unbalanced datasets, model bias, and a lack of interpretability.

Furthermore, a lot of AI-powered prediction models function as "black boxes," offering little visibility into how they make decisions, which makes it challenging for clinicians to trust and follow their advice. The medical community finds it difficult to successfully apply AI-driven solutions in the absence of a strong, scalable, and explicable predictive framework that can reliably estimate stroke risk while preserving openness and flexibility. Millions of people are left vulnerable to the devastation caused by stroke in the absence of such a system, which raises mortality rates, causes long-term disability, raises hospitalization expenses, and lowers quality of life. These issues must be resolved immediately by creating a predictive system that guarantees dependability, equity, and simplicity of interpretation for medical practitioners in addition to improving accuracy and efficiency.

4. Methodology:

In order to facilitate early diagnosis and intervention, the main goal of this project is to create a sophisticated machine learning (ML) framework for automated stroke prediction. Stroke is a serious illness that can cause permanent impairments or even death. Therefore, minimizing its effects requires early discovery and control. The goal of this project is to develop a highly accurate and interpretable stroke prediction system by utilizing a variety of machine learning methods. Data collection, preprocessing, feature engineering, model training, performance assessment, interpretability analysis, ensemble optimization, and model deployment for clinical use are all steps in the process. In order to guarantee that the predictive model is strong, dependable, and practical in real-world applications, each of these processes is essential. Refer to Figure 1, which graphically depicts the entire workflow and highlights each stage from data availability checks to model deployment, to gain a better understanding of the stroke prediction process step-by-step. Each stage is explained in full below.

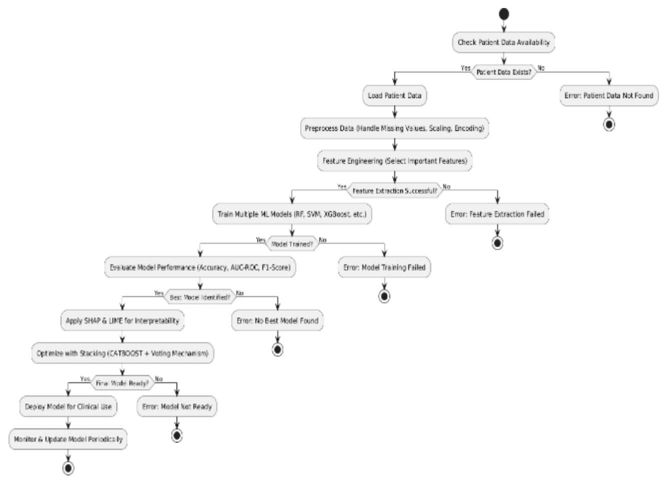

Figure.1 Systematic Flow of Work

1.1. Check Patient Data Availability

Verifying the availability of patient data is the initial step in the stroke prediction procedure. Data availability is crucial before moving forward since machine learning models use past medical records to determine stroke risk factors. To make sure patient datasets exist, the system runs a preliminary check. The procedure advances to the following stage if the required data is available. The system presents an error, though, stating that data gathering is necessary before proceeding if patient records are lacking. This stage guarantees that there is a solid basis of adequate and pertinent patient data upon which the entire predictive procedure is based.

1.2. Load and Preprocess Data

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 1,2025 , https://doi.org/10.63328/IJRDES-V7RI1P8

Following confirmation, patient data is entered into the system and undergoes preprocessing to guarantee precision and coherence. Model performance may be impacted by missing values, discrepancies, or scale changes that are frequently present in medical data. Imputation techniques, which fill in the gaps in the dataset with approximated values, are used in preprocessing to handle missing values. In order to preserve consistency between records, characteristics like blood pressure, cholesterol, and BMI are also standardized through scaling and normalization. In order to transform non-numeric variables, such as smoking status and gender, into a numerical format that machine learning algorithms can use, category encoding is used in medical datasets. By improving the quality of the data, this stage enables the model to train efficiently and produce precise predictions.

1.3. Feature Engineering

Following preprocessing, the most important attributes influencing stroke prediction are found and extracted using feature engineering. Choosing the most pertinent features enhances model accuracy and efficiency because not all patient attributes are equally significant. The factors that have the biggest impact on stroke risk, such as age, hypertension, or lifestyle choices, are identified using a variety of methods, including correlation analysis and recursive feature elimination. An error is raised and the dataset must be modified before continuing if feature extraction is unsuccessful because of irrelevant attributes or poor data quality. In order to help the model concentrate on important health indicators for stroke prediction, feature engineering makes sure that only relevant data is included.

1.4. Train Multiple ML Models

Several machine learning models are developed to predict the risk of stroke after key features have been extracted. The dataset is separated into subsets for testing and training; the testing data assesses the model's prediction power, while the training data aids in the model's learning of patterns from previous medical cases. Various algorithms are used to evaluate patient records and categorize people according to their risk levels, including Random Forest, Support Vector Machine (SVM), and XGBoost. Before attempting the training process again, modifications must be made if the model training procedure is unsuccessful because of inadequate data, incorrect parameter tuning, or computing constraints. Models that have been successfully trained move on to the evaluation phase, where they are tested for their ability to generalize well to new patient instances.

1.5. Evaluate Model Performance

An evaluation phase is carried out after model training to gauge the learned algorithms' dependability and forecast accuracy. The effectiveness of the algorithm in identifying stroke cases is assessed using a number of performance criteria. While the AUC-ROC score assesses the model's capacity to distinguish between patients at high and low risk, accuracy gauges the overall correctness of predictions. Because it strikes a balance between precision and recall, the F1-score is also taken into account, especially when working with unbalanced datasets where stroke cases may be less common. Alternative machine learning approaches or further optimizations are investigated to improve a model's performance if it has low predictive power. Only the most dependable models are chosen for additional development thanks to this stage.

1.6. Apply SHAP & LIME for Interpretability

Model interpretability is essential for medical applications because medical practitioners must comprehend the reasoning behind a model's predictions. SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) are used in this step to shed light on how the chosen model was chosen. SHAP provides clarity in risk assessment by identifying the most important factors that influence stroke prediction, such as age, high blood pressure, and prior medical disorders. LIME, on the other hand, examines local decision limits inside the model to produce explanations that are accessible by humans. By ensuring that the model's predictions are not only accurate but also comprehensible and practical, this interpretability phase fosters confidence among medical practitioners.

1.7. Optimize with Stacking (CATBOOST + Voting Mechanism)

Following the selection of an appropriate model, stacking strategies and ensemble learning are used to carry out additional optimization. To increase forecast accuracy, several algorithms are merged rather than depending on a single model. In order to take use of their combined advantages, the stacking process combines many models, such as CATBOOST, XGBoost, and Logistic Regression. In order to reach a final judgment, projections from several models are combined using a voting method. This approach lowers the possibility of inaccurate classifications while increasing reliability. Before moving on, hyperparameter tweaking and other adjustments are undertaken if the model is still unable to achieve the necessary accuracy levels. The ultimate model will have the best level of accuracy and resilience thanks to this optimization procedure.

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 1,2025 , https://doi.org/10.63328/IJRDES-V7RI1P8

1.8. Deploy Model for Clinical Use

The model is implemented in a clinical setting for practical application when it has been thoroughly refined and verified. Doctors can enter patient data and get immediate stroke risk assessments because to the stroke prediction system's integration with hospital software. During the deployment phase, a user-friendly interface must be developed so that medical professionals may easily interact with the model. With this stage, the stroke prediction model moves from a research prototype to a useful clinical tool that supports risk management and early detection. The concept promotes prompt medical intervention and proactive stroke prevention by making it available to healthcare practitioners.

1.9. Monitor & Update Model Periodically

To preserve the accuracy and efficacy of the model after deployment, ongoing monitoring and updates are necessary. The model is periodically retrained to accommodate changing medical trends as fresh patient data becomes available. To make sure the system is dependable and does not lose accuracy over time, performance measurements are examined on a regular basis. Updates to the algorithm are also made to take into account the most recent developments in stroke prediction research. By providing constantly accurate and current risk assessments, this continuous maintenance guarantees that the model will continue to be a useful tool in the healthcare industry.

5. Experimental Results

In order to focuses on Accuracy, Precision, Recall, and F1-score to evaluate the effectiveness of several machine learning models for stroke prediction. Several models were tested, including Random Forest, Logistic Regression, SVM, K Nearest Neighbors (KNN), Naïve Bayes, XGBoost, CatBoost, and a Stacking Classifier. The findings show that ensemble models perform well, especially CatBoost and the Stacking Classifier. The Stacking Classifier is the best model for stroke prediction, with nearly flawless scores (0.999) on all measures are mentioned in below Fig 2. Simpler models, such as Naïve Bayes and Logistic Regression, on the other hand, are less accurate and precise, which makes them less dependable when managing complex medical data.

A thorough comparison of many models across important performance indicators is shown in this Figure 3. With an accuracy, precision, recall, and F1-score of 0.999, the Stacking Classifier and CatBoost models exhibit the best performance.

ML Model	Accuracy	Precision	F1_score	Recall
Random Forest	0.952	0.952	0.952	0.952
Logistic Regression	0.777	0.777	0.777	0.778
SVM	0.809	0.808	0.808	0.812
KNN	0.927	0.926	0.927	0.932
Naive Bayes	0.752	0.751	0.751	0.752
XGBoost	0.895	0.895	0.895	0.897
Extension CatBoost	0.960	0.960	0.960	0.960
Extension Stacking Classifier	0.999	0.999	0.999	0.999

Figure.2 Algorithmic Analysis

This indicates that nearly flawless predictions are being made by the model. With an accuracy of 0.96, CatBoost outperforms conventional classifiers by a wide margin. Although Random Forest and KNN also exhibit good accuracy (over 0.90), the stacking strategy outperforms them. However, Naïve Bayes and Logistic Regression fare the lowest, highlighting their shortcomings when dealing with complicated datasets.

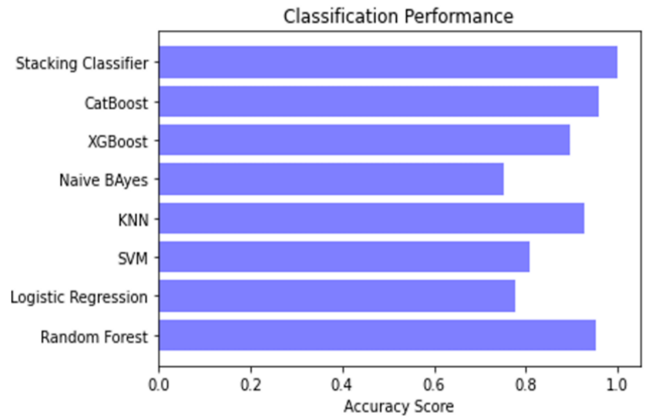

Figure.3 Classification Performance Analysis

Each model's accuracy scores are displayed visually in the bar chart. It is evident that while Logistic Regression and Naïve Bayes have the lowest accuracy, Stacking Classifier and CatBoost have the greatest. According to the difference between ensemble models and traditional models, adding several classifiers increases prediction accuracy overall. Although they still perform well, XGBoost, KNN, and Random Forest are not as effective as the stacking approach.

The precision chart assesses each model's ability to prevent false positives. When a model has a high precision score, it may accurately identify stroke instances without incorrectly classifying an excessive number of non-stroke cases as strokes. With the greatest precision (0.999) once more, the Stacking Classifier shows incredibly low false positive rates. With a precision of 0.96, CatBoost comes in second, while Naïve Bayes and Logistic Regression have

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 1,2025 , https://doi.org/10.63328/IJRDES-V7RI1P8

lower precision and are hence more likely to provide false positive results.

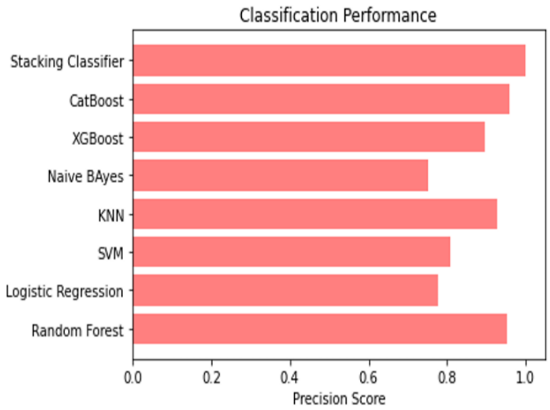

Figure.4 Result Analysis Data Classification

6. Conclusion

This study effectively used machine learning techniques to create a predictive model for brain stroke detection, showcasing the potential of data-driven healthcare solutions. The algorithm successfully identifies people who are at high risk of stroke by examining important variables like age, body mass index, smoking status, heart disease, hypertension, and other pertinent parameters. The most accurate and dependable method for stroke prediction was chosen as a result of performance comparison and evaluation made possible by the application of many machine learning algorithms. The findings demonstrate how important early identification is for averting serious health issues and enhancing patient outcomes. By helping medical practitioners make well-informed judgments, the predictive model can facilitate prompt intervention and individualized treatment regimens. Furthermore, this study highlights how crucial it is to include machine learning into medical diagnostics, demonstrating how it might improve conventional healthcare procedures.

Even if the model's accuracy was encouraging, there is still room for improvement. Predictions could be improved by expanding the dataset to include patient history, lifestyle characteristics, and real-time clinical data. Performance could also be further optimized by adjusting hyperparameters, using deep learning methods, and utilizing sophisticated ensemble models. Better transparency and confidence in model predictions may also result from the application of explainable AI approaches, which would increase the model's suitability for practical medical applications. This study represents a major advancement in the use of artificial intelligence for stroke prediction. By facilitating early diagnosis, lowering death rates, and lessening the long- term effects of strokes on patients and healthcare systems, these models have the potential to completely transform healthcare with more development and practical applications.

References

[1]. Senjuti Rahman, Mehedi Hasan, and Ajay Krishno Sarkar. “Prediction of Brain Stroke using Machine Learning Algorithms and Deep Neural Network Techniques”. Researchgate.2023.

[2]. Jyotismita Chaki, Marcin Wozniak. “Deep Learning and Artificial Intelligence in Action (2019–2023): A Review on Brain Stroke Detection, Diagnosis, and Intelligent Post-Stroke Rehabilitation Management”. IEEE.2024.

[3]. Sathya Sundaram, Pavithra, Poojasree, Priyadharshini. “STROKE PREDICTION USING MACHINE LEARNING”. Researchgate.2022.

[4]. Priyanka Bathla, Rajneesh Kumar. “A hybrid system to predict brain stroke using a combined feature selection and classifier”. ScienceDirect.2024.

[5]. A. Srinivas, Joseph Prakash Mosiganti. “A brain stroke detection model using soft voting based ensemble machine learning classifier”. ScienceDirect.2023.

[6].Venkata Sravan Telu, Vinay Padimi and Devarani Devi Ningombam. “Optimizing Predictions of Brain Stroke Using Machine Learning”. Researchgate.2022.

[7].Sayyed Saleh Sayyed Mousavi, Mohammad Saeed Majedi. “A Quantitative Microwave Imaging Approach for Brain Stroke Classification Based on the Generalized Tikhonov Regularization”. IEEE.2024.

[8]. Rishi Raj, Dayananda Pruthviraja, Ayush Gupta, Jimson Mathew, Santhosh Kumar Kannath, Adity Prakash, Jenyrajan.