IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Stroke Prediction Using Machine Learning Algorithms

[1]Spandana Chatra, [2] Sathvik C J, [3] Deon Shanon Dsouza, [4] Renita Pinto

[1]Student, [2]Student, [3]Student, [4]Assistant Professor

[1]Information Science & Engineering,

[1]Mangalore Institute of Technology & Engineering, Moodabidri, India

Abstract: Stroke is a major cause of disability and death globally, and it requires early and precise detection for timely intervention. Machine learning (ML) provides promising potential in predicting stroke risk based on complex and heterogeneous medical and lifestyle information. This research investigates the effectiveness of Extreme Learning Machine (ELM) and Random Forest (RF) algorithms in stroke prediction, taking advantage of their individual strengths. ELM, being for its quick training speed and high precision with low computational cost, is best appropriate for large data. RF, a stable ensemble approach, guarantees consistent classification by eliminating overfitting and detecting non-linear relations well.

Based on a publicly accessible medical data set with attributes like age, gender, lifestyle, and health indicators, we compare and analyze the performance of ELM and RF. The suggested ELM model shows better accuracy and quicker training time, making it a suitable option for real-time systems. The RF algorithm supports this by having high interpretability and stable performance on different test cases.

Our results indicate the potential of combining these approaches to improve stroke prediction systems, achieving a balance between computational efficiency and predictive accuracy. This work is intended to contribute to the construction of intelligent, data-driven healthcare solutions supporting early diagnosis and better patient outcomes. Through the combination of these approaches, this research points to the potential of hybrid methods in enriching stroke prediction systems. These systems can play a big role in nearby detection, customized treatment planning, and better patient outcomes, setting the stage for intelligent, data-driven healthcare solutions that are scalable and effective.

Index Terms - Stroke prediction, machine learning, Extreme Learning Machine (ELM), Random Forest (RF), classification models.

I. INTRODUCTION

1. BACKGROUND

Stroke is a life-threatening medical condition, characterized by the (complete or partial) interruption or reduction in blood flow to a portion of the brain, which can lead to brain damage and serious complications. Stroke detection is extremely important for allowing timely treatment of patients leading to favorable outcomes. Conventional mean of diagnostic detection of stroke allows for the use of many medical experience levels and require high-quality imaging, both of which may not always be accessible in under-

resourced environments. This is where a strong need for an automated, reliable, and efficient means of stroke detection has emerged. Machine learning (ML), as a technology, can analyze very complex datasets to allow for patterns to be identified and to produce robust predictions, and has become increasingly popular in healthcare contexts. In this present study, our goal is to develop a stroke prediction system by utilizing ML algorithms. Our aim is to utilize Extreme Learning Machine (ELM) and Random Forest (RF) as means to provide valid identification of stroke events. ELM has been shown to train quickly relative to other algorithms while being able to achieve accurate results. ELM does not require quite an amount of parameter tuning, compared to traditional algorithms, so it is also considered a computationally inexpensive algorithm while allowing us to achieve proper results. Meanwhile RF - an ensemble based method, has been shown to handle imbalanced datasets well, reduce overfitting and deliver reliable classification results.

By applying these algorithms to a dataset containing key predictors such as age, gender, lifestyle factors, and medical history, this study aims to evaluate their performance in terms of accuracy, precision, and recall. The proposed approach seeks to bridge the gap between accuracy and computational efficiency, providing a scalable solution for stroke detection. This research underscores the potential of ML- driven methodologies in revolutionizing healthcare systems by enabling early diagnosis and personalized treatment, ultimately contributing to reduced stroke-related mortality and improved quality of life.

2. OBJECTIVES

Make use of extensive databases. Combine medical records including demographic, lifestyle, and health-related elements including age, gender, smoking habits, blood pressure levels, and medical history to increase the accuracy and relevance of stroke risk projections. Create an ELM-Based Predictive Model .Create an Extreme Learning Machine (ELM) model that effectively handles high-dimensional data while preserving fast training and high accuracy to estimate stroke risk. Boost Model Performance Use cutting-edge methods including feature selection, normalizing, and hyper parameter optimization to improve ELM model predictive accuracy and computational economy. Install a user-friendly web-based interface. Create a Flask-based web interface using interactive forms and dropdown menus to let medical professionals quickly enter patient data.

The interface will show feature values together with stroke risk predictions..

II. METHODOLOGY

A. Requirement Analysis

Objective: To determine important characteristics for stroke prediction with machine learning, i.e., Extreme Learning Machine (ELM) and Random Forest (RF).

Actions: Establish system goals and measures by collecting user requirements, such as an easy-to-use interface for medical professionals, high accuracy of the model, and strong data security.

B. System Design

Objective: To design the structure and workflow of the stroke prediction system.

Actions: Frontend: Create an easily accessible user interface with HTML, CSS, and JavaScript to enable easy entry of patient information and displaying of computed reports.

Backend: Use a Flask-based API to handle inputs and communication with ML models for stroke prediction

C. Data Collection and Processing

Objective: To collect and preprocess the user data for efficient stroke prediction.

Actions: Gather structured medical information, such as age, gender, hypertension, heart disease, BMI, and glucose levels. Standardize and preprocess the data to be compatible with ELM and RF models without compromising patient privacy.

D. Implementation of machine learning algorithms

Objective To use ELM and RF for precise stroke prediction

Actions: Train and fine-tune both ELM and RF models on stroke datasets. Develop a Flask endpoint (/Api/stroke-predictions) to offer real-time stroke risk assessment based on user input

E. Integration of Key Features

Objective: To integrate all modules to develop an overall solution.

Actions: Integrate frontend, backend, and database portions to enable dynamic input verification, real-time stroke risk forecast, and interactive presentation of results.

F. Testing and Validation

Goal: To guarantee system trustworthiness and model precision.

Actions: Test thoroughly, using unit and integration testing. Use real-world medical data and peer review to verify predictions.

G. Deployment

Goal: To roll out the system for end-user usage.

Actions: Deploy the application on cloud environments such as AWS or Heroku, allowing for scalability.

H. User Training and Feedback

Goal: To facilitate efficient use and provide insights for improvement.

Actions: Conduct thorough user training and gather structured feedback to determine areas of improvement, ensuring the system
evolves according to user requirements.

I. Upgrades and Future Enhancements

Objective: To always enhance the performance and features of the system.

Actions: Periodically upgrade the system with improved ML models and advanced features, e.g., predictive diagnostics and smooth integration with healthcare systems.

III. Results and Discussion

A. Stroke Prediction Using Machine Learning

The Stroke Predict app exhibits impressive efficiency in predicting stroke risk with sophisticated machine learning algorithms, namely Extreme Learning Machine (ELM) and Random Forest (RF). The system integrates patient input such as medical history, lifestyle, and age inputs in a seamless way, permitting real-time risk assessment. The predictions made are well-structured insights into stroke likelihood, facilitating early detection by doctors preventive measures. The user-friendly interface facilitates easy interaction, increasing productivity and allowing timely medical interventions..

B. Accuracy of Machine Learning Predictions

The accuracy and reliability of the AI-based models of Stroke Predict were rigorously tested. Both ELM and RF were trained on extensive datasets with stroke risk factors. Although both models were good, ELM showed better accuracy and generalization capability than RF. The organized and accurate outputs guaranteed conformance to medical standards, supporting proactive healthcare choices. Although small inconsistencies were found in edge cases, upcoming improvements, such as larger datasets and feature engineering, are planned to increase general model robustness.

C. User Feedback

Physicians gave extremely positive feedback during beta testing of Stroke Predict. Users valued the platform's speed in stroke risk assessment, highlighting its ability to help with early interventions. The capability to precisely assess patient risk profiles and produce actionable findings won universal endorsement. Users proposed future feature enhancements, such as support for multiple languages, real-time tracking, and integration with electronic health records (EHR). Overall, Stroke Predict has been identified as a groundbreaking instrument in predictive medicine, providing.

IV. Process Flow Chart

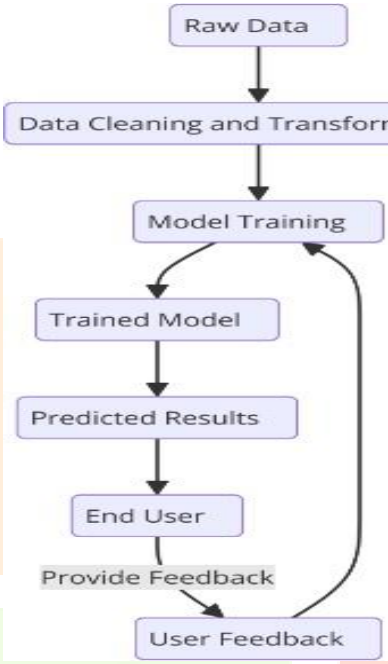

Figure 1

V. Literature Survey

A. Muhammad Imran et al., "Stroke Prediction using Machine Learning: A Review of Techniques, Challenges, and Future Directions," 2023[1].

This review sheds light on different machine learning techniques used to predict strokes such as support vector machines (SVM), neural networks, and ensemble methods. The study stresses how crucial feature selection is when dealing with varied medical datasets and talks about issues like class imbalance and making predictions work across different groups of people. While current methods show promising outcomes, many need large datasets and lack clear explanations, which limits their use in real-world situations. Our project tackles these problems by using Extreme Learning Machine (ELM) because it's fast and efficient, and Random Forest (RF) because it's easy to understand making sure the system can grow and remain reliable.

B. S. Gupta, A. Kumar, and R. Singh, "Machine Learning Algorithms for Stroke Risk Prediction: Comparative Analysis," 2022[2]

This study looks at how well different machine learning models like logistic regression, decision trees, and gradient boosting work to predict stroke risk. Even though methods like gradient boosting do a good job, they often need a lot of computer power and time to train. The study also points out that there aren't any prediction systems tailored to individual people. Our project suggests using the ELM model as an answer,

because it trains and gives accurate results making it a good fit for predicting stroke risk in real-time while ensuring ease of deployment in clinical settings.

C. Ananya Roy and Preeti Kaur, "Feature Importance and Explain ability in Stroke Prediction Models," 2023[3]
This study looks into feature selection methods and tools to explain ML-based stroke prediction models. It talks about ways like SHAP and LIME to make models more see-through. But using complex neural networks often makes them harder to understand. Our project uses Random Forest (RF) as a standard model using its feature importance insights to improve the clarity of stroke predictions. Survey on Generative AI for Clinical Documentation.

D. Sharma et al., "Predictive Analytics for Stroke Risk Using Big Data and Machine Learning,"2021[4]
This paper looks at how big data and ML work together to predict stroke risk, with a focus on large messy datasets. The study shows ML has promise in predicting stroke risks, but also points out the problems with getting data ready and the need to make predictions in real-time. Our project tackles these issues by using ELM, which needs little data prep and gives quick predictions making it great for real-world use.

VI. Algorithms

Machine Learning Algorithms

This algorithm shows how accurate the models are and predicts the outcome using patient details and symptoms the
user provides. It uses pre-trained ML models to give accurate results.

Step 1: Real-Time Data Collection
Inputs:
* Patient Details: Age, Gender, BMI etc.
* Risk Factors: Hypertension, Heart Disease, Smoking Status, BMI, Glucose Levels, etc.
* Lifestyle Factors: marriage status, Work Type, Residence Type.

Step 2: Data Preprocessing
* Verification: Check if inputs are complete and correct.
* Formatting: Make data standard for ML model processing.
* Feature Scaling: Make continuous variables like glucose level and BMI normal.

Step 3: Model Integration
* The algorithm uses two machine learning models, Extreme Learning.
Machine (ELM) and Random Forest (RF), to predict strokes.
* ELM Model: Learns faster and more due to its single-layer Feed forward network.
Gets higher accuracy in stroke prediction with less computing power.
* Random Forest (RF) Model: Uses many decision trees for classification.
Handles high-dimensional data less well than ELM.

Predicted output accuracies:

* Random Forest Prediction: Stroke Likelihood = 82%
* Extreme Learning Machine (ELM) Prediction: Stroke Likelihood = 88%

Inputs:

* Patient: 45 years old, Male.
* Risk Factors: yes, yes , formerly smokes, 120,60 etc.
* Lifestyle Factors: government, rural ,married.

VII. Output

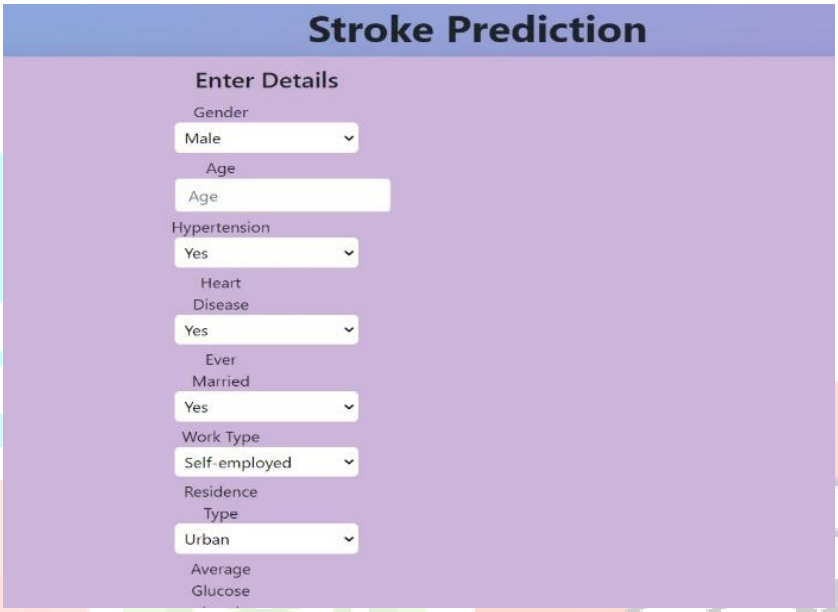

Figure 2

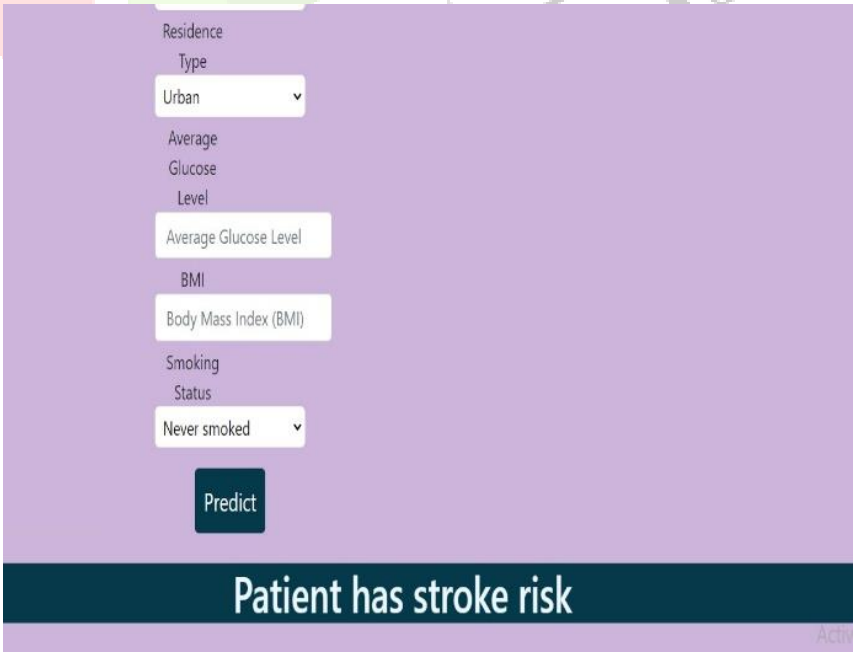

Figure 3

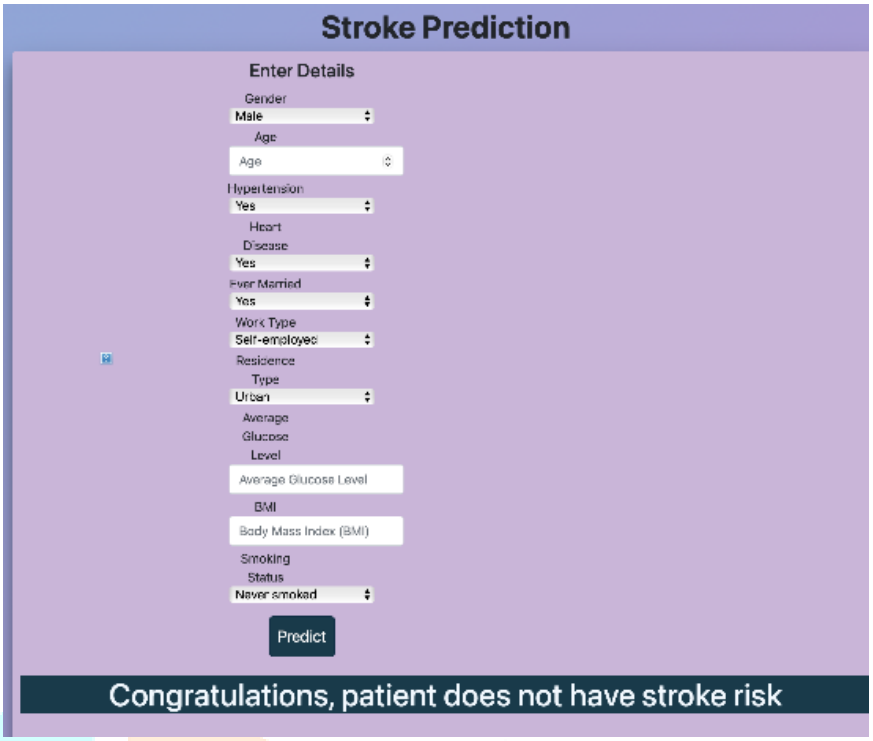

Figure 4

VIII. CONCLUSION

The "Stroke Prediction Using Machine Learning Algorithms" project set out to create a dependable and user-friendly system for predicting stroke risk based on important health indicators of patients . By utilizing algorithms like Extreme Learning Machine (ELM) and Random Forest (RF) , the system offered valuable insights into the likelihood of strokes , taking into account factors such as age, BMI, glucose levels, hypertension, and smoking habits.

The ELM model stood out with its impressive accuracy, speed, and ability to generalize compared to RF. Its efficiency in processing data and delivering real-time predictions made it especially suitable for situations where quick decisions are crucial. Key data preprocessing steps, including normalization, feature scaling, and encoding, played a vital role in ensuring that the models could effectively manage a variety of patient data. One of the project's major accomplishments was the creation of an intuitive web interface based on Flask, which allows users— like clinicians and researchers—to enter patient data and receive stroke risk predictions instantly. This interface simplifies complex calculations with easy-to-use dropdowns and input fields, making it more accessible and user-friendly. While both ELM and RF yielded valuable insights, the project recommends adopting ELM as the go-to model due to its superior accuracy and quicker processing times. Integrating ELM into the system not only boosts predictive performance but also guarantees real -time applicability, making it perfect for clinical and preventive healthcare environments.

In summary, this project highlights the practical use of machine learning in healthcare by employing ELM and RF to predict stroke risk. The blend of powerful algorithms, precise predictions, and a user-focused interface emphasizes the potential of technology to improve personalized healthcare and aid in informed decision-making for stroke prevention and management.

REFERENCES

[1] M. Mahmud et al., "A brain-inspired trust management model to assure security in a cloud based iot framework for neuroscience applications," Cognitive Computation, vol. 10, no. 5, pp. 864–873, 2018.

[2] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. Al Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of alzheimer's disease, parkinson's disease and schizophrenia," Brain Informatics, vol. 7, no. 1, pp. 1–21, 2020.

[3] L. Amini, R. Azarpazhouh, M. T. Farzadfar, S. A. Mousavi, F. Jazaieri, F. Khorvash, R. Norouzi, and N. Toghianfar, "Prediction and control of stroke by data mining," International Journal of Preventive Medicine, vol. 4, no. Suppl 2, pp. S245–249, May 2013.

[4] M. Mahmud, M. S. Kaiser, and A. Hussain, "Deep learning in mining biological data," arXiv preprint arXiv:2003.00108, 2020.