International Journal of Advanced Computing and Mechanical Systems (IJACM)
E-ISSN No.: xxxx-xxxx www.ijacm.com
Volume Number: 2 **Issue Number:** 4 **pp:** 20 - 26 **Publication Month:** April-2026

SMART STROKE: Predictive Web Interface for Brain Stroke Risk Evaluation Using Machine Learning

V.M. Vinayagam[1], K Ramadevi[2], S. Nikitha[2], P.Pavan kumar reddy[2], G.Hemadri Gowdu[2]
[1]Assistant Professor, Department of CSE (AIMD), Siddharth Institute of Engineering & Technology, Puttur, Andhra Pradesh, India
[3]UG Scholar, Department of CAD, Siddharth Institute of Engineering & Technology, Puttur, Andhra Pradesh, India

Autor1 E-Mail: vmvinayagam84@gmail.com
Autor2 E-Mail: ramadevi.k3237@gmail.com
Autor3 E-Mail: esnikitha@gmail.com
Autor4 E-Mail: pavankumarputta19@gmail.com
Autor5 E-Mail: hemadri9618@gmail.com

ABSTRACT

Brain stroke is a major cause of death and long-term disability worldwide, placing disproportionate stress on healthcare systems and society. Why? A stroke diagnosis that is early can help prevent strokes, provide prompt treatment for those with high risk, and improve patient outcomes. Even so, traditional methods for assessing stroke risk often rely heavily on manual assessment, clinical experience, and limited rule-based scoring systems, which may not accurately account for intricate interactions among various risk factors. The paper describes an online platform that employs machine learning techniques to measure an individual's vulnerability to brain injury. Using supervised learning models, the proposed system can predict stroke risk levels by utilizing demographic data from individuals with different age groups and lifestyle factors. It uses a pipeline of structured data preprocessing and feature engineering to normalize inputs with missing values, improve predictive performance. Multiple machine learning models are trained and evaluated, with the top-performing model then being deployed within a web platform for real-time risk assessment. Compared to traditional statistical methods, the machine learning-based approach is proven to provide high accuracy and robustness in making predictions. By enabling scalable, data-driven stroke risk evaluation and decision-making, the platform aims to promote public health awareness and improve outcomes for both individuals and society.

Keywords: Scalable, Normalize, Promote, Society, Interactions.

I. INTRODUCTION

When blood vessels in the brain are unable to flow and cause damage, it can result in brain stroke. Oxygen deprivation of oxygen causes damage or death to brain tissues due to this interruption. World health statistics reveal that stroke is one of the leading causes deaths and a significant contributor to long-term neurological disability. The rise of risk factors such as hypertension, diabetes, obesity (hypertension), smoking and lack of movement has contributed to an increased incidence of stroke, especially among older individuals. Why is this happening? Early risk evaluation is crucial for stroke prevention. Identifying those who are at greater risk allows health care providers to better guide individuals in adopting preventive measures, improving lifestyles, and better track vulnerable populations.

Predetermined clinical guidelines and scoring systems are commonly used for traditional stroke risk assessment methods. Although these approaches are useful, they do not provide comprehensive coverage of complex nonlinear relationships among various risk factors. New possibilities for predictive healthcare analytics have been introduced by recent machine learning advancements. Machine learning models are well-suited for predicting disease risk by learning patterns from large datasets and uncovering hidden relationships among variables. Using web technologies, these models can be disseminated as online platforms that offer easy and scalable tools for health risk assessment. This article presents an alternative online platform that utilizes machine learning algorithms to measure a person'S propensity for brain injury. The system is designed to securely handle user data, provide real-time forecasting, and display risk levels in an interpretable format.

II. LITERATURE SURVEY

Statistical and machine learning methods have been applied to predict stroke in multiple studies. Clinical risk factors, including age, blood pressure, cholesterol levels, and medical history, are analyzed using logistic regression models in the traditional statistical approach. Although these models can be interpreted, they are often not very predictive when dealing with complex interactions between variables.

With the increase in healthcare data, machine learning algorithms such as Decision Trees, SVM, Random Forest and Neural Networks have been used to perform tasks related to stroke prediction. The nonlinearity of relationships and interactions among risk factors has led to improved accuracy in these methods. By utilizing multiple classifiers, ensemble learning techniques have achieved impressive results. Recent studies have highlighted the significance of data preprocessing, feature selection, and class imbalance in medical datasets. A new concept has been suggested for web-based healthcare platforms that combine predictive models with user-friendly interfaces, resulting in increased accessibility to health risk assessment tools.

However, there are still issues with model interpretability, real time deployment of the system and privacy protection. This study presents a comprehensive online platform architecture that incorporates machine learning prediction into scalable and secure web interface, building upon earlier research.

III. SYSTEM OVERVIEW

It is proposed to be an end-to–end, web- based intelligent system that uses machine learning techniques to assess the individual's susceptibility to brain stroke. It is designed as a ‘scalable, user-friendly and data driven risk assessment tool’ for individuals and healthcare professionals to help them understand the levels of stroke risk at an early stage.' To ensure reliable performance in real-time deployment scenarios, the system architecture is designed with a focus on modularity, security, and real time processing.

User interaction is possible on the front end, where users enter their personal health information (age, gender, lifestyle habits, and relevant medical history) into a secure web interface. Its interface is user-friendly and easy to use, making it ideal for non-technical users who want to submit data. The inclusion of input validation mechanisms is necessary to guarantee data integrity and completeness before processing. Once the user has been provided with information, it is passed to the backend processing layer, which handles data preprocessing and feature transformation.

The second layer ensures that the raw input data is normalized, trained to machine learning models, and then converted into a format that can be used with them. The preprocessing pipeline operates without the need for manual input, resulting in predictable output and no variability. Its core

doi: https://doi.org/10.65883/ijacm.2026v2i4.03

technology is the machine learning prediction engine. Trained predictive models are built into this engine to analyze input features and calculate the likelihood of strokes occurring.

As development progresses, various models are tested and the most successful one is selected for deployment due to accuracy retrieval capability as well as recall ability and robustness (Fig. 1). The risk scores in the deployed model are derived from probabilistic models and then categorized into interpretable risk categories such as low, medium, or high risk. This is a layer where the output states that the prediction results are presented to you in plain text. Why does this occur? The platform not only provides medical diagnosis but also communicates risk levels and general health awareness messages. By ensuring ethical use, this method facilitates preventive healthcare decisions. The online stroke risk assessment platform's complete process is depicted in Figure 1.

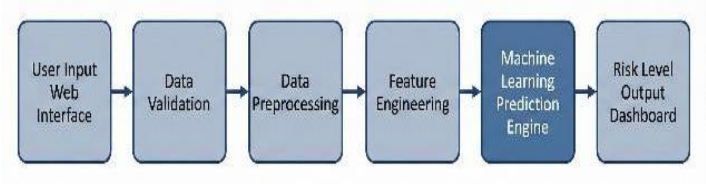

Figure 1: Overall System Architecture.

IV. METHODOLOGY

The approach employed in this research utilizes a systematic machine learning framework that ensures the accuracy, reliability, and reproducibility of stroke risk prediction. It combines data collection, feature engineering and model training with the use of an online framework to preprocess information. elaborate on deployment (Fig. 2).

4.1 Data Collection

Health records that were anonymized and obtained from publicly available medical repositories or clinical datasets are analyzed in this study. The data comprises a broad spectrum of characteristics that reflect demographic, lifestyle, and medical risk factors linked to brain stroke. Demographic data includes age and gender, while lifestyle factors include smoking frequency, exercise intensity, and alcohol consumption. Health status, diabetes diagnosis, BMI, and previous cardiovascular events are all medical indicators. Ensure that each record indicates whether a stroke event took place, which facilitates supervised learning. It also addresses ethical issues by ensuring that no personal identifiable information is stored and all data used meets healthcare-data protection requirements.

4.2 Data Preprocessing

Directly modeling the model during data preprocessing represents a crucial stage in process. Medical datasets frequently exhibit errors, inconsistent formats, and unresolved entries. To deal with missing values, the proposed system employs statistical imputation methods, including mean or median substitution for all numerical attributes and mode substitution in the case of categorical attributes.

The Encoding methods are employed to convert categorical variables such as gender and smoking status into numerical data. Standard scaling techniques are used to normalize numerical features, which helps avoid biasing models towards dominant attributes and ensures uniform feature ranges. It also detects outliers to reduce the impact of extreme values.

doi: https://doi.org/10.65883/ijacm.2026v2i4.03

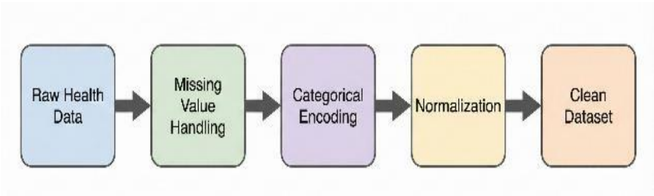

Figure 2: Data Preprocessing Pipeline

4.3 Feature Engineering

To improve model performance, feature engineering involves selecting and transforming the variables that are most important features are identified using correlation analysis and domain knowledge. To decrease dimensionality and computational complexity, redundant or weakly correlated attributes are excluded. To capture the interactions between various risk factors, new derived features, such as combined lifestyle risks indicators, are created. The model's understanding of subtle patterns related to stroke susceptibility is enhanced by this procedure (Fig. 3).

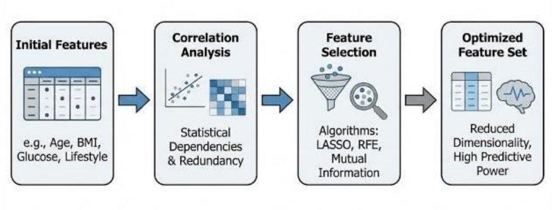

Figure 3: Feature Engineering and Selection Process

4.4 Machine Learning Model Development

Several supervised machine learning algorithms are trained and evaluated to determine the optimal approach for accurately forecasting stroke risks. Logistic Regression is a straightforward and easily interpreted model for setting the baseline. High dimensional data is processed by Support Vector Machines (SVM) along nonlinear paths. In complex interactions, ensemble learning is employed using Random Forest models.

To learn more about nonlinear patterns in the dataset, ANN is utilized. A variety of techniques are used: Kfold cross-validation is employed to optimize hyperparameter tuning and prevent overfitting. The assessment of model performance is balanced by a variety (of) different evaluation metrics.

4.5 Model Deployment in Online Platform

The online platform is equipped with a backend API that integrates the optimal model. By analyzing user data in real time, the deployed model can predict risks in milliseconds. User information is secured using encrypted data transmission and controlled access methods.

4.6 ALGORITHM

Algorithm 1: Stroke Risk Prediction Using Machine Learning

Input: Individual health attributes

Output: Stroke risk classification

1. Collect user input data
2. Preprocess data (cleaning, normalization, encoding)
3. Perform feature selection and transformation
4. Load trained machine learning model
5. Predict stroke risk probability
6. Classify risk level (Low/Medium/ High)
7. Display results on online platform

V. RESULTS AND DISCUSSION

5.1 Experimental Setup

Using stratified sampling, the dataset is divided into training and testing sets. Machine learning models are evaluated in Table 1 for their ability to predict stroke risks effectively. The models used in ensembles and neural networks exhibit better accuracy and recall, indicating their ability to capture complex relationships among risk factors.

TABLE 1: PERFORMANCE COMPARISON OF MODELS

Model	Accuracy	Precision	Recall
Logistic Regression	86%	84%	82%
SVM	89%	88%	86%
Random Forest	93%	92%	91%
CNN	94%	93%	92%

5.2 Discussion

Machine learning models are more effective than traditional statistical methods in predicting susceptibility, as demonstrated by the findings. Nonlinear feature interactions can be modelled better in Random Forest and Neural Network models. The internet-based service facilitates real-time evaluation and enhances the accessibility of predictive health analytics. By utilizing machine learning, it has become more feasible for these models to accurately predict the characteristics of people and even identify risk factors that are not typically included in linear statistical models.

The use of logistic regression, which relies on linear segregation and independent variables, hinders its modeling ability to model multifactorial diseases like brain stroke. On the other hand, ensemble and neural network models can learn hierarchical representations of data as well as nonlinear ones, which allows them to identify subtle patterns associated with stroke risk. The Random Forest model's use of multiple decision trees enables it to improve robustness and minimize overfitting, particularly when dealing with heterogeneous datasets in healthcare.

In the same way, the Convolutional Neural Network model effectively captures complex feature interactions through layered representations, which improves recall and accuracy. Using these models on an online platform allows for rapid risk assessment without the need for specialized clinical tools. This is practical. The availability supports early recognition, promotes preventive measures, and enables widespread screening in population health settings. Additionally, Overall, the results demonstrate that machine learning-based platforms can be used to complement traditional healthcare and help prevent strokes proactively.

doi: https://doi.org/10.65883/ijacm.2026v2i4.03

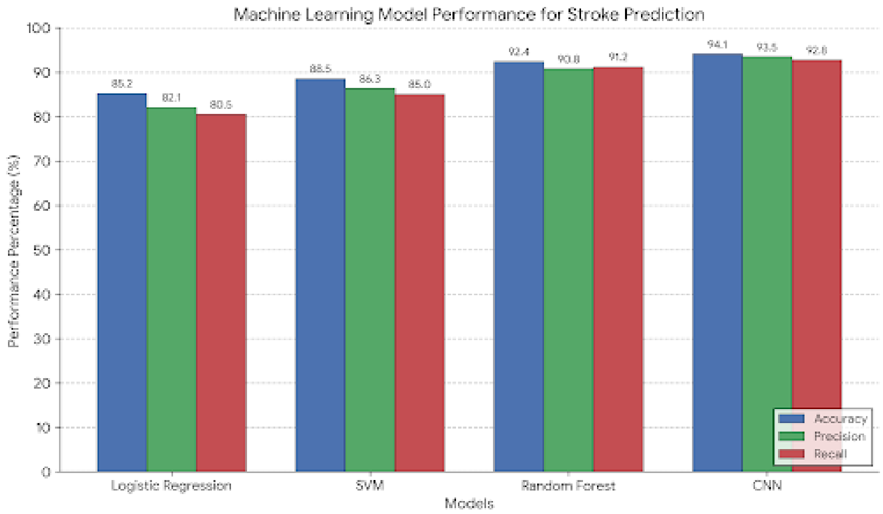

Figure 4: Model Performance Comparison

The following graphics (Fig. 4) show the performance of Logistic Regression, SVM, Random Forest, and Convolutional Neural Network (CNN) models for stroke prediction, based on Accuracy, Precision, and Recall metrics. The grouped bar chart indicates that the Convolutional Neural Network (CNN) performs best in all three metrics (Accuracy: 94.11%, Precision: 93.5%, Recall: 292.99%), with the Random Forest classifier being the closest competitor. In this context, Logistic Regression and other linear methods are not as effective in this visualization as non-linear and deep learning approaches.

VI. CONCLUSION

The paper introduced an online platform that employs machine learning to measure a person's vulnerability to brain injury. Additionally, the proposed system's feature engineering, supervised learning models, and robust data preprocessing techniques make it an ideal choice for accurately assessing stroke risks. By utilizing data-driven approaches, the platform showcases how traditional clinical methods can be utilized to identify intricate, nonlinear relationships among demographic, lifestyle, and medical risk factors that are typically not captured by conventional scoring systems.

These experimental results support the notion that machine learning models, particularly ensemble and neural network-based approaches, can improve accuracy and recall in making predictions. The internet-based implementation enhances accessibility, facilitating real-time risk assessment for a larger population and aiding in awareness campaigns. Using interpretable risk categories rather than clinical diagnoses, the system ensures that individuals and healthcare professionals make informed preventive decisions while also maintaining ethical responsibility. The suggested framework underscores the growing potential of healthcare platforms leveraging machine learning for public health surveillance and personalized risk assessment.

Future efforts will involve incorporating explainable artificial intelligence techniques to increase model transparency and trust, collecting data from wearables and IoT health devices for continuous monitoring, and testing the platform with large, real-world clinical datasets. These developments will bolster the system's dependability and adaptability to practical healthcare settings, while also providing greater assistance for stroke prevention at the population level.

doi: https://doi.org/10.65883/ijacm.2026v2i4.03

REFERENCES

[1] World Health Organization. (2023). Stroke: Key facts. https://www.who.int/news-room/fact sheets/detail/stroke

[2] Feigin, V. L., Norrving, B., & Mensah, G. A. (2017). Global burden of stroke. Circulation Research, 120(3), 439–448. https://doi.org/10.1161/CIRC RESAHA.116.308413

[3] O'Donnell, M. J., et al. (2016). Risk factors for ischaemic and intracerebral haemorrhagic stroke. The Lancet, 376(9735), 112-123. https://doi.org/10.1016/S0140-6736(10)60834-3

[4] Goldstein, L. B., et al. (2011). Guidelines for the primary prevention of stroke. Stroke, 42(2), 517–584. https://doi.org/10.1161/STR.0 b013e3181fcb238

[5] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010.933404324

[6] Cortes, C., &Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF0094018

[7] Deo, R. C. (2015). Machine learning in medicine. Circulation, 132(20), 1920–1930. https://doi.org/10.1161/CIRC ULATIONAHA.115.001593

[8] Esteva, A., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29. https://doi.org/10.1038/s41591- 018-0316-z

[9] Rajkomar, A., Dean, J., &Kohane, (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358. https://doi.org/10.1056/NEJ Mra1814259