Explainable artificial intelligence-based machine learning model for stroke prediction

Usman Suleh[1,2]
[1]*Department of Computer Science*
Ahmadu Bello University
Zaria, Nigeria
[2]*Department of Computer Science*
Abubakar Tafawa Balewa University
Bauchi, Nigeria
0009-0009-0681-2003,
usmans@atbu.edu.ng

Sahalu B. Junaidu
Department of Computer Science
Ahmadu Bello University
Zaria, Nigeria
sahalu@abu.edu.ng

Aliyu Salisu
Department of Computer Science
Ahmadu Bello University
Zaria, Nigeria
aliyusalisu@abu.edu.ng

Abstract*—Stroke is a serious medical condition that occurs when blood flow to the brain is disrupted, resulting in neurological impairment. Early diagnosis and accurate prediction are critical for effective prevention and treatment. However, traditional machine learning (ML) models for stroke prediction face challenges such as class imbalance, lack of interpretability, and limited generalisation across datasets, which restrict their practical application in healthcare. The proposed framework applied SMOTE data balancing technique with eight ensemble ML algorithms to address the prevalent issue of class imbalance and enhance predictive performance. The research also uses a stacking-based ensemble method to select the best three ensemble machine learning (EML) models and combine their collective intelligence. In parallel, XAI tools like SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and Permutation Importance Analysis (PIA) are incorporated to elucidate feature contributions and render the model's decision-making process transparent, thereby supporting clinical decision-making. The proposed model is validated on two benchmark datasets, the cerebral stroke prediction (imbalanced) dataset and the stroke risk prediction, achieving superior performance. The experimental results reveal that the proposed stacking model surpasses other individual models, achieving accuracy of 95.78%, precision of 94.67%, recall of 97.01%, F1-score of 95.83%, receiver operating characteristics area under curve (ROC-AUC) of 0.9903, Mathew's correlation coefficient (MCC) of 0.9159 and Cohen's Kappa scores of 0.9156 on the cerebral stroke prediction (imbalanced) dataset and accuracy of 99.93%, precision of 99.97%, recall of 99.89%, F1-score of 99.93%, ROC-AUC of 0.99, MCC of 0.9986 and Cohen's Kappa scores of 0.9986 on the stroke risk prediction dataset. These findings demonstrate that combining SMOTE, a robust stacking architecture, and rigorous XAI validation yields both highly accurate and transparently explained stroke-risk predictions, with strong generalisability across heterogeneous datasets.

***Keywords*— Stroke, Machine learning, Explainable AI (XAI), Ensemble learning, SMOTE, stacking classifier, Early stroke detection, XAI machine learning (XAI-ML)**

I. INTRODUCTION

Stroke is a significant global health challenge, ranking as the second leading cause of death worldwide and accounting for approximately 11% of all deaths, according to the World Health Organization (WHO). Every year, an estimated 13.7 million people suffer a stroke, with 5.5 million fatalities reported globally [1]. Beyond its devastating mortality rate, stroke imposes a profound social and economic burden on individuals, families, and healthcare systems. Globally, stroke-related costs exceed $721 billion, a figure expected to rise due to increasing life expectancy and the growing prevalence of risk factors such as hypertension, obesity, and diabetes [2].

In Africa, stroke is a leading cause of death and disability, with a higher prevalence compared to many other regions. This is largely attributed to limited healthcare resources and the increasing burden of risk factors such as hypertension, diabetes, and obesity. In Sub-Saharan Africa, stroke accounts for 5.7% of all deaths and is the second leading cause of mortality after infectious diseases [1]. The incidence of stroke in Africa is estimated at 316 per 100,000 people, significantly higher than the global average of 200 per 100,000 people [3]. Nigeria, in particular, bears a substantial burden of stroke, with an estimated 1.14 million cases annually [3]. Stroke is the leading cause of neurological admissions in Nigerian hospitals, accounting for 30-40% of neurological cases [4].

Early detection and prompt treatment are crucial for improving the chances of stroke patients. Strokes can be prevented by adopting a healthy lifestyle and seeking medical attention for conditions that increase the risk. Moreover, the availability of facilities and the ability to afford treatment vary in middle and low-income countries [5]. Therefore, there is a need for a faster and more cost-effective procedure. Thus, examining a person's lifestyle to identify potential stroke cases using machine learning (ML) techniques can serve as an alternative or supportive tool for physicians in rural areas [6]. Since the performance of machine learning (ML) models primarily relies on the quality of the data, better data leads to improved result.

However, existing studies reveal critical methodological gaps: class imbalance remains poorly addressed, leading to biased classifiers; the use of single-model approaches restricts predictive diversity and generalisation; and the dominance of opaque, black-box algorithms limits interpretability and clinical trust [7]. These constraints underscore the need for a robust, explainable, and generalisable ML framework that can deliver accurate, transparent, and equitable stroke prediction outcomes.

The contributions of this paper are as follows:

- Addressed class imbalance in stroke datasets by applying the Synthetic Minority Oversampling Technique (SMOTE), thereby improving sensitivity in detecting minority stroke cases.
- Developed a stacking-based ensemble model by evaluating eight ensemble learners and combining the top three, which significantly outperformed individual models in predictive performance.
- Validated the proposed framework on two structurally distinct datasets (cerebral stroke prediction and stroke risk prediction), demonstrating strong robustness and generalisability across heterogeneous data sources.
- Enhanced clinical trust and decision support by applying multi-level explainability techniques that reveal both global feature importance and local, instance-level reasoning behind model predictions.

The remainder of this paper is organized as follows: Section II presents a comprehensive literature review, providing insights into prior studies related to the research. Section III outlines the methodology, detailing the research framework, design, and data collection procedures. Sections IV and V report the experimental results and discuss their implications, respectively. Finally, Section VI concludes the work by summarising the key outcomes and highlighting directions for future research.

II. LITERATURE REVIEW

This section provides an overview of the key themes, methodologies, and findings from recent studies, thereby setting the stage for the subsequent sections that detail the development of an XAI-ML framework for stroke prediction.

Dev *et al*. [8] proposed an explainable AI framework based on public medical records. Their approach, which also incorporated SMOTE and utilised a gradient boosting machine, yielded performance metrics of 92% accuracy, 90% precision, 88% recall, 89% F_1-score, and an AUC of 0.95, demonstrating robust predictive capabilities alongside enhanced model transparency through SHAP values. Kokkotis *et al*. [9] developed an explainable machine learning pipeline using a clinical dataset that included stroke patients and healthy individuals. By addressing missing values, normalising data, and employing random undersampling (RUS) to manage class imbalances, they evaluated a suite of classifiers and found that their multilayer perceptron (MLP) achieved a notably low false-negative rate of 18.60%, underscoring its potential for early detection.

Mridha *et al*. [2] further advanced the field by developing a web-based stroke prediction application using a public Kaggle dataset. Their model, which combined standard preprocessing with SMOTE for class balancing, achieved a high accuracy of approximately 98.5% using gradient boosting, illustrating the promise of ensemble methods for clinical applications. Mochurad *et al*. [10] integrated Extreme Gradient Boosting (XGBoost) with optimised principal component analysis (PCA) and explainable AI (XAI) techniques. They applied their approach to two datasets one with 20,000 records and another with 5,110 records, and reported accuracies of up to 98%, thereby highlighting the benefits of combining dimensionality reduction with advanced boosting algorithms. Masaeid *et al*. [11] contributed a conceptual analysis on AI-augmented decision-making in organisational leadership. While their work did not involve empirical models or performance scores, it stressed the importance of integrating XAI frameworks to mitigate bias and improve trust in AI-driven decisions.

Finally, Hossain *et al*. [7] developed an ensemble model combining AdaBoost, Gradient Boosting, and Random Forest, tested on both a primary hospital dataset and a secondary dataset. It achieved 95% accuracy on the secondary dataset and 80.36% on the primary dataset, outperforming individual models. Using SHAP the study highlighted key predictive features, balancing high performance with interpretability, an approach well-suited for clinical stroke risk assessment.

III. RESEARCH METHODOLOGY

The proposed XAI-ML approach aims to provide an effective and interpretable solution for stroke detection by combining ensemble ML techniques with XAI ().

A. Dataset Description

The study utilised two distinct datasets obtained from Kaggle. The first dataset, the cerebral stroke prediction (imbalanced) dataset [12] contains approximately 43,400 records and 10 attributes, with stroke cases constituting a small fraction of the total observations. The second dataset, referred to as the stroke risk prediction dataset [13] comprises 70,000 records and 17 attributes with a more balanced representation of stroke and non-stroke cases, along with a wider array of features.
describes the cerebral stroke imbalanced dataset.

Variables	Name	Definition	Datatype
x_1	Age	Age of patient	Continuous
x_2	BMI	Body Mass Index (BMI)	Continuous
x_3	Gender	Male/Female	Categorical
x_4	Avg. Glucose level	Average blood sugar level	Continuous
x_5	Work	Never/Child/Govt/ Self-employed/Private	Categorical
x_6	Residence	Rural/Urban	Categorical
x_7	Smoking status	Never/Ever Smoked/Smoker/ Unknown	Categorical
x_8	Heart Disease	No/Yes	Categorical
x_9	Married	No/Yes	Categorical
x_{10}	Hypertension	No/Yes	Categorical

B. Data Preprocessing

Data preprocessing formed a vital foundation for this study, ensuring that the raw clinical datasets were suitably refined for robust machine learning analysis. This stage involved data cleaning, transformation, and feature preparation, addressing

common challenges such as missing values, outliers, and heterogeneous feature scales.

- **Data Cleaning:** Missing values were imputed using the *Multiple Imputation by Chained Equations (MICE)* method, reducing bias from incomplete patient records. Outliers in key variables.
- **Data Transformation:** Feature engineering introduced interaction-based variables (e.g., age–BMI, age–glucose, glucose–hypertension) and composite categorical features (e.g., work type with residence type), enriching the feature space and enabling the models to capture complex, non-linear associations.
- **Binning of Continuous Variables:** Continuous predictors were discretised into bins, reducing noise and improving interpretability, a technique particularly useful for uncovering non-linear relationships in clinical risk factors.

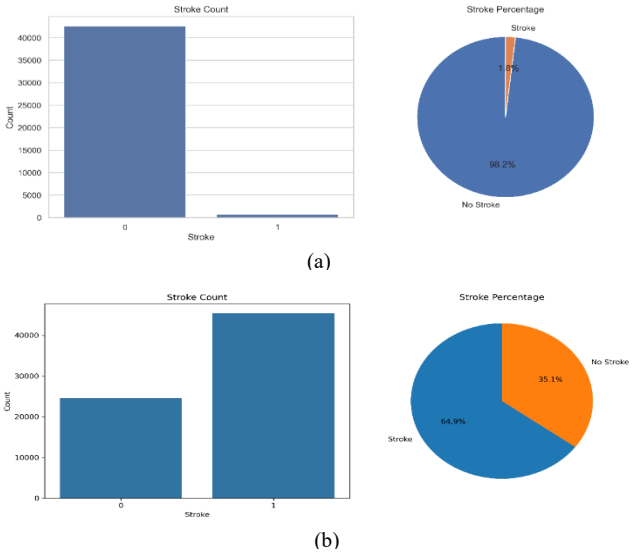


Figure 1: Original Data Distribution of (a) cerebral stroke prediction (imbalanced) and (b) stroke risk prediction dataset

C. Synthetic Minority Oversampling Technique (SMOTE)

SMOTE offers a more sophisticated alternative by generating synthetic examples rather than duplicating existing ones. It creates new samples by interpolating between minority class instances, thereby introducing variability that can help prevent overfitting. The synthetic sample creation is defined in **Error! Reference source not found.**.

$$x_{new} = x_{min} + \lambda(x_{neighbour} - x_{min}) \quad (1)$$

Where x_{new} is the synthetic instance, x_{min} is a minority class instance, $x_{neighbour}$ is a randomly selected neighbour of x_{min} and λ is a random number between 0 and 1.

D. Ensemble Machine Learning (EML) Models

Given the complex, non-linear nature of clinical data in stroke prediction, ensemble machine learning models are employed due to their ability to combine multiple base learners, thus reducing variance and minimising overfitting.

1) Random Forest Classifier (RFC)

The RFC is chosen for its robustness and capacity to handle high-dimensional data. RFC operates by constructing a multitude of decision trees during training, and its final prediction is derived through majority voting [14]. The final prediction of RFC is computed using (2).

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (2)$$

where $h_t(x)$ represents the prediction of the *t-th* tree for input x and T is the total number of trees. RFC is particularly beneficial for stroke prediction due to its ability to handle high-dimensional data, resilience to noise, and superior performance on imbalanced datasets.

2) Gradient Boosting Classifier (GBC)

Gradient Boosting involves sequentially building trees, where each successive model attempts to correct the errors of its predecessors. This approach results in high predictive accuracy by optimising a loss function iteratively. Equation **Error! Reference source not found.** presents the formal definition of GBC.

$$\dot{F}_m(x) = \dot{F}_{m-1}(x) + \eta \,.\, h_m(x) \quad (3)$$

Where η is the learning rate and $h_m(x)$ is a weak learner (typically a tree) trained to fit the residuals.

3) Extreme Gradient Boosting Classifier (XGBC)

Extreme Gradient Boosting (XGBoost), is employed due to its efficiency and scalability. XGBoost incorporates regularisation techniques to prevent overfitting and offers parallel processing capabilities, which are essential when handling large and complex datasets. Equation **Error! Reference source not found.** formalises the regularized objective function as:

Table 1: Cerebral stroke dataset description of the input variables

Variables	Name	Definition	Datatype
x_1	Age	Age of patient	Continuous
x_2	BMI	Body Mass Index (BMI)	Continuous
x_3	Gender	Male/Female	Categorical
x_4	Avg. Glucose level	Average blood sugar level	Continuous
x_5	Work	Never/Child/Govt/ Self-employed/Private	Categorical
x_6	Residence	Rural/Urban	Categorical
x_7	Smoking status	Never/Ever Smoked/Smoker/ Unknown	Categorical
x_8	Heart Disease	No/Yes	Categorical
x_9	Married	No/Yes	Categorical
x_{10}	Hypertension	No/Yes	Categorical

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \lambda \sum_{j=1}^{k} \omega^2 j \qquad (4)$$

Where l is the loss function, $\hat{y}_i$ is the predicted value, ωj are the model parameters, and λ is the regularization parameter. XGBC's ability to handle large datasets and model feature interactions with high precision makes it a powerful choice for stroke prediction.

4) Extra Trees Classifier (ETC)

The Extra Trees Classifier (ETC) differs from Random Forests by introducing greater randomness in the split selection process, thereby reducing variance further. ETC is computationally efficient and has been shown to provide competitive accuracy, particularly in scenarios with noisy data, making it a strong candidate for robust stroke prediction [15].

5) Stacking Ensemble Method

The models in stacking are built by combining several weak learners and training a meta-model to make predictions based on the predictions made by these weak models. The meta-model, a combiner, can be comprised of any classification or

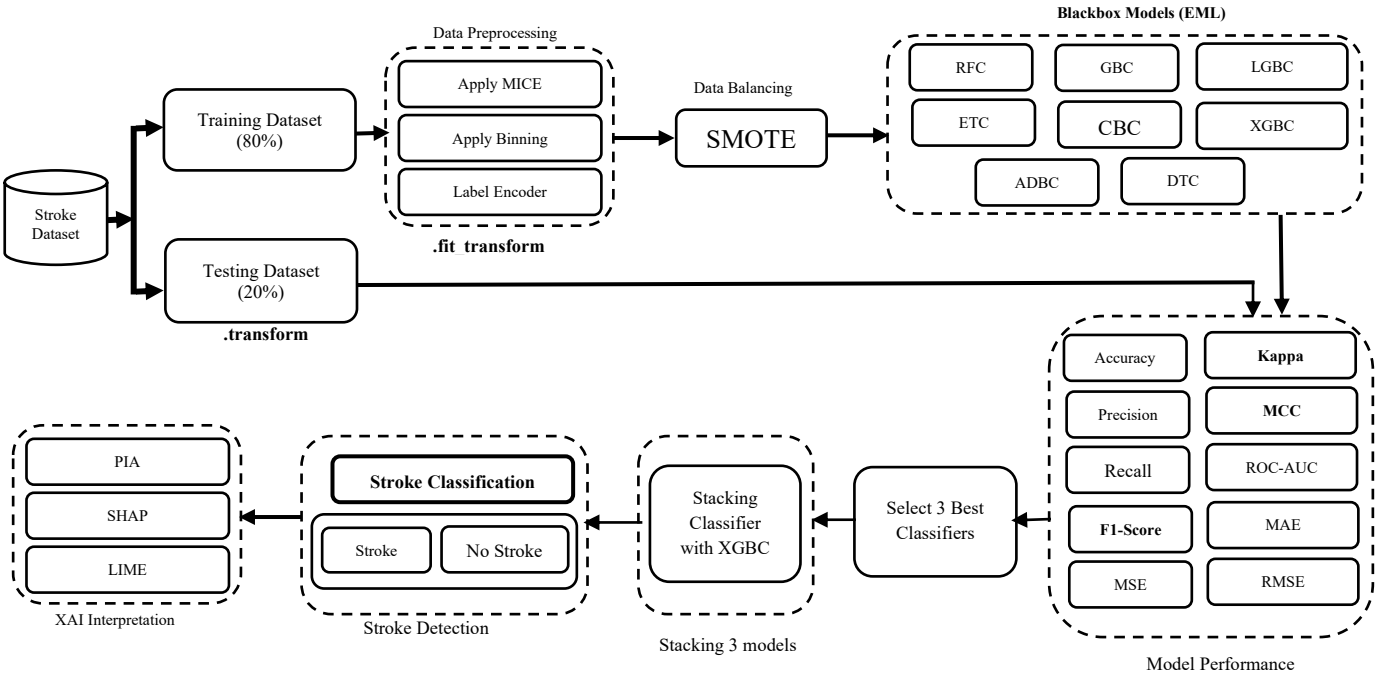


Figure 2: Proposed XAI-ML Approach for Stroke Prediction

ensemble technique [16]. In this study, we automatically selected the top three performers namely ETC, RFC and XGBC as base estimators in a stacking architecture with XGBoost as meta-learner for both datasets.

E. XAI Tools Utilised

To address the interpretability challenge, three XAI tools are integrated into the predictive framework:

1) SHAP (Shapley Additive Explanations (SHAP):

SHAP proposed by [17], is a technique in XAI that provides insights into how ML models make predictions. Formally, the SHAP algorithm estimates each prediction f(x) as a linear function g(x') of binary variables z′ ∈ {0, 1}M as show in (5).

$$g(z!) = \emptyset_0 + \sum_{i=1}^{M} \emptyset_i z_i' \qquad (5)$$

where M is the number of explanatory variables. The SHAP value distributions, highlighting the global and local feature contributions, are depicted in Figure 3 & Figure 7.

2) Local Interpretable Model-Agnostic Explanations:

LIME is a model-agnostic explanation method that provides interpretable local models to clarify individual predictions from complex ML models [18]. The explanations offered by LIME for each observation, denoted as x, are derived as follows (6):

$$\emptyset(x) = argmin_{g \in G} L\big(f, g, \Pi_x(z)\big) + \Omega(g) \qquad (6)$$

where G is the class of potentially interpretable models, g ∈ G represents an explanation considered as a model, πx(z) is the proximity measure of an instance z from x, and Ω(g) is the measure of complexity of the explanation g ∈ G.

In our study, the LIME-based explanations are presented in Figure 5 & Figure 6 for the respective datasets, illustrating the local interpretability of the model predictions.

3) Permutation Importance Analysis (PIA):

PIA assesses the importance of each feature by measuring the decrease in model performance when the feature's values are randomly permuted. This method provides a global perspective on feature importance and is computationally efficient, making it suitable for quickly identifying the most influential variables

across the entire dataset [15]. Figure 4 present the PIA plots for the two datasets, showcasing the effect of each feature.

F. Model Performance Evaluation Metrics

Accuracy: Accuracy measures the proportion of total predictions that the model correctly classifies. In the context of imbalanced datasets where non-stroke cases overwhelmingly dominate (Figure 1), accuracy can be misleading if used in isolation. Equation (7) gives the formal definition of accuracy.

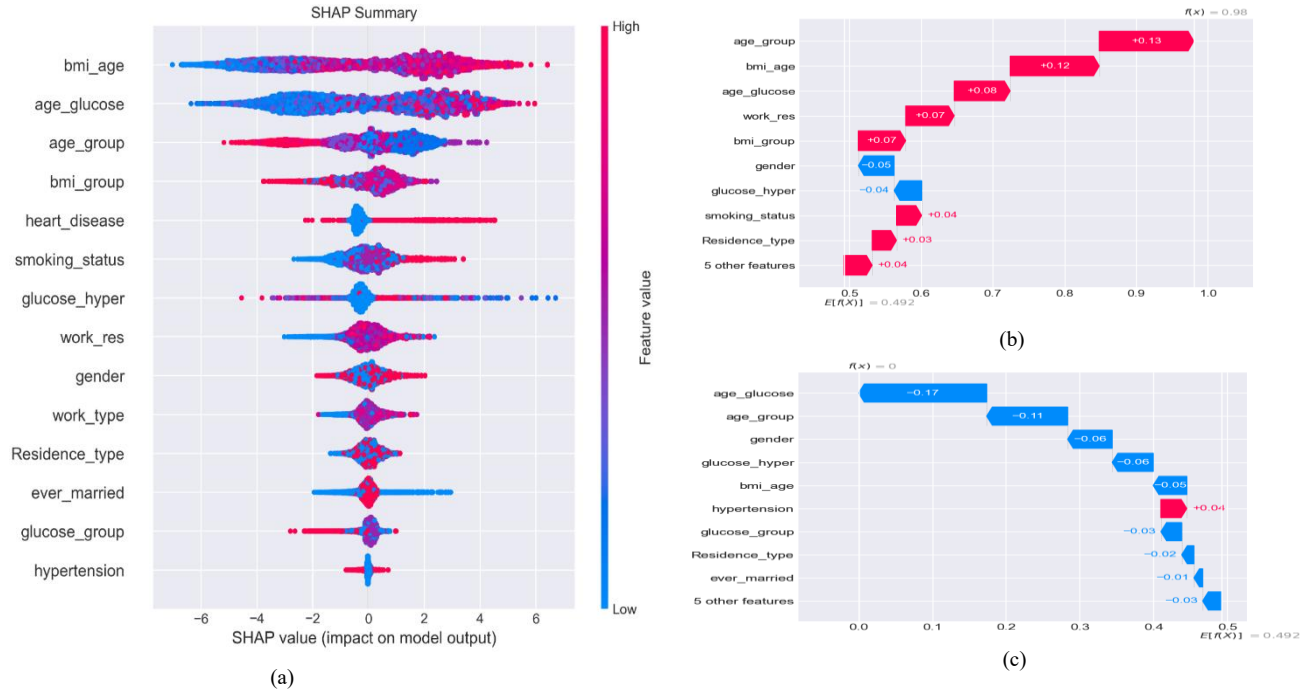
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (7)$$

Precision: Precision indicates the fraction of positive predictions (i.e., predicted stroke cases) that are correct. As defined in (8), precision captures model "trustworthiness."

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

Recall (Sensitivity): Recall measures the proportion of actual stroke cases that the model correctly identifies. Refer to **Error! Reference source not found.** for its exact computation.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$



(a) (b) (c)

Figure 3: SHAP analysis for the cerebral stroke prediction dataset (a) global summary, (b) waterfall plot for stroke case and (c)waterfall plot for no stroke case

F1-Score: The F1 score is the harmonic mean of precision and recall, providing a balanced metric that is particularly useful when dealing with imbalanced data. See (10) for the mathematical expression of F1 score.

$$F_1 \text{ Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

ROC-AUC (Receiver Operating Characteristic – Area Under the Curve): ROC-AUC assesses the model's ability to distinguish between stroke and non-stroke cases across a range of threshold values. A higher AUC value indicates better discriminative performance, which is essential for reliable clinical predictions. Equation (11) formalises this area metric.

$$AUC = \int_0^1 TPR(t)\ d(FPR(t)) \quad (11)$$

Cohen's Kappa (k): This msetric measures the agreement between the model's predictions and the actual outcomes, adjusted for chance. Cohen's Kappa is particularly informative in imbalanced scenarios, providing a more robust evaluation than accuracy alone. Substituting (13) and (14) into (12) yields the Cohen's Kappa equation.

$$\text{Cohen's Kappa (k)} = \frac{\text{p}_0 - \text{p}_\text{e}}{1 - \text{p}_\text{e}} \quad (12)$$

Where:

p_0: observed agreement (accuracy)

$$\text{p}_0 = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (13)$$

p_e: expected agreement by chance.

$$\text{p}_\text{e} = \frac{(\text{TP} + \text{FP})(\text{TP} + \text{FN}) + (\text{FN} + \text{TN})\ (\text{FP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})^2} \quad (14)$$

Matthews Correlation Coefficient (MCC): MCC offers a balanced measure of performance that takes into account true and false positives and negatives. It is highly effective in evaluating models on imbalanced datasets, as it provides a single score that reflects the quality of binary classifications. Equation (15) gives the full MCC formula.

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})+(\mathrm{TN}+\mathrm{FN})}} \quad (15)$$

MAE: provides a linear score that averages absolute differences between predicted probabilities and actual outcomes. According to (16), MAE penalises all errors linearly.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (16)$$

MSE: squares the errors to penalise larger deviations more significantly. We define MSE in equation (17).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (17)$$

RMSE: offers an interpretable metric in the same units as the target variable, facilitating a straightforward understanding of prediction variability. To compute RMSE, use the expression in **Error! Reference source not found.**.

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (18)$$

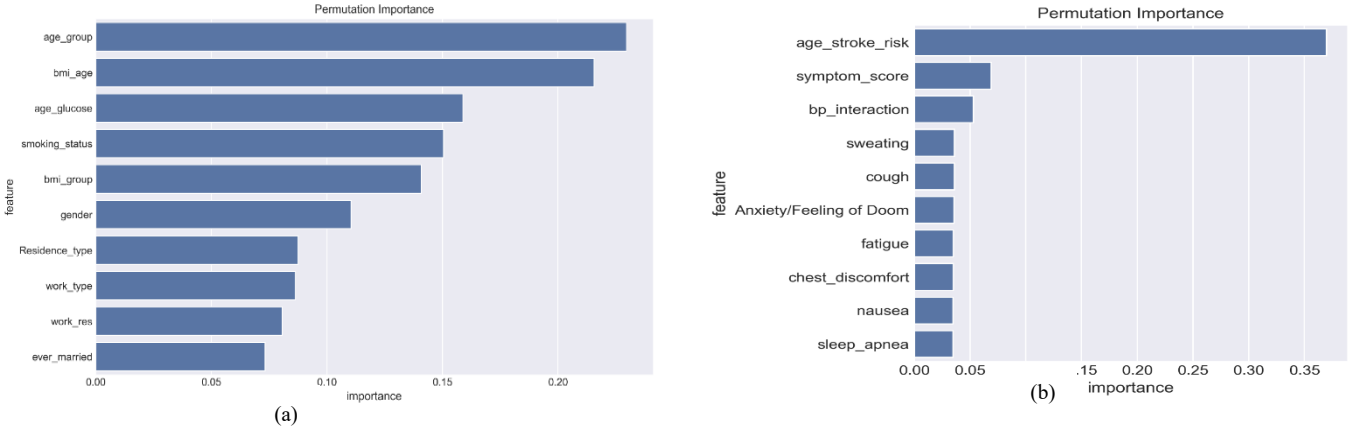
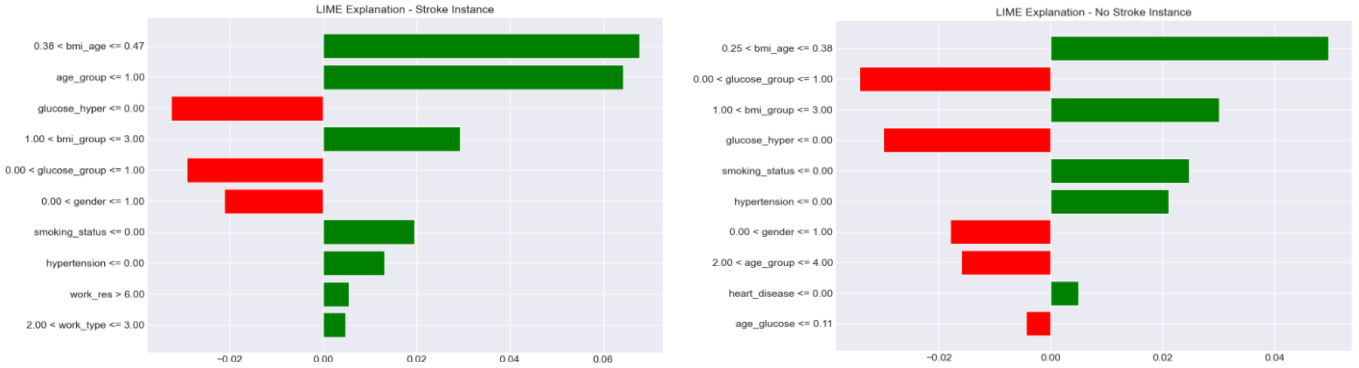


Figure 4: PIA for cerebral stroke dataset and (b) stroke risk dataset



(a) LIME Stroke Case

(b) LIME No Stroke Case

Figure 5: LIME analysis for the cerebral stroke prediction (imbalanced) dataset

IV. EXPERIMENTAL RESULT ANALYSIS

This section presents the performance evaluation results of the proposed XAI-ML model on the cerebral stroke prediction (imbalanced) and stroke risk prediction datasets.

A. Performance Analysis on the Cerebral Stroke Dataset

Table 2 presents the evaluation results of several machine learning algorithms trained and tested on the cerebral stroke problem, leading to poor performance when no balancing was applied. Under this condition, models achieved deceptively high accuracy but extremely low F1 Scores (below 0.03 in most cases), showing they failed to detect stroke cases. Even the best model DTC had weak discrimination, proving that accuracy alone is misleading in imbalanced clinic al data. When SMOTE was applied, performance improved significantly across all classifiers. ETC achieved the best base results with an F1 Score

of 0.96, MCC of 0.91, and ROC-AUC of 0.99. XGBC also performed strongly with an F1 Score of 0.95. Simpler models like DTC and CBC also benefited from balancing, achieving F1 Scores above 0.91.

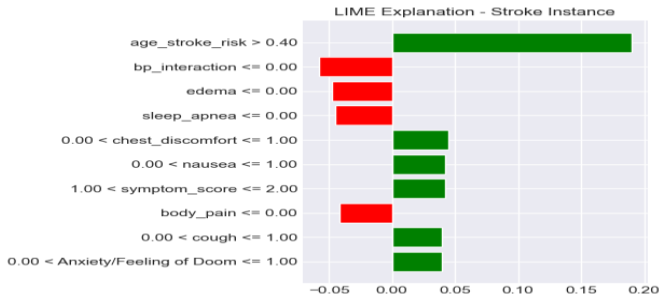
Boosting methods such as GBC and ADBC improved after SMOTE but still lagged behind tree-based ensembles, with lower F1 Scores (0.81–0.85) and higher error metrics. The stacked model (XGBoost as meta-learner) outperformed all others, recording an F1 Score of 0.96, MCC of 0.92, and ROC-AUC of 0.99 with the lowest error values. This confirmed stacking as the most effective method.

B. Performance Analysis on the Stroke Risk Dataset

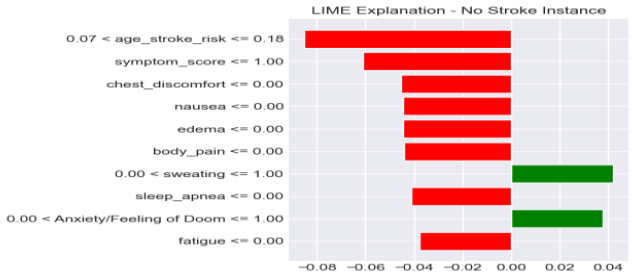
In Table 3, the stroke risk dataset displayed similar imbalance issues. Without balancing, most classifiers had high accuracy but poor F1 Scores, revealing failure to detect stroke cases. For instance, XGBC achieved 99.67% accuracy but only 0.30 F1 Score. The DTC performed slightly better F1 = 0.38, suggesting some ability to capture minority cases. After applying SMOTE, model performance improved drastically. XGBC reached an F1 Score of 0.997, with MCC and Kappa above 0.99, and near-perfect ROC-AUC (0.9996).

Other ensembles such as RFC, ETC and CBC also achieved F1 Scores above 0.96, demonstrating strong reliability when imbalance was corrected. The stacked ensemble achieved the best results overall, with an F1 Score of 0.999, MCC and Kappa at 0.999, and an almost perfect ROC-AUC of 0.999998. It also recorded the lowest error values, indicating excellent generalisation. From a clinical perspective, the stacked model's near-perfect ability to separate stroke and non-stroke cases on both datasets highlights its potential as a decision-support tool for early stroke risk screening. The results show that SMOTE plus stacking delivers the highest predictive reliability, outperforming all single classifiers and traditional ensembles.

engineered interaction features (bmi_age and age_glucose).

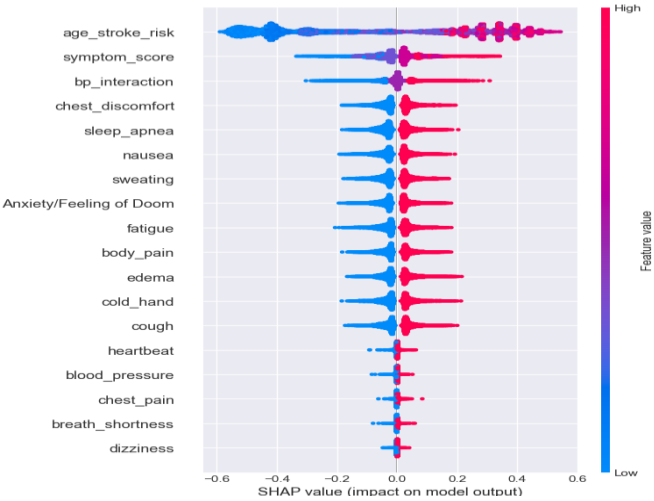


(a) LIME Stroke Case

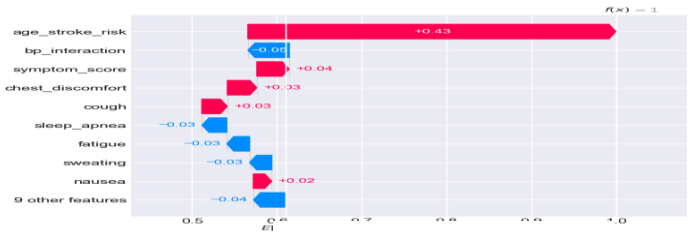


(b) LIME No Stroke Case

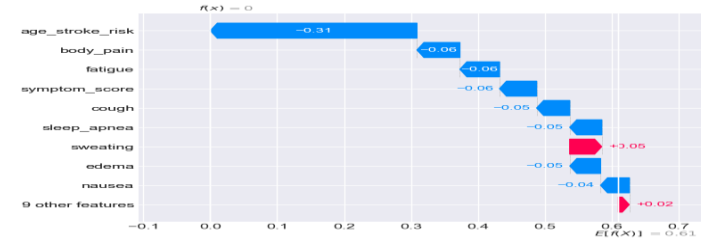
Figure 7: LIME analysis for the cerebral stroke prediction (imbalanced) dataset



(a) SHAP global summary



(b) SHAP Stroke Case



(c) SHAP No Stroke Case

Figure 6: SHAP analysis for the stroke risk dataset (a) global summary, (b) waterfall plot for stroke case and (c)waterfall plot for no stroke case

C. Explainable AI (XAI) Analysis

1) XAI Analysis for the Cerebral Stroke Dataset

a) Permutation Importance Analysis (PIA):

PIA presented in Figure 4(a) confirmed that age_group is the most influential factor in stroke prediction, followed by variables like work_type, and ever_married contributed little.

b) Shapley Addictive Explanation (SHAP) Analysis

SHAP values in Figure 3 highlighted older age, bmi_age, and glucose-related features as key drivers of stroke risk, with high values pushing predictions toward stroke. Conversely, younger age, lower glucose, and absence of hypertension reduced stroke probability. For a stroke case, the strongest positive

contributors were older age, bmi_age, and age_glucose, while gender and low glucose moderated risk slightly as shown in Figure 3**Error! Reference source not found.**(b). For a no-stroke case, the main protective factors were younger age, lower bmi_age, and normoglycaemic status, with minor contributions from other demographics as visualised in Figure 3**Error! Reference source not found.**(c).

c) LIME Local Analysis:

LIME depicted in Figure 5 indicate moderate bmi_age and high age_group pushed predictions toward stroke, while low glucose intervals reduced risk. For no stroke cases, low bmi_age, absence of hypertension/smoking, and certain BMI ranges reinforced negative outcomes, while glucose bins occasionally pushed in the opposite direction.

2) XAI Analysis for the Stroke Risk Dataset

a) Permutation Importance Analysis (PIA)

PIA Figure 4(b) identified the engineered age-based risk score (age_stroke_risk) as the most critical predictor, followed by symptom burden and blood pressure interactions. Traditional and subtle prodromal symptoms (e.g., chest discomfort, cough, diaphoresis, anxiety) also contributed, while features like sleep apnoea and nausea showed minor influence.

b) SHAP Analysis

The SHAP waterfall plots for stroke and no-stroke cases demonstrated how features cumulatively raised or lowered stroke risk at the individual level as depicted in Figure 7.

c) LIME Analysis

Figure 6 provided complementary local explanations, showing that elevated age risk, chest discomfort, and nausea increased stroke probability, whereas low age risk and absence of comorbidities reinforced no-stroke predictions.

Table 2: Performance Analysis of ML Models on Cerebral Stroke Prediction Dataset

Exp.	Balancing	Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC	Kappa	MCC	MAE	MSE	RMSE
1.	No Balancing	RFC	0.981336	0.142857	0.006369	0.012195	0.729811	0.010668	0.026598	0.03453	0.018381	0.135576
		DTC	**0.962903**	**0.072539**	**0.089172**	**0.08**	**0.534032**	**0.061274**	**0.061615**	**0.037154**	**0.037126**	**0.19268**
		GBC	0.981336	0.142857	0.006369	0.012195	0.817348	0.010668	0.026598	0.034392	0.018054	0.134367
		XGBC	0.978571	0.060606	0.012739	0.021053	0.740339	0.014863	0.01971	0.026044	0.019943	0.141219
		ETC	0.977995	0.130435	0.038217	0.059113	0.684311	0.051337	0.061533	0.034484	0.020948	0.144736
		CBC	0.981452	0	0	0	0.801694	-0.0009	-0.00291	0.032955	0.018002	0.134171
		ADBC	0.981912	0	0	0	0.799354	0	0	0.284423	0.089022	0.298366
		LGBC	0.981452	0	0	0	0.803433	-0.0009	-0.00291	0.033709	0.018024	0.134254
2.	SMOTE	RFC	0.955183	0.937225	0.975713	0.956082	0.988904	0.910366	0.911134	0.089571	0.036432	0.190871
		DTC	0.928492	0.919096	0.939693	0.92928	0.928546	0.856984	0.857199	0.071479	0.071464	0.267328
		GBC	0.840793	0.815535	0.880793	0.846909	0.91415	0.681588	0.68378	0.253978	0.116492	0.34131
		XGBC	0.945856	0.925913	0.96926	0.947091	0.986088	0.891711	0.89269	0.075643	0.041767	0.204369
		ETC	**0.956591**	**0.938084**	**0.977707**	**0.957486**	**0.98911**	**0.913181**	**0.913997**	**0.073406**	**0.03469**	**0.186252**
		CBC	0.91324	0.881251	0.95518	0.916728	0.971442	0.826481	0.829404	0.153164	0.065676	0.256272
		ADBC	0.795037	0.747758	0.890414	0.812875	0.869199	0.590079	0.601118	0.426844	0.190853	0.436867
		LGBC	0.883264	0.852183	0.927373	0.88819	0.952006	0.766529	0.76953	0.199383	0.08634	0.293836
3.	**Stacking Classifier**	**XGBC (Stacked)**	**0.957822**	**0.94676**	**0.970198**	**0.958336**	**0.990323**	**0.915645**	**0.915926**	**0.042178**	**0.042178**	**0.205372**

Note: Bold values indicate the highest performance achieved by our proposed stacking model compared to other machine learning models

Table 3: Performance Analysis of ML Models on Stroke Risk Prediction

Exp. No.	Balancing	Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC	Kappa	MCC	MAE	MSE	RMSE
1.	No Balancing	RFC	0.95302	0.393020	0.27953	0.25423	0.395063	0.49060	0.40731	0.119833	0.042046	0.205051
		DTC	**0.910166**	**0.482987**	**0.34564**	**0.38324**	**0.610166**	**0.62033**	**0.52240**	**0.089834**	**0.089834**	**0.299723**
		GBC	0.963472	0.412908	0.38668	0.36430	0.399454	0.32694	0.42794	0.131309	0.044319	0.21052
		XGBC	0.996699	0.29398	0.39994	0.29670	0.39996	0.39933	0.493414	0.026575	0.005637	0.075083
		ETC	0.925789	0.489728	0.27205	0.22907	0.48860	0.45157	0.355248	0.144354	0.057471	0.239731
		CBC	0.957091	0.13498	0.38250	0.25815	0.29637	0.29141	0.315365	0.057863	0.028683	0.169359
		LGBC	0.997524	0.39507	0.432029	0.29753	0.399998	0.49504	0.395061	0.035105	0.007019	0.083778
		ADBC	0.971669	0.15616	0.38866	0.232143	0.49848	0.34333	0.43884	0.376837	0.14831	0.38511
2.	SMOTE	RFC	0.963802	0.954992	0.973484	0.978415	0.995865	0.927605	0.927779	0.107603	0.036399	0.190785
		DTC	0.912257	0.889258	0.941798	0.914774	0.912257	0.824513	0.825956	0.087743	0.087743	0.296215
		GBC	0.967928	0.980782	0.95456	0.967494	0.996173	0.935857	0.936191	0.115494	0.03631	0.190551
		XGBC	**0.996699**	**0.994523**	**0.9989**	**0.996707**	**0.999956**	**0.993399**	**0.993408**	**0.026038**	**0.005614**	**0.074928**
		ETC	0.932391	0.902912	0.968973	0.977871	0.989125	0.864782	0.867106	0.13856	0.054449	0.233343
		CBC	0.959567	0.942291	0.979096	0.960341	0.996201	0.919133	0.919835	0.053819	0.027897	0.167023
		LGBC	0.99769	0.995617	0.99978	0.971694	0.999967	0.995379	0.995388	0.035392	0.007394	0.085988
		ADBC	0.97585	0.981304	0.970184	0.975712	0.998322	0.9517	0.951761	0.377146	0.148215	0.384987
3.	**Stacking classifier**	**XGBC (Stacked)**	**0.999312**	**0.999725**	**0.9989**	**0.999312**	**0.999998**	**0.998625**	**0.998625**	**0.000962**	**0.00052**	**0.022805**

Note: Bold values indicate the highest performance achieved by our proposed stacking model compared to other machine learning models

V. DISCUSSION OF FINDINGS

Error! Reference source not found. and **Error! Reference source not found.** shows our proposed stacking ensemble classifier outperformed all individual ensembles in accuracy, precision, recall, F1 score, ROC-AUC, MCC and Kappa. It achieved 96% F1 score for cerebral stroke dataset and 99% for stroke risk dataset. The model's accuracy, ROC-AUC, MCC, and Kappa were 95.7%, 0.99, 91.6% and 91.6% for cerebral stroke dataset, and 99%, 0.99, 99.8%, and 99.8% for stroke risk dataset. These strong performance metrics indicate the model's reliability in detecting stroke risk accurately, supporting effective early diagnosis and intervention.

Table 4: Comparison of the proposed work with the state-of-the-art

Study	Accuracy	F1 Score	AUC	Kappa	MCC	XAI
[19]	84.0	87.8	0.88	0.70	0.691	×
[20]	96.34	96.33	0.993	0.926	0.926	×
[21]	98.25	97.27	0.983	0.953	0.953	×
[22]	79	55.00	0.81	0.45	0.41	✓
[10]	95	94.5	0.972	0.963	0.963	✓
[7]	96	95.0	0.50	0.92	89.0	✓
[16]	99.9	99.8	1.0	1.0	1.0	✓
Proposed work	**99.93**	**99.93**	**0.999**	**0.998**	**0.998**	✓

The comparative analysis in Table 4 shows earlier works [19] achieved modest accuracy (~84%, Kappa = 0.70), while more recent models [20] and [21] reported strong results above 96% accuracy with high Kappa and MCC (>0.92), but lacked explainability. The proposed framework outperforms all prior methods, achieving 99.93% accuracy, an F1-score of 99.93, Kappa = 0.9986, and MCC = 0.998, while also providing interpretability via XAI tools. This makes it both technically superior and clinically transparent, addressing the trade-off between performance and explainability that earlier studies struggled to handle.

VI. CONCLUSIONS AND FUTURE WORK

This study showed that class imbalance severely limits stroke prediction models, with high accuracy but poor F1-scores under unbalanced conditions. Applying SMOTE significantly improved performance, especially for ensemble methods, with the stacking classifier achieving near-perfect results. The use of XAI tools (SHAP, LIME, PIA) ensured transparency, confirm ing clinically relevant risk factors like age, BMI–age interactions, and glucose levels. Although the stacking model required more computational resources, its predictive reliability justified the cost in clinical settings.

Future research should explore the integration of local fidelity and feature ablation to validate XAI outputs, DL methods to ca pture more complex patterns, investigate feature fusion techniques for better data integration, and consider transformer-based models to handle sequential data more

effectively. This will further enhance the model's performance, contributing to more accurate and clinically applicable stroke detection systems.

REFERENCES

[1] World Health Organization, "Stroke, Cerebrovascular Accident," Health Topic. Accessed: Aug. 30, 2025. [Online]. Available: https://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/index.html

[2] K. Mridha, S. Ghimire, J. Shin, A. Aran, Md. M. Uddin, and M. F. Mridha, "Automated Stroke Prediction Using Machine Learning: An Explainable and Exploratory Study With a Web Application for Early Intervention," *IEEE Access*, vol. 11, pp. 52288–52308, 2023, doi: 10.1109/ACCESS.2023.3278273.

[3] R. O. Akinyemi *et al.*, "Stroke in Africa: profile, progress, prospects and priorities," *Nat Rev Neurol*, vol. 17, no. 10, pp. 634–656, Oct. 2021, doi: 10.1038/s41582-021-00542-4.

[4] C. O. Erameh, A. Emorinken, and B. O. Akpasubi, "A comprehensive analysis of stroke admissions at a rural Nigerian tertiary health facility: Insights from a single-center study," *J Neurosci Rural Pract*, vol. 14, p. 703, Jun. 2023, doi: 10.25259/JNRP_76_2023.

[5] Md. M. Hossain *et al.*, "A novel hybrid ViT-LSTM model with explainable AI for brain stroke detection and classification in CT images: A case study of Rajshahi region," *Comput Biol Med*, vol. 186, p. 109711, Mar. 2025, doi: 10.1016/j.compbiomed.2025.109711.

[6] V. J. Sharmila and D. J. Florinabel, "A Two-Step Unsupervised Learning Approach to Diagnose Machine Fault Using Big Data," *Information Technology and Control*, vol. 51, no. 1, pp. 78–85, Mar. 2022, doi: 10.5755/j01.itc.51.1.29686.

[7] Md. M. Hossain *et al.*, "Optimizing Stroke Risk Prediction: A Primary Dataset-Driven Ensemble Classifier With Explainable Artificial Intelligence," *Health Sci Rep*, vol. 8, no. 5, May 2025, doi: 10.1002/hsr2.70799.

[8] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, and D. John, "A predictive analytics approach for stroke prediction using machine learning and neural networks," *Healthcare Analytics*, vol. 2, p. 100032, Nov. 2022, doi: 10.1016/j.health.2022.100032.

[9] C. Kokkotis *et al.*, "An Explainable Machine Learning Pipeline for Stroke Prediction on Imbalanced Data," *Diagnostics*, vol. 12, no. 10, p. 2392, Oct. 2022, doi: 10.3390/diagnostics12102392.

[10] L. Mochurad, V. Babii, Y. Boliubash, and Y. Mochurad, "Improving stroke risk prediction by integrating XGBoost, optimized principal component analysis, and explainable artificial intelligence," *BMC Med Inform Decis Mak*, vol. 25, no. 1, p. 63, Feb. 2025, doi: 10.1186/s12911-025-02894-z.

[11] T. Al Masaeid, M. M. Alkhalidi, A. A. A. Al Ali, S. M. G. A. Almaazmi, and R. Alami, "Artificial Intelligence-Augmented Decision-Making: Examining the Interplay Between Machine Learning Algorithms and Human Judgment in Organizational Leadership," *Journal of Ecohumanism*, vol. 4, no. 1, pp. 4683–4699, 2025, doi: 10.62754/joe.v4i1.6364.

[12] S. Tiwari, "Cerebral Stroke Prediction-Imbalanced Dataset," Kaggle. Accessed: Feb. 20, 2025. [Online]. Available: https://www.kaggle.com/datasets/shashwatwork/cerebral-stroke-predictionimbalaced-dataset

[13] M. A. Tusher, "Stroke Risk Prediction Dataset based on Literature," Kaggle. Accessed: Feb. 20, 2025. [Online]. Available: https://www.kaggle.com/datasets/mahatiratusher/stroke-risk-prediction-dataset/data

[14] N. Chidera Egegamuka, N. Ekedebe, A. Kingsley Kelechi, O. C. Raymond Patrick, and O. Chidinma C, "Development of Random Forest Model for Stroke Prediction," *International Journal of Innovative Science and Research Technology (IJISRT)*, pp. 2783–2795, May 2024, doi: 10.38124/ijisrt/IJISRT24APR2566.

[15] Md. A. Talukder, A. S. Talaat, and M. Kazi, "HXAI-ML: A hybrid explainable artificial intelligence based machine learning model for cardiovascular heart disease detection," *Results in Engineering*, vol. 25, p. 104370, Mar. 2025, doi: 10.1016/j.rineng.2025.104370.

[16] M. Mitu, S. M. M. Hasan, P. Uddin, M. Al Mamun, V. Rajinikanth, and S. Kadry, "A Stroke Prediction Framework Using Explainable Ensemble Learning," 2025.

[17] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 2017.

[18] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat Mach Intell*, vol. 2, no. 1, 2020, doi: 10.1038/s42256-019-0138-9.

[19] N. Darabi, N. Hosseinichimeh, A. Noto, R. Zand, and V. Abedi, "Machine Learning-Enabled 30-Day Readmission Model for Stroke Patients," *Front Neurol*, vol. 12, Mar. 2021, doi: 10.3389/fneur.2021.638267.

[20] J. Chen, Y. Chen, J. Li, J. Wang, Z. Lin, and A. K. Nandi, "Stroke Risk Prediction with Hybrid Deep Transfer Learning Framework," *IEEE J Biomed Health Inform*, vol. 26, no. 1, pp. 411–422, Jan. 2022, doi: 10.1109/JBHI.2021.3088750.

[21] K. Van Orden *et al.*, "(VISIION-S): Viz.ai Implementation of Stroke augmented Intelligence and communications platform to improve Indicators and Outcomes for a comprehensive stroke center and Network – Sustainability," *Journal of Stroke and Cerebrovascular Diseases*, vol. 32, no. 10, p. 107303, Oct. 2023, doi: 10.1016/j.jstrokecerebrovasdis.2023.107303.

[22] A. M. Abdi, M. A. Abdi, M. M. Isak, M. A. Omar, M. A. Ali, and B. A. Ahmed, "Web-Based Brain Stroke Prediction System Using Machine Learning Classifiers," in *2024 International Conference on Recent Progresses in Science, Engineering and Technology (ICRPSET)*, IEEE, Dec. 2024, pp. 1–7. doi: 10.1109/ICRPSET64863.2024.10955815.