DESIGN AND IMPLEMENTATION OF A BRAIN STROKE PREDICTION AND PATIENT–DOCTOR COMMUNICATION SYSTEM USING MACHINE LEARNING

Shriniwas Dubbewar[1], Harsh Bhidekar[2], Prof. Leena Raut[3]
[1,2]PG Scholar, [3]Assistant Professor Department of Computer Application
K.D.K.College of Engineering, Nagpur,Maharashtra, India
Shriniwasdubbewarmca24f@kdkce.edu.in, harshbhidekar.mca24f@kdkce.edu.in, leena.raut@kdkce.edu.in

Abstract

Brain stroke is one of the leading causes of mortality and long-term disability worldwide, where early prediction and timely medical intervention play a critical role in improving patient outcomes. Traditional clinical assessment methods often rely heavily on manual evaluation and may fail to detect early warning signs accurately. In recent years, machine learning techniques have demonstrated significant potential in medical risk prediction; however, many existing stroke prediction systems focus primarily on accuracy and lack explainability and real-world deployment readiness. This paper presents the design and implementation of a Brain Stroke Prediction and Patient–Doctor Communication System using machine learning and web technologies. The proposed system employs supervised learning models such as Logistic Regression, Random Forest, and XGBoost for stroke risk prediction, along with data preprocessing and imbalance handling techniques. Explainability is incorporated using feature-importance analysis to support clinical understanding. The system is deployed as a web-based application with an integrated patient–doctor communication module, enabling real-time alerts and consultation. Experimental evaluation demonstrates improved prediction performance and practical usability, making the system suitable for real-world healthcare support.

I. INTRODUCTION

Brain stroke is a severe medical condition caused by the interruption of blood supply to the brain, leading to irreversible neurological damage if not treated promptly. According to global health statistics, stroke remains a major contributor to mortality and long-term disability, particularly in developing regions where access to specialized healthcare is limited. Early prediction of stroke risk can significantly reduce fatality rates by enabling preventive care and timely medical intervention.

Conventional stroke diagnosis primarily depends on clinical tests, imaging techniques, and physician expertise. While effective, these methods are reactive in nature and often identify stroke only after symptoms become critical. With the growing availability of medical datasets and advances in machine learning, predictive systems have emerged as a promising solution for early risk assessment. Despite these advancements, many existing systems emphasize predictive accuracy while overlooking explainability, usability, and integration with real clinical workflows.

This paper proposes a Brain Stroke Prediction System that combines machine learning-based risk assessment with an integrated patient–doctor communication framework. The system not only predicts stroke risk but also facilitates timely interaction between patients and healthcare professionals, thereby bridging the gap between prediction and intervention.

II. LITERATURE REVIEW AND MOTIVATION

A) Machine Learning in Stroke Prediction

Several studies have explored the application of machine learning techniques in predicting brain stroke using clinical and demographic data. Algorithms such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and Neural Networks have been widely evaluated. Tree-based ensemble models often demonstrate superior accuracy due to their ability to capture nonlinear relationships among risk factors.

b) Limitations of Existing Approaches

Although high accuracy is reported in many studies, practical adoption remains limited. Most models operate as black boxes, offering little interpretability to medical professionals. Additionally, many research efforts stop at model evaluation and do not address deployment, usability, or patient–doctor interaction.

c) Motivation

The motivation behind this work is to develop a system that balances predictive performance with explainability and real-world usability. By integrating machine learning prediction with a web-based *communication mechanism, the proposed system aims to support preventive healthcare rather than isolated prediction.*

d) Research Gap

Existing stroke prediction systems largely focus on classification accuracy and dataset optimization. There is limited emphasis on explainable predictions, system deployment, and clinical communication support. Furthermore, most solutions lack mechanisms for alerting doctors and enabling patient interaction following risk detection. The proposed system addresses this gap by combining accurate prediction, explainability, and patient–doctor communication within a unified framework.

III. PROPOSED SYSTEM ARCHITECTURE AND DESIGN

a) System Overview

The proposed system is a web-based healthcare application designed to predict stroke risk and facilitate medical communication. It consists of machine learning models for prediction, a Flask-based backend for model integration, and a role-based web interface for patients and doctors.

b) System Modules

Stroke Prediction Module: *This module processes patient health data and applies trained machine learning models to predict stroke risk. Input features include age, hypertension, heart disease, glucose level, BMI, and lifestyle factors.*

Explainability Module: *Feature-importance analysis is used to explain model predictions, helping doctors understand contributing risk factors.*

Patient–Doctor Communication Module: *A secure messaging system enables patients and doctors to exchange messages, receive alerts, and track message delivery and read status.*

Database Management Module: *All patient records, predictions, and communication logs are stored securely using a centralized database.*

c) Architecture Layers

The system follows a layered architecture consisting of a user interface layer, application logic layer, and data storage layer, ensuring modularity and scalability.

d) Technical Stack and Implementation Details

The system is implemented using Python and Flask for backend development, machine learning libraries such as Scikit-learn and XGBoost for model training, and PHP-based modules for secure messaging. HTML, CSS, and JavaScript are used for frontend development, while MySQL is employed for data storage.

IV. METHODOLOGY AND SYSTEM DEVELOPMENT

a) Development Methodology

An iterative development approach was adopted, beginning with data analysis and model experimentation, followed by system integration and interface design.

b) Algorithm Flow

Input Data → Preprocessing → Data Balancing (SMOTE) → Train/Test Split → Model Training → Model Evaluation → Prediction
→ Output (Risk Result and Recommendations)

V. EXPERIMENTAL EVALUATION AND RESULTS

a) Dataset and Experimental Setup

The experimental evaluation was conducted using a publicly available brain stroke dataset containing demographic, clinical, and lifestyle attributes such as age, gender, hypertension status, heart disease history, average glucose level, body mass index (BMI), smoking status, and work type. Prior to model training, extensive data preprocessing was performed, including handling missing values, categorical feature encoding, and normalization of numerical attributes. As the dataset exhibited class imbalance between stroke and non-stroke cases, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to achieve balanced class distribution.

The dataset was divided into training and testing sets using an 80:20 split. Model performance was evaluated using standard classification metrics including accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

b) Model Performance Evaluation

Multiple machine learning models were implemented and evaluated to assess their suitability for stroke prediction. Logistic Regression was used as a baseline model due to its interpretability and simplicity. Tree-based ensemble models, namely Random Forest and XGBoost, were employed to capture nonlinear relationships among features and improve predictive performance.

Experimental results indicate that the Random Forest and XGBoost models significantly outperformed Logistic Regression across all evaluation metrics. The Random Forest model achieved high classification accuracy with improved recall, indicating better detection of high-risk stroke cases. XGBoost demonstrated comparable accuracy with superior precision, making it effective in reducing false-positive predictions.

c) Explainability and Feature Importance Analysis

To enhance clinical trust and interpretability, feature-importance analysis was performed using model-based importance measures. The analysis revealed that age, hypertension, average glucose level, BMI, and heart disease were the most influential features contributing to stroke prediction. These findings align with established medical knowledge, thereby validating the reliability of the proposed model.

The integration of explainability enables healthcare professionals to understand the rationale behind predictions, supporting informed decision-making rather than relying solely on black-box outputs.

VI. COMPARATIVE ANALYSIS WITH EXISTING SOLUTIONS

The system depends on the quality and size of available datasets. Additionally, cross-device synchronization and real-time sensor data integration are not currently supported.

VII. TECHNICAL IMPLEMENTATION DETAILS

The proposed Brain Stroke Prediction and Patient–Doctor Communication System is implemented using a modular and scalable technology stack to ensure reliability, maintainability, and ease of deployment in real-world healthcare environments.

Machine Learning Layer:

The prediction engine is developed using Python, leveraging well-established machine learning libraries such as Scikit-learn and XGBoost. The dataset is first loaded into a Pandas data frame for preprocessing. Missing values are handled using statistical imputation techniques, while categorical attributes are encoded using label encoding and one-hot encoding where appropriate. Feature scaling is applied to numerical attributes to improve model convergence. Class imbalance is addressed using the SMOTE algorithm, which synthetically generates minority class samples to prevent biased learning.

Multiple supervised learning models are trained, including Logistic Regression as a baseline model and Random Forest and XGBoost as advanced ensemble models. Hyperparameter tuning is performed using grid search techniques to optimize model performance. The trained models are evaluated using standard classification metrics, and the best-performing model is serialized using joblib for deployment.

Backend Integration Layer:

The trained machine learning model is integrated into a Flask-based backend application. Flask provides RESTful API endpoints that accept patient health parameters as input and return stroke risk predictions as output. This API layer acts as a bridge between the frontend interface and the prediction engine, ensuring secure and efficient data exchange. The backend also manages user authentication, request validation, and logging of prediction results.

Patient–Doctor Communication Module:

The communication subsystem is implemented using PHP scripts integrated with a MySQL database. Separate role-based modules are maintained for patients and doctors to ensure controlled access. The messaging functionality supports message creation, delivery confirmation, and read acknowledgements, thereby improving reliability of communication. Automated email notifications are triggered for high-risk predictions to alert doctors promptly.

Frontend Layer:
The user interface is developed using HTML, CSS, and JavaScript, focusing on usability and clarity. Patients can input health data, view prediction results, and communicate with doctors, while doctors can review patient risk reports and respond through a dedicated dashboard. The frontend interacts with the backend APIs asynchronously to provide a responsive user experience.

Data Storage and Security:
All patient information, prediction results, and communication logs are stored in a centralized MySQL database. Basic security measures such as input validation, role-based access control, and server-side processing are implemented to protect sensitive medical data. This layered technical implementation ensures seamless integration between prediction, communication, and data management components.

VIII. Limitations and Considerations

The system depends on the quality and size of available datasets. Additionally, cross-device synchronization and real-time sensor data integration are not currently supported.

IX. Future Enhancements and Extensions

Future work includes integrating real-time wearable data, enhancing explainability using advanced XAI techniques, and deploying the system as a mobile application.

X. Conclusion

This paper presented a Brain Stroke Prediction and Patient–Doctor Communication System using machine learning. By combining predictive accuracy, explainability, and practical deployment, the system provides a comprehensive solution for early stroke risk assessment and preventive healthcare support.

References

1. Manisha Sirsat, Eduardo Ferme, and Joana Câmara, “Machine Learning for Brain Stroke: A Review,” *Journal of Stroke and Cerebrovascular Diseases*, vol. 29, no. 10, pp. 1–12, 2020.
2. Harish Kamal, Victor Lopez, and Sunil A. Sheth, “Machine Learning in Acute Ischemic Stroke Neuroimaging,” *Frontiers in Neurology*, vol. 9, pp. 1–9, 2018.
3. R. Kumar and P. Singh, “A Comparative Study of Machine Learning Techniques for Stroke Prediction,” *IEEE Access*, vol. 8, pp. 202345–202356, 2020.
4. Kaggle, “Stroke Prediction Dataset,” Kaggle Datasets Repository, 2023.
5. S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 4765–4774, 2017.
6. IEEE, *IEEE Editorial Style Manual*, IEEE Publishing, 2023.