IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Brain Stroke Prediction Using Machine Learning

Dr. Mrs. S. P. Khandait
K.D.K college of Engineering Nagpur
Department of Information Technology

Bhuvneshwari Janbandhu
K.D.K college of Engineering Nagpur
Department of Information Technology

Meghana Parate
K.D.K college of Engineering Nagpur
Department of Information Technology

Mrunali Barsagade
K.D.K college of Engineering Nagpur
Department of Information Technology

Nandini Peshne
K.D.K college of Engineering Nagpur
Department of Information Technology

I.Abstract

Brain stroke prediction using machine learning is an emerging field that aims to identify the individuals at high risk of stroke before its onset, enabling timely avoidance and intervention. Strokes, a primary cause of death and disability, often result from interruption flow of blood due to clots or vessel rupture. Traditional diagnostic methods can be slow and may miss early warning signs. Machine learning offers a data-driven method by analyzing the important health indicators, daily life factors, and physiological parameters including age, blood pressure, cholesterol, and medical history. Algorithms like logistic regression, ensemble models (e.g., XGBoost, LightGBM) can detect complex patterns to accurately classify patients by risk level. These predictive models support early detection, personalized care strategies, and improved outcomes. Future advancements may integrate real-time applications with wearable devices and hospital systems, enhancing stroke prevention efforts and transforming proactive healthcare delivery.

II.Introduction

Brain stroke is a major serious medical condition that occurs when the blood supply to the part of brain is interrupted or reduced, can cause brain cells to die due to lack of oxygen and the nutrients. It is one of the leading causes of death and long-term disability of a worldwide, and early detection plays a vital role in reducing its impact. Stroke is generally classified into the two major types such as: cerebrovascular accident, which is cause of flow in the blood vessels, an hemorrhagic stroke, which is caused due to the break or facture of a blood vessels in the brain. The early identification of the risk of stroke can help in timely treatment and preventive care, which ultimately saves lives. Traditional diagnostic methods often rely on clinical assessment and imaging techniques, which may not always provide early warning or take longer to analyze.

In this context, machine learning (ML) provides an efficient and intelligent solution by analyzing large amounts of patient health data to display the hidden patterns and risk factors associated with stroke. Parameters such as age, gender, blood pressure, cholesterol levels, diabetic status, smoking habits, obesity, heart disease and lifestyle choices can be processed using ML algorithms to predict whether a person is at risk of stroke or not. By training models such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), Random Forests and advanced combination methods such as XG Boost or Light GBM, accurate predictions can be made based on historical medical data. Such predictive systems can classify patients into risk groups, allowing healthcare providers to take preventive measures in advance.
Combination of machine learning models in stroke prediction has the potential to transform health care by enabling early detection, personalized treatment.

III. Objectives

The primary objective of this research is to develop a robust and accurate brain stroke prediction using machine learning techniques. The specific objectives of this study are listed and explained below:

1. **To Predict the Risk of Brain Stroke Using Machine Learning Models**
   Develop predictive models capable of accurately classifying individuals as stroke-prone or not based on clinical and demographic data.
2. **To Enable Early Detection and Prevention**
   Facilitate early identification of high-risk individuals, allowing for timely medical attention and preventive measures, thereby reducing the severity or occurrence of stroke-related complications.
3. **To Provide a Scalable and Automated Decision Support System**
   Design a system that can be easily integrated into healthcare infrastructures, including telemedicine platforms, for automated stroke risk assessment, particularly in rural or underserved areas with limited access to neurologists or diagnostic tools.
4. **To Improve Accuracy Over Traditional Methods**
   Leverage the capabilities of machine learning to outperform traditional statistical or rule-based systems in terms of accuracy, precision, recall, and overall reliability.
5. **To Ensure Model Interpretability and Clinical Relevance**
   Incorporate explainable AI (XAI) techniques to make the prediction outcomes interpretable for healthcare professionals, thus increasing trust and facilitating informed clinical decision-making.
6. **To Support Personalized Healthcare Recommendations**
   Provide individual-level risk assessments that can inform lifestyle changes, medical treatments, or further diagnostic testing, thereby promoting personalized preventive healthcare strategies.

IV. Literature Survey

"Brain Stroke Prediction Using Machine Learning Approach":

Previous studies have applied various machine learning algorithms for stroke prediction. Jeena et al. used Support Vector Machine (SVM) for analyzing stroke risk factors. Govindarajan et al. applied ANN, Logistic Regression, and Decision Tree, achieving 95% accuracy with ANN. Sung et al. used clinical data and found Boosted Tree models gave the highest accuracy of 96.6%. Choudhary and Singh implement PCA with ANN, reaching up to 97.7% accuracy. Selma et al. compared Decision Tree and KNN, finding Decision Tree performed better for ischemic stroke diagnosis. These studies highlight the effectiveness of ML models like ANN, SVM, and Decision Trees in predicting brain stroke risk.

“Brain Stroke prediction using machine learning algorithms”:

Previous studies show that machine learning (ML) greatly improves stroke prediction accuracy compared to traditional methods. Researchers like Bandi et al. (2020) and sirsat et al. (2020) highlighted ML's role in predicting stroke severity and reviewing its applications. Sailasya and Kumari (2021) and Emon et al. (2020) compared different ML algorithms, finding them effective for stroke risk assessment. Overall, literature indicates that ML enables early detection and better prevention of strokes.

“Brain stroke prediction by using machine learning”:

Previous studies show that machine learning (ML) techniques greatly enhance the accuracy of stroke prediction. Researchers such as Venketasubramanian et al. (2018) and Chin et al. (2018) highlighted the global impact of stroke and developed early detection models using deep learning. Jalayondeja (2018) found Decision Tree and Neural Network effective for stroke prediction, while Kim et al. (2019) and Rohit Ghosh et al. (2019) used ML and CNN models for classifying MRI and CT scans. Overall, literature suggests that ML methods like Random Forest and CNN improve early diagnosis and prevention of brain stroke.

“Brain OK: Brain stroke prediction using machine learning”:

Previous research highlights the growing use of machine learning (ML) for brain stroke prediction and diagnosis.

Sirsat et al. (2020) classified ML techniques for stroke and reviewed their performance, while Kamal et al. (2018) discussed ML's role in acute ischemic stroke imaging and treatment. Kim et al. (2019) used NLP and ML to classify MRI reports for stroke detection, and Lakshmi et al. (2017) showed that SVM achieved higher accuracy than Random Forest in brain lesion segmentation.

Yu et al. (2019) and Sailasya & Kumari (2021) demonstrated that ML algorithms like Decision Tree and Naïve Bayes effectively predict stroke risk.

Overall, literature indicates that ML models-especially Random Forest and SVM-provide high accuracy and reliability in predicting and classifying brain strokes.

“Brain Stroke Prediction”:

Amrutha and Kini (2025) developed a machine learning-based model for predicting brain strokes using patient health data such as age, hypertension, heart disease, BMI, and glucose level. The study compared algorithms including Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and Random Forest, finding that the Random Forest model achieved the highest accuracy and AUC score (0.89). Similarly, Aslam et al. (2021) and Kaur and Singh (2021) also highlighted the effectiveness of ensemble learning methods for medical risk prediction. These studies demonstrate that data-driven models can assist in early stroke detection, enabling preventive healthcare and improving clinical decision-making.

“Brain Stroke Detection Using Machine Learning”

Singh, Kaur, and Larhgotra (2024) explored the use of various machine learning algorithms such as Support Vector Machines (SVM), Random Forests, K-Nearest Neighbors (KNN), and Convolutional Neural Networks (CNNs) for detecting brain strokes from medical imaging data. The study emphasized preprocessing, feature extraction, and transfer learning to enhance model performance. Results indicated that deep learnings models, particularly CNNs, achieved superior accuracy in identifying stroke regions. The authors concluded that integrating machine learning with medical imaging could significantly improve diagnostic speed and precision, supporting better clinical decision-making.

“Prediction of Stroke Using Machine Learning”:

Kunder et al. (2020) proposed a stroke prediction system using machine learningalgorithms such as Decision Tree, Naïve Bayes, and Artificial Neural Network (ANN) to predict the likelihood of stroke based on modifiable risk factors like age, hypertension, heart disease, BMI, glucose level, and smoking status. Their study utilized data preprocessing, feature encoding, and model training to compare classifier performance using ROC-AUC metrics. The results indicated that ANN achieved the highest accuracy and reliability in identifying stroke-prone individuals. The authors concluded that integrating machine learning models with health monitoring systems can provide early warnings and support preventive healthcare decisions.

V.Methodology

The proposed system aims to predict the possibility of stroke risk based on various health and lifestyle parameters provided by the user through a web-based application. The complete methodology involves data collection, model training, testing, and result visualization, as described below:

1. Data Collection and Input Parameters

The system collects user details such as:
Gender, Age, Hypertension, Heart Diseases, Marital Status, and Work Type
Residence Type, Average Glucose level, Body Mass Index (BMI), and Smoking status
These parameters are entered by the user through a web application interface.

2. Data Processing

The collected data is preprocessed to handle missing values, normalize numerical features, and encode categorical attributes. This ensures the data is suitable for model training and prediction.

3. Model Training

Multiple models are developed and trained using the prepared dataset to analyze the relationship between user parameters and stroke occurrence.
The models used include:
Logistic Regression
Random Forest Classifier
XGBoost Classifier
LightGBM (Light Gradient Boosting Machine)
Each of the model is evaluated based on accuracy, precision, recall, and F1-score.

4. Model Selection

The model with the highest accuracy and best performance metrics is selected as the final model for stroke prediction.

5. **User Interaction and Prediction Process**

The user inputs their information through the web interface. The system processes the input and tests it against the selected trained model. Based on the model's outcome, the system predicts whether the user is at risk of stroke or not at risk.

6. Result Display

The final prediction result is displayed on the web interface:
If the output indicates "Stroke Risk Diagnosed," the user is notified accordingly.
If the output indicates "No Stroke Risk Diagnosed," it is also displayed to the user.

7. System Output

The web interface thus provides an interactive and accessible way for users to understand their stroke risk based on personal health and lifestyle data.

VI.Workflow:

The Brain Stroke Prediction System helps people know their stroke risk early. The user enters health details like age, blood pressure, sugar level, weight, and medical history on a web page. This information is sent to a Flask-based system, which checks the data and passes it to a trained machine learning model. The model then predicts whether the stroke risk is Low, Moderate, or High. The system shows this result on the user's screen and also gives health advice. For example, if the risk is high, the user may be told to see a doctor and improve their lifestyle. The system is fast, easy to use, and helps in stroke prevention. In the future, it can be improved using smartwatches and better AI models for real-time monitoring.

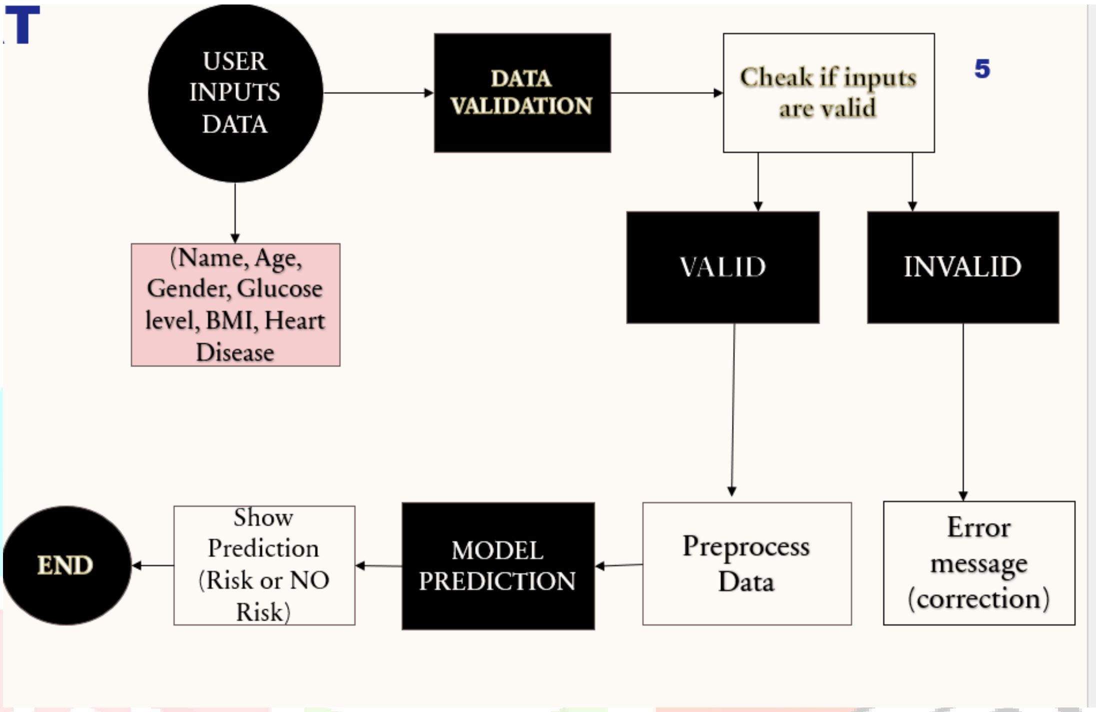

Fig. Workflow of System

VII.Conclusion

In this study, we proposed a machine learning-based system for the early prediction of brain stroke, utilizing various health and lifestyle-related parameters to identify individuals at high risk. By leveraging algorithms such as Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM), the system demonstrated the potential to offer accurate and efficient stroke risk classification.

The results highlight the promise of data-driven models in enhancing clinical decision-making, especially in scenarios where timely diagnosis is critical. The integration of such predictive tools into healthcare infrastructure can not only support physicians in preventive care but also extend medical assistance to underserved and remote areas lacking specialized resources. Furthermore, the emphasis on model interpretability and personalization makes the system a practical candidate for real-world deployment, where transparency and patient-specific recommendations are crucial. Future work will focus on incorporating real-time health data, expanding the dataset, and employing deep learning techniques to further improve prediction performance.

The research aims to create strong machine learning models that can reliably identify strokes from medical imaging data with high sensitivity and specificity by means of comprehensive experimentation and evaluation. To evaluate the trained models' performance in real-world scenarios, independent datasets will be used for validation, and real-world clinical data will be used for testing. In the end, this project's successful completion will result in the creation of a dependable and effective diagnostic tool for brain stroke detection, which may help medical professionals diagnose and treat stroke patients in a timely manner. Through the increased precision and efficacy in stroke detection, the initiative hopes to improve patient outcomes and clinical setting healthcare delivery.

VIII. References

[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.