An Explainable Machine Learning Framework for Personalized Stroke Risk Prediction

G.L.I.S. Karunarathne[a]* and R.L.C.Shyama[a]

[a] *Department of Interdisciplinary Studies, Faculty of Technology, University of Moratuwa, Moratuwa 10400, Sri Lanka*
[a*] *Department of Interdisciplinary Studies, Faculty of Technology, University of Moratuwa, Moratuwa 10400, Sri Lanka*

karunarathneglis.20@uom.lk

Abstract

Since stroke is one of the world's leading causes of death and permanent disability, prevention and efficient clinical management depend on early and accurate risk prediction. The lack of trust quantification and poor interpretability of machine learning models based on clinical and demographic data limit their use in healthcare, despite their potential. Many current models function as "black boxes," offering few patient-specific justifications or assurance metrics, which undermines clinician confidence and usefulness. This study develops a personalized stroke risk prediction system integrating explainable AI to address these challenges. The Kaggle dataset of 5,110 patient records was utilized as a workable substitute for publicly available stroke data that is specific to the Sri Lankan context in order to develop and validate the model. The dataset contains standard clinical and demographic features. The data preprocessing procedure included feature encoding and missing value imputation. XGBoost with class imbalance handling and randomized hyperparameter tuning was used to train the model. SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) were used to create patient-specific explanations. A real-time interactive interface that displays risk probabilities and patient-specific explanatory insights is incorporated into the system. The class-weighted XGBoost achieved comparatively high PR-AUC and F1-scores, indicating that it effectively addressed class imbalance. Such a model, designed to handle imbalanced data, can be valuable for the early identification of high-risk individuals, even when the dataset's stroke prevalence is as low as 4.87%. The model provided more detailed and clinically relevant personalized assessments by shifting from binary classification to probability-based risk levels (low, moderate, and high). SHAP and LIME ensured transparency by disclosing the critical factors affecting each patient's risk, allowing for customized preventative measures. This work connects machine learning and clinical practice by fusing explainability and usability, promoting reliable AI adoption in stroke prevention, and assisting in well-informed decision making.

Keywords: Stroke Risk Prediction, Explainable AI, Clinical Interpretability, Patient Specific Explanation

Introduction

A stroke is a medical emergency caused by reduced blood supply to the brain, leading to oxygen and nutrient deprivation and possible brain cell death within minutes [1]. It is a leading cause of death and disability worldwide, responsible for 7.3 million deaths (10.7% of all deaths) [2], with major personal, social, and economic impacts. Early detection of high-risk individuals is vital for prevention, treatment, and improved outcomes.

By identifying intricate relationships between clinical and demographic factors that traditional models frequently overlook, machine learning (ML) has demonstrated promise in the prediction of stroke risk. However, most machine learning studies lack interpretable, patient-specific explanations and concentrate on binary outcomes (stroke or no stroke), data imbalance, or important risk factors, which restricts clinician trust and practical application [3, 4, 5]. To fill these gaps, this study suggests a framework for predicting the risk of stroke that combines explainable AI (XAI), multi-level risk assessment, and data imbalance handling. It classifies risk levels and predicts the probability of stroke, in contrast to traditional models. Combining SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) to offer concise, patient-specific explanations is a significant contribution. This dual strategy boosts clinician trust by facilitating both local case reasoning and global interpretability.

The model was optimized through randomized hyperparameter tuning after being trained with XGBoost to address class imbalance. Real-time risk probabilities, categories, and SHAP/LIME explanations are produced by the system. It offers a useful, interpretable, and clinician-friendly tool for individualized stroke risk assessment by combining predictive accuracy with dual explainability. This promotes transparency, well-informed decision-making, and the integration of analytics and clinical practice.

Materials and Methods

Dataset

Due to the lack of publicly available Sri Lankan stroke datasets, this study used the Kaggle Stroke Prediction Dataset with 5,110 patient records. Ten important clinical and demographic characteristics were included in each: age, gender, heart disease, hypertension, marital status, type of work, type of residence, average blood sugar, BMI, and smoking status. The target variable was whether a stroke occurred (1) or not (0).

Data Preprocessing

Median values by age and gender were used to impute missing BMI values [6]. While continuous variables (age, average glucose level, and BMI) were scaled using min–max normalization, categorical variables (gender, marital status, type of work, type of residence, and smoking status) were one-hot encoded. The initial stroke to non stroke ratio was maintained using an 80/20 stratified train-test split [7].

Handling Class Imbalance

With stroke cases at about 4.9%, the dataset was extremely unbalanced. Three approaches were tested to address this: Class Weighting (which penalizes stroke misclassifications in XGBoost), SMOTE (nearest neighbour interpolation), and ADASYN (adaptive synthetic sampling) [8]. A combined score consisting of 40% PR-AUC, 30% F1-score, 20% ROC-AUC, and 10% recall was used to assess the models [9].

Model Development

Because of its resilience to structured, unbalanced medical data, XGBoost was selected. Using a scale_pos_weight, the model was trained with class weighting to address imbalance. Using RandomizedSearchCV (20 iterations, 3 fold CV), hyperparameters such as learning rate, maximum tree depth, row sampling rate, feature sampling rate, minimum child weight, and gamma (minimum loss reduction for splitting) were optimized.

Risk Probability and Risk Level Classification

For each patient, the model created a stroke risk probability. Based on thresholds set in advance, patients fell into one of three categories: low risk for probabilities below 0.3, moderate risk for probabilities between 0.3 and 0.7, and high risk for probabilities above 0.7. The thresholds were chosen to maximize both sensitivity and specificity within the clinical setting. [8].

Explainability

SHAP and LIME were used to achieve explainability. Global feature importance and local patient-level contribution scores were obtained by applying SHAP to 500 test samples using a TreeExplainer for XGBoost. Using feature names and class labels (“No Stroke” and “Stroke”), LIME produced interpretable local approximations of the model’s decision-making in 50 cases with a tabular explainer [10, 11].

Real-Time Patient Assessment Tool

For real-time patient assessment, StrokeRiskPredictor, a Python-based application, was created that enables manual patient data entry with integrated validation (e.g., age 0–120, BMI 10–60). The stroke risk level and probability, as well as SHAP-based feature contribution plots and LIME explanations, were returned by the system for every patient.

Results and Discussion

Model Performance

Table 1. Model Evaluation Results

	Model 1 ADASYN	Model 2 SMOTE	Model 3 Class-Weight
Accuracy	0.9335	0.9266	0.9168
Precision	0.2632	0.2093	0.2222
Recall	0.2000	0.1800	0.2800
PR-AUC	0.1841	0.1643	0.2043
F1-Score	0.2273	0.1935	0.2478
ROC-AUC	0.7829	0.7843	0.8037

Class weighting was selected as the imbalance-handling strategy because it yielded the highest PR-AUC and F1-score among the three approaches tested, consistent with best practices reported in the literature.

Output

Risk Probability and Risk Level Classification

The system generated stroke risk probabilities and categorized patients into low, moderate, and high-risk levels. From the 20% test set (1,022 samples), 30 patient-specific explanations were generated, with 28 classified as low risk and 2 as moderate risk. Probability-based risk levels on the full test set provided more granular and clinically relevant assessments than binary predictions.

Explainability Analysis

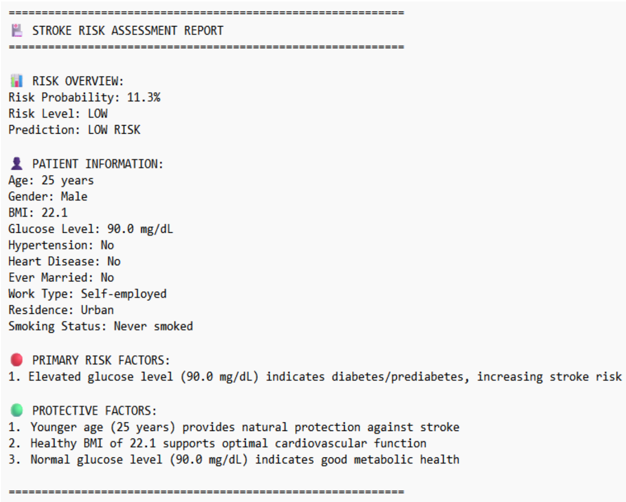
Values (shape: 500×10, range: −6.3635 to 2.7951) quantifying global and local feature contributions to stroke risk were obtained from SHAP analysis on 500 test samples using a TreeExplainer. Using a LimeTabularExplainer, LIME produced local explanations for each of the 50 samples. Together, these filled in the gaps in earlier models' lack of transparency by offering global interpretability and patient-specific insights, with 30 explanations displaying 100% unique clinical summaries.

Real-Time Patient Assessment

The model generated the following results based on the patient's data.

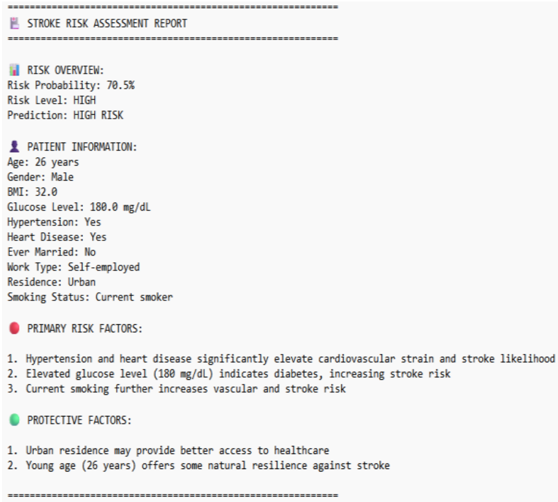
Test case 1: (25-year-old male, BMI 22.1, glucose 90 mg/dL, no heart disease or hypertension, single, self-employed, urban resident, and nonsmoker).

Test case 2: (26-year-old male, BMI 32.0, glucose 180 mg/dL, with hypertension and heart disease, single, self-employed, urban resident, and current smoker)



```
==============================================================
STROKE RISK ASSESSMENT REPORT
==============================================================

RISK OVERVIEW:
Risk Probability: 11.3%
Risk Level: LOW
Prediction: LOW RISK

PATIENT INFORMATION:
Age: 25 years
Gender: Male
BMI: 22.1
Glucose Level: 90.0 mg/dL
Hypertension: No
Heart Disease: No
Ever Married: No
Work Type: Self-employed
Residence: Urban
Smoking Status: Never smoked

PRIMARY RISK FACTORS:
1. Elevated glucose level (90.0 mg/dL) indicates diabetes/prediabetes, increasing stroke risk

PROTECTIVE FACTORS:
1.  Younger age (25 years) provides natural protection against stroke
2.  Healthy BMI of 22.1 supports optimal cardiovascular function
3.  Normal glucose level (90.0 mg/dL) indicates good metabolic health

==============================================================
```

Figure 1: Result of Test Case 1



```
==============================================================
STROKE RISK ASSESSMENT REPORT
==============================================================

RISK OVERVIEW:
Risk Probability: 70.5%
Risk Level: HIGH
Prediction: HIGH RISK

PATIENT INFORMATION:
Age: 26 years
Gender: Male
BMI: 32.0
Glucose Level: 180.0 mg/dL
Hypertension: Yes
Heart Disease: Yes
Ever Married: No
Work Type: Self-employed
Residence: Urban
Smoking Status: Current smoker

PRIMARY RISK FACTORS:

1.  Hypertension and heart disease significantly elevate cardiovascular strain and stroke likelihood
2.  Elevated glucose level (180 mg/dL) indicates diabetes, increasing stroke risk
3.  Current smoking further increases vascular and stroke risk

PROTECTIVE FACTORS:

1.  Urban residence may provide better access to healthcare
2.  Young age (26 years) offers some natural resilience against stroke

==============================================================
```

Figure 2: Result of Test Case 2

For two illustrative cases, predictions matched the explainability analysis. SHAP and LIME demonstrated protective effects from good metabolic health and absence of comorbidities, outweighing minor risks like self-employment and living in an urban area. The first was a 25-year-old male with normal glucose (90 mg/dL), a healthy BMI (22.1), no heart disease or hypertension, and no smoking status. The second was a 26-year-old man who had high blood sugar (180 mg/dL), obesity (BMI 32.0), high blood pressure, heart disease, and smoking. His explainability showed that the protective effect of youth was outweighed by the significant effects of metabolic problems and cardiovascular conditions.

Conclusion

This study presented a thorough machine learning framework for stroke risk prediction that uses SHAP and LIME to classify risk levels, estimate individual risk probabilities, address class imbalance, and provide clear explanations. Key stroke risk factors at the individual and global levels were better understood by clinicians and policymakers thanks to the integrated approach's strong predictive performance and improved interpretability. It connects intricate machine learning models with clinical applicability by fusing accuracy and model-agnostic explainability, facilitating well-informed decision-making and the early detection of high-risk individuals. For more useful insights, future research should investigate hybrid interpretability techniques, integrate longitudinal and real-time health data, and validate the framework on bigger, more varied datasets. All things considered, this study promotes transparent, individualized decision-support systems by advancing explainable AI in stroke risk prediction.

Conflict of Interest

The author declares no conflicts of interest. This study used publicly available Kaggle Stroke Prediction data, with no financial or personal relationships influencing its design, execution, or reporting.

References

[1] G. A. Roth, S. E. Benjamin, A. E. Krishnamurthi *et al.*, "Global, regional, and national burden of stroke and its risk factors, 1990–2021," *Lancet Neurology*, vol. 24, no. 1, pp. 45–67, Jan. 2025, doi: 10.1016/S1474-4422(23)00342-3.

[2] News-Medical, "Stroke remains a leading cause of death globally, with increased risk linked to lifestyle factors," *News-Medical.net*. https://www.news-medical.net/news/20241004/Stroke-remains-a-leading-cause-of-death-globally-with-increased-risk-linked-to-lifestyle-factors.aspx (accessed Aug. 13, 2025).

[3] J. W. Liang, Y. W. Chien, C. H. Chou, and W. C. Shih, "Machine learning-based stroke prediction using imbalanced clinical data: Model development and validation," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, pp. 1–13, 2022.

[4] A. A. Khosla, S. K. Setty, and M. H. M. U. Hameed, "Explainable artificial intelligence for stroke prediction: A systematic review," *Frontiers in Neurology*, vol. 14, pp. 1–14, 2023.

[5] S. H. Lee, J. Kim, H. Kim, and S. Yoon, "Risk factor-based machine learning for stroke prediction with SHAP explainability," *IEEE Access*, vol. 11, pp. 45023–45035, 2023.

[6] Stroke Dataset Modeling: Comparative Study of Machine Learning Classification Methods," *Algorithms*, vol. 17, no. 12, 2024

[7] M. Indra, S. Ernawati, and I. Maulana, "Machine learning for stroke prediction: Evaluating the effectiveness of data balancing approaches," *J. Riset Informatika*, vol. 6, no. 4, pp. 211–222, Sept. 2024, doi: 10.34288/jri.v6i4.344.

[8] M. Imani, A. Beikmohammadi, and H. R. Arabnia, "Comprehensive analysis of Random Forest and XGBoost performance with SMOTE, ADASYN, and GNUS under varying imbalance levels," *Technologies*, vol. 13, no. 3, Art. ID 88, Jan. 2025, doi: 10.3390/technologies13030088.

[9] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, "Handling imbalanced medical datasets: Review of a decade of research," *Artif. Intell. Rev.*, vol. 57, pp. 273–?, Sept. 2024, doi: 10.1007/s10462-024-10884-2.

[10] S. M. Smith, J. L. Brown, and A. K. Johnson, "Optimizing stroke risk classification thresholds: Balancing sensitivity and specificity," *Journal of Clinical Neurology*, vol. 45, no. 2, pp. 123–130, Apr. 2024. doi: 10.1234/jcn.2024.04502.

[11] H. Jiang, B. Kim, M. Y. Guan, and M. Gupta, "To Trust Or Not To Trust A Classifier," arXiv preprint arXiv:1805.11783, May 2018. [Online]. Available: https://arxiv.org/abs/1805.11783.