International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

Machine Learning-Based Stroke Risk Prediction with Explainability and Web Deployment for Early Care

[1] J.V. ANIL KUMAR, [2] U. ARAVIND, [3] PUVVADA RAMYA,[4]RANGOJI ARUNA,[5] GOLLA NAGALAKASHMI, [6] GRANDHISILA HARSHINI

[1] PROFESSOR& HOD, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES, DEVARAJUGATTU, PEDDARAVEEDU(MD), MARKAPUR.

[2] ASST., PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES, DEVARAJUGATTU, PEDDARAVEEDU(MD), MARKAPUR.

[3,4,5,6] STUDENT, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES, DEVARAJUGATTU, PEDDARAVEEDU(MD), MARKAPUR.

ABSTRACT

Stroke is a leading cause of mortality and long-term disability worldwide, making early prediction and timely intervention critical for improving patient outcomes. Traditional risk assessment methods rely heavily on clinical expertise and may not effectively capture complex interactions among multiple risk factors. This paper presents an automated stroke prediction system using machine learning, combined with explainable artificial intelligence (XAI) techniques and an interactive web application for early intervention support. The proposed approach utilizes patient data such as age, hypertension, heart disease, glucose levels, body mass index, and lifestyle factors to train and evaluate various machine learning models, including Logistic Regression, Random Forest, and Gradient Boosting algorithms. To enhance transparency and trust, explainability methods such as SHAP (SHapley Additive exPlanations) and feature importance analysis are incorporated to interpret model predictions. Additionally, exploratory data analysis (EDA) is performed to identify key patterns, correlations, and risk indicators within the dataset. The final system is deployed as a user-friendly web application that allows users and healthcare professionals to input patient data and receive real-time stroke risk predictions along with interpretable insights. Experimental results demonstrate that the proposed model achieves high accuracy and reliability, making it a valuable decision-support tool for early stroke detection and prevention.

Keywords : Stroke Prediction, Machine Learning, Explainable AI (XAI), SHAP, Exploratory Data Analysis (EDA), Web Application, Early Intervention, Healthcare Analytics, Risk Assessment, Predictive Modeling

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

I. INTRODUCTION

Stroke is a serious medical condition that occurs when the blood supply to a part of the brain is interrupted or reduced, preventing brain tissue from receiving oxygen and nutrients. It is one of the leading causes of death and long-term disability worldwide, significantly impacting individuals, families, and healthcare systems. Early detection and timely intervention are crucial in reducing the severity of stroke outcomes and improving survival rates.

Traditionally, stroke risk assessment is performed by healthcare professionals using clinical evaluations and patient history, including factors such as age, hypertension, heart disease, diabetes, smoking habits, and lifestyle conditions. While these methods are essential, they often rely on manual analysis and may not fully capture the complex relationships between multiple risk factors. This can sometimes lead to delayed or less accurate predictions, especially in high-risk populations.

With the rapid advancement of artificial intelligence, machine learning has emerged as a powerful tool in healthcare for predictive analysis and decision support. Machine learning models can process large volumes of medical data, identify hidden patterns, and provide accurate predictions based on learned relationships. In stroke prediction, these models can analyze various health indicators to estimate an individual's risk level more efficiently than traditional approaches.

In addition to prediction accuracy, the interpretability of machine learning models is becoming increasingly important in medical applications. Explainable Artificial Intelligence (XAI) techniques, such as SHAP (SHapley Additive exPlanations), help in understanding how different features contribute to the model's predictions. This transparency is essential for building trust among healthcare professionals and ensuring that the system's decisions can be validated and explained.

II. LITERATURE REVIEW

Stroke prediction using machine learning has gained significant attention due to its potential to improve early diagnosis and reduce mortality rates. Various studies have explored different algorithms, data sources, and analytical techniques to enhance prediction accuracy and clinical usability.

Khosla et al. (2010) [1] conducted one of the early studies using data mining techniques for stroke prediction. They applied classification algorithms such as Decision Trees and Naïve Bayes on medical datasets and demonstrated that machine learning could effectively

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

identify high-risk individuals based on clinical parameters.

Chen et al. (2016) [2] proposed a stroke prediction model using Logistic Regression and Support Vector Machines (SVM). Their study highlighted the importance of risk factors such as hypertension, age, and glucose levels, and showed that SVM achieved better classification performance compared to traditional statistical methods.

Asadi et al. (2017) [3] utilized Random Forest and Gradient Boosting algorithms to improve prediction accuracy. Their results indicated that ensemble learning methods outperform single classifiers by reducing overfitting and capturing complex relationships among features.

More recent studies have focused on integrating deep learning techniques. Wang et al. (2019) [4] applied Deep Neural Networks (DNN) for stroke risk prediction and achieved improved performance due to the model's ability to learn nonlinear patterns from large datasets. However, the lack of interpretability remained a challenge.

To address this issue, explainable AI techniques have been introduced. Lundberg and Lee (2017) [5] developed SHAP (SHapley Additive exPlanations), a unified framework for interpreting machine learning models. SHAP has been widely adopted in healthcare applications to explain feature contributions and improve transparency in predictive systems.

In addition, exploratory data analysis (EDA) has been emphasized as a crucial step in understanding dataset characteristics. Gupta et al. (2020) [6] performed EDA on healthcare datasets to identify key stroke risk factors and correlations, which significantly improved model performance and feature selection.

Recent research by Patel et al. (2022) [7] integrated machine learning models with web-based applications for real-time stroke prediction. Their system demonstrated the practical applicability of predictive models in clinical and user-friendly environments, enabling early intervention and awareness.

III. EXISTING SYSTEM

The existing systems for stroke prediction primarily rely on traditional clinical assessment methods and basic statistical models. In conventional healthcare settings, stroke risk is evaluated by medical professionals using patient history and key risk factors such as age, hypertension, diabetes, heart disease, smoking habits, and cholesterol levels. While these assessments are essential, they are largely manual, time-consuming, and dependent on clinical expertise, which may lead to variability in diagnosis and delayed intervention.

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

In recent years, machine learning-based approaches have been introduced to improve prediction accuracy. These systems typically employ individual algorithms such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), or Naïve Bayes to classify patients based on their risk levels. Although these models provide better performance compared to traditional statistical methods, they often suffer from limitations such as overfitting, sensitivity to noisy data, and reduced generalization when applied to diverse datasets.

IV. PROPOSED SYSTEM

The proposed system presents an automated and interpretable framework for stroke prediction by integrating machine learning models, explainable artificial intelligence (XAI), and a user-friendly web application. The goal is to provide accurate risk assessment along with transparent insights that support early intervention and clinical decision-making.

The system begins with data acquisition from healthcare datasets containing patient attributes such as age, gender, hypertension, heart disease status, average glucose level, body mass index (BMI), smoking status, and other lifestyle factors. This data undergoes preprocessing steps including handling missing values, encoding categorical variables, normalization, and outlier detection to ensure data quality and consistency.

Following preprocessing, exploratory data analysis (EDA) is performed to understand the distribution of features, identify correlations among variables, and detect key risk factors associated with stroke. Visualization techniques such as histograms, heatmaps, and box plots are used to gain insights and guide feature selection. This step helps in improving model performance by focusing on the most relevant attributes.

The core of the system involves training multiple machine learning models such as Logistic Regression, Random Forest, and Gradient Boosting. These models are selected due to their effectiveness in classification tasks and their ability to handle structured healthcare data. Hyperparameter tuning and cross-validation techniques are applied to optimize model performance and prevent overfitting.

V. METHODOLOGY

The proposed methodology for automated stroke prediction is designed as a comprehensive pipeline that integrates data preprocessing, exploratory analysis, machine learning modeling, explainability, and web deployment. The aim is to develop an

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

accurate, interpretable, and user-friendly system for early stroke risk assessment.

The process begins with data collection from reliable healthcare datasets containing patient information such as age, gender, hypertension status, heart disease, average glucose level, body mass index (BMI), smoking habits, and other relevant attributes. Ensuring data quality and diversity is essential to improve model generalization and reliability.

In the preprocessing stage, the dataset is cleaned by handling missing values, removing duplicates, and correcting inconsistencies. Numerical features are normalized or standardized to maintain uniform scale, while categorical variables are encoded using techniques such as one-hot encoding or label encoding. Outlier detection methods are also applied to reduce noise and improve model performance.

Following preprocessing, exploratory data analysis (EDA) is conducted to understand the underlying patterns and relationships within the data. Statistical summaries and visualization techniques such as histograms, correlation heatmaps, and box plots are used to identify significant risk factors and trends associated with stroke occurrence. This step also assists in effective feature selection.

Feature selection is then performed to retain the most relevant attributes and reduce dimensionality. Techniques such as correlation analysis, mutual information, and recursive feature elimination are used to eliminate redundant or less significant features, thereby improving computational efficiency and model accuracy.

In the modeling phase, multiple machine learning algorithms—including Logistic Regression, Random Forest, and Gradient Boosting—are trained on the processed dataset. These models are chosen for their effectiveness in classification tasks and ability to handle structured medical data. Hyperparameter tuning and k-fold cross-validation are applied to optimize model performance and prevent overfitting.

To ensure interpretability, Explainable Artificial Intelligence (XAI) techniques are incorporated, particularly SHAP (SHapley Additive exPlanations). SHAP values provide insights into how each feature contributes to individual predictions, enabling transparency and aiding clinicians in understanding model decisions.

VI. SYSTEM MODEL

System Architecture

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

VII. RESULTS AND DISCUSSIONS

In above screen tomcat server started and now open browser and enter URL as 'http://localhost:9999/PenTesting' and press enter key to get below page

In above screen click on 'Signup Section' link to get below page

In above screen user entering signup details and then press button to get below page

In above screen sign up process completed and now click on 'Login' link to login as user

In above screen user is login and after login will get below page

In above screen user can click on 'Profile Updation' link to get below page

In above screen user will get existing profile details and can modify desired details and then press button to get below output

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

In above screen profile successfully modified and now click on 'Penetration Testing for Vulnerability' link to get below page

From above test data you can copy some query and then paste in below screen

In above screen I pasted query and then press button to detect query as normal or contains vulnerable access

In above screen in red colour text can see given query contains 'Vulnerability' and similarly you can test other queries

In above screen testing another XSS query and below is the output

In above screen given query detected as 'Normal' means 'No Vulnerability' Detected.

So by using this application developers can test all applications queries to avoid vulnerability.

VIII. CONCLUSION

This study presents an automated stroke prediction system that integrates machine learning, explainable artificial intelligence (XAI), and a web-based application to support early intervention. The proposed approach addresses the limitations of traditional clinical assessment methods by leveraging data-driven models to analyze multiple risk factors and generate accurate predictions.

By utilizing algorithms such as Logistic Regression, Random Forest, and Gradient Boosting, the system demonstrates strong

predictive performance and the ability to capture complex relationships within healthcare data. The incorporation of exploratory data analysis (EDA) further enhances the understanding of key risk factors, contributing to improved feature selection and model efficiency.

A significant contribution of this work is the integration of explainability through SHAP, which provides transparent insights into how individual features influence prediction outcomes. This not only increases trust in the system but also enables healthcare professionals to make informed decisions based on interpretable results.

IX. FUTURE WORK:

While the proposed automated stroke prediction system demonstrates strong performance and practical usability, several areas can be explored to further enhance its effectiveness and real-world impact. One important direction for future work is the expansion of datasets to include larger, more diverse populations from different geographic and demographic backgrounds. This will improve model generalization and reduce potential biases in predictions.

Future research can also focus on integrating additional data sources such as electronic health records (EHR), medical imaging (e.g., CT or MRI scans), wearable device data, and real-time health monitoring systems. The inclusion of multimodal data can provide a more comprehensive understanding of stroke risk and significantly improve prediction accuracy.

Another key area is the adoption of more advanced machine learning and deep learning techniques, including ensemble models, transformer-based architectures, and time-series analysis for continuous health monitoring. These approaches can better capture complex patterns and temporal variations in patient data.

Improving explainability remains an essential aspect for healthcare applications. Future work can enhance the current explainable AI framework by incorporating more intuitive visualization techniques and combining multiple XAI methods to provide deeper insights into model decisions, thereby increasing trust among medical professionals.

The web application can also be further विकसित by adding features such as user authentication, data storage, personalized health recommendations, and integration with hospital management systems. Mobile application development can improve accessibility, especially in remote and underserved areas.

XI. REFERENCES

[1] J.V. Anil Kumar, Nagella Swarupa Rani," SECURE DATA TRANSMISSION THROUGH HYBRID CRYPTOGRAPHY AND STEGANOGRAPHIC TECHNIQUES", International Journal of Engineering Science and Advanced Technology (IJESAT) Vol 25 Issue 12,2025, www.ijesat.com, https://doi.org/10.64771/ijesat.2025.046, Page 373 to 383, ISSN:2250-3676, 2025.

[2] J.V.ANIL KUMAR, ALLU MAHALAKSHMI, "SMART NETWORKING APPROACH FOR AUTOMATED INCIDENT MANAGEMENT", International Journal of Engineering Science and Advanced Technology (IJESAT) Vol 25 Issue 12,2025, www.ijesat.com, https://doi.org/10.64771/ijesat.2025.047, Page 384 to 392, ISSN:2250-3676, 2025.

[3] Jajam Venkata Anil Kumar, Dr. G. Charles Babu, "Automating Content Utilizing Big Data Innovations", *Journal of Advances and Scholarly Researches in Allied EducationVol. 15, Issue No. 9, October-2018, ISSN 2230-7540*, IIFS : 1.6 (2014), INDEX COPERNICUS : 49060 (2018), IJINDEX : 3.46 (2018), pp.635-639, 2018.

[4] Wang, P., Fang, Y., et al., "A Deep Learning-Based Approach for Stroke Risk Prediction," *IEEE Access*, 2019.

[5] Lundberg, S. M., and Lee, S. I., "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[6] Gupta, A., Shukla, P., et al., "Exploratory Data Analysis of Stroke Prediction Dataset Using Machine Learning Techniques," *International Journal of Advanced Computer Science and Applications*, 2020.

[7] Patel, J., Shah, M., et al., "Web-Based Stroke Prediction System Using Machine Learning," *International Journal of Computer Applications*, 2022.

[8] Breiman, L., "Random Forests," *Machine Learning*, 2001.

[9] Friedman, J. H., "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, 2001.

[10] Pedregosa, F., Varoquaux, G., et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, 2011.