As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Issue 06, pages 295-300.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue06.html
DOI: https://doi.org/10.5281/zenodo.20739557

Design of a Predictive Analytics Model for Brain Stroke Risk Estimation Using Patient Historical Data

V. Swami Naidu, Muthireddi Naga Iswarya, Posimsetti Navya Sri, Vallepu Hema

Department of Information Technology Swarnandhra College of Engineering & Technology, Seetharamapuram, Narsapur, Andhra Pradesh, INDIA.

To Cite this Article
V. Swami Naidu, Muthireddi Naga Iswarya, Posimsetti Navya Sri & Vallepu Hema (2026). Design of a Predictive Analytics Model for Brain Stroke Risk Estimation Using Patient Historical Data. International Journal for Modern Trends in Science and Technology, 12(06), 295-300. https://doi.org/10.5281/zenodo.20739557

Article Info
Received: 16 May 2026; Revised: 07 May 2026; Accepted: 12 June 2026.

KEYWORDS

Brain Stroke Prediction, Logistic Regression, Machine Learning in Healthcare, Predictive Analytics, Patient Historical Data, Data Preprocessing, Streamlit Web Application, Rule-Based Decision System, BMI, Glucose Level, Hypertension, Smoking Status, Scikit-learn, Python.

ABSTRACT

Stroke is a serious medical condition and one of the leading causes of death and long-term disability worldwide. Early prediction and prevention of strokes can significantly reduce its impact on human health. This project, titled "Design of a Predictive Analytical Model for Brain Stroke Risk Estimation Using Patient Historical Data," aims to develop an intelligent system that predicts the likelihood of stroke based on patient information.The proposed system utilizes machine learning techniques, specifically Logistic Regression, to analyze various patient attributes such as age, gender, hypertension, heart disease, body mass index (BMI), glucose level, and smoking status. The dataset is preprocessed using techniques like data imputation, feature scaling, and one-hot encoding to ensure accuracy and consistency. The trained model is capable of providing both prediction results and probability scores indicating the risk level of stroke.In addition to the machine learning model, a rule-based decision system is incorporated to enhance interpretability by considering key medical factors such as smoking habits, chronic diseases, and age. This hybrid approach improves reliability and provides more meaningful insights to users.The system is implemented using Python and deployed through a Streamlit web application, allowing users to input patient details and obtain real-time predictions in an interactive and user-friendly interface.

Overall, this project demonstrates how predictive analytics and machine learning can be effectively used in healthcare to assist in early diagnosis, improve decision-making, and support preventive measures for stroke risk management.

1. INTRODUCTION

1.1 Brief Information

The healthcare sector has undergone significant advancements over the years, with the integration of modern technologies enabling improved diagnosis, treatment, and patient care. Traditionally, medical decisions were based on generalized observations and clinical experience. While effective to some extent, this approach often fails to consider individual variations in patient health conditions, lifestyle, and medical history, which are critical in predicting diseases such as stroke.

Stroke is a major health concern worldwide and is one of the leading causes of death and long-term disability [3]. It occurs when the blood supply to the brain is interrupted, leading to severe neurological damage. Early identification of individuals at risk of stroke is essential for timely intervention and prevention. However, conventional diagnostic methods are often reactive rather than proactive, focusing on treatment after the occurrence of the condition rather than predicting it beforehand [8].

With the emergence of Artificial Intelligence (AI) and Machine Learning (ML), predictive healthcare has gained significant importance. Machine learning algorithms have the capability to analyze large volumes of patient data and identify hidden patterns that are not easily detectable through traditional methods. These techniques enable the development of predictive models that can estimate the likelihood of diseases based on historical patient data.

This project focuses on the development of a predictive analytical model for stroke risk estimation using patient historical data. The proposed system utilizes machine learning techniques to analyze various patient attributes such as age, gender, hypertension, heart disease, glucose level, body mass index (BMI), and smoking status. By processing these factors, the system aims to predict the probability of stroke occurrence and assist in early risk assessment.

1.2 Purpose

The purpose of this project is to develop an AI-powered interview preparation portal that provides a structured and efficient approach to learning for students and job seekers. The system is designed to integrate personalized question recommendations, real-time coding practice, and instant feedback within a single platform. By leveraging AI and code execution technologies, it aims to enhance users' problem-solving abilities, improve coding skills, and make the overall interview preparation process more effective, accessible, and user-friendly.

1.3 Motivation

The motivation behind this project stems from the common challenges faced by candidates during stroke prediction — the lack of a unified platform that combines machine learning predictions with interpretable outputs. Many existing systems either focus solely on accuracy or lack real-time user interaction. This project is motivated by the need to create a centralized, intelligent system that offers personalized predictions, rule-based explanations, and a seamless user experience, helping healthcare professionals make better-informed decisions.

1.4 Problem Statement

The current stroke risk prediction process poses several challenges due to lack of structured approaches, limited practical tools, and absence of interpretable AI guidance. Traditional methods are often time-consuming, fragmented, and insufficient for early risk detection.

The major problems include:

- Lack of a centralized platform for both theoretical and practical stroke analysis
- Limited opportunities for real-time prediction and execution
- Absence of personalized feedback and risk-level explanations
- Difficulty in identifying patient-specific risk factors
- Time-consuming and unorganized preparation of diagnostic systems

2. LITERATURE SURVEY

2.1 Introduction

The rapid advancement of Artificial Intelligence (AI), machine learning, and data-driven systems has significantly transformed the landscape of digital healthcare, predictive modeling, and clinical decision support. Stroke prediction platforms, in particular, have benefited from these advancements by integrating AI-driven personalization, real-time feedback mechanisms, and automated risk analysis. This chapter presents a comprehensive review of existing literature related to stroke prediction, machine learning for healthcare, data preprocessing, and rule-based decision systems that form the foundation of the proposed system.

2.2 Machine Learning for Disease Prediction

Machine learning has become an essential component in the field of healthcare, particularly in disease prediction and risk assessment. Various machine learning algorithms have been applied to predict diseases such as diabetes, heart disease, and strokes with considerable accuracy [5, 6]. Supervised learning techniques, including Logistic Regression, Decision Trees, and Support Vector Machines, are widely used for classification tasks in healthcare. Despite their effectiveness, challenges such as data imbalance, missing values, and variability in patient data still require careful handling.

2.3 Stroke Prediction Using Machine Learning

Stroke prediction has gained significant attention in recent years due to its critical impact on public health [1, 3]. Several studies have focused on developing predictive models that analyze risk factors such as age, hypertension, heart disease, glucose levels, and lifestyle habits. Logistic Regression is commonly used due to its simplicity and interpretability, while Random Forest and Neural Networks have also been explored to improve accuracy [7]. However, many existing systems focus primarily on accuracy without providing clear explanations for their results, limiting practical healthcare deployment.

2.4 Data Preprocessing and Feature Engineering

Data preprocessing plays a vital role in building effective machine learning models [4]. Medical datasets often contain missing values, inconsistent data formats, and categorical variables requiring transformation. Common preprocessing techniques include handling missing values using imputation methods, encoding categorical variables, and normalizing data. Feature engineering further involves selecting and transforming relevant attributes such as BMI, glucose levels, and smoking status that are critical in determining stroke risk levels.

2.5 Limitations of Existing Systems

Although various machine learning-based systems have been developed for disease prediction, they have several limitations. Many systems rely solely on historical data without incorporating real-time patient input, and lack user-friendly interfaces. Additionally, some models act as black boxes, providing predictions without clear explanations. Existing systems often do not combine predictive modeling with rule-based logic, which could enhance interpretability and decision-making [2, 9].

Relevance of Literature to the Proposed System

The proposed Brain Stroke Risk Estimation System builds upon existing research by: applying machine learning classification principles [5, 7]; incorporating data preprocessing best practices [4]; integrating rule-based decision logic for interpretability [6]; and utilizing real-time execution concepts aligned with modern healthcare tools [3]. By synthesizing these research contributions, the proposed system offers a practical and novel solution to modern stroke prediction challenges.

3. PROPOSED SYSTEM

The proposed system introduces a Smart Brain Stroke Risk Prediction Portal with Integrated Machine Learning and Rule-Based Decision Making. It reduces dependency on multiple platforms by providing a single, user-friendly solution for analysis, prediction, and result interpretation. The system utilizes machine learning techniques, real-time prediction through the trained Logistic Regression model, and rule-based performance analysis to create an interactive and personalized healthcare experience.

3.1 System Architecture

The system architecture consists of multiple components: the User Interface (Streamlit), Data Preprocessing Module, Machine Learning Model, Rule-Based Decision Engine, and Output Display. Patient data flows from input through preprocessing, to prediction and rule-based evaluation, and finally to result visualization.

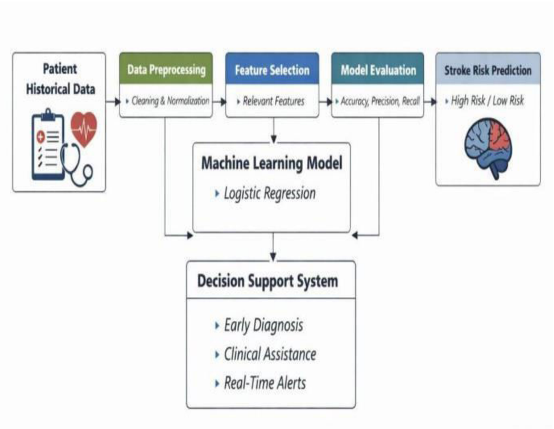

Fig 1: System Architecture

3.2 Use Case Diagram

The use case diagram provides a high-level representation of the system functionalities by illustrating the interaction between users (actors) and the system features (use cases). The primary actor is the user, who interacts with functionalities such as data entry, stroke prediction, viewing results, and receiving explanations. The diagram captures and organizes system requirements in a simple and understandable way.

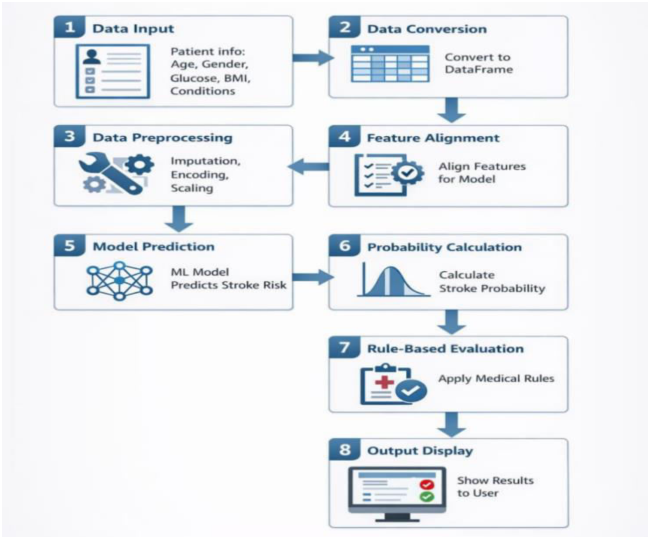

Fig 2: Use Case Diagram

3.3 Class Diagram

The class diagram illustrates key classes such as User, PatientData, PredictionModel, and RuleEngine, along with their properties and functionalities. It defines how data is structured and how different classes interact with each other through relationships. For example, a User provides PatientData, which the PredictionModel processes to generate a risk classification and probability score.

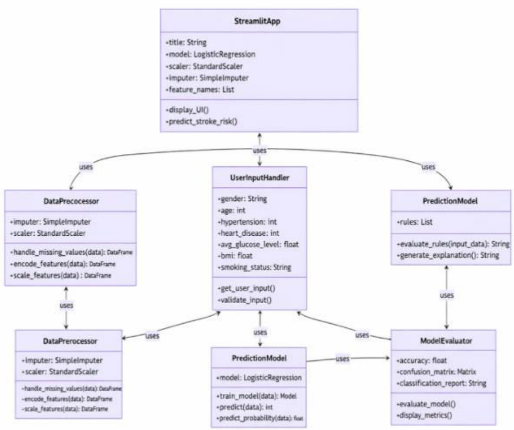

Fig 3: Class Diagram

3.4 Advantages of Proposed System

• Provides early prediction of stroke risk using patient historical data
• Improves accuracy through the use of Logistic Regression machine learning techniques
• Handles large and complex datasets efficiently
• Enhances interpretability by integrating rule-based logic
• Offers real-time prediction through an interactive Streamlit interface
• User-friendly design enables easy access and usability
• Supports data-driven decision-making in healthcare ediction results.

4. RESULTS

The primary objective of this project was to design and develop a Smart Brain Stroke Risk Prediction System using patient historical data. The system successfully fulfills its requirements by integrating modern web technologies, machine learning, and rule-based decision logic into a unified platform.

Input Panel

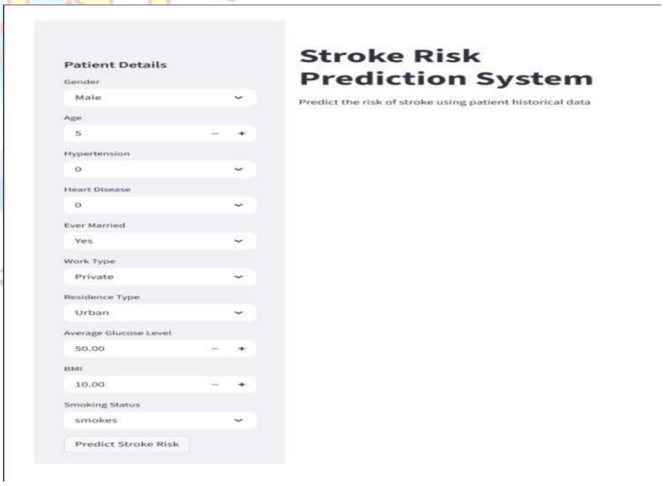

Fig 4: Input Panel

4.1 PREDICTION RESULTS AND ANALYSIS

The prediction results are generated based on user-provided patient data. The system classifies each input into either high risk or low risk categories and provides a probability percentage indicating the likelihood of stroke.

For patients with multiple risk factors such as high glucose levels, elevated BMI, presence of hypertension, or smoking habits, the system tends to predict a higher risk. Conversely, patients with fewer risk factors, such as younger age and absence of chronic conditions, are classified as low risk.

Prediction Output – Low Risk

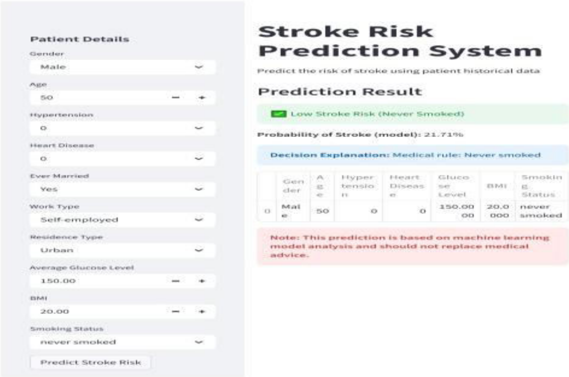

Fig 5: Low Stroke Risk Result

Prediction Output – Moderate Risk

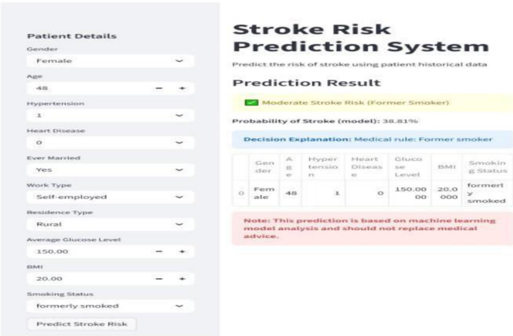

Fig 6: Moderate Stroke Risk Result

Prediction Output – High Risk

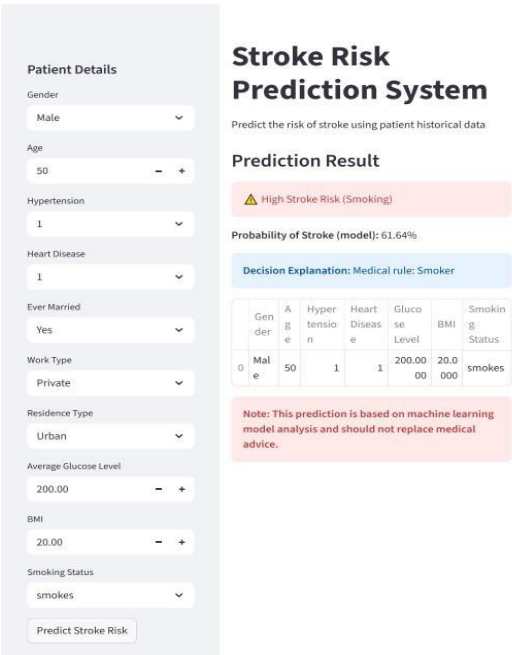

Fig 7: High Stroke Risk Result

5. FUTURE SCOPE

The Stroke Risk Prediction System has been successfully designed and implemented. There are several areas for future enhancement:

- Add support for more patient parameters such as cholesterol levels and genetic history
- Implement advanced machine learning models such as Random Forest and Neural Networks for improved accuracy
- Enhance personalized recommendations using deep learning algorithms
- Integrate video tutorials and detailed solution explanations for healthcare awareness
- Develop a mobile application for better accessibility
- Add discussion forums for peer learning and doctor consultations
- Include performance analytics dashboards for tracking patient history over time
- Introduce integration with Electronic Health Records (EHR) systems
- Improve UI/UX for a more interactive user experience
- Enable real-time health monitoring data feeds from wearable devices

6. CONCLUSION

The Brain Stroke Risk Prediction System has been successfully designed and implemented to estimate the likelihood of stroke using patient historical data. The system utilizes machine learning techniques, specifically Logistic Regression, to analyze various risk factors such as age, gender, hypertension, heart disease, glucose level, body mass index (BMI), and smoking status. By processing these attributes through a structured data-driven approach, the system provides reliable and real-time stroke risk predictions.

The implementation of data preprocessing techniques, including handling missing values through imputation, encoding categorical variables, and scaling numerical features, has significantly improved the accuracy and consistency of the model. The trained machine learning model effectively learns patterns from historical patient data and generates both a risk classification and a probability score, offering a comprehensive assessment of individual stroke risk.

In addition to the machine learning model, the integration of a rule-based decision mechanism enhances the interpretability of predictions by applying predefined medical conditions such as smoking habits, hypertension, and age-related factors. This hybrid approach ensures that the system is not only accurate but also transparent and understandable for healthcare professionals and general users alike.

The development of a user-friendly Streamlit web interface further improves the accessibility and usability of the system, allowing users to input patient details and receive real-time predictions without requiring technical expertise. The system displays risk classification, probability scores, and decision explanations, making it a practical tool for early stroke risk assessment.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Learn about Stroke. Available online: https://www.world-stroke.org/world-stroke-day-campaign/why-stroke-matters/learn-about-stroke (accessed on 25 May 2022).

[2] Elloker, T.; Rhoda, A.J. The relationship between social support and participation in stroke: A systematic review. Afr. J. Disabil. 2018, 7, 1–9. [CrossRef] [PubMed]

[3] Katan, M.; Luft, A. Global burden of stroke. In Seminars in Neurology; Thieme Medical Publishers: New York, NY, USA, 2018; Volume 38, pp. 208–211.

[4] Bustamante, A.; et al. Blood biomarkers to differentiate ischemic and hemorrhagic strokes. Neurology 2021, 96, e1928–e1939. [CrossRef] [PubMed]

[5] Xia, X.; Yue, W.; Chao, B.; et al. Prevalence and risk factors of stroke in the elderly in Northern China. J. Neurol. 2019, 266, 1449–1458. [CrossRef] [PubMed]

[6] Alloubani, A.; Salem, A.; Abdelhafiz, I. Hypertension and diabetes mellitus as a predictive risk factors for stroke. Diabetes Metab. Syndr. Clin. Res. Rev. 2018, 12, 577–584. [CrossRef]

[7] Boehme, A.K.; Esenwa, C.; Elkind, M.S. Stroke risk factors, genetics, and prevention. Circ. Res. 2017, 120, 472–495. [CrossRef]

[8] Mosley, I.; Nicol, M.; Donnan, G.; Patrick, I.; Dewey, H. Stroke symptoms and the decision to call for an ambulance. Stroke 2007, 38, 361–366. [CrossRef]

[9] Lecouturier, J.; et al. Response to symptoms of stroke in the UK: A systematic review. BMC Health Serv. Res. 2010, 10, 1–9. [CrossRef]

[10] Gibson, L.; Whiteley, W. The differential diagnosis of suspected stroke: A systematic review. J. R. Coll. Physicians Edinb. 2013, 43, 114–118. [CrossRef]

[11] Rudd, M.; Buck, D.; Ford, G.A.; Price, C.I. A systematic review of stroke recognition instruments in hospital and prehospital settings. Emerg. Med. J. 2016, 33, 818–822. [CrossRef]

[12] Delpont, B.; et al. Pain after stroke: A review. Rev. Neurol. 2018, 174, 671–674. [CrossRef]

[13] Kumar, S.; Selim, M.H.; Caplan, L.R. Medical complications after stroke. Lancet Neurol. 2010, 9, 105–118. [CrossRef]

[14] Ramos-Lima, M.J.M.; et al. Quality of life after stroke: Impact of clinical and sociodemographic factors. Clinics 2018, 73, e418. [CrossRef]

[15] Gittler, M.; Davis, A.M. Guidelines for adult stroke rehabilitation and recovery. JAMA 2018, 319, 820–821. [CrossRef]