RESEARCH PAPER

AI-Based Stroke Risk Prediction Using Ensemble Machine Learning

[1]Umme Summaiyya and [2]Dr. Mohammed Abdul Waheed

[1]PG Student, Department of CSE, Visvesvaraya Technological University, Regional office Kalaburagi, Kalaburagi, India
[2]Associate Professor, Department of CSE,Visvesvaraya Technological University, Regional office Kalaburagi, Kalaburagi, India
[1]ummesummaiyya090@gmail.com and [2]mawaheed@vtu.ac.in

Received: 24th April, 2026; Revised: 20th May 2026; Accepted: 30th June, 2026; Available Online: 10th July, 2026

ABSTRACT
Stroke is main cause of death around the world, which highlights the importance of prediction and prevention of this illness. As for now, traditional stroke detection relies on the clinical symptoms and the experience of the doctor, which delays the process of recognizing the patients at high risk. In order to address this problem, the project "Stroke Prediction Using Ensemble Learning" suggests developing a novel intelligent healthcare prediction system based on machine learning to assess the possibility of stroke using health parameters of the patient. Thus, the project considers various medical attributes, to identify patients' risk of suffering from strokes. The proposed solution makes use of such techniques as Ensemble Learning, Random Forest, Gradient Boosting, AdaBoost, and Voting Classifier to improve the overall results and reduce the effect of the weaknesses associated with separate methods. In addition, the project includes such stages as preprocessing of data that involves handling of missing values, normalization, feature selection, and class balancing to increase the quality of data for predictions. Thanks to the training of the suggested ensemble, healthcare professionals will be able to detect complicated relationships in the medical data to predict patients' risk of strokes with maximum efficiency. Moreover, the system will have a convenient UI that will allow entering necessary parameters to obtain predictions immediately. Accuracy, precision, recall, F1-score, and a confusion matrix are used as indicators of effectiveness of the model. Developed system will help healthcare professionals to make decisions in the field and provide patients with better preventative care services to reduce mortality rates related to the stroke problem.

Keywords: Random Forest, Gradient Boosting, AdaBoost, age, hypertension, heart disease, glucose level, body mass index (BMI), smoking status, brain stroke

How to cite this article: Summaiyya U, Waheed MA, AI-Based Stroke Risk Prediction Using Ensemble Machine Learning. Int J Drug Deliv Technol. 2026;16(71s): 852-857. DOI: 10.25258/ijddt.16.71s.99
Source of support: Nil.
Conflict of interest: None

1. INTRODUCTION

Topic chosen for my healthcare project involves stroke – a life-threatening condition characterized by an obstruction or slow blood flow to the brain depriving brain tissues of oxygen and vital nutrients. It can lead to permanent damage of brain functions and disabilities. Stroke is among the top causes of mortality and disability on a global scale, according to healthcare statistics. Timely identification and treatment of strokes could significantly mitigate their effects [1].

Over the past decade, due to rapid advancements in machine learning and artificial intelligence, modern healthcare became more effective at predicting diseases and assisting doctors with decision-making. Using machine learning, one could analyze large amounts of medical data and identify patterns beyond the capabilities of a clinician. There are many different models of machine learning; however, the method called Ensemble Learning has shown promising results due to the ability to combine multiple algorithms [3].

The main purpose of the project called "Stroke Prediction Using Ensemble Learning" is developing an intelligent system that would predict the probability of stroke using patient information. Parameters such as age, hypertension, heart diseases, blood sugar levels, BMI, smoking habit, and related factors will be used in order to determine the stroke's likelihood. [5].

Among the main steps of the project are collecting health-related data, pre-processing data (handling missing values, normalizing data, balancing the dataset), selecting features, creating and training a model, testing the model, and analyzing its performance. The pre-processing stage will increase the efficiency of the process of developing a model, whereas evaluating the performance of the model will help ensure precision predictions [4].

Ultimate goal of the project is to help clinicians and patients determine stroke risks based on their personal characteristics and take necessary precautions in order to prevent stroke. Using ensemble learning and medical data analysis, it would be possible to provide a cost-effective, fast, and efficient solution to predicting and preventing strokes in patients. [6].

**Author for Correspondence: ummesummaiyya090@gmail.com*

LITERATURE SURVEY

M. Amrutha and K. S. Shashidhar [1], 2025, proposed a The model used machine learning algorithms to create predictive models with a precision of 92%. However, limitations include the use of small datasets and lack of clinical validation.

Brain Stroke Detection model uses machine learning techniques to detect stroke disease. V. K. Singh et al., [2] performed multiple machine learning approaches' comparative analysis for stroke detection, which yielded a 92% accuracy.

S. Bosubabu et al., [4] created Brain Stroke Prediction framework that can evaluate the probability of developing stroke diseases by machine learning and factoring analysis. This paper achieved an accuracy of 92%, although challenges related to generalization among different population groups remain an issue.

The approach of brain stroke prediction by A. K. Kadam et al., [5] involving usage of supervised ML techniques and the use of medical data for forecasting the chances of stroke occurrence with a precision rate of 91%. Some of th limitations of the paper include a small dataset and few optimizations of the features.

Brain Stroke Prediction utilising ML approach uses classification algorithms to predict the stroke disease. Real-time implementation of the algorithm and lack of Graphical User Interface were significant limitations.

A. M. Kunder et al., [7] have made attempts at predicting stroke risk using machine learning techniques along with clinical healthcare data with an accuracy of 88%. Lack of prediction accuracy and algorithm comparisons have been identified as two main limitations of the paper.

METHODOLOGY

As is evident from the title, the "Stroke Prediction Using Ensemble Learning" methodology follows a systematic process designed for the prediction of the likelihood of a stroke utilising ML techniques. The processes involve collecting and preparing the data, selecting features, training the models, employing ensemble methods, making predictions, and evaluating their performance.

1. Data Collection

To begin with, one requires a dataset consisting of patients' medical details useful for predicting a stroke. The attributes included within the dataset are:

- Age
- Gender
- Hypertension
- Heart disease
- Average glucose level
- Body Mass Index (BMI)
- Smoking status
- Work type
- Marital status
- Residence type
- Stroke status (target variable)

2. Data Preprocessing

Unprocessed medical datasets have several imperfections such as missing values, noise, duplicates, and inconsistences. Preprocessing helps transform the data from raw form into clean data to improve its quality. The processes involved in preprocessing include:

- Eliminating duplicated data entries

- Imputation of missing values using averages or medians

- Categorical encoding using Labeling techniques

- Normalization of numerical data for better data quality

3. Feature Selection

Feature selection refers to highlighting the variables that are critical in determining the likelihood of stroke occurrence. This process allows us to eliminate redundancy in the calculations and enhance accuracy of the model. Techniques like correlation analysis and importance ranking help in determining which features should be selected.

4. Ensemble Learning Model Development

Ensemble learning makes up the core of this system. This process involves combining different machine learning models to achieve greater accuracy than what could be achieved with just one model. Some of the models that have been combined include:

- Random Forest
- Gradient Boosting
- AdaBoost
- Voting Classifier

These models are trained on the well-structured health-care data set. Each machine learning model makes a prediction about stroke risk and combines all predictions to produce the final prediction.

5. Model Training and Testing

Training data is utilised by the models to learn from the patterns and correlations present in patient's health information. In contrast, the testing dataset is kept aside to measure the performance of the predictions made. Training of the models helps in discovering hidden relationships.

6. Stroke Prediction

After the system receives its required training, it will receive the patient's data and predict if the patient is likely to have a stroke. This prediction will be made as follows:

- Stroke Risk Detected
  or
- No Stroke Risk Detected

The system can also give probabilities that will give the likelihood of a stroke.

7. Performance Evaluation

The ensemble learning algorithm is evaluated using various methods of measurement:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

8. Result Analysis and Visualization

The last stage involves the visualization of the results of the predictions in graphs and charts such as:

- Accuracy comparison graphs
- Confusion matrix visualization
- Feature importance graphs
- ROC Curve

Graphical representation shows how the predictive model works as well as medical parameters that significantly affect the prediction of strokes.

This implies that the presented approach provides a practical and intelligent system for predicting strokes with high efficiency based on ensemble methods.

System Architecture

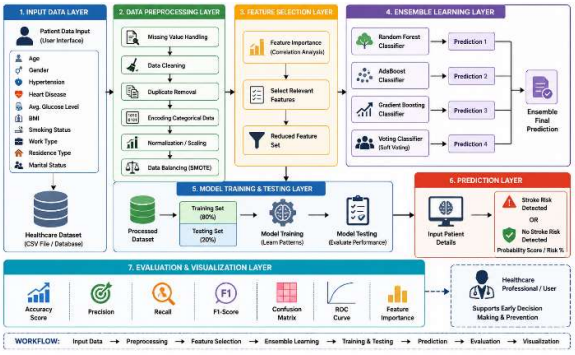

Figure 1: System Architecture

The suggested framework "Prediction of Stroke Using Ensemble Learning" follows a consistent narrative throughout the process of transforming health data into stroke prediction by combining health data and machine learning into one story. In the first place, preprocessing of the data is conducted which includes data cleaning, dealing with missing values, normalizing and encoding, and balancing. Then, the top features in the medical domain are determined and provided as inputs to an ensemble learning method that combines various techniques, such as random forest, AdaBoost, Gradient Boosting, and Voting Classifier. This framework uses learning models to train and make predictions based on the healthcare data associated with strokes.

RESULTS AND DISCUSSIONS

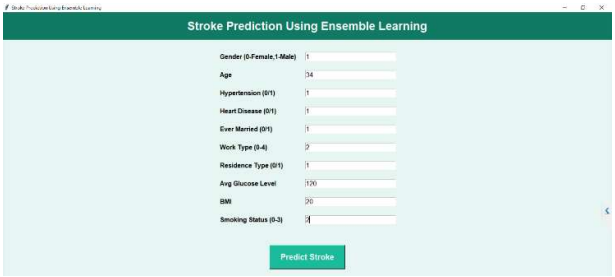

Figure 2: Input patient Data

Figure 3: Prediction

Classification Report

	precision	recall	f1-score	support
0	0.94	1.00	0.97	960
1	0.00	0.00	0.00	62
accuracy			0.94	1022
macro avg	0.47	0.50	0.48	1022
weighted avg	0.88	0.94	0.91	1022

Figure 4: Random Forest Classification Report

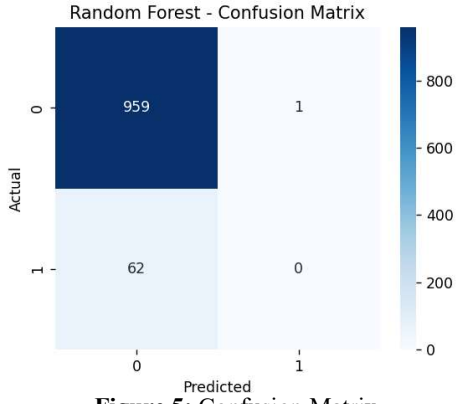

Figure 5: Confusion Matrix

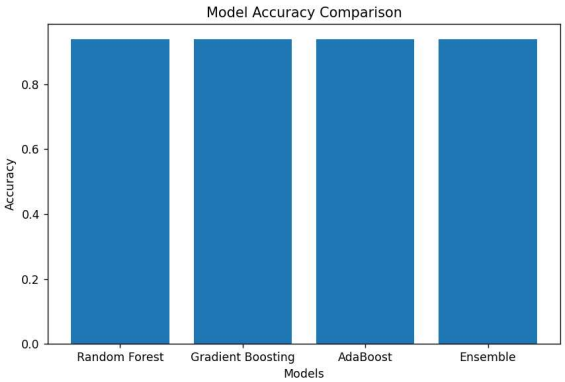

Figure 6: Model Accuracy comparison

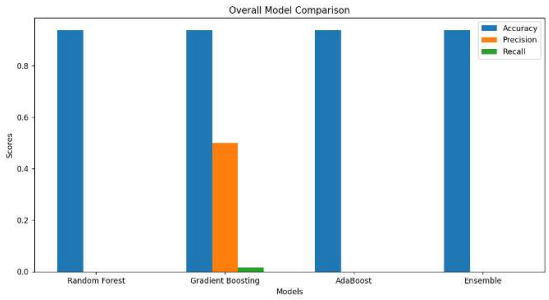

Figure 7: Performance Metrics

CONCLUSION

An idea of implementing the system named "Stroke Prediction Using Ensemble Learning" can be considered as an effective way to estimate a patient's stroke risk via machine learning. By considering a number of factors, including age, hypertension, cardiovascular diseases, glucose level, BMI, and smoker status, the system evaluates the patient's health state and detects his/her probability of having a stroke. The utilization of Ensemble Learning approaches like Random Forest, AdaBoost, Gradient Boosting, and Voting Classifier will help increase prediction accuracy and efficiency. It includes data preprocessing, feature selection, model training and testing, and careful evaluation. This system gives medical professionals an opportunity to make fast and well-informed preventive decisions, showing great importance of AI in contemporary medicine. Moreover, this system can ensure reliable stroke anticipation and reduce the mortality rate in future. The further studies might involve bigger datasets, more sophisticated image processing, and IoT technologies.

REFERENCES

[1] A. K. Kadam, P. Agarwal, Nishtha, and M. Khandelwal, "Brain Stroke Prediction Using Machine Learning Approach," IRE Journals, vol. 6, no. 1, pp. 273-277, Jul. 2022.

[2] A. Chavan, K. Yadav, C. A. Sonawane, and D. K. Sharma, "Brain Stroke Prediction using Machine Learning Algorithms," International Journal of Enhanced Research in Management & Computer Applications, vol. 12, no. 11, pp. 39-45, Nov. 2023.

[3] S. Bosubabu, S. Vasavi, M. Padmavathi, P. Aswini, and P. Sai Rohit, "Brain Stroke Prediction by Using Machine Learning," International Journal of Current Science (IJCSPUB), vol. 13, no. 3, pp. 815-821, Jul. 2023.

[4] N. Saxena, A. Choudhary, D. S. Bhamra, and P. Maru, "BrainOK: Brain Stroke Prediction using Machine Learning," Journal of Emerging Technologies and Innovative Research (JETIR), vol. 9, no. 4, pp. 143-148, Apr. 2022.

[5] M. Amrutha and K. S. Shashidhar, "Brain Stroke Prediction: A Phase 1 Report on Developing a Predictive Model for Brain Stroke Prediction," Journal of Emerging Technologies and Innovative Research (JETIR), vol. 12, no. 7, pp. 1184–1187, 2025.

[6] V. K. Singh, A. Kaur, and A. Larhgotra, "Brain Stroke Detection Using Machine Learning," International Journal of Novel Research and Development (IJNRD), vol. 9, no. 4, pp. 355–359, Apr. 2024.

[7] A. M. Kunder, H. N. Shashank, S. Srikanth, and A. M. Thejas, "Prediction of Stroke Using Machine Learning," Conference Paper, CMR Institute of Technology, Bangalore, June 2020.

[8] Soumyabrata Dev, Hewei Wang, Chidozie Shamrock Nwosu, Nishtha Jain, Bharadwaj Veeravalli, Deepu John. "A predictive analytics approach for stroke prediction using machine learning and neural networks."Healthcare Analytics, Volume 2, 2022, 100032.

[9] R. Pitchai, Bhasker Dappuri, P. V. Pramila, M. Vidhyalakshmi, S. Shanthi, Wadi B. Alonazi, Khalid M. A. Almutairi, R. S. Sundaram, Ibsa Beyene. "An Artificial Intelligence-Based Bio-Medical Stroke Prediction and Analytical System Using a Machine Learning Approach."Computational Intelligence and Neuroscience, Volume 2022, Article ID 5489084, 9 pages, 2022.

[10] Sathya Sundaram M., Pavithra K., Poojasree V., Priyadharshini S."Stroke Prediction Using Machine Learning."International Advanced Research Journal in Science, Engineering and Technology (IARJSET),Vol. 9, Issue 6, June 2022, pp. 141-148.

[11] Gangavarapu Sailasya, Gorli L. Aruna Kumari."Analyzing the Performance of Stroke Prediction using ML Classification Algorithms."International Journal of Advanced Computer Science and Applications (IJACSA),Vol. 12, No. 6, 2021.

[12] Nikita, Gaurav Parashar."Brain Stroke Detection and Prediction Using Machine Learning Approach: A Cloud Deployment Perspective."2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), IEEE, August 2023.

[13] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, "Utilization of machine learning algorithms for the classification of stroke diseases," Neural Computing and Applications, vol. 32, pp. 817–828, 2020.

[14] T. Tazin, M. N. Alam, N. N. Dola, M. S. Bari, S. Bourouis, M. Monirujjaman Khan et al, "Detection and prediction of stroke disease using advanced learning methods," Journal of Healthcare Engineering, vol. 2021, 2021