International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(12s) 2026 pp. 1174-1183
https://doi.org/10.70917/ijcisim-2026-3989
Article

Stroke prediction using Ensemble Learning

Bhagyashri Patil[1], Priyadarshini C Patil[2], Soumya M A[3], Sayyada Fahmeeda Sultana[4], Amareshwari Patil[5], Jyothi[6], Shashank[7]

[1]Department of Computer Network Engineering, PDA College of Engineering, Kalaburagi, Karnataka, India
Email: bhagyapatil749@gmail.com
[2]Department of Computer Science Engineering, PDA College of Engineering Kalaburagi, Karnataka, India
Email: priyadarshinicpatil@gmail.com
[3]Department of Computer Science Engineering, PDA College of Engineering Kalaburagi, Karnataka, India
Email: smanakal@gmail.com
[4]Department of Computer Science Engineering, PDA College of Engineering Kalaburagi, Karnataka, India
Email: sayyadafahmeeda@gmail.com
[5]Department of Computer Science Engineering, PDA College of Engineering Kalaburagi, Karnataka, India
Email: aspatilpdacse@gmail.com
[6]Department of Computer Network Engineering, PDA College of Engineering Kalaburagi, Karnataka, India
Email: jyothic0531@gmail.com
[7]Department of Computer Network Engineering, PDA College of Engineering Kalaburagi, Karnataka, India
Email: puranikshashank08@gmail.com

Abstract: Stroke is one of the leading causes of death and long-term disability worldwide, making early prediction and preventive healthcare extremely important. Traditional methods of stroke diagnosis mainly depend on clinical observations and medical expertise, which may sometimes delay timely identification of high-risk patients. To address this challenge, the proposed project "Stroke Prediction Using Ensemble Learning" introduces an intelligent healthcare prediction system that utilizes machine learning techniques to predict the likelihood of stroke occurrence based on patient health parameters. The system analyzes important medical attributes such as age, hypertension, heart disease, glucose level, body mass index (BMI), smoking status, and other lifestyle-related factors to provide accurate stroke risk assessment. The proposed model employs Ensemble Learning techniques, which combine multiple machine learning algorithms to improve prediction accuracy and robustness. Algorithms such as Random Forest, Gradient Boosting, AdaBoost, and Voting Classifier are integrated to enhance overall system performance and reduce the limitations of individual models. Data preprocessing techniques including missing value handling, normalization, feature selection, and class balancing are applied to improve the quality of the dataset and ensure reliable predictions. The trained ensemble model is capable of identifying complex patterns in medical data and classifying patients into stroke-risk categories with high efficiency. The system is designed with a user-friendly interface that allows healthcare professionals or users to input patient information and obtain instant prediction results. Performance evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix are used to assess the effectiveness of the proposed model. By providing early stroke risk prediction, the system can support doctors in decision-making, promote preventive healthcare, and reduce mortality rates associated with stroke. The proposed approach demonstrates how ensemble learning and artificial intelligence can contribute significantly to modern healthcare systems by enabling faster, more accurate, and cost-effective medical predictions.

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

1. Introduction

Stroke is a serious medical condition that occurs when the blood supply to the brain is interrupted or reduced, preventing brain tissues from receiving oxygen and nutrients. This condition can lead to permanent brain damage, disability, or even death if not diagnosed and treated at an early stage. According to global health reports, stroke is one of the leading causes of mortality and long-term disability worldwide. Factors such as hypertension, diabetes, heart disease, obesity, smoking, stress, and unhealthy lifestyle habits significantly increase the risk of stroke. Early

prediction and timely medical intervention can greatly reduce the severity and complications associated with this disease [1][2].

In recent years, the rapid growth of artificial intelligence and machine learning technologies has transformed the healthcare sector by enabling intelligent disease prediction and decision-support systems. Machine learning algorithms are capable of analyzing large volumes of medical data and identifying hidden patterns that may not be easily recognized through traditional clinical methods. Among various machine learning approaches, Ensemble Learning has gained significant attention due to its ability to combine multiple models and produce more accurate and reliable predictions than individual algorithms [3][4].

The proposed project, "Stroke Prediction Using Ensemble Learning," aims to develop an intelligent prediction system that can estimate the likelihood of stroke occurrence using patient health information. The system utilizes medical parameters such as age, hypertension, heart disease, glucose level, body mass index (BMI), smoking status, and other related health conditions to predict stroke risk. Ensemble learning techniques such as Random Forest, AdaBoost, Gradient Boosting, and Voting Classifier are employed to improve prediction accuracy and reduce model errors [5][6].

The system follows several important stages including data collection, preprocessing, feature selection, model training, testing, and performance evaluation. Data preprocessing techniques such as handling missing values, normalization, and balancing the dataset help improve the efficiency of the prediction model. The trained ensemble model is evaluated using performance metrics like accuracy, precision, recall, F1-score, and confusion matrix to ensure reliable prediction results [2][4].

The main objective of this project is to assist healthcare professionals and individuals in identifying stroke risk at an early stage so that preventive measures and medical treatments can be taken promptly. By integrating ensemble learning techniques with healthcare data analysis, the proposed system provides a cost-effective, fast, and efficient solution for stroke prediction and supports the advancement of intelligent healthcare systems [6][7].

2. Literature Survey

M. Amrutha and K. S. Shashidhar [1], 2025, proposed a Brain Stroke Prediction system using machine learning techniques to analyze patient medical attributes and predict stroke risk at an early stage. The study developed predictive

models using classification algorithms and achieved an accuracy of 92%. However, the system was limited by the use of a small dataset size and lacked real-time clinical validation.

V. K. Singh, A. Kaur, and A. Lahrgotra [2], 2024, presented a Brain Stroke Detection system using Machine Learning algorithms for detecting stroke-related conditions. The study implemented different machine learning approaches and performed comparative analysis for stroke detection, achieving an accuracy of 92%. The limitation of the work was that it did not include deep learning techniques or medical imaging-based analysis.

A. Chavan, K. Yadav, C. A. Sonawane, and D. K. Sharma [3], 2023, developed a Brain Stroke Prediction system using Machine Learning algorithms such as Decision Tree, Random Forest, and Logistic Regression for stroke risk prediction. The proposed approach achieved an accuracy of 91%. However, the prediction performance highly depended on proper data preprocessing and dataset quality.

S. Bosubabu, S. Vasavi, M. Padmavathi, P. Aswini, and P. Sai Rohit [4], 2023, proposed a Brain Stroke Prediction by Using Machine Learning system for predicting stroke risk using healthcare datasets and risk factor analysis. The proposed machine learning framework achieved an accuracy of 92%. However, the system had limitations in generalization for diverse patient populations.

A. K. Kadam, P. Agarwal, Nishtha, and M. Khandelwal [5], 2022, developed a Brain Stroke Prediction Using Machine Learning Approach using supervised machine learning techniques to predict stroke occurrence based on medical parameters. The proposed model achieved an accuracy of 91%. The limitation of the work was the use of a small dataset and fewer feature optimization techniques.

N. Saxena, A. Choudhary, D. S. Bhamra, and P. Maru [6], 2022, introduced BrainOK: Brain Stroke Prediction using Machine Learning, which focused on predicting stroke risk using machine learning and classification algorithms. The proposed system achieved an accuracy of 90%. However, the system lacked real-time implementation and GUI integration.

A. M. Kunder, H. N. Shashank, S. Srikanth, and A. M. Thejas [7], 2020, proposed Prediction of Stroke Using Machine Learning for early stroke risk prediction using machine learning models and clinical healthcare data. The system achieved an accuracy of 88%. The limitations included lower prediction accuracy and limited algorithm comparison.[8] Soumyabrata Dev et al., (2022) Soumyabrata Dev, Hewei Wang, Chidozie Shamrock Nwosu, Nishtha Jain, Bharadwaj Veeravalli, and Deepu John proposed a predictive analytics approach for stroke prediction using machine learning and neural networks. The study compared traditional machine learning algorithms with deep learning models for stroke risk assessment. Various patient health parameters were used to train predictive models. Neural network-based approaches demonstrated improved prediction accuracy compared to conventional methods. The research highlighted the importance of predictive analytics in early stroke detection and healthcare decision-making.

[9] R. Pitchai et al., (2022) R. Pitchai, Bhasker Dappuri, P. V. Pramila, M. Vidhyalakshmi, S. Shanthi, Wadi B. Alonazi, Khalid M. A. Almutairi, R. S. Sundaram, and Ibsa Beyene developed an AI-based biomedical stroke prediction and analytical system using machine learning techniques. The proposed system utilized patient medical data to identify stroke risk factors and generate predictions. Multiple machine learning algorithms were evaluated to determine the most effective model. The system provided analytical insights to support healthcare professionals in diagnosis. Results showed that AI-based prediction models can enhance the accuracy and efficiency of stroke risk assessment.

[10] Sathya Sundaram M. et al., (2022) Sathya Sundaram M., Pavithra K., Poojasree V., and Priyadharshini S. presented a machine learning-based stroke prediction system. The study focused on preprocessing healthcare datasets and applying classification algorithms for stroke prediction. Different machine learning techniques were analyzed to identify the best-performing model. The proposed approach improved prediction reliability by considering multiple patient health indicators. The findings demonstrated that machine learning can serve as an effective tool for early stroke detection and prevention strategies..

3. Methodology

The proposed “Stroke Prediction Using Ensemble Learning” system follows a systematic methodology to accurately predict the possibility of stroke occurrence using machine learning techniques. The methodology consists of several stages including data collection, data preprocessing, feature selection, model training, ensemble learning implementation, prediction, and performance evaluation. Each stage plays an important role in improving the efficiency and reliability of the prediction system.

3.1. Data Collection

The first step in the methodology is collecting a healthcare dataset containing patient medical information related to stroke prediction. The dataset includes important attributes such as:

Age

Gender

Hypertension

Heart disease

Average glucose level

Body Mass Index (BMI)

Smoking status

Work type

Marital status

Residence type

Stroke status (target variable)

The dataset is obtained from reliable medical or healthcare sources and stored in CSV format for further processing.

3.2. Data Preprocessing

Raw medical datasets may contain missing values, noise, duplicate records, and inconsistent data. Therefore, preprocessing is performed to improve data quality. The preprocessing stage includes:

Removing duplicate entries

Handling missing values using mean or median replacement

Converting categorical data into numerical format using Label Encoding or One-Hot Encoding

Normalizing numerical features for better model performance

Splitting the dataset into training and testing datasets

Data balancing techniques such as SMOTE (Synthetic Minority Oversampling Technique) may also be applied to handle class imbalance between stroke and non-stroke cases.

3.3. Feature Selection

Feature selection is performed to identify the most important attributes influencing stroke prediction. Selecting relevant features helps reduce computational complexity and improve model accuracy. Techniques such as correlation analysis and feature importance ranking are used to select meaningful medical parameters.

3.4. Ensemble Learning Model Development

The core component of the system is the Ensemble Learning model. Ensemble learning combines multiple machine learning algorithms to achieve better prediction accuracy than a single model. The proposed system utilizes algorithms such as:

Random Forest

Gradient Boosting

AdaBoost

Voting Classifier

These models are trained using the processed healthcare dataset. Each algorithm independently predicts stroke risk, and the ensemble method combines their outputs to produce the final prediction result.

3.5. Model Training and Testing

The dataset is divided into training and testing sets, commonly using an 80:20 ratio. The training dataset is used to train the ensemble models, while the testing dataset is used to evaluate prediction performance. During training, the algorithms learn hidden patterns and relationships among patient health parameters.

3.6. Stroke Prediction

After successful training, the system accepts patient medical details as input and predicts whether the patient is at risk of stroke. The prediction output is displayed as:

Stroke Risk Detected

or

No Stroke Risk Detected

The system may also provide probability scores representing the level of stroke risk.

3.7. Performance Evaluation

The performance of the ensemble learning model is evaluated using various evaluation metrics such as:

Accuracy

Precision

Recall

F1-Score

Confusion Matrix

These metrics help measure the effectiveness, reliability, and prediction capability of the proposed system.

3,8. Result Analysis and Visualization

Finally, the prediction results are analyzed and visualized using graphs and charts such as:

Accuracy comparison graphs

Confusion matrix visualization

Feature importance graphs

ROC Curve

Visualization helps in understanding model performance and identifying the most influential medical features affecting stroke prediction.

Thus, the proposed methodology provides an efficient and intelligent framework for early stroke prediction using ensemble learning techniques, helping healthcare professionals make faster and more accurate medical decisions.

System Architecture

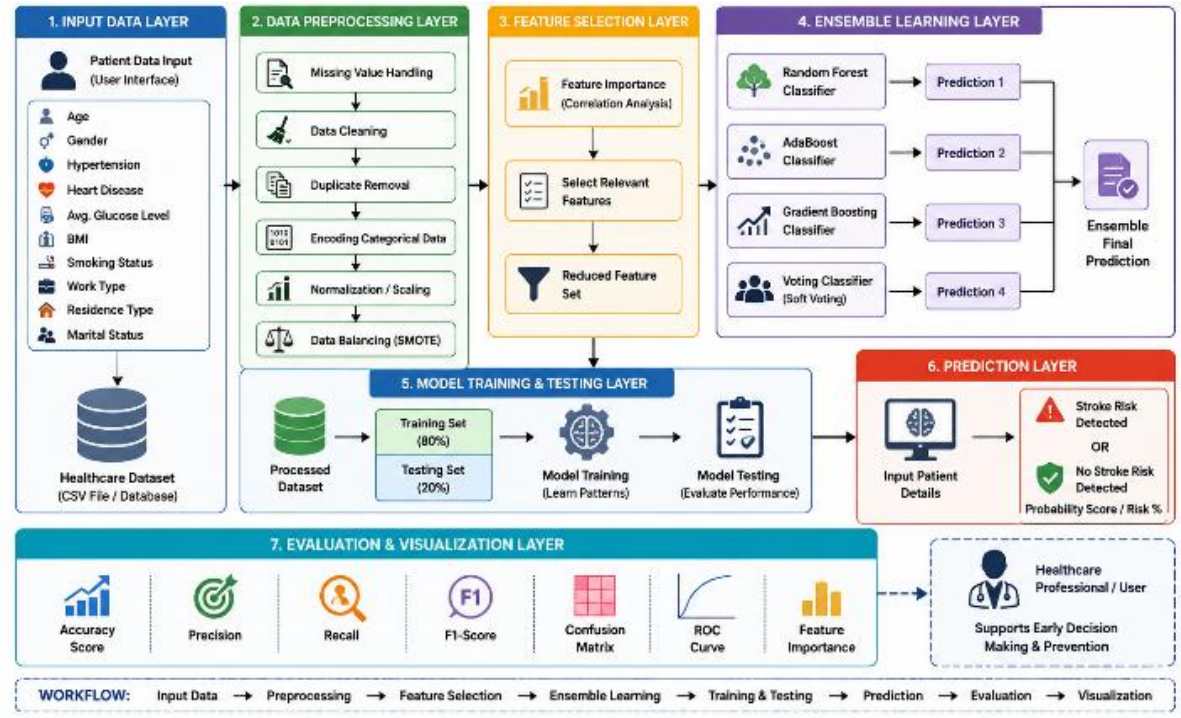

Figure 4.1: System Architecture

The system architecture of the proposed “Stroke Prediction Using Ensemble Learning” model explains the complete workflow of predicting stroke risk using healthcare data and machine learning techniques. Initially, patient details such as age, hypertension, heart disease, glucose level, BMI, and smoking status are collected as input data. The data then undergoes preprocessing steps like cleaning, handling missing values, normalization, encoding, and dataset balancing to improve data quality. After preprocessing, important medical features are selected and provided to ensemble learning algorithms such as Random Forest, AdaBoost, Gradient Boosting, and Voting Classifier. These

models are trained and tested using healthcare datasets to learn patterns related to stroke occurrence. Finally, the trained system predicts whether a patient has a risk of stroke and evaluates the model performance using metrics like accuracy, precision, recall, F1-score, confusion matrix, and ROC curve. The architecture helps provide accurate, fast, and reliable stroke prediction for better healthcare decision-making..

4. Results And Discussions

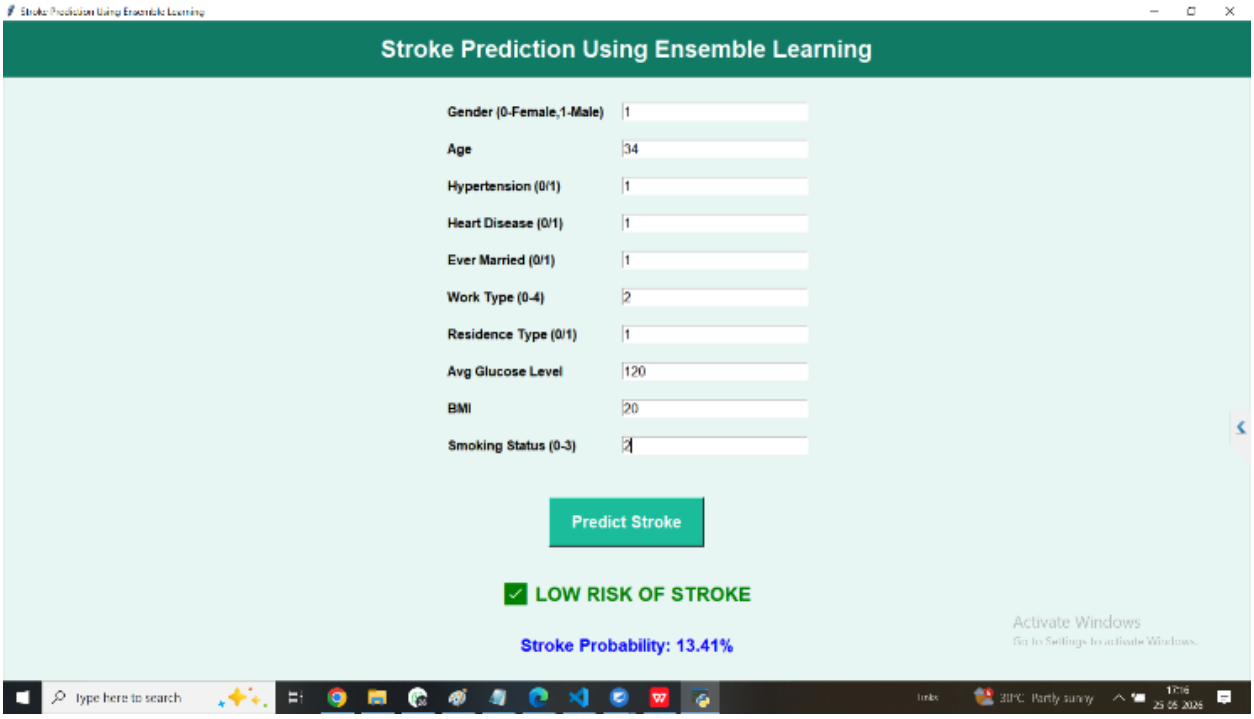

Figure 5.1: Input patient Data

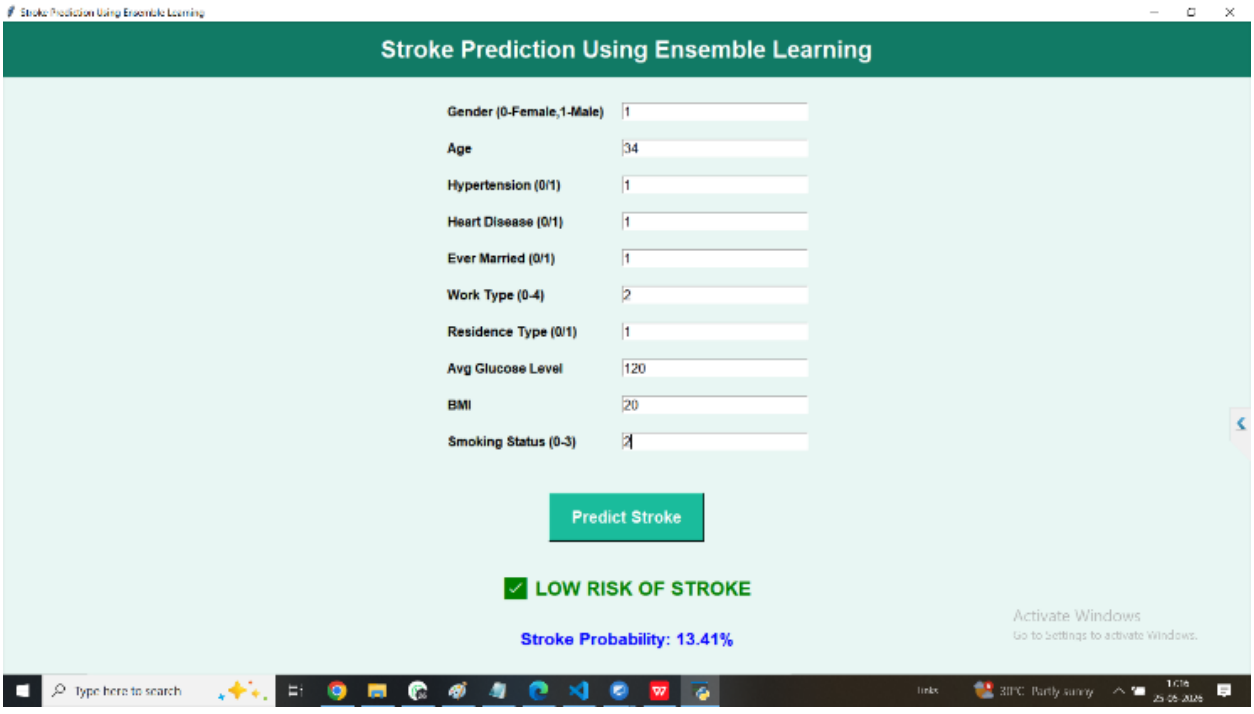

Figure 5.2: Prediction

```
================================================
MODEL: Random Forest
================================================

Accuracy : 0.9383561643835616
Precision: 0.0
Recall   : 0.0
F1 Score : 0.0

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 5.3: Random Forest Classification Report

\

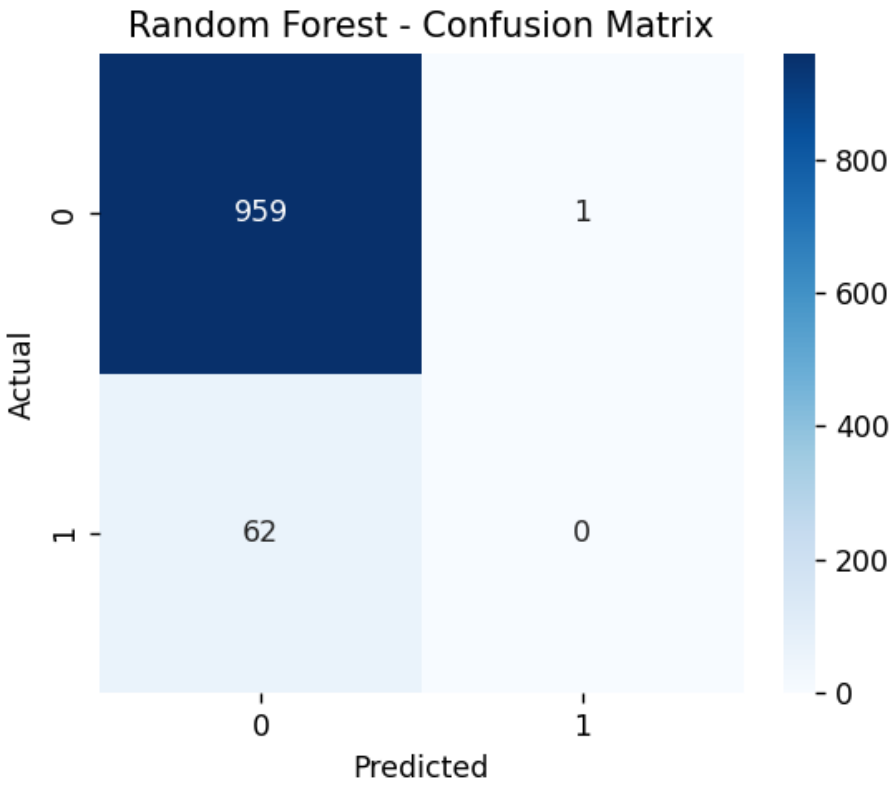

Figure 5.4: Confusion Matrix

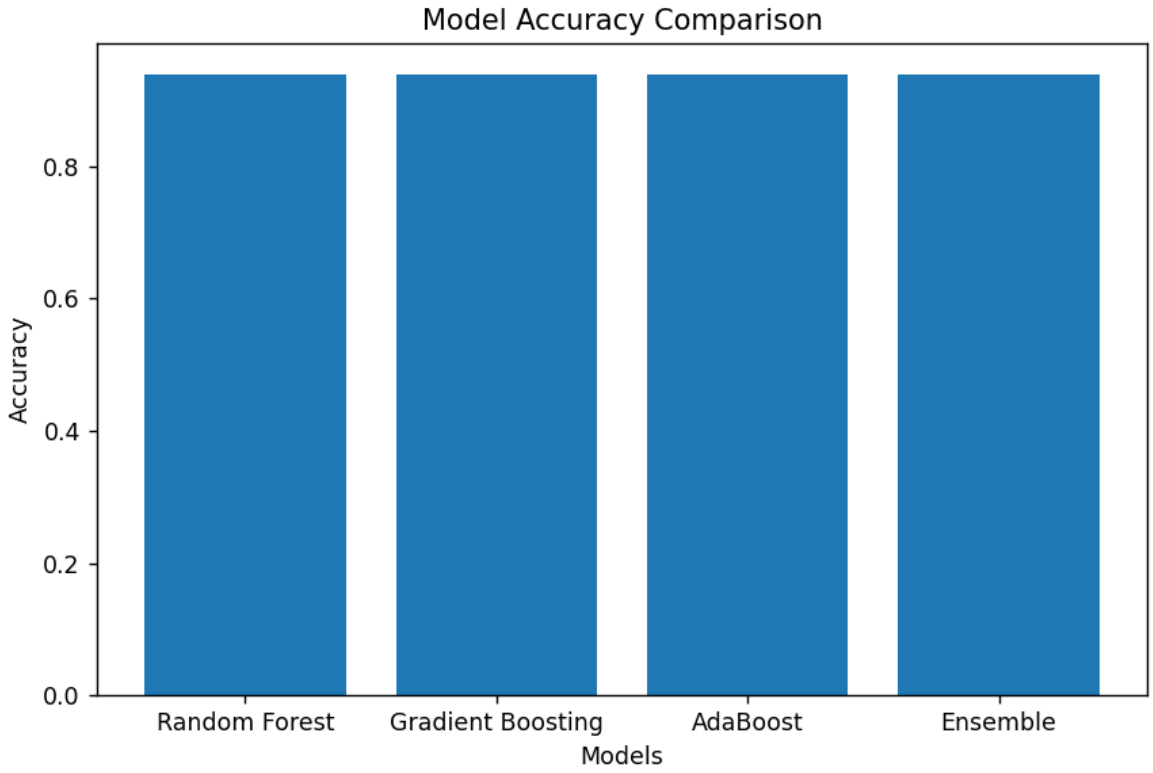

Figure 5.5: Model Accuracy comparison

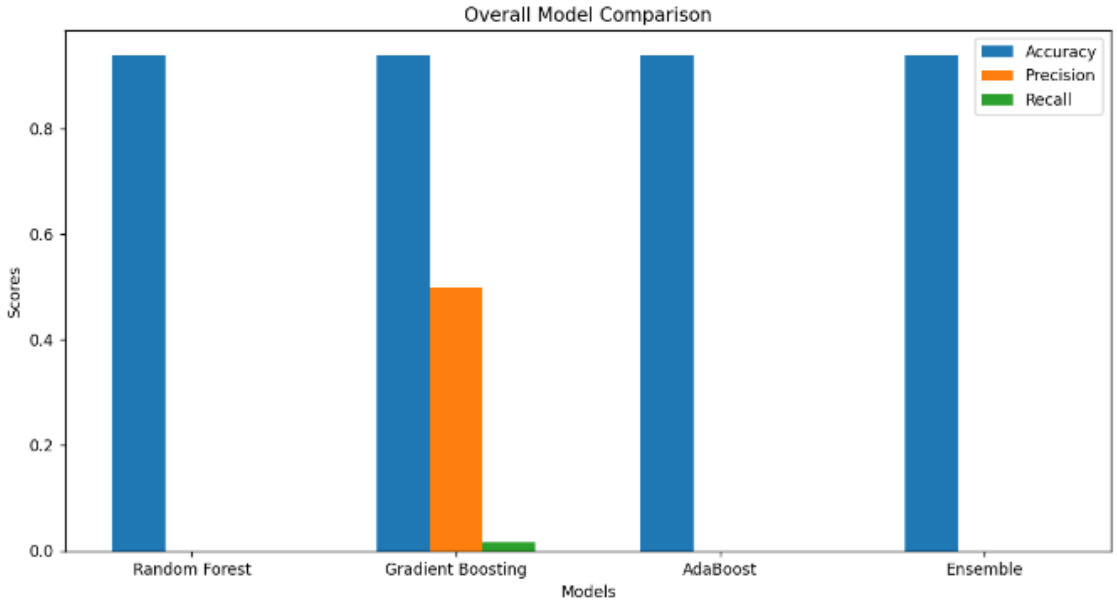

Figure 5.6: Performance Metrics

5. Conclusion

The proposed "Stroke Prediction Using Ensemble Learning" system provides an effective and intelligent solution for the early detection of stroke risk using machine learning techniques. By utilizing important healthcare parameters such as age, hypertension, heart disease, glucose level, BMI, and smoking status, the system can accurately

analyze patient conditions and predict the likelihood of stroke occurrence. The use of Ensemble Learning algorithms such as Random Forest, AdaBoost, Gradient Boosting, and Voting Classifier improves prediction accuracy, reliability, and overall model performance compared to individual machine learning models.The system performs essential processes including data preprocessing, feature selection, model training, testing, and performance evaluation to ensure efficient and accurate predictions. Evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix demonstrate the effectiveness of the proposed approach in identifying stroke risk. The developed system can assist healthcare professionals in making faster and better clinical decisions while also supporting preventive healthcare measures. The proposed model highlights the significant role of artificial intelligence and ensemble learning in modern healthcare applications. The system offers a cost-effective, fast, and reliable approach for stroke prediction, which can help reduce mortality rates and improve patient care through early diagnosis and timely medical intervention.

Future work

Future developments may focus on expanding datasets, refining image

processing methods, and incorporating IoT-based health monitoring solutions for proactive stroke detection.

Reference

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
15. Priyadarshini C Patil, “An Efficient Food Image Classification By Inception-V3 based CNNs”, International journal of scientific & technology research (IJSTR), ISSN 2277-8616, vol 9,issue 03, march 2020, pp-6986-6992.