(RESEARCH ARTICLE)

Brain stroke detection using neural networks: A deep learning approach for early diagnosis and prevention

Vishal Sahani * and Jasleen Chaudhary

Department of Biomedical Sciences, Student, Thapar Institute of Engineering and Technology, Punjab, India.

Abstract

This study investigates the effectiveness of three machine learning algorithms—Naive Bayes, Decision Tree, and Random Forest—for stroke detection based on patient data. We utilize a publicly available dataset from Kaggle that includes clinical features such as age, gender, hypertension, and heart disease history. Each model is evaluated on its Area under the Receiver Operating Characteristics (AUROC), accuracy, precision, recall, and F1-score to determine the best performer in predicting stroke risk.

1. Introduction

A stroke is a critical medical condition characterized by a disruption in blood flow to the brain, resulting in brain cell damage[1-4]. The two primary types of strokes are ischemic, which occur when blood vessels are blocked, and hemorrhagic, which arises from bleeding into or around the brain[3-6]. Symptoms can manifest quickly and may include weakness on one side of the body, difficulties with speech, dizziness, and vision impairment. A transient ischemic attack (TIA), or mini stroke, presents similar symptoms but lasts for a shorter duration. Several factors increase the risk of experiencing a stroke, with high blood pressure being the most significant. Additional risk factors include elevated cholesterol levels, obesity, diabetes, smoking, and a history of TIAs or heart conditions like atrial fibrillation. Diagnosis often involves a thorough physical examination alongside imaging tests such as CT or MRI scans, which help determine the type of stroke and its severity. Prompt diagnosis and treatment are essential for minimizing potential long-term complications, including mobility issues and loss of cognitive functions.

Stroke is a leading cause of morbidity and mortality worldwide, making early detection critical for effective intervention and improved patient outcomes. With the increasing availability of healthcare data, machine learning techniques have emerged as valuable tools for predicting stroke risk based on clinical and demographic features. Based on Mondal et al.'s work which discusses the prediction of aneurysm occlusion using neural networks, we design our study where we take into account different demographic features and predict brain stroke[7]. This paper explores the performance of three prominent algorithms—Naive Bayes, Decision Tree, and Random Forest—in detection of stroke using a dataset comprising various patient characteristics[5, 8, 9]. By comparing these algorithms, we aim to identify the most effective method for stroke detection, ultimately contributing to better clinical decision-making and enhanced patient care.

2. Material and Methods

The study utilizes a publicly available dataset from Kaggle containing various clinical features related to stroke risk, including demographic information, medical history, and lifestyle factors.

---

* Corresponding author: Vishal Sahani

We use three classical machine learning models for detection of stroke and do a comparative study.

Naive Bayes: This probabilistic classifier applies Bayes' theorem with the assumption that the features are independent given the class label. It calculates the posterior probability of each class based on the likelihood of the observed features, making it particularly effective for large datasets.

Decision Tree: This model uses a tree-like structure to make decisions based on feature values. Each node represents a feature, branches represent decision rules, and leaves represent outcomes. Decision Trees are intuitive and easy to interpret, making them useful for understanding the factors influencing stroke risk.

Random Forest: An ensemble method that combines multiple Decision Trees to improve predictive accuracy. It operates by constructing numerous trees during training and outputs the mode of their predictions. Random Forest mitigates overfitting issues associated with individual Decision Trees and enhances robustness.

Each model is implemented using the Python library scikit-learn. The dataset is split into training and testing subsets, allowing model training and validation. Performance metrics such as accuracy, precision, recall, and F1-score are used to evaluate each model's effectiveness in detecting stroke risk[10]. Cross-validation techniques are applied to ensure the reliability of the results. Figure 1 describes the dataset and Figure 2 gives the workflow for the study.
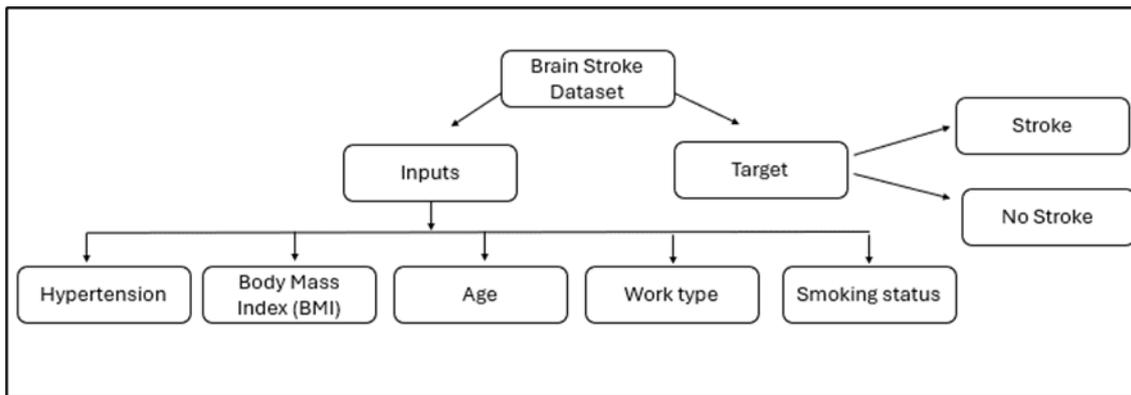
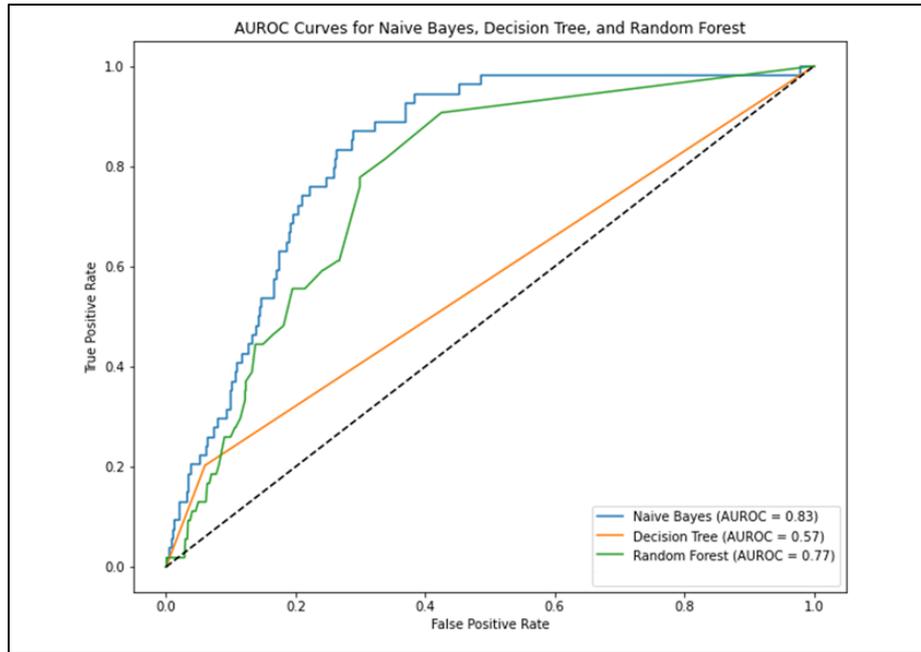**Figure 1** Demonstrates the dataset.
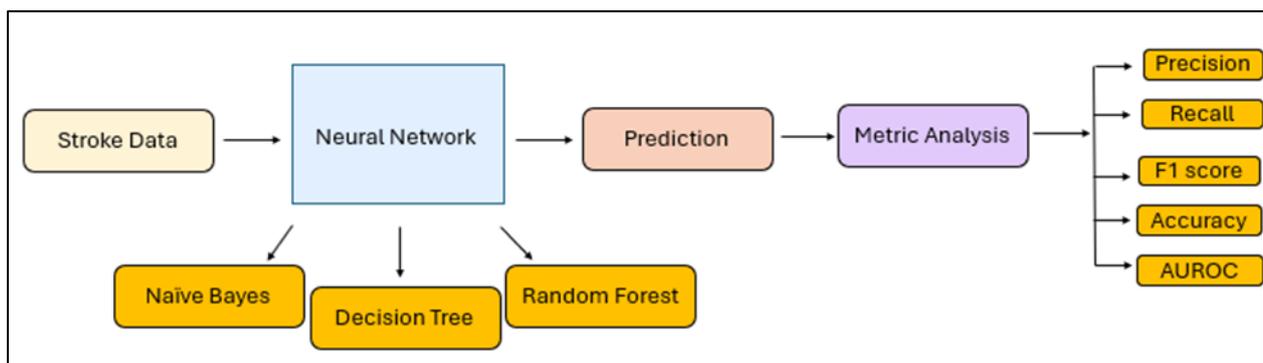
Figure 2 Demonstrates the workflow for the study.

3. Results

Figure 3 gives the AUROC for different neural networks. Table 1 shows the performance evaluation metrics for different neural networks.

AUROC Curves for Naive Bayes, Decision Tree, and Random Forest

Figure 3 Demonstrates the AUROC for (A) Naïve Bayes (B) Decision Tree and (C) Random Forest respectively

Table 1 Demonstrates performance evaluation metrics.

Neural network	Accuracy	Precision	Recall	F1 score	AUROC
Naïve Bayes	0.88	0.75	0.70	0.72	0.83
Decision Tree	0.89	0.69	0.60	0.66	0.57
Random Forest	0.93	0.84	0.82	0.80	0.77

4. Conclusion

In conclusion, this study highlights the effectiveness of machine learning algorithms in stroke detection, comparing Naive Bayes, Decision Tree, and Random Forest models. These findings underscore the potential of leveraging data-driven approaches to enhance early stroke detection and inform timely medical interventions, ultimately improving patient outcomes. Future work may focus on incorporating additional features and refining these models to further enhance their predictive capabilities

Compliance with ethical standards

Acknowledgments

The Authors would like to thank the heads of the institution, teachers, staff and students for their co-operation and support.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

Statement of ethical approval

The study was approved by the Institutional Ethics Committee

Statement of informed consent

Informed consent was obtained from all participants included in the study

References

[1]   Chin C-L, Lin B-J, Wu G-R, Weng T-C, Yang C-S, Su R-C, Pan Y-J: An automated early ischemic stroke detection system using CNN deep learning algorithm. In: *2017 IEEE 8th International conference on awareness science and technology (iCAST): 2017*: IEEE; 2017: 368-372.

[2]   Gaidhani BR, Rajamenakshi R, Sonavane S: Brain stroke detection using convolutional neural network and deep learning models. In: *2019 2nd International conference on intelligent communication and computational techniques (ICCT): 2019*: IEEE; 2019: 242-249.

[3]   Gupta S, Mishra A, Menaka R: Ischemic stroke detection using image processing and ANN. In: *2014 IEEE international conference on advanced communications, control and computing technologies: 2014*: IEEE; 2014: 1416-1420.

[4]   Pereira DR, Reboucas Filho PP, de Rosa GH, Papa JP, de Albuquerque VHC: Stroke lesion detection using convolutional neural networks. In: *2018 International joint conference on neural networks (IJCNN): 2018*: Ieee; 2018: 1-6

[5]   Karthik R, Menaka R, Johnson A, Anand S: Neuroimaging and deep learning for brain stroke detection-A review of recent advancements and future prospects. *Computer Methods and Programs in Biomedicine* 2020, 197:105728.

[6]   Qureshi AA, Zhang C, Zheng R, Elmeligi A: Ischemic stroke detection using EEG signals. In: *CASCON: 2018*; 2018: 301-308.

[7]   Mondal P, Shiraz Bhurwani MM, Setlur Nagesh SV, Lai PMR, Davies JM, Levy EI, Vakharia K, Levitt M, Siddiqui AH, Ionita CN: Minimizing human-induced variability in quantitative angiography for a robust and explainable AI-based occlusion prediction in flow diverter-treated aneurysms. *Journal of NeuroInterventional Surgery* 2025:jnis-2025-023416.

[8]   Kumar GR, Vyshnavi P, Prasanna S, Reddy TH, Charanya C, Chandrababu P: Brain Stroke Detection Using Machine Learning. *International Journal of Research in Engineering, Science and Management* 2022, 5(3):34-36.

[9]   Mamatha B, Sreelatha R, Saravanamuthu M: Brain stroke detection using machine learning. *Int J Creative Res Thoughts (IJCRT)* 2022, 10(8):287-297.

[10]   Parisa N, Mohammad Mahdi Shiraz B, Ahmad R, Parmita M, Michael HU, Kyle AW, Swetadri Vasan Setlur N, Ciprian NI: Exploring multimodal data fusion and augmentation strategies for imbalanced datasets for intraoperative aneurysm occlusion prognosis. In: *ProcSPIE: 2025*; 2025: 134100U.