Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

# A Survey on Intelligent Healthcare System for Stroke Prediction using Machine Learning

[1] Sharvari Manjunath, [2] Reefa Naz, [3] Spoorthi Jain G B, [4] Rashmi V Shetty
Department of Computer Science and Engineering Malnad College of Engineering, Hassan, Karnataka, India

Kavyasri M N
Associate Professor
Department of Computer Science and Engineering Malnad College of Engineering, Hassan, Karnataka, India

***Abstract*** **- Stroke remains a foremost contributor to mortality and lasting physical impairment globally, placing immense pres-sure on modern healthcare infrastructure. Timely and accurate identification of stroke risk holds the potential to substantially decrease fatality rates and enhance patient recovery trajecto-ries. Over recent years, the application of Machine Learning (ML) and Artificial Intelligence (AI) within predictive medical systems has garnered considerable research interest, primarily for their capacity to examine complex clinical data and flag individuals at risk before a cerebrovascular event occurs. A range of computational methods — including Random Forest (RF), Artificial Neural Networks (ANN), Logistic Regression (LR), Support Vector Machines (SVM), and advanced Deep Learning architectures — have been systematically explored for this purpose. These frameworks leverage patient-specific attributes such as age profile, hypertension status, fasting glucose readings, Body Mass Index (BMI), tobacco use history, and prior cardiac conditions to generate stroke risk assessments. To address common data quality challenges, preprocessing workflows incor-porating normalization, feature engineering, and oversampling strategies like SMOTE are frequently employed. This survey critically examines existing ML-driven methodologies for stroke risk forecasting, evaluating their design choices, comparative strengths, inherent limitations, and reported findings, while underscoring the central role of intelligent clinical systems in advancing early-stage detection and preventive care.**

***Index Terms*****—Stroke Risk Forecasting, Machine Learning, Artificial Neural Networks, Random Forest, Logistic Regression, Clinical Decision Support, SMOTE, Predictive Healthcare.**

## I. INTRODUCTION

A stroke is a life-threatening neurological episode that arises when the cerebral blood supply is abruptly disrupted, either through arterial obstruction (ischemic) or vessel rupture (hemorrhagic). It stands among the most consequential causes of both premature death and persistent functional disability worldwide. The ability to detect stroke susceptibility at an early stage is of paramount clinical importance, as rapid therapeutic intervention is directly linked to reduced neurolog-ical damage and improved survival outcomes. Conventional diagnostic approaches, which rely heavily on physician ex-pertise and manual chart reviews, are often inadequate when processing the scale and complexity of contemporary patient datasets. The rapid maturation of AI and ML technologies has given rise to a new generation of clinically intelligent systems capable of transforming disease detection and risk stratification. These data-driven methods excel at uncovering latent relationships within multi-dimensional health records, enabling predictions that would be difficult for traditional sta-tistical models to achieve. Within the stroke prediction domain, architectures such as ANN, RF, LR, SVM, and various deep learning frameworks have been explored by researchers as viable clinical decision-support tools. Key clinical indicators that inform stroke prediction models include demographic variables (age), physiological measures (blood pressure, BMI, serum glucose), lifestyle factors (smoking behaviour), and existing comorbidities (diabetes mellitus, ischaemic heart dis-ease). Despite notable advancements, persistent obstacles such as class imbalance in clinical datasets, opaque model decision-making, the absence of real-time predictive capability, and limited deployment in actual clinical environments continue to hinder widespread adoption. This survey evaluates exist-ing ML-based stroke prediction frameworks, contrasting their technical approaches and practical implications to identify gaps and future directions.

## II. LITERATURE SURVEY

The application of computational intelligence to stroke risk assessment has attracted growing scholarly attention, with numerous studies introducing varied algorithmic strategies to enhance diagnostic precision and support clinicians in proac-tive patient management. Kumari and Garg [1] investigated a suite of classifiers — including Neural Networks, SVM, RF, Decision Trees, and Gradient Boosting — for cerebrovascular event forecasting. To address the black-box nature of these models, the authors incorporated eXplainable AI (XAI) tools, specifically LIME and SHAP, to render model outputs inter-pretable to end users. Among the evaluated approaches, RF demonstrated superior performance, attaining a prediction ac-curacy in the vicinity of 99%, while XAI mechanisms provided clinically meaningful justifications for model decisions. S and Shastri [2] conducted a systematic appraisal of both classical ML and contemporary deep learning techniques for brain stroke classification. Their review encompassed KNN, SVM,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

RF, LR, ANN, and Voting Classifier ensembles. The synthesis revealed that RF, SVM, and ANN consistently produced com-petitive predictive results across heterogeneous datasets, while also emphasising that unbalanced class distributions in medical data represent a recurring methodological challenge. Neeluku-mari et al. [3] developed an IoT-integrated health surveillance framework for continuous stroke risk assessment. Their archi-tecture connected ECG, pulse rate, and body temperature sen-sors to an ESP32 microcontroller interfaced with a cloud back-end. Predictive intelligence was provided by an ANN model, which achieved 92% accuracy and outperformed a parallel Generative AI module, demonstrating the suitability of ANN for real-time physiological monitoring environments. Jain and Mangal [4] assessed the comparative predictive performance of LR, Decision Trees, and RF through carefully structured preprocessing and feature importance analysis. Their findings indicated that LR, despite its relative simplicity, yielded the highest classification accuracy at 95.34%, underscoring the value of feature-level optimisation even in computationally lightweight models. Masamha et al.[5] framed stroke detection as a supervised classification task, employing the CRISP-DM framework to structure their analytical workflow. SMOTE was applied to counteract the skewed class distribution commonly observed in clinical datasets. Their model identified hyper-tension, patient age, BMI, smoking history, and HIV status as prominent stroke predictors. RF once again yielded state-of-the-art performance at approximately 99% accuracy, with a notably low false alarm rate, supporting its clinical utility. Collectively, the reviewed literature demonstrates a strong con-sensus around RF, ANN, and LR as the leading ML strategies for stroke prediction. Nevertheless, several unresolved chal-lenges — including dataset imbalance, the need for transparent model reasoning, absence of real-time deployment pipelines, and limited clinical validation — continue to motivate further research toward more robust and deployable intelligent stroke prediction systems.

## III. PROPOSED SYSTEM

The proposed system uses machine learning techniques to predict stroke risk using patient healthcare data such as age, hypertension, glucose level, BMI, heart disease, and smoking habits. The system implements Artificial Neural Network (ANN), Random Forest (RF), and Logistic Regression (LR) algorithms to improve prediction accuracy and reliability. Preprocessing techniques such as data cleaning, normalization, and SMOTE are used to improve data quality and handle imbalanced datasets. The system supports healthcare profes-sionals in early diagnosis and preventive healthcare planning.

## IV. PRELIMINARY DESIGN METHODOLOGY

The proposed methodology consists of the following stages:

- **Data Collection:** Collect patient healthcare data from medical records and monitoring systems.
- **Data Preprocessing:** Apply data cleaning, normalization, feature selection, and SMOTE.
- **Model Training:** Train ANN, Random Forest, and Lo-gistic Regression algorithms using processed data.
- **Stroke Prediction:** Predict whether the patient has high or low stroke risk.
- **Decision Support:** Provide prediction results to health-care professionals for early diagnosis and treatment plan-ning.

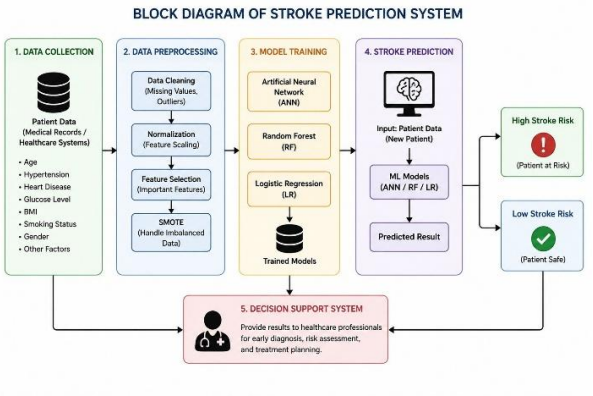


Fig. 1: Block Diagram of Stroke Prediction System

## V. COMPARATIVE ANALYSIS

Table I presents a structured comparison of the five reviewed ML-based stroke prediction systems, evaluating each against key dimensions such as methodology, strengths, limitations, and reported accuracy.

## VI. CONCLUSION

This survey has examined the role of ML-driven compu-tational frameworks in advancing stroke risk prediction and clinical decision support. The reviewed systems demonstrate that algorithms such as RF, ANN, and LR are capable of delivering high predictive reliability across a variety of healthcare datasets, particularly when complemented by ef-fective preprocessing workflows including feature selection, data normalisation, and class balancing via SMOTE. Notwith-standing these achievements, several fundamental limitations persist across the field. Most existing systems rely on static, retrospective datasets and lack the infrastructure for real-time clinical integration. Interpretability remains a concern for many high-performing models, and few systems have undergone rigorous validation in live healthcare settings. The computational demands of deep learning architectures also re-strict their viability in resource-limited clinical environments. Future development efforts should prioritise the construction of lightweight, explainable, and real-time predictive systems that leverage IoT-enabled sensor data and can be seamlessly embedded within point-of-care clinical workflows. Greater emphasis on multi-site dataset diversity, federated learning approaches, and prospective clinical trials will be critical for bridging the gap between research prototypes and deployable stroke prevention tools.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

## REFERENCES

[1] R. Kumari and H. Garg, "Interpretation and Analysis of Machine Learning Models for Brain Stroke Prediction," in Proc. 2023 6th Int. Conf. on Information Systems and Computer Networks (ISCON), IEEE, 2023.

[2] B. S. and S. K. Shastri, "A Review on Application of Machine Learning Algorithms to Predict the Brain Stroke," in Proc. 2025 IEEE Int. Conf. on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), IEEE, 2025.

[3] K. S. Neelukumari, R. Rithika, L. Murali, S. E., and N. N., "A Predictive Machine Learning Analysis for Stroke Prediction," in Proc. 2025 3rd Int. Conf. on Artificial Intelligence and Machine Learning Applications (AIMLA), IEEE, 2025.

[4] V. Jain and A. Mangal, "Performance Enhancement of Machine Learning Algorithms for Predicting Stroke," in Proc. 2024 Int. Conf. on Circuit Power and Computing Technologies (ICCPCT), IEEE, 2024.

[5] T. Masamha, F. J. Kiwa, S. Chinofunga, and K. Mujaji, "A Recurrent Neural Networks Model for Early Stroke Detection in Adults," in Proc. 2023 2nd Zimbabwe Conf. on Information and Communication Technologies (ZCICT), IEEE, 2023.

TABLE I: Comparative Analysis of Machine Learning-Based Stroke Prediction Systems

| Title | Proposed Approach | Advantages | Limitations | Research Result |
|---|---|---|---|---|
| Interpretation and Analysis of Machine Learning Models for Brain Stroke Prediction | Neural Network, SVM, RF, Decision Tree, and Gradient Boosting integrated with XAI methods (LIME & SHAP) for enhanced model transparency | Strong predictive accuracy; explainability of results; robust feature importance analysis | Small dataset size; risk of overfitting; no real-time deployment. | RF delivered approximately 99% accuracy, outperforming other models tested |
| A Review on Application of ML Algorithms to Predict Brain Stroke | Systematic review of KNN, SVM, RF, ANN, LR, and Voting Classifiers across diverse stroke datasets. | Wide algorithmic coverage; useful benchmarking of methods; highlights best performers. | Purely theoretical; inconsistent cross-study results limit generalisability. | SVM, RF, and ANN consistently reached accuracies approaching 99%. |
| Predictive Machine Learning Analysis for Stroke Prediction | ANN and Generative AI paired with ECG, pulse, and temperature sensors via ESP32 and cloud infrastructure. | Enables continuous remote monitoring; supports proactive clinical alerts. | Performance tied to sensor reliability; hardware constraints; Generative AI underperformed. | ANN attained 92% accuracy, surpassing the Generative AI counterpart. |
| Performance Enhancement of ML Algorithms for Predicting Stroke | Comparative evaluation of LR, Decision Tree, and RF using preprocessing and feature-level analysis | Straightforward pipeline; improved model output via preprocessing; results are interpretable. | Narrow algorithm scope; deep learning approaches were not explored. | LR produced the top accuracy of 95.34% after feature optimization. |
| A Recurrent Neural Networks Model for Early Stroke Detection in Adults | Supervised ML with SMOTE-based class balancing and CRISP-DM methodology for structured early detection. | Effective handling of class imbalance; minimal false positives; reliable detection rates. | Lacks real-time operational capability; dataset diversity is limited. | RF achieved approximately 99% accuracy with significantly reduced false prediction rates. |