Stroke prediction using data balancing method and extreme gradient boosting

Abd Mizwar A. Rahim[1], Anna Baita[1], Firman Asharudin[1], Wahid Miftahul Ashari[1], Walidy Rahman Hakim[2], Andriyan Dwi Putra[1], Supriatin[1], Eko Pramono[1]
[1]Informatics Study Program, Faculty of Computer Science, Amikom University of Yogyakarta, Yogyakarta, Indonesia
[2]Department of Pharmaceutical Sciences and Technology, Al-Irsyad University, Cilacap, Indonesia

Article Info

Article history:

Received Aug 13, 2024
Revised Jan 3, 2026
Accepted Jan 22, 2026

Keywords:

Data balancing
Data preprocessing
Extreme gradient boosting
Feature selection
Stroke prediction

ABSTRACT

Stroke is one of the leading causes of death worldwide, creating an urgent need for effective early detection systems, particularly because conventional methods often struggle with class imbalance and produce biased evaluations. Previous studies have primarily focused on accuracy while overlooking model consistency, data pre-processing quality, and probability-based evaluation. This study evaluates model performance under three conditions: original data using extreme gradient boosting (XGBoost) with scale_pos_weight, original data using the easy ensemble classifier, and class-balanced data generated using random oversampling (ROS), adaptive synthetic sampling (ADASYN), and synthetic minority over-sampling technique (SMOTE). Each model underwent missing value handling, normalization, feature preparation, and hyperparameter optimization using grid search. Performance was assessed using area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), confidence intervals, calibration curves, Shapley additive explanations (SHAP), decision curve analysis (DCA), and external validation. The results demonstrate that data resampling significantly improves performance, with the XGBoost-SMOTE combination achieving the best results, including an accuracy of 0.99, AUROC of 0.998, and AUPRC of 0.986, outperforming the other approaches. This method provides more consistent and balanced predictions, supporting the application of artificial intelligence for early stroke risk identification.

Corresponding Author:

Abd Mizwar A. Rahim
Department of Computer Science, Amikom University of Yogyakarta
Depok, Sleman, Yogyakarta 55281, Indonesia
Email: abdulmizwar@amikom.ac.id

1. INTRODUCTION

Every year 13.7 million people experience stroke, and more than 5.8 million of them die from this disease [1]. According to the World Health Organization (WHO), stroke is the second cause of death globally, contributing to around 11% of total deaths [2]. Disabilities that often occur after a person experiences a stroke include speech problems, physical limitations, weakness or paralysis on one side of the body, difficulty in grasping or holding objects, and decreased communication abilities [3].

Research on stroke shows that this condition requires serious attention because it can have a significant impact on a country's economic growth. If not treated quickly and appropriately, stroke can cause serious complications such as dementia [4]. Dementia is a medical term that refers to a number of symptoms associated with a significant decline in cognitive function, causing disruption in a person's daily activities [5].

Symptoms that often occur in dementia include disturbances in memory, judgment skills, problem-solving, language, and independence in daily activities [6].

There are various ways that can help medical staff quickly identify whether someone is experiencing stroke symptoms, one of which is using machine learning technology. The use of this technology has proven effective in classifying and optimizing the development of the health service system [7]–[9]. For example, treating patients infected with heart disease can be predicted from data generated by the health industry so that it can help and save someone's life in the long term, at least it can shorten the time it takes to find out if a patient is diagnosed with the disease because it is helped by the machine learning method used [10], [11].

There have been several previous studies with the same case, namely the prediction of stroke. Existing research has applied several machine learning methods for classification, including the random forest (RF) classifier method, artificial neural network (ANN), support vector machine (SVM), C4.5, and naïve Bayes (NB). The best result from previous research was 98% accuracy.

Previous research related to stroke prediction, this research used the extreme gradient boosting (XGBoost) method and also implemented data pre-processing techniques including labelEncoder, and dealing with empty values using the technique of changing empty values with average values. This research achieved an accuracy rate of 96% [12]. Other studies also predict stroke but use different methods, namely XGBoost, k-nearest neighbor (KNN), NB, RF, SVM, and logistic regression (LR). The implementation of the pre-processing technique of this research is to overcome missing values and normalize data. The results of this research reached 91% accuracy [13]. The next research has the same topic, namely stroke classification using machine learning methods, namely ANN, SVM, decision tree (DT), LR, and bagging and boosting. implementation of techniques before entering the classification process, namely cleaning data, includes dealing with missing values and deleting duplicate data. The results of this research state that the best level of accuracy is 95% [14]. Next, research on the same topic also uses several machine learning methods including NB, RF, LR, KNN, stochastic gradient descent (SGD), DT, and multilayer perceptron (MLP). This research also applies several pre-processing techniques to first overcome missing values in the processing dataset, and overcome data imbalance in the dataset using synthetic minority over-sampling technique (SMOTE). The results of this research obtained an accuracy rate of 98% [15]. The latest research is on stroke prediction machine learning algorithms, development and evaluation of prediction models. This research carries out a comparative analysis of machine learning methods using datasets with balanced and unbalanced data conditions. The results of this research have the best accuracy of 96% using the RF method using balanced data [16].

There are shortcomings in that previous research firstly has not addressed the condition of unbalanced data, this can cause the model to be biased towards the majority class, causing inaccurate evaluations such as misleading accuracy, and the potential for overfitting on majority data. As a result, the model may fail to recognize or predict occurrences of minority classes effectively, reducing the general ability of the model to adapt, and producing suboptimal solutions in the relevant application context [17]. Apart from that, data normalization techniques have not been implemented, and differences in scale between features can significantly affect the performance and stability of the model. Features with a larger range of values tend to have a more dominant influence in the learning process, while features with a smaller range of values may play less of a role or be ignored in determining model predictions [18]. The last technique that is not applied is k-fold cross-validation in evaluating machine learning models, there is a risk that the evaluation of model performance will be inconsistent and subjective. By only doing one division of training data and validation data, evaluation results can be too optimistic or pessimistic depending on how the data is randomly divided, thus not providing an accurate picture of how well the model can predict unseen data [19].

To get good accuracy, rely on one of the data pre-processing techniques, namely feature selection when using this method in the classification process. Recent research on the effect of feature selection on the accuracy of machine learning models has made a major contribution to the identification process [20]. Apart from that, the accuracy of the model achieves good results by applying the data balancing method, which has been proven in research regarding the impact of the data balancing approach with a case study [21]. Another technique that can be applied to achieve good evaluation scores is data normalization. This technique has also been proven to be able to increase accuracy in the classification process. This has been done in research investigating the impact of data normalization on classification performance [22]. The last one is the implementation of the k-fold cross-validation technique. This technique is not a function that increases accuracy directly, but rather an evaluation technique that helps in validating model performance better [23].

This study classified stroke risk through a series of stages consisting of data pre-processing (imputation of missing body mass index (BMI) values, handling outliers using the interquartile range (IQR) method, encoding categorical variables using LabelEncoder, and normalizing numerical features with min-max scaling), dividing the data into training and test data, and applying various balancing techniques to the training data using SMOTE, adaptive synthetic sampling (ADASYN), and random oversampling (ROS). The model was then developed using the XGBoost algorithm with hyperparameter tuning, including variations of the scale_pos_weight parameter, and the EasyEnsemble method as a comparison. Performance evaluation was conducted using the area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) metrics, calibration plots, Shapley additive explanations (SHAP) interpretability analysis, bootstrap confidence intervals, and external validation to determine the optimal model for predicting stroke risk.

2. METHOD

This section explains the research flow regarding disease classification, starting from data acquisition, data pre-processing, use of the XGBoost model, and evaluation of model performance as in Figure 1. Figure 1 clearly illustrates the overall process in a visual manner. Thus, readers can easily follow the methodology.

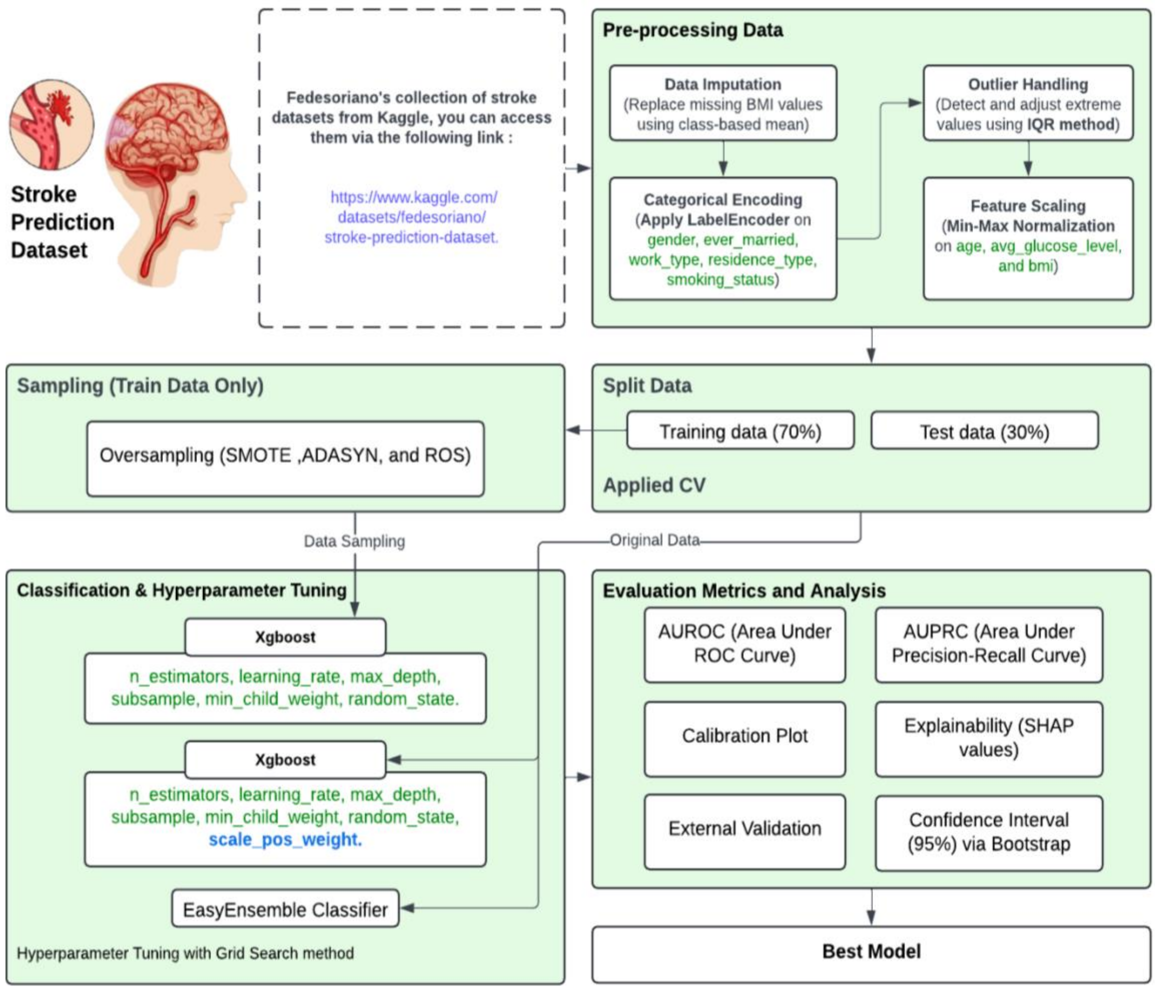

Figure 1. Research flow

2.1. Dataset

This research dataset was taken from the Kaggle dataset [24]. This dataset has 5,110 data points, consists of 12 attributes, 10 independent variables as features, and one dependent variable as a class label. The 10 independent variables in question are gender, age, hypertension, heart_disease, ever_married, work_type, residence_type, avg_glucose_level, bmi, and smoking_status. The label is the stroke attribute in this dataset. The class has two values: 0, which means there is no indication of stroke, and 1, which means there is an indication of stroke. Table 1 dataset describes the dataset feature information.

Table 1. Dataset

No	Feature	Information
1	Gender	Female Male
2	Age	Age
3	Hypertensi	1 hypertension 0 does not have hypertensi
4	heart_disease	1 Have heart disease 0 does not have heart disease
5	ever_married	1 means married 0 means not married
6	work_type	Children Personal Never work Government work Entrepreneur
7	Residence_type	Rural Urban
8	avg_glucose_level	Average glucose level
9	bmi	body mass index
10	smoking_status	Never smoked Used to smoke
11	Stroke	0 (Class does not indicate stroke) 1 (Class indicated stroke)

2.2. Pre-processing data

In this study, the data preprocessing stage consists of four main steps. These steps include data imputation, categorical encoding, outlier handling, and feature scaling. These steps aim to ensure that the data used in the training model is clean, consistent, and ready for processing by machine learning algorithms.

2.2.1. Data imputation

Dataset features that have empty values, namely the BMI feature, we chose a technique to overcome this by changing the empty values to the average value of BMI having a stroke and BMI not having a stroke. The empty values of the entire dataset features can be seen in Table 2. Table 2 shows that the empty value is 201 data, if the empty value in the BMI feature has class 0, then the empty value is changed to the average BMI value of class 0, and vice versa if the empty value in the BMI feature has class 1, then the empty value is changed to the average BMI value of class 1. The average BMI values of classes 0 and 1 can be seen in Table 3.

Table 2. Features that have an empty value condition

No.	Feature	Number of empty values
1	Gender	0
2	Age	0
3	Hypertensi	0
4	heart_disease	0
5	ever_married	0
6	work_type	0
7	Residence_type	0
8	avg_glucose_level	0
9	bmi	201
10	smoking_status	0
11	Stroke	0

Table 3. Average BMI class 1 and BMI class 0

No.	Feature	Average value
1	BMI indicates stroke (1)	30.47
2	BMI does not indicate stroke (0)	28.82

Table 3 shows a comparison of the average BMI values between two classes: class 1 (indicating stroke) and class 0 (not indicating stroke). Based on the calculation results, individuals who experienced a stroke had an average BMI of 30.47, while individuals who did not experience a stroke had an average BMI of 28.82. All stages of this process can be seen in Figure 2.

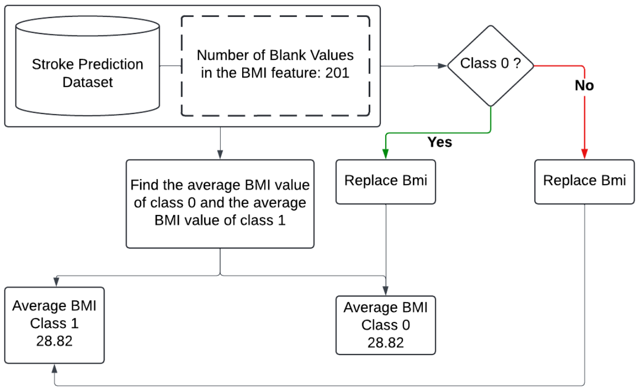

Figure 2. Visualization of the experiment setup for filling in blank values in the BMI feature

2.2.2. Categorical encoding

Applying the LabelEncoder technique to convert categorical features into numeric data, only a few attributes are changed in this labeling, including gender, ever_married, work_type, residence_type, and smoking_status. The results of this process can be seen in Table 4. Figure 3 shows the complete process of converting categorical values to numeric values using LabelEncoder. Since machine learning algorithms like XGBoost can only process numeric data, label encoding techniques are used to transform those features.

Table 4. LabelEncoder results

Gender	Age	Hypertensi	Heart_ disease	Ever_ married	Work_ type	Residence_ type	Avg_Glucose_ level	bmi	Smoking_ status	Stroke
1	67	0	1	1	2	1	228.69	36.6	1	1
0	61	0	0	1	3	0	202.21	30.47	2	1
1	80	0	1	1	2	0	105.92	32.5	2	1
0	9	0	0	1	2	1	171.23	34.4	3	1
0	79	1	0	1	3	0	174.12	24	2	1
0	82	0	1	0	2	1	215.6	24.9	2	0
0	54	0	0	1	0	0	91.61	25.2	2	0
0	49	0	0	1	2	0	138.16	19.4	2	0
0	24	0	0	1	2	1	75.23	29	2	0
0	37	0	0	1	2	0	75.18	48.2	1	0
1	34	0	1	1	2	1	106.23	28.82	1	0

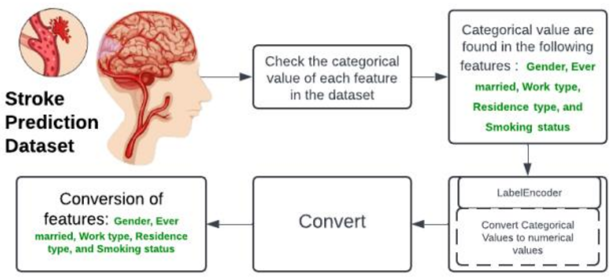

Figure 3. Visualization of the experiment setup for converting categorical values to numeric values with LabelEncoder

2.2.3. Outlier handling

To avoid the extreme influence of unrepresentative values [25], outlier detection and handling are performed on numeric features such as avg_glucose_level and BMI. This process uses the *IQR* method, determining the lower and upper bounds based on the (1).

$$IQR = Q3 - Q1 \tag{1}$$

Information: $Q1$ is first quartile (25th percentile) → lower limit of the first 25% of data. $Q3$ is third quartile (75th percentile) → upper limit of the last 25% of data. IQR is interquartile range, shows the spread of the middle data (50% of the data). After the *Q1* and *Q3* values are obtained, the lower and upper limits for detecting outliers are calculated as (2) and (3).

$$Lower\ Bound = Q1 - 1.5 \times IQR \tag{2}$$

$$Upper\ Bound = Q3 + 1.5 \times IQR \tag{3}$$

Information: *lower bound* is minimum value that is still considered normal, *upper bound* is the maximum value that is still considered normal, and values outside the range lower bound, upper bound are considered outliers.

2.2.4. Feature scaling uses the min-max normalization technique

To ensure the values of widely different dataset features have a uniform scale when building machine learning models, we need to standardize. One of the standardization techniques used is min-max normalization, which changes the original data linearly so that the values between attributes are balanced. With this normalization, attribute values are transformed such that they range between 0 and 1, allowing for consistent comparisons across attributes [26]. Numeric features with widely varying value ranges, such as age, average glucose level, and body weight, were normalized using the min-max normalization method to keep each attribute's value within the range of 0 to 1, preventing any single feature from dominating the model's learning process due to differences in scale. Only these features require normalization because they are continuous numeric values with significant variations in values. Other features, such as gender, smoking status, occupation, residence type, and sex, were categorized using LabelEncoder, and the hypertension and heart disease features were already binary data sets of 0 and 1 and therefore did not require additional normalization [27]. This method can use the (4).

$$N = \frac{MinRange(x - minValue)(maxRange - minRange)}{MaxValue - MinValue} \tag{4}$$

Where *N* is the normalized value (the new value after transformation); *x* is the original value of an attribute before normalization; *minValue* is the maximum value of the attribute, and *maxValue* is the normalized value in the range 0, 1. *minRange* is the lower bound of the desired new scale range (usually 0); and *maxRange* is the upper bound of the desired new scale range (usually 1). The entire normalization process can be seen in Figure 4.

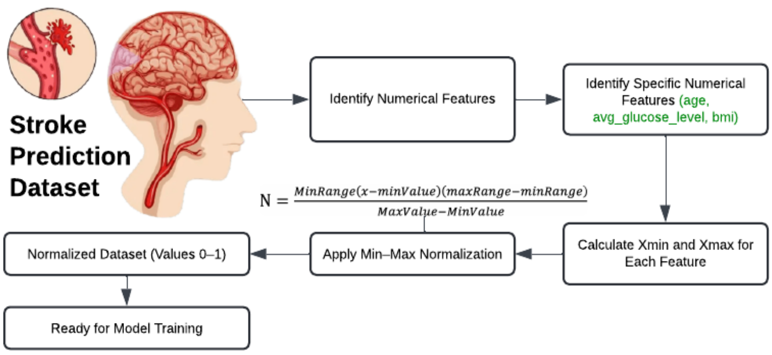

Figure 4. Visualization of the experiment setup for numeric features that have very different value ranges using min-max normalization

2.3. Split data

Training/test data split divides the dataset into training data and test data. The split in this study divides the training data and test data into 70/30 using a stratified split to maintain class proportion balance. The stratified 5-fold cross-validation process is applied only to the training data to evaluate and optimise model performance. In contrast, the test data is used once in the final stage to measure the model's generalisation ability without the risk of data leakage. With this division, the primary objective is to evaluate the model's performance when making predictions using a total of 1,536 test data points. In general, machine learning models can produce good accuracy if they have a small amount of test data [28]. Therefore, in this study, we increased the amount of test data and evaluated whether the model produced satisfactory results. Table 5 illustrates the data distribution.

Table 5. Results of feature selection using the backward elimination method

Fold=k1–k5		
Amount of training data	Test data amount	Amount of data
3,582	1,536	5,118

2.3.1. Sampling

This research dataset experienced unbalanced data conditions, we applied several methods, one of which was the use of the SMOTE method to overcome this. This method can handle a class imbalance in the dataset by balancing the number of minority classes so that it is comparable to the majority class. The result of this technique is synthetic data created based on the KNN [29] as defined in (5).

$$X_{syn} = X_i + (X_{knn} - X_i) \cdot \delta \quad (5)$$

Next, applying the ROS method, this method can also perform data balancing. This method works by randomly copying the minority class (fewer classes) until the number is comparable to the majority (more classes) [30]. Furthermore, the implementation of the random undersampling (RUS) method works by randomly reducing the number of majority classes so that it is proportional to the number of minority classes. This approach reduces the majority representation in the dataset [31].

The latter implements the ADASYN method, this method operates by identifying the relative difficulty level of each minority example in the dataset, this is done by calculating the ratio between the number of majority neighbors and the total number of neighbors (majority and minority) for each minority example. Minority examples that have lower ratios are considered more difficult and more important to expand. ADASYN then creates synthetic samples for these examples by extending the line between the minority example and its neighbors in feature space, focusing on the examples that are most difficult for the model to identify. This approach ensures that the resulting dataset has a better representation of minority classes, improving model performance in cases with significant class imbalance [32]. An overview of the entire data balancing process with all the methods used can be seen in Figure 5.

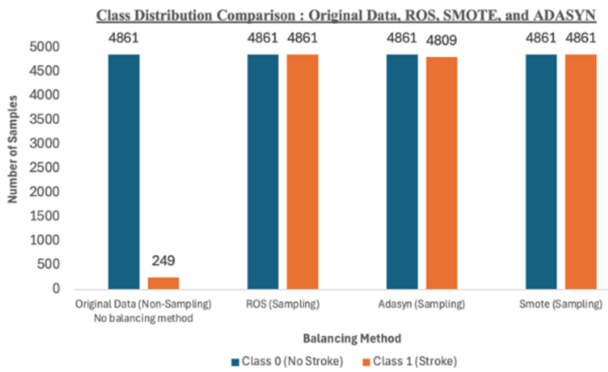

Figure 5. Data balancing and non-balancing results

2.4. Classification with XGBoost algorithm, EasyEnsemble classifier, and hyperparameter tunning

This study employs three modelling methods to evaluate algorithm performance under various data conditions, including oversampled data and original unbalanced data, during both the classification and hyperparameter tuning stages. In the first method, the XGBoost algorithm was trained on training data that had undergone oversampling with SMOTE, ADASYN, and ROS. Since the class distribution in this data was balanced, the settings were made without the scale_pos_weight parameter. In the second method, the characteristics of class imbalance were maintained by using XGBoost on the original data without sampling. This method allows the scale_pos_weight parameter to be included in the hyperparameter search space because it serves to impose a greater penalty on prediction errors in minority classes. This enables the model to learn from imbalanced class distributions more proportionally.

XGBoost combines boosting and gradient boosting methods. In boosting, XGBoost is used to classify errors from previous models, and it is use of gradient descent helps minimize errors during the creation or development of new models [33]. XGBoost requires several parameters to obtain an optimal model called hyperparameters which are used to adjust various aspects of machine learning so that they can influence the performance of the method in processing datasets, several parameters are used to improve classification using the XGBoost method [34], can be seen in Table 6.

Table 6. Parameters in XGBoost method

Parameter	Information
max_depth	Maximum depth of the tree.
eta (learning_rate)	Prevents overfitting by reducing size
min_child_weight	Minimum weight of child_node
n_estimators	Number of trees
subsample	Randomly sampling from training data before constructing the tree.
random_state	internal random number generator initialization

Hyperparameter tuning was performed using the grid search method, which tests all parameter combinations in a predetermined search space. The range of values used included max_depth with five variations (8, 10, 11, 13, 15), learning_rate with five values (0.01, 0.02, 0.05, 0.07, 0.1), min_child_weight with two values (0.5 and 1.0), and n_estimators with two variations (150 and 300). The subsample parameter is locked at a value of 0.5 to maintain consistency in the proportion of samples used in each tree, while random_state is set to 42 to ensure reproducibility. With this configuration, the total number of hyperparameter combinations tested is 100 (5×5×2×2×1). Each combination is evaluated using k-fold cross-validation (k=5), resulting in a total of 500 model trainings. This approach ensures that hyperparameter selection is stable, consistent, and capable of representing the model's performance in a generalizable manner on unseen data.

The third method uses an EasyEnsemble classifier on the initial data to address data imbalance. This method works by undersampling the majority class to form several balanced subsets. Then, in each subgroup, several weak models are trained, and a pooling mechanism is used to combine their predictions. The result is a model that is more resistant to class imbalance and more stable. Additionally, we perform parameter tuning on the model using EasyEnsemble. The parameters we use are n_estimators, base_estimator (a DT with two max_depth), and substitution; the best values for each can be found by testing the method, specifically through a grid search view.

2.5. Evaluation and interpretation

Model performance evaluation is conducted comprehensively by combining several key metrics. It also includes uncertainty analysis, calibration measurements, model interpretability, and external validation to ensure generalisation capabilities. This evaluation approach is designed in accordance with best practices in machine learning-based predictive modelling in the health domain and for handling imbalanced data.

2.5.1. AUROC and AUPRC

AUROC and AUPRC is used to assess the discriminatory performance of a model. AUROC assesses the model’s ability to distinguish between positive and negative classes at various decision thresholds. In contrast, AUPRC assesses unbalanced datasets more accurately because it focuses on the relationship between precision and recall for minority classes. Using these two metrics ensures an unbiased and balanced evaluation, particularly when predicting the risk of rare events [35].

2.5.2. Confidence intervals

To illustrate the statistical uncertainty of the evaluation results, each performance metric is accompanied by a confidence interval. Confidence intervals are calculated through repeated bootstrapping on the test data, and the estimates obtained reflect the variability of model performance across different samples [36]. Combining CIs improves the reliability of interpretation and allows for better comparisons between models.

2.5.3. Calibration plots

To assess calibration, calibration plots and additional calibration scores, such as the Brier score, are used. Calibration determines the level of likelihood of the model's predictions compared to the actual probability of events. Calibration plots are used to assess whether the model tends to be overconfident in its predictions. Clinical applications and decision support systems require good models because they can distinguish classes and generate well-calibrated probabilities [37].

2.5.4. Explainability

Using the explainable method, SHAP, the aspect of interpretability was examined. This study provides an understanding of the contribution of each feature to model predictions at both the global level (across the entire dataset) and the local level (for individual predictions). The explainability approach makes the model clear and facilitates stakeholders, especially those working in the medical or public policy fields [38].

2.5.5. Decision analysis

In addition to standard metrics, a decision analysis was performed to evaluate the model's value in real-world decision-making. Decision curve analysis (DCA), which assesses the net benefit of the model at various risk thresholds, is included in this analysis. DCA serves to evaluate whether the model truly has clinical or operational advantages over basic methods such as treat-all or treat-none. Therefore, the evaluation not only considers statistical performance but also the value of the model in real-life situations [39].

2.5.6. External validation

The trained and evaluated model is then tested through external validation with data from various sources or time periods. External validation shows the model's performance in situations outside the initial training data distribution. This step is crucial for assessing the generalisation and strength of the model and is essential in predictive research aimed at wider application [40].

3. RESULTS AND DISCUSSION

This section presents the results of the modelling and evaluation process of this study. To fulfil the research objectives, a comprehensive analysis was conducted to evaluate the performance of the XGBoost model under three different conditions: oversampling data, original data with weight adjustment using scale_pos_weight, and original data processed using the EasyEnsemble classifier. To evaluate the practical benefits of the model in decision making, decision analysis included performance evaluation using AUROC and AUPRC metrics, prediction calibration, uncertainty levels through confidence intervals, model interpretability using SHAP, and decision analysis. To ensure the model's generalisation ability to data outside the training distribution, external validation was also performed. The following are the test results for the original data with weight adjustment using scale_pos_weight, as shown in Figure 6.

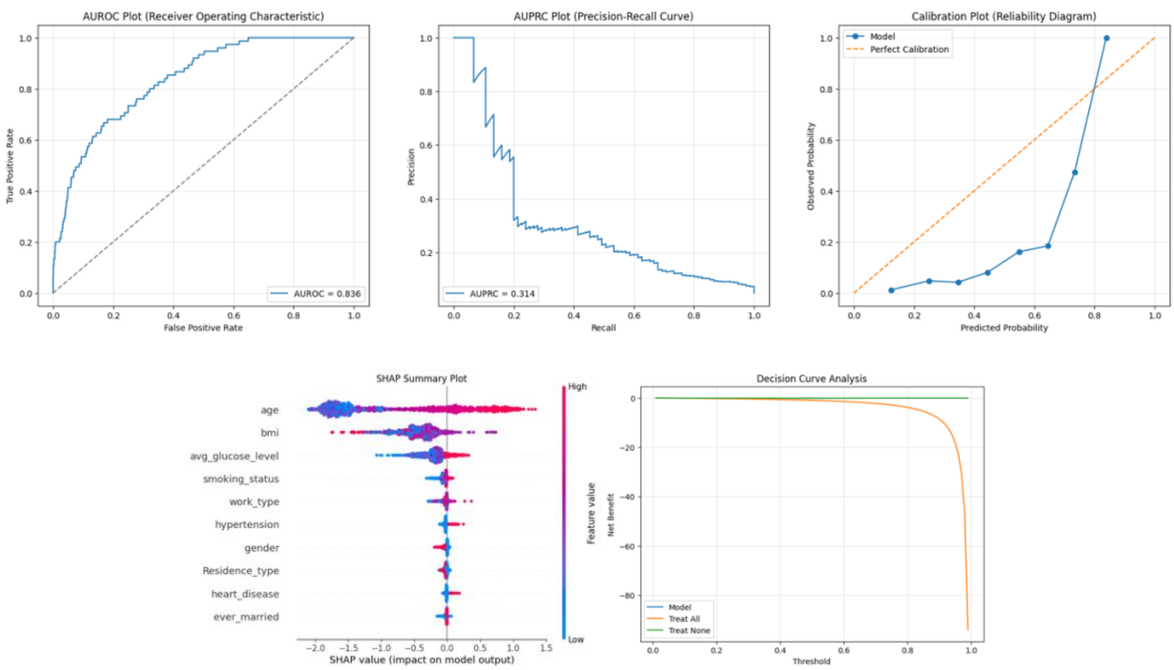

Figure 6. Evaluation results of AUROC and AUPRC, confidence intervals, calibration plots, explainability, decision analysis, and external validation in testing original data with the XGBoost method using the scale_pos_weight parameter

Figure 6 shows that the XGBoost model calibrated using scale_pos_weight exhibits solid predictive performance on the test data, with the best combination of hyperparameters being learning_rate=0.01, max_depth=8, min_child_weight=1.0, and n_estimators=150. The model achieved an AUROC value of 0.836, indicating a strong ability to distinguish between stroke and non-stroke classes. The AUPRC value of 0.314 remained above the baseline in conditions of high class imbalance. The calibration graph indicates that the model tends to estimate probabilities conservatively in the low range but is more accurate at high probabilities. SHAP analysis confirmed that age, BMI, and glucose levels are the most influential predictors, in line with clinical literature. In addition, DCA shows that the model provides better decision benefits than treat-all or treat-none strategies, proving that this class-weighting approach is practical and relevant for stroke risk prediction scenarios. Furthermore, the results of testing on the original data using the EasyEnsemble classifier method are shown in Figure 7.

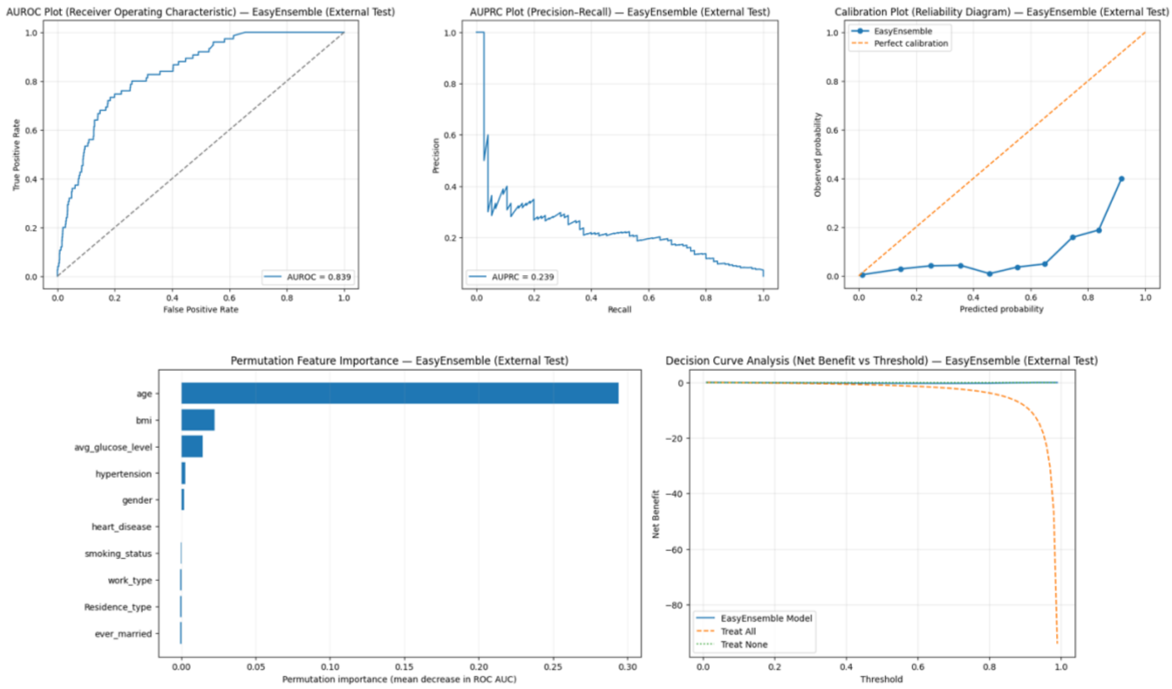

Figure 7. Evaluation results of AUROC and AUPRC, confidence intervals, calibration plots, explainability, decision analysis, and external validation in testing original data using the EasyEnsemble Classifier method

Figure 7 shows the best performance results for the EasyEnsemble classifier method using the original data with the DT configuration depth=6, n_estimators=30, and replacement=false. In the external test set, the model achieved an AUROC of 0.839, indicating strong discriminatory power; however, the AUPRC value of 0.239 suggests challenges in predicting minority classes. The calibration curve shows that the prediction probabilities are not fully aligned with the actual probabilities, especially in the middle probability range. Feature importance analysis identifies age as the most influential predictor, followed by BMI and average glucose level, while other features contribute minimally. Meanwhile, DCA shows that the model provides small but still positive decision benefits at most thresholds. Overall, this model is suitable for use as a solid baseline, but improvements in accuracy for minority classes are still needed. Next are the results of testing on sampled data using the ADASYN technique, which were then classified using the XGBoost method, as shown in Figure 8.

Figure 8 shows the results of testing the sampling data (adasyn) using the XGBoost method, indicating that the best configuration, with max_depth=15, learning_rate=0.05, min_child_weight=1.0, and n_estimators=300, exhibits fairly good classification capabilities. The AUROC value of 0.768 indicates that the model can distinguish between negative and positive classes moderately. In contrast, the AUPRC value of 0.107 suggests that performance on minority classes remains limited, which may be attributed to data imbalance. Although the probability predictions are not yet fully aligned with the actual distribution, the calibration curve gives a Brier score of 0.0921. This indicates that the probability predictions are relatively accurate. According to SHAP analysis, the variables of age, BMI, and average glucose level are those that most influence the model's predictions. These findings are consistent with previous studies. However, DCA shows that the model does not provide a greater net benefit compared to the “serve all” and “serve none” strategies. As a result, the model cannot be used for threshold-based decision making. The following are the results of classification using the XGBoost method in the second data sampling condition with the ROS technique, as shown in Figure 9.

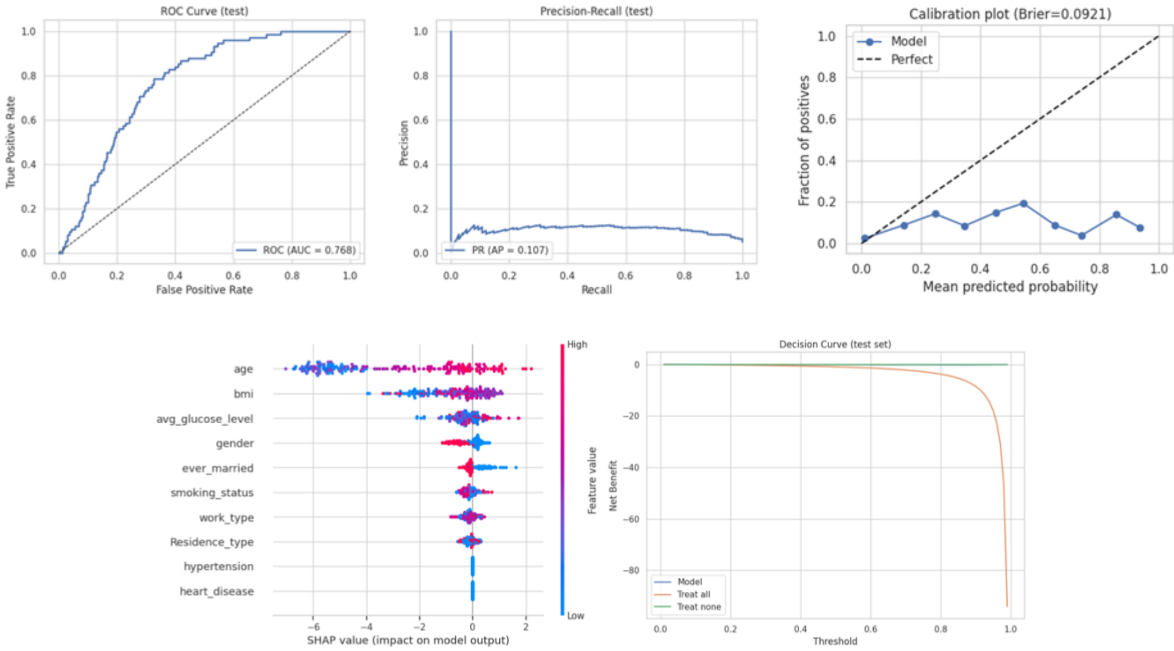

Figure 8. Evaluation results of AUROC and AUPRC, confidence intervals, calibration plots, explainability, decision analysis, and external validation in sampling data testing (adasyn) using the XGBoost method

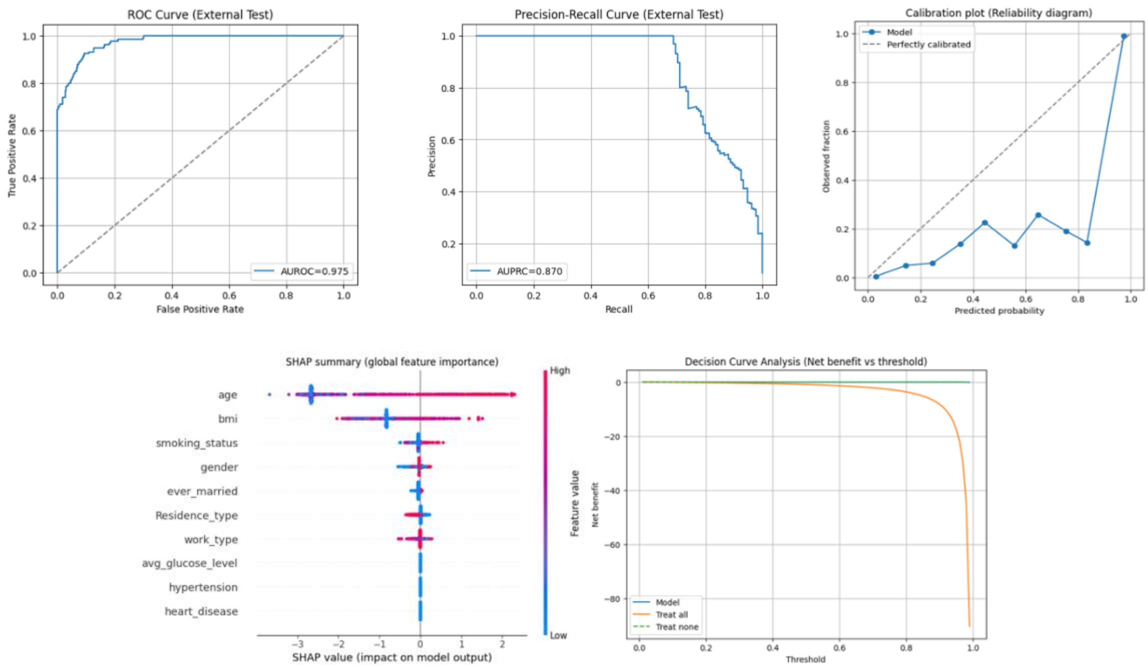

Figure 9. Evaluation results of AUROC and AUPRC, confidence intervals, calibration plots, explainability, decision analysis, and external validation in sampling data testing (ROS) using the XGBoost method

Figure 9 shows the results of hyperparameter optimization, indicating that the optimal configuration for the XGBoost model was found with a combination of max_depth=8, learning_rate=0.01, min_child_weight=1.0, and n_estimators=300. With these settings, the model performed very well in classifying the test data. The strong ability to distinguish between positive and negative classes is demonstrated by an AUROC of 0.975. The results are consistent with the precision-recall curve, which shows an AUPRC value of 0.870, indicating the model's stability in unbalanced data conditions. Although the general trend remains on the ideal line, calibration evaluation shows that the predicted probabilities tend to be less confident across most of the prediction range. According to explainability analysis using SHAP, age, BMI, and smoking status are the most significant factors affecting predictions. Other features, such as gender, marital status, and hypertension, have a negligible influence. According to the decision curve test, the model consistently outperforms the treat-none and treat-all strategies, providing a positive net benefit at various decision thresholds. Finally, the results of classification using the XGBoost method in the third data sampling condition, with the SMOTE technique, are presented in Figure 10.

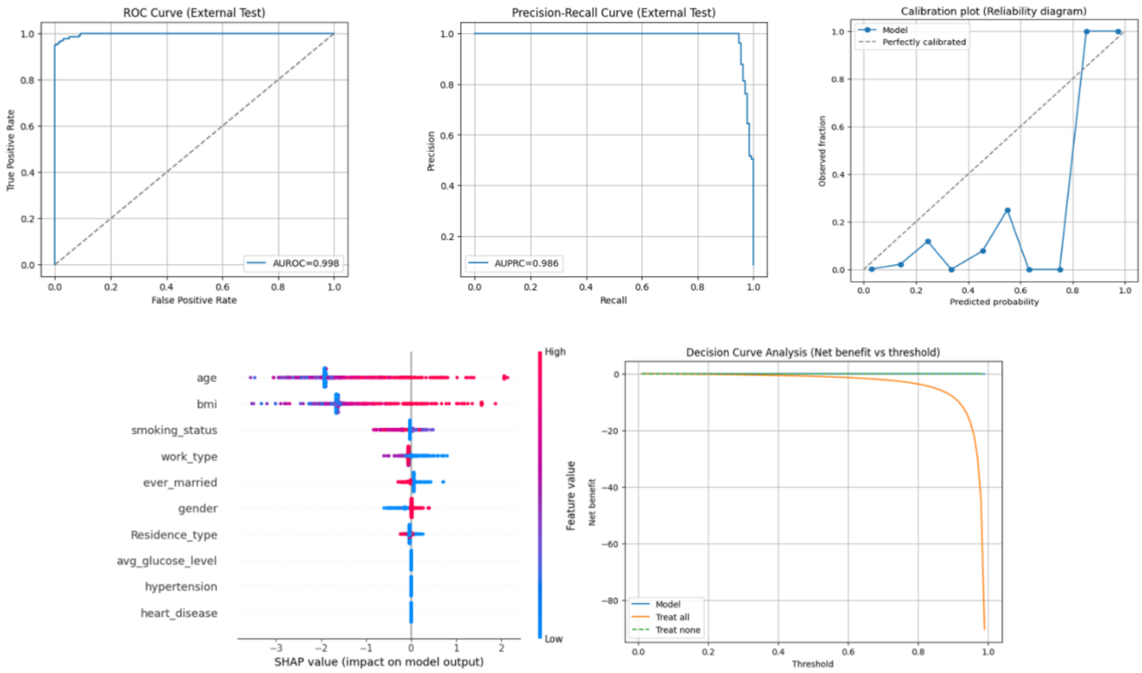

Figure 10. Evaluation results of AUROC and AUPRC, confidence intervals, calibration plots, explainability, decision analysis, and external validation in sampling data testing (SMOTE) using the XGBoost method

Figure 10 shows that the results of sampling data testing (SMOTE) with the XGBoost method have the best hyperparameter optimisation for the XGBoost model sampling data testing, which can be achieved with a combination of max_depth=10, learning_rate=0.02, min_child_weight=1.0, and n_estimators=150. An excellent ability to distinguish between positive and negative classes has been demonstrated in the evaluation of external test data, as indicated by an AUROC value of 0.998, which shows an excellent capability. The model's robustness in conditions of an imbalanced class distribution is demonstrated by the precision-recall curve, which yields an AUPRC of 0.986. Although the calibration results show a deviation in the middle probabilities from the ideal line, the model still shows an acceptable prediction tendency. According to SHAP analysis, age, BMI, and smoking status are the main factors in the model's decision-making. Factors such as occupation and marital status follow these factors. However, the results of the DCA show that the model provides a greater net benefit than the treat-all and treat-none strategies at various decision threshold values. Overall, the results demonstrate that the XGBoost model optimised with SMOTE exhibits excellent performance, is stable, and is practically relevant for prediction tasks on imbalanced data. Table 7 shows the results of the overall classification comparison of the proposed tests. The results presented are the best tests for each data condition.

Table 7 presents the test results on the original data, indicating that XGBoost, with settings of learning_rate 0.01, max_depth 8, min_child_weight 1.0, and n_estimators 150, exhibits moderate performance, with AUROC 0.836, AUPRC 0.314, and accuracy 0.88. It is reported that age, BMI, and blood glucose concentration level are the most effective predictors. On the original data, the EasyEnsemble Classifier model exhibits comparable performance, with an AUROC of 0.839, but a lower AUPRC (0.239) and an accuracy of only 0.69. This suggests that sensitivity to minority classes is lacking, even though the essential features that emerge are essentially the same. Model performance showed significant differences in the sampled data scenario. ADASYN produced an AUROC of 0.768, AUPRC of 0.107, and an accuracy of 0.87, indicating that this method was unable to improve classification capabilities optimally. In contrast, ROS shows a significant improvement, with an AUROC of 0.975, AUPRC of 0.870, and accuracy of 0.94. The main features affecting the prediction are age, BMI, and smoking status. With the configuration max_depth 10, learning_rate 0.02, min_child_weight 1.0, and n_estimators 150, SMOTE provided the best approach; achieving an AUROC of 0.998, AUPRC of 0.986, and accuracy of 0.998, making it the most superior model in separating classes and handling data imbalance. Overall, these findings show that, compared to traditional sampling and ensemble approaches, oversampling techniques, particularly SMOTE, significantly improve model performance.

Table 7. Comparison of the overall classification results of the proposed test

Method	Best parameters	AUROC	AUPRC	SHAP-top features	Accuracy
Testing original data with the XGBoost method using the scale_pos_weight parameter	learning_rate=0.01, max_depth=8, min_child_weight=1.0, n_estimators=150	0.836	0.314	age, bmi, avg_glucose_level.	0.88
Testing original data with the EasyEnsemble classifier method	DecisionTree depth=6, n_estimators=30, replacement=False	0.839	0.239	age, bmi, avg_glucose_level.	0.69
Data testing sampling with the XGBoost method			ADASYN		
	max_depth=15, learning_rate=0.05, min_child_weight=1.0, dan n_estimators=300	0,768	0.107	age, bmi, avg_glucose_level.	0.87
			ROS		
	max_depth=8, learning_rate=0.01, min_child_weight=1.0, n_estimators=300.	0,975.	0,870	age, bmi, smoking_status.	0.94
			SMOTE		
	max_depth=10, learning_rate=0.02, min_child_weight=1.0, n_estimators=150	0,998	0.986	age, bmi, smoking_status.	0.99

4. CONCLUSION

This study demonstrates that the performance of stroke prediction models can be significantly enhanced by employing data preprocessing methods, addressing class imbalance, and optimising hyperparameters. Testing conducted under various data conditions confirms that XGBoost is the most consistent model, with an AUROC of 0.998, AUPRC of 0.986, and accuracy of 0.998. This model demonstrates exceptional discrimination capabilities and strong prediction stability, despite the original class distribution being imbalanced. ADASYN and ROS were able to improve performance, but did not consistently outperform SMOTE. For now, EasyEnsemble classifier and XGBoost work well on raw data with scale_pos_weight, but are less effective at identifying minority classes. The findings indicate that the combination of XGBoost and SMOTE is most effective in supporting the early identification of stroke risk. Further evaluation results, including AUROC, AUPRC, confidence intervals, calibration plots, SHAP-based interpretability analysis, and DCA, are reinforced. Overall, this study demonstrates that employing proper balancing techniques and systematically tuning model parameters are crucial steps in enhancing prediction quality for medical problems with imbalanced class distributions. The recommended method is not only more accurate but also easier to interpret and has relevant practical benefits to aid clinical decision-making.

ACKNOWLEDGMENTS

A big thank you to the Institute for Research and Community Service (LPPM), Amikom University, Yogyakarta, for the support, guidance, and facilities provided during the implementation of this research. This research also received funding support from the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) of the Republic of Indonesia through a research grant program, which plays a vital role in supporting the smoothness and success of this research activity.

FUNDING INFORMATION

This research was funded by a research grant from the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia (Kemendikbudristek) through the Research Grant Program managed by the Institute for Research and Community Service (LPPM) of Amikom University Yogyakarta, under the Early Career Faculty Research Scheme and grant/contract number 107/E5/PG.02.00.PL/2024.

AUTHOR CONTRIBUTIONS STATEMENT

The Contributor Role Taxonomy (CRediT) is used in this journal to prevent authorship disputes and increase collaboration in the research and publication process by providing a clear explanation of each author's contribution.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Abd Mizwar A. Rahim	✓	✓		✓	✓	✓			✓	✓	✓			✓
Anna Baita		✓	✓	✓	✓	✓		✓	✓			✓		✓
Firman Asharudin		✓	✓	✓			✓		✓			✓	✓	✓
Wahid Miftahul Ashari			✓		✓		✓		✓		✓		✓	✓
Walidy Rahman Hakim		✓		✓		✓		✓		✓		✓	✓	
Andriyan Dwi Putra			✓		✓		✓		✓		✓		✓	✓
Supriatin		✓	✓		✓		✓			✓	✓		✓	✓
Eko Pramono			✓		✓		✓		✓		✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

No informed consent was required as this research did not involve personal data of identifiable individuals or human subjects.

ETHICAL APPROVAL

This research did not involve human participants or animals, so no ethical approval was required. The entire research process adheres to the standards and policies established by the institution and relevant national regulations.

DATA AVAILABILITY

The stroke prediction dataset, which supports this research, is openly and publicly available on fedesoriano's Kaggle platform. You can access it via the following link: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset. The analysis and model testing process in this research relies on this dataset.

REFERENCES

[1] V. L. Feigin *et al.*, "World stroke organization: global stroke fact sheet 2025," *International Journal of Stroke*, vol. 20, no. 2, pp. 132–144, Feb. 2025, doi: 10.1177/17474930241308142.

[2] M. O. Owolabi *et al.*, "The state of stroke services across the globe: report of world stroke organization–world health organization surveys," *International Journal of Stroke*, vol. 16, no. 8, pp. 889–901, Oct. 2021, doi: 10.1177/17474930211019568.

[3] C. Hazelton *et al.*, "Interventions for perceptual disorders following stroke," *Cochrane Database of Systematic Reviews*, vol. 2022, no. 11, Nov. 2022, doi: 10.1002/14651858.CD007039.pub3.

[4] V. L. Feigin *et al.*, "Pragmatic solutions to reduce the global burden of stroke: a world stroke organization–lancet neurology commission," *The Lancet Neurology*, vol. 22, no. 12, pp. 1160–1206, Dec. 2023, doi: 10.1016/S1474-4422(23)00277-6.

[5] G. Cipriani, S. Danti, L. Picchi, A. Nuti, and M. D. Fiorino, "Daily functioning and dementia," *Dementia and Neuropsychologia*, vol. 14, no. 2, pp. 93–102, Jun. 2020, doi: 10.1590/1980-57642020dn14-020001.

[6] L. Cavanagh and J. S. Paulsen, "Neuropsychology and vascular cognitive impairment and dementia," *Neurologic Clinics*, vol. 42, no. 4, pp. 809–820, Nov. 2024, doi: 10.1016/j.ncl.2024.05.006.

[7] 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, Nov. 2022, doi: 10.1016/j.health.2022.100032.

[8] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. Al Mamun, and M. S. Kaiser, "Performance analysis of machine learning approaches in stroke prediction," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Nov. 2020, pp. 1464–1469. doi: 10.1109/ICECA49313.2020.9297525.

[9] N. Biswas, K. M. M. Uddin, S. T. Rikta, and S. K. Dey, "A comparative analysis of machine learning classifiers for stroke prediction: a predictive analytics approach," *Healthcare Analytics*, vol. 2, Nov. 2022, doi: 10.1016/j.health.2022.100116.

[10] S. Srinivasan, S. Gunasekaran, S. K. Mathivanan, B. A. M. M. B, P. Jayagopal, and G. T. Dalu, "An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database," *Scientific Reports*, vol. 13, no. 1, Aug. 2023, doi: 10.1038/s41598-023-40717-1.

[11] T. B. Sasongko and A. M. Arrahim, "A novel EfficientMobileDenseNet for Monkeypox detection and classification," *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 4, pp. 897–912, May 2025, doi: 10.22266/ijies2025.0531.58.

[12] A. Gupta *et al.*, "Predicting stroke risk: an effective stroke prediction model based on neural networks," *Journal of Neurorestoratology*, vol. 13, no. 1, Feb. 2025, doi: 10.1016/j.jnrt.2024.100156.

[13] K. Mridha, S. Ghimire, J. Shin, A. Aran, M. 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.

[14] P. O. Akinwumi, S. Ojo, T. I. Nathaniel, J. Wanliss, O. Karunwi, and M. Sulaiman, "Evaluating machine learning models for stroke prediction based on clinical variables," *Frontiers in Neurology*, vol. 16, Sep. 2025, doi: 10.3389/fneur.2025.1668420.

[15] E. Dritsas and M. Trigka, "Stroke risk prediction with machine learning techniques," *Sensors*, vol. 22, no. 13, Jun. 2022, doi: 10.3390/s22134670.

[16] E. M. Alanazi, A. Abdou, and J. Luo, "Predicting risk of stroke from lab tests using machine learning algorithms: development and evaluation of prediction models," *JMIR Formative Research*, vol. 5, no. 12, Dec. 2021, doi: 10.2196/23440.

[17] V. W. de Vargas, J. A. S. Aranda, R. d. S. Costa, P. R. d. S. Pereira, and J. L. V. Barbosa, "Imbalanced data preprocessing techniques for machine learning: a systematic mapping study," *Knowledge and Information Systems*, vol. 65, no. 1, pp. 31–57, Jan. 2023, doi: 10.1007/s10115-022-01772-8.

[18] L. B. V. de Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," *Applied Soft Computing*, vol. 133, Jan. 2023, doi: 10.1016/j.asoc.2022.109924.

[19] D. Wilimitis and C. G. Walsh, "Practical considerations and applied examples of cross-validation for model development and evaluation in health care: tutorial," *JMIR AI*, vol. 2, Dec. 2023, doi: 10.2196/49023.

[20] A. P. M. S. Hamdard and A. P. H. Lodin, "Effect of feature selection on the accuracy of machine learning model," *International Journal of Multidisciplinary Research and Analysis*, vol. 6, no. 9, Sep. 2023, doi: 10.47191/ijmra/v6-i9-66.

[21] P. Mooijman, C. Catal, B. Tekinerdogan, A. Lommen, and M. Blokland, "The effects of data balancing approaches: a case study," *Applied Soft Computing*, vol. 132, Jan. 2023, doi: 10.1016/j.asoc.2022.109853.

[22] E. Alshdaifat, D. Alshdaifat, A. Alsarhan, F. Hussein, and S. M. F. S. El-Salhi, "The effect of preprocessing techniques, applied to numeric features, on classification algorithms' performance," *Data*, vol. 6, no. 2, Jan. 2021, doi: 10.3390/data6020011.

[23] H. Moayedi, A. Osouli, H. Nguyen, and A. S. A. Rashid, "A novel Harris hawks' optimization and k-fold cross-validation predicting slope stability," *Engineering with Computers*, vol. 37, no. 1, pp. 369–379, Jan. 2021, doi: 10.1007/s00366-019-00828-8.

[24] Fedesoriano, "Stroke prediction dataset," Kaggle. Accessed: Dec. 20, 2025. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

[25] J. Garlits *et al.*, "Statistical approaches for establishing appropriate immunogenicity assay cut points: impact of sample distribution, sample size, and outlier removal," *The AAPS Journal*, vol. 25, no. 3, Apr. 2023, doi: 10.1208/s12248-023-00806-5.

[26] M. Shantal, Z. Othman, and A. A. Bakar, "A novel approach for data feature weighting using correlation coefficients and min–max normalization," *Symmetry*, vol. 15, no. 12, Dec. 2023, doi: 10.3390/sym15122185.

[27] S. Sinsomboonthong, "Performance comparison of new adjusted min-max with decimal scaling and statistical column normalization methods for artificial neural network classification," *International Journal of Mathematics and Mathematical Sciences*, vol. 2022, pp. 1–9, Apr. 2022, doi: 10.1155/2022/3584406.

[28] A. Bustillo, R. Reis, A. R. Machado, and D. Y. Pimenov, "Improving the accuracy of machine-learning models with data from machine test repetitions," *Journal of Intelligent Manufacturing*, vol. 33, no. 1, pp. 203–221, Jan. 2022, doi: 10.1007/s10845-020-01661-3.

[29] J. Miao and W. Zhu, "Precision–recall curve (PRC) classification trees," *Evolutionary Intelligence*, vol. 15, no. 3, pp. 1545–1569, Sep. 2022, doi: 10.1007/s12065-021-00565-2.

[30] Y. Zhang, L. Deng, and B. Wei, "Imbalanced data classification based on improved random-SMOTE and feature standard deviation," *Mathematics*, vol. 12, no. 11, May 2024, doi: 10.3390/math12111709.

[31] J. Sadaiyandi, P. Arumugam, A. K. Sangaiah, and C. Zhang, "Stratified sampling-based deep learning approach to increase prediction accuracy of unbalanced dataset," *Electronics*, vol. 12, no. 21, Oct. 2023, doi: 10.3390/electronics12214423.

[32] M. Zakariah, S. A. AlQahtani, and M. S. Al-Rakhami, "Machine learning-based adaptive synthetic sampling technique for intrusion detection," *Applied Sciences*, vol. 13, no. 11, May 2023, doi: 10.3390/app13116504.

[33] Z. Arif Ali, Z. H. Abduljabbar, H. A. Tahir, A. B. Sallow, and S. M. Almufti, "Extreme gradient boosting algorithm with machine learning: a review," *Academic Journal of Nawroz University*, vol. 12, no. 2, pp. 320–334, May 2023, doi: 10.25007/ajnu.v12n2a1612.

[34] D. Tarwidi, S. R. Pudjaprasetya, D. Adytia, and M. Apri, "An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach," *MethodsX*, vol. 10, 2023, doi: 10.1016/j.mex.2023.102119.

[35] W. Li and Q. Guo, "Plotting receiver operating characteristic and precision–recall curves from presence and background data," *Ecology and Evolution*, vol. 11, no. 15, pp. 10192–10206, Aug. 2021, doi: 10.1002/ece3.7826.

[36] K. Takahashi, K. Yamamoto, A. Kuchiba, and T. Koyama, "Confidence interval for micro-averaged F1 and macro-averaged F1 scores," *Applied Intelligence*, vol. 52, no. 5, pp. 4961–4972, Mar. 2022, doi: 10.1007/s10489-021-02635-5.

[37] T. S. Filho, H. Song, M. P.-Nieto, R. S.-Rodriguez, M. Kull, and P. Flach, "Classifier calibration: a survey on how to assess and improve predicted class probabilities," *Machine Learning*, vol. 112, no. 9, pp. 3211–3260, Sep. 2023, doi: 10.1007/s10994-023-06336-7.

[38] P. K. Kanti, P. Sharma, V. V. Wanatasanappan, and N. M. Said, "Explainable machine learning techniques for hybrid nanofluids transport characteristics: an evaluation of Shapley additive and local interpretable model-agnostic explanations," *Journal of Thermal Analysis and Calorimetry*, vol. 149, no. 21, pp. 11599–11618, Nov. 2024, doi: 10.1007/s10973-024-13639-x.

[39] Q. Yu, Z. Hou, and Z. Wang, "Predictive modeling of preoperative acute heart failure in older adults with hypertension: a dual perspective of SHAP values and interaction analysis," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, Nov. 2024, doi: 10.1186/s12911-024-02734-6.

[40] F. Maleki, K. Ovens, R. Gupta, C. Reinhold, A. Spatz, and R. Forghani, "Generalizability of machine learning models: quantitative evaluation of three methodological pitfalls," *Radiology: Artificial Intelligence*, vol. 5, no. 1, Jan. 2023, doi: 10.1148/ryai.220028.

BIOGRAPHIES OF AUTHORS

Abd Mizwar A. Rahim he obtained his bachelor's degree in 2019 and master's degree in 2022 in Computer Science from Amikom University Yogyakarta. Since March 2023, he has been a permanent lecturer in the Informatics Study Program, Faculty of Computer Science, Amikom University Yogyakarta. His research areas include data mining, machine learning, and deep learning, with a focus on algorithm optimization and the development of intelligent systems to support decision making. In addition to teaching, he is actively involved in various research projects and has published scientific papers in national and international journals in the fields of artificial intelligence, data analysis, and information systems. He also plays a role in academic activities related to research and community service. His research interests include the application of artificial intelligence in the fields of health, industry, and public services. He can be contacted at email: abdulmizwar@amikom.ac.id.

Anna Baita she is a lecturer and researcher at the Informatics Study Program, Faculty of Computer Science, Amikom University Yogyakarta, Indonesia. Her research interests include machine learning, deep learning, natural language processing, and the application of data-driven approaches in the development of intelligent systems. She is actively involved in various institutional research activities, academic collaborations, and student research project supervision. Her works demonstrate a commitment to integrating artificial intelligence techniques with real-world problems, particularly in the fields of public communication, digital literacy, and technology-based community empowerment. In addition, his research focuses on the development and optimization of artificial intelligence models to improve the performance, reliability, and applicability of systems in various fields. She can be contacted at email: anna@amikom.ac.id.

Firman Asharudin he has actively conducted research in the fields of web design and web development, with a strong focus on creating effective and user-centered digital solutions. His research interests include object detection, web-based application development, decision support systems, and UI/UX design using the design thinking approach. He has led and participated in various collaborative and interdisciplinary research projects. Among these is a DIKTI-funded project that integrated IoT, computer networks, and information systems, resulting in the development of a public street lighting monitoring system. Additionally, he received further DIKTI funding through the BIMA program for the development of a waste monitoring and sorting system. He can be contacted at email: firman_asharudin@amikom.ac.id.

Wahid Miftahul Ashari he earned his master of Engineering degree from Institut Teknologi Bandung in 2017, specializing in cybersecurity. His academic background has strengthened his expertise in protecting digital assets and securing information systems. He is actively involved in research and development of cybersecurity solutions, with particular emphasis on data protection, information system security, and improving digital security awareness within the community. In addition to his research activities, He has completed several professional certifications in the field of cybersecurity, reflecting his strong commitment to professional excellence and continuous skill development. Through research, training, and practical implementation, he consistently contributes to the advancement of cybersecurity practices and the promotion of a safer digital environment. He can be contacted at email: wahidashari@amikom.ac.id.

Walidy Rahman Hakim He earned his bachelor's degree in 2019 from a recognized institution, completing a thesis titled analysis and optimization of bandwidth management based on QoS at Bjongngopi Cafe, which explored practical strategies for enhancing network performance in a commercial environment. Building on this foundation, he pursued and successfully completed his master's degree in 2022 at Amikom University Yogyakarta, with a thesis entitled evaluation of IT infrastructure governance implementation using COBIT 5, analyzing frameworks to optimize IT governance and organizational efficiency. Since March 2025, he has been serving as a permanent lecturer in the Informatics Study Program within the Faculty of Pharmacy, Science, and Technology at Al-Irsyad University Cilacap, where he engages in teaching, curriculum enhancement, and student guidance. His research interests primarily encompass software engineering, computer networks, and the internet of things (IoT), with applications in modern technological challenges. He remains actively involved in various research projects, regularly publishes scholarly articles in national journals, and contributes as an editor for the campus research journal, promoting academic discourse. For inquiries or collaboration. He can be contacted at email: walidy.hakim@universitasalirsyad.ac.id.

Andriyan Dwi Putra He obtained his master's degree in Information Technology from Amikom University Yogyakarta in 2015. He is currently a lecturer at Amikom University Yogyakarta, where his research focuses on information security, software development, and internet of things (IoT). He has published various scientific articles in areas such as mikrotik network security, Android forensics, and machine learning–based intelligent systems. In addition to his teaching and research activities, he actively contributes to the development of appropriate technology through the Kedaireka research scheme and industry partnerships. His work reflects a strong commitment to applied research and technological innovation that addresses real-world problems. He can be contacted at email: andriyan@amikom.ac.id.

Supriatin she is a lecturer at Amikom University Yogyakarta. She earned her master's degree in Informatics Engineering from Amikom University Yogyakarta in 2014. Her research interests focus on data mining, decision support systems (DSS), and expert systems. She has published various scientific articles in the field of information systems in reputable journals and conference proceedings. In addition to teaching and conducting research, she is actively involved in community service programs funded through the BIMA scheme. Her academic work reflects a strong commitment to applied research and knowledge dissemination. For academic and professional correspondence. She can be contacted at email: supriatin@amikom.ac.id.

Eko Pramono he earned his bachelor's degree in Electrical Engineering in 1995 with a thesis titled control of nitrogen plants by PLC and his master's degree in 2006 with a thesis titled design of a code generator for 'basic stamp' microcontrollers based on state diagrams in sequential systems from Universitas Gadjah Mada, Yogyakarta. Since 2008, he has been a permanent lecturer at the Informatics Study Program, Faculty of Computer Science, Amikom University Yogyakarta, Indonesia. His research focus and academic interests include the internet of things (IoT), edge computing, computer networks, and radio over IP. He is active in various research projects and has published scientific papers in national journals. In addition, he teaches courses related to IoT hardware and electronics. He can be contacted at email: eko.p@amikom.ac.id.