Original Article

DOI: 10.5582/bst.2025.01013

Investigating perioperative pressure injuries and factors influencing them with imbalanced samples using a Synthetic Minority Over-sampling Technique

Yiwei Zhou[1,2,3], Jian Wu[1], Xin Xu[4], Guirong Shi[4], Ping Liu[4,*], Liping Jiang[4,*]

[1] Business School, University of Shanghai for Science and Technology, Shanghai, China;
[2] School of Intelligent Emergency Management, University of Shanghai for Science and Technology, Shanghai, China;
[3] Smart Urban Mobility Institute, University of Shanghai for Science and Technology, Shanghai, China;
[4] Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China.

SUMMARY: This study investigates the use of machine learning (ML) models combined with a Synthetic Minority Over-sampling Technique (SMOTE) and its variants to predict perioperative pressure injuries (PIs) in an imbalanced dataset. PIs are a significant healthcare problem, often leading to prolonged hospitalization and increased medical costs. Conventional risk assessment scales are limited in their ability to predict PIs accurately, prompting the exploration of ML techniques to address this challenge.We utilized data from 7,292 patients admitted to a tertiary care hospital in Shanghai between May 2017 and July 2023, with a final dataset of 2,972 patients, including 158 with PIs. Seven ML algorithms—Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Extra Trees (ET), K-Nearest Neighbors (KNN), and Decision Trees (DT)—were used in conjunction with SMOTE, SMOTE+ENN, Borderline-SMOTE, ADASYN, and GAN to balance the dataset and improve model performance.Results revealed significant improvements in model performance when SMOTE and its variants were used. For instance, the XGBoost model hadan AUC of 0.996 with SMOTE, compared to 0.800 on raw data. SMOTE+ENN and Borderline-SMOTE further enhanced the models' ability to identify minority classes. External validation indicatedthat XGBoost, RF, and ET exhibited the highest stability and accuracy, with XGBoost having an AUC of 0.977. SHAP analysis revealed that factors such as anesthesia grade, age, and serum albumin levels significantly influenced model predictions.In conclusion, integrating SMOTE with ML algorithms effectively addressed a data imbalance and improved the prediction of perioperative PIs. Future work should focus on refining SMOTE techniques and exploring their application to larger, multi-center datasets to enhance the generalizability of these findings, and especially for diseaseswith a lowincidence.

Keywords: machine learning, pressure injuries, SMOTE, predictive modelling, data imbalance

1. Introduction

Machine learning (ML) is playing an increasingly important role in data processing in clinical care, showing great potential for improving patient prognosis and optimizing healthcare management. However, a data imbalance is prevalent in clinical healthcare data, which is mainly evidentin the uneven distribution of case samples, where the number of samples for common diseases far exceeds that for rare diseases. This imbalance may lead to bias in the construction of ML models, making the models tend to predict categories with a higher frequency of occurrence while ignoring categories with smaller sample sizes. This bias not only affects the recall and accuracy of the model, but also limits the effective application of conventional classification algorithms in disease diagnosis, and especially in healthcare domains that require precise identification of a small number of cases (*1*).

An imbalance in clinical care data is a pervasive challenge in healthcare analytics, and especially for datasets with skewed class distributions. This issue presents significant hurdles in training classifiers for predictive modeling tasks, as highlighted by Kumar *et al.*(*2*). To address this, researchers have delved into a variety of solutions, with the Synthetic Minority Oversampling Technique (SMOTE) being a prominent one. SMOTE aims to enhance the predictive performance of ML models by rectifying class imbalances in clinical outcome prediction. In a pivotal study by Ishaq *et al.*,

the emphasis was on refining the survival prediction of heart failure patients through the useof SMOTE and sophisticated data mining. This studyuseda suite of nine classification models, underscoring the critical role of mitigating adata imbalanceto bolster the predictive accuracy of clinical datasets (*3*). Parallel to this, Goorbergh *et al.* conducted a case study on prediction modeling for ovarian cancer diagnosis. They scrutinized the repercussions of imbalance correction on the effectivenessof logistic regression models. Their findings underscored the potential detrimental effects of class imbalance corrections on risk prediction models, underscoring the necessity for judicious useof balancing techniques toanalyzeclinical data (*4*). Ridwan *et al*. usedML techniques and SMOTE to uncover patterns and risk factors within the Pima Indian diabetes dataset. By adeptly using SMOTE to detect diabetes, their study significantly advanced the scientific understandingof usingML algorithms to analyze clinical data (*5*). In a related vein, Javid *et al.*tackledthe issue of a class imbalance in the early diagnosis of Alzheimer's disease based on MRI images. They usedSMOTE to ensure an equitable distribution of samples across each class, thereby enhancing the performance of deep learning models in medical imaging data for disease diagnosis (*6*). This study highlightsthe significance of balancing techniques like SMOTE in increasing the effectiveness of deep learning models in this area.

In summary, a data imbalance is a pervasive challenge in ML, and particularly in healthcare datasets where minority class samples (*e.g*., patients with pressure injuries, or PIs) are often critical for accurate predictions. Various methods of data enhancement have been developed to address this issue, each with its own strengths and limitations. One of the most widely used techniques is SMOTE, which generates synthetic samples by interpolating between existing minority class samples. While SMOTE has shown significant success in improving model performance, other methods such as Adaptive Synthetic Sampling (ADASYN), Borderline-SMOTE, SMOTE+ENN (Edited Nearest Neighbors), and Generative Adversarial Networks(GAN) techniques have also emerged as promising alternatives. The current study aims to explore the effectiveness of these methods in predicting perioperative PIs and to compare their performance to conventional SMOTE.

PIs, also known as pressure ulcers, are a common and serious healthcare problem worldwide, leading to prolonged hospitalization, poor quality of life, increased mortality, and higher medical costs. They are defined as localized injuries to the skin and underlying tissue, usually over a bony prominence, resulting from pressure or a combination of pressure and shear(7). Given the preventable nature of these injuries, there is a critical need for effective early warning models to assist clinicians and nurses in making timely predictions and taking preventive action. SMOTE is used to enhance the predictive power of the model by balancing the class distribution by adding a few class samples to the training data, thus improving the model's prediction accuracy for PIs (*8,9*).

Conventional risk assessment scales (*e.g*., the Braden, Norton, and Waterlow scales) have been widely used but are limited in performance and are workload-intensive(*10*). As a result, artificial intelligence algorithms have been explored as they can capture patterns in complex data and have advantages in predicting time-to-event data, which is a common occurrence in clinical practice. ML models for predicting various medical outcomes, including PIs, have been developed by utilizing large datasets and algorithmic learning.

Nowadays, there are a growing number of instances whereMLis used in medicine, but the small amount of data has been a limitation in the aspects related to disease prediction, so the main aimof the current study was to evaluate the usefulness and effectiveness of the various resampling techniques in the prediction of PIs.Thegoal is to construct a ML model that can effectively predict PIs in emergency patients. To achieve this goal, seven ML algorithms were used in combination with the SMOTE algorithm and related methodsof extension to deal with the data imbalance problem.

2. Materials and Methods

2.1. Model selection

SMOTE is a technique for dealing with imbalanced datasets by generating synthetic samples of a few classes to balance the category distribution and was first proposed by Chawla *et al.* in 2002. This method creates new sample points by interpolating between the minority class samples and their k-nearest neighbors, thereby increasing sample diversity and reducing the risk of overfitting(*11*). In the current study, the SMOTE algorithm was used to enhance the model's ability to recognize the minority category (*i.e*., patients with PIs).

The basic steps of the underlying logic are as follows:

i) Select a minority sample X as the "root sample"for synthesizing a new sample.

ii) Find by Euclidean distance the k nearest neighboring samples (usually k is odd, *e.g.*, 5) of that sample, which also belong to the minority category. For two points X("x1,y1,z1,...") and O("x2,y2,z2,...") coordinates in n-dimensional space, the Euclidean distance d between them can be calculated withthe following formula.

$$d = \sqrt{(x2 - x1)^2 + (y2 - y1)^2 + (z2 - z1)^2 + \cdots} \quad (1)$$

iii) For each nearest-neighbor sample O, perform the following steps to generate a new sample point O_{new}. Calculate the root sample X and its nearest neighbor

samples O: dif = O -X; generate a random number between [0, 1] λ: and Use this formula to synthesize the value of each attribute of the new sample O_{new}.

$$O_{new} = O + \lambda \times (X - O)\,\lambda \quad (2)$$

iv) Repeat step 3 to produce the required number of new samples.

The key to the SMOTE algorithm is that instead of simply copying existing minority class samples, it creates new sample points by interpolating between the minority class samples, which increases the diversity of the samples and reduces the risk of overfitting. This approach is particularly useful in situations where the number of minority samples is small but each sample is important.A basic diagram of SMOTE is shown in Figure 1.

Similar to SMOTE, ADASYN generates synthetic samples but focuses more on the difficult-to-learn regions of the minority class, potentially improving model performance(*12*).

SMOTE+ENN is a hybrid technique that combines SMOTE with the ENN technique to efficiently deal with imbalanced datasets. First, a large amount of oversampled data is generated using the SMOTE method described above, and then ENN is used to clean the dataset by removing noisy samples, ENN works by identifying samples whose nearest neighbors belong to a different class and removing them. This helps toreduce noise and improve the quality of the dataset. This approach helps reduce overfitting and enhances the model's generalization ability (*13*).

Borderline-SMOTE is an enhanced version of SMOTE that generates synthetic samples specifically from minority class samples near the decision boundary to improve classification performance by targeting the most informative samples. Minority samples are categorized into three types: Safe (surrounded mostly by minority class samples), Danger (surrounded mostly by majority class samples and considered to be on the decision boundary), and Noise (surrounded entirely by majority class samples). Only Danger samples are used to create synthetic samples by selecting neighboring minority samples and interpolating between them using the same formula as SMOTE. There are two variants: Borderline-SMOTE1 generates synthetic samples using only minority class neighbors, whereas Borderline-SMOTE2 uses any neighbor (regardless of class) to introduce more diversity. The key advantage of Borderline-SMOTE is its focus on the decision boundary, which reduces the risk of generating noisy samples and enhances the effectiveness of synthetic samples (*14*).

GANs are advanced generative models that use a generator network to create synthetic samples and a discriminator network to distinguish between real and synthetic samples. GANs can produce high-quality synthetic data, potentially improving model performance by increasing the diversity of the minority class(*15*).

The current study used seven different ML models to predict PIs in emergency patients, each of which has its own unique strengths that make them perform well when dealing with specific types of data and problems.Support Vector Machine (SVM) is effective in dealing with high-dimensional spatial data and non-linear problems, being able to find hyperplanes that maximize the class interval. In PI prediction, SVM can help identify complex patterns, and especially when the feature space is large(*16*).Random Forest (RF), as an integrated learning method, improves the stability and accuracy of the model by constructing multiple decision trees and is very resistant to overfitting. When faced with imbalanced datasets, RF provides robust predictions and reduces the variance of predictions by integrating multiple models(*17*).Extreme Gradient Boosting (XGBoost) is an efficient gradient boosting framework that is capable of handling large-scale datasets and typically

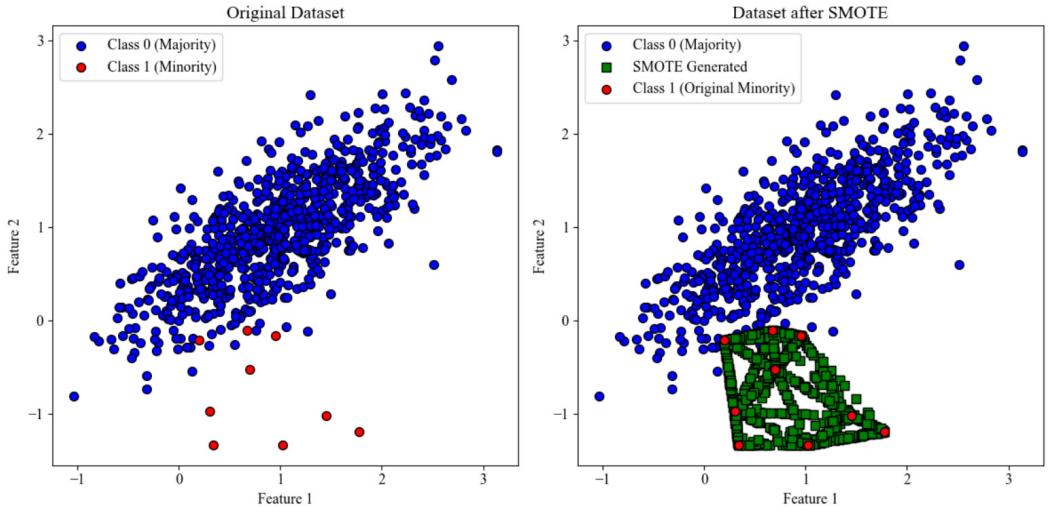

Figure 1. The basic working principle of SMOTE. Modelling SMOTE workings using randomly generated data.

outperforms conventional gradient boosting methods in terms of prediction performance. XGBoost performs well when dealing with datasets with a large number of features, which makes it suitable for extraction of key information from a large amount of patient datato predict PIs(*18*).Extra Trees (ET) is able to effectively deal with non-linear relationships and imbalanced datasets and improve the recognition of a few classes through its high stochasticity and integrated learning(*17*). K-Nearest Neighbors (KNN) is a simple instance-based learning algorithm that does not require a training phase and can directly use training data for prediction. KNN performs well on small datasets, so it is suitable when the sample size is not particularly large, as in the current study, and especially when SMOTE processingis used (*19*).Logistic regression (LR) is a linear model that is suitable for binary classification problems and can provide a probabilistic interpretation of the prediction results. When predicting PIs, LR can provide a direct interpretation of the probability of a patient developing a pressure injury, which is useful for clinical decision-making(*20*). Decision trees (DTs) are intuitive models that are easy to understand and interpret and can clearly demonstrate the relationship between features and target variables. DTs can help to identify the most important risk factors and can be used as a baseline for comparison to more complex models(*17*).

In the current study, the main challenge faced was the problem of a data imbalance, *i.e.*, the number of patients with PIs (positive sample) was much smaller than the number of patients without PIs (negative sample). To address this issue, the SMOTE algorithm was used to balance the dataset and the seven ML models described earlier were usedto construct predictive models. These models were chosen based on their extensive use and history of success in dealing with imbalanced datasets, handling high-dimensional data, providing predictive explanations, and in medical predictive modelling. Comparing the performance of these models enables the identification of the most appropriate model for the currentdata and problem, thereby improving the accuracy and reliability of predicting PIs. In addition, the diversity of these models allows evaluationand validation of predictions from different perspectives, ensuring that the findings are robust and reliable.

2.2. Participants

Data from a total of 7,292 patients consisting of7,171 indicators were selected from all recorded inpatient data ata tertiary care hospital in Shanghai during the period from May 2017 to July 2023 (numerous interfering items in data during the COVID-19 epidemic were not selected), and a total of 549 patients with PIs (7.53%) served as the initial screening subjects. After data processing, data from the remaining 2,972 patientsserved as the final data for this study and included 158patients with PIs (5.32%).

2.3. Data preprocessing

When dealing with the huge number of 7,171 feature variables, the XGBoost model was usedto identify the features that contribute most to the model performance. The advantage of XGBoost is that it is able to filter the features efficiently when there are missing values in the data, enabling the initialselectionof the top 32 feature variables that have the greatest impact on the model. Through further in-depth analyses, those features that were not strongly associated with PIswere eliminated and 27 key feature variables were ultimately selected, laying a solid foundation for building an accurate prediction model. In order to maintain the high quality of the dataset and reduce the noise interference in model training, a key decision was made to eliminate sets of data with more than 8 missing values among the 27 key feature variables. This strategy helps to maintain the integrity of the dataset while avoiding the uncertainty introduced by too many missing values, ensuring the reliability of the data and the stability of model training. After completing the screening of feature variables and the reduction of the dataset, in-depth data preprocessing was performedon the remaining data. This includes meticulous treatment of missing values, outliers, and duplicate records, steps that are critical to ensuring the quality of the data and the smooth running of subsequent experiments.

Data preprocessing consisted mainly of the following: *i*) Categorical variables. Missing values for characteristic variables in the data involving categorical variables are uniformly filled in using plurality in the current study; *ii*) Continuous variables. Missing values for continuous variables in this study were filled in using the mean of the age groups. Age groups were every 10 years, and 0-9 and10-19 were each averaged and populated within their age range.

These comprehensive data preprocessing measures ensured the cleanliness and consistency of the dataset, providing a solid data foundation for subsequent model training and analysis.

2.4. Evaluation metrics

In the model training phase, the datasetwas divided into training and validation sets at a ratio of 7:3, and multiple ML models were used to predict whether PIs occurred in emergency patients. In the model evaluation phase, two key evaluation metrics were used: the Confusion Matrix and ROC_AUC.

ROC curveswere also plotted and AUC values were calculated for each model; ROC curves demonstrate the model's performance under different thresholds, while AUC values quantify the model's ability to distinguish between positive and negative categories, with higher AUC values indicating better classification performance.

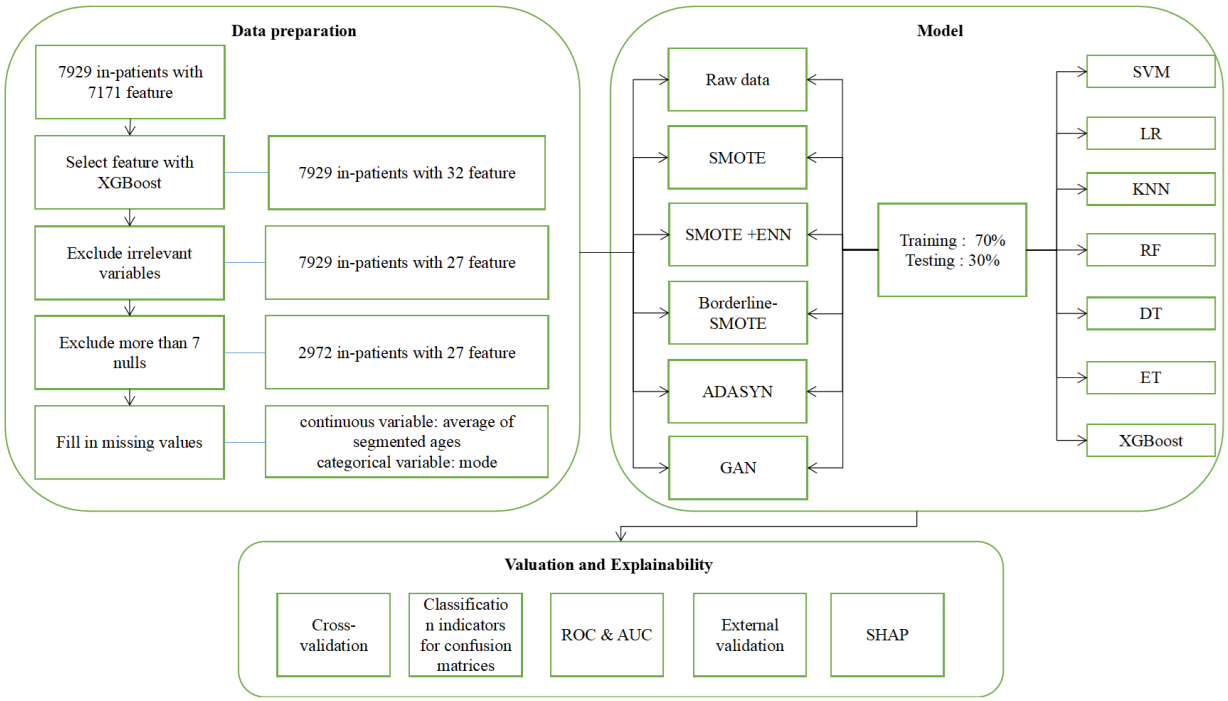

Figure 2. Flowchart for this study.

Finally, the confusion matrices and AUC values of the different modelswere compared to determine which model performed best in predicting PIs. This comprehensive assessment approach alloweda full understanding and comparison of the performance of each model in order to select the most appropriate model to aid inclinical decision-making.

2.5. Experimental design

The flow of this study is shown in Figure 2.

3. Results

3.1. Participants' characteristics

The distribution and comparison of several basic characteristics of patients with PIs (PI) and patients without PIs (Non-PI) is shown in Table 1. The table lists characteristics including sex, age, hypertension (HTN), hyperlipidemia, diabetes mellitus (DM), cardiovascular disease (CVD), history of malignancy, smoking status, drinking status, body temperature, pulse rate (PR), respiratory rate (RR), diastolic blood pressure (DBP), systolic blood pressure (SBP), body mass index (BMI), serum albumin, operating time, intra-operative blood transfusion, intra-operative hypotension (IH), surgical position (thisrefers to the specific position of the patient during surgery. 1-3 are supine, prone, and lateral positions, respectively. 0 is an undefined position), surgical dressing, dressing site, anesthesia grade, method of anesthesia, oxygen saturation (SpO_2), self-care competency grade, and blood glucose (BG). Categorical variables are expressed as the number (percentage) andcontinuous variables are expressed asthe mean (range). Comparison of these variables revealed significant differences in these characteristics between the two groups, with variables such as age, pulse rate, body mass index and method of anesthesia differing significantly between the two groups while variables such as sex, hypertension, and diabetes mellitus did not.

Given that the original dataset is multidimensional, visually depicting the newly generated positive samples presents a challenge. To overcome this, all variables wereprojected onto a single axis, thereby facilitating a clear visualization of the samples created by the SMOTE algorithm. For further details, refer to Figure 3.
Data after different methods of enhancement are shown in Tables 2-6.

3.2. Comparison of ML-based models

In this study, the confusion matrix of the model after using SMOTE and its variants revealed significant improvements as shown in Table 7. For example, the SMOTE-enhanced XGBoost model hadextremely high TP and TN values in internal validation while minimizing FP and FN values, indicating that the model performed well in identifying a small class of samples (patients with PIs). In addition, methods such as SMOTE+ENN and Borderline-SMOTE, although slightly inferior to SMOTE in some models, further improved the model's ability to identify minority classes by optimizing the sample quality or focusing on the borderline region.

Table 1. Basic characteristics of patients

Variables	Non-PI (*n*=2,814)	PI (*n*=158)	*p* value
Sex, *n* (%)	1,614 (57.4%)	69 (43.6%)	0.625
Age, years	54 [0-98]	74 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	420 (14.9%)	26 (17.5%)	0.105
Hyperlipidemia, n (%)	12 (0.05%)	0 (0%)	0.445
Diabetes Mellitus (DM), *n* (%)	114 (4.1%)	14 (8.9%)	0.604
Cardiovascular Disease (CVD), *n* (%)	4 (0.2%)	0 (0%)	0.700
History of Malignant Tumor, *n* (%)	116 (4.1%)	14 (8.9%)	0.460
Smoking Status, *n* (%)	280 (10.0%)	13 (8.2%)	0.089
Alcohol Consumption Status, *n* (%)	2 (0.1%)	1 (0.6%)	0.709
Body Temperature, °C	36.6 [35.3-40.7]	36.6 [35.2-40.3]	0.684
Pulse Rate (PR), bpm	88 [38-198]	84 [52-165]	0.003
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	0.187
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	81 [54-134]	0.291
Systolic Blood Pressure (SBP), mmHg	124 [47-277]	135 [100-195]	0.278
Body Mass Index (BMI), kg/m²	21.7 [5.4-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.7 [16.5-59]	32.9 [18.4-45.7]	0.488
Operating Time, h	2.90 [0.15-7.29]	2.82 [0.25-10.02]	0.586
Intraoperative Blood Transfusion, *n* (%)	262 (9.3%)	12 (7.6%)	0.693
Intraoperative Hypotension (IH), *n* (%)	246 (8.7%)	6 (3.8%)	0.912
Surgical Position	0.8 [0-3]	0.4 [0-3]	0.532
Surgical Dressing, *n* (%)	2,351 (83.5%)	125 (79.1%)	0.746
Dressing Site, *n* (%)	464 (16.5%)	33 (20.9%)	0.823
Anesthesia Grade	2.3 [0-5]	1.8 [0-5]	0.011
Anesthesia Method, *n* (%)	2,678 (95.2%)	26 (17.5%)	< 0.001
Oxygen Saturation (SpO2), %	97 [65-100]	96 [47-100]	0.567
Self-Care Ability Grade	2.8 [1-3]	2.9 [2-3]	0.006
Blood Glucose (BG)	7.9 [1.4-28.0]	8.0 [3.8-17.6]	0.243

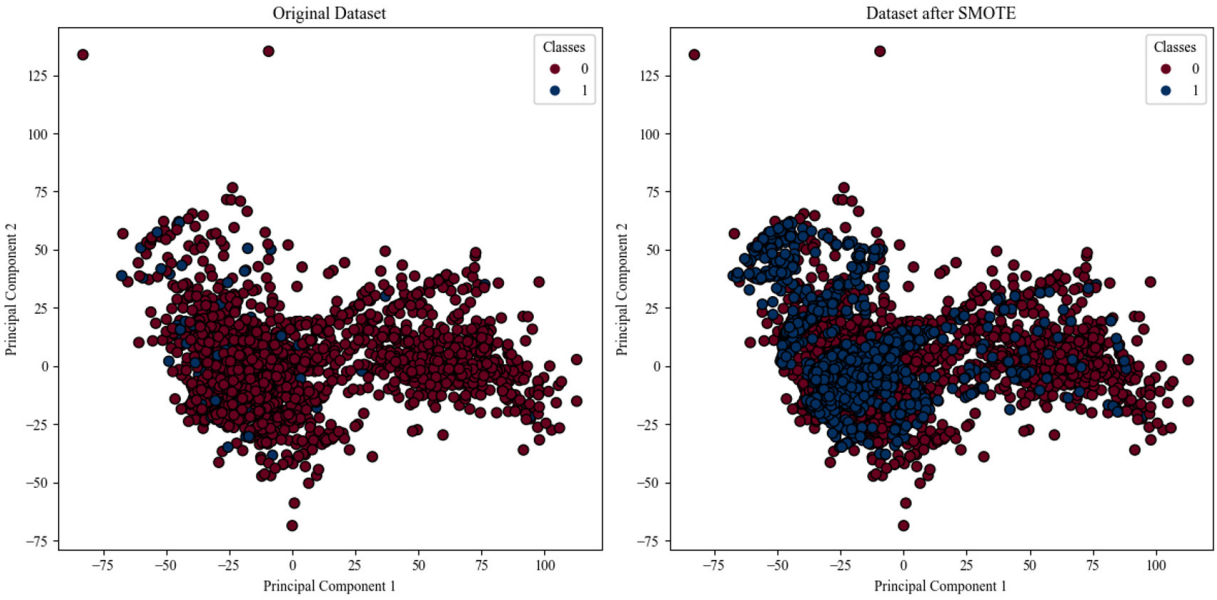

Figure 3. Status of data generated by SMOTE.

ADASYN and GAN also showed good performance in the confusion matrix, although they may face some challenges withhigh-dimensional data.

The analysis of the confusion matrix allows a more intuitive view of the impact of different methods of data enhancement on model performance. For example, SMOTE results in ahigh recall and precision in most models, while SMOTE+ENN performs well in removing noise, albeit possibly at the expense of some sample diversity.Borderline-SMOTE and ADASYN, in contrast, display better recognition of minority classes in specific models, although they have limited overall performance gains.GAN generated high-quality minority class samples, but its generated samples may be too close to the original samples, leading to an increased risk of overfitting.

As can be seen from Table 8, the performance metrics (*e.g.*, precision, recall, F1 score, accuracy, and AUC) of

Table 2. Supplementary data-enhanced dataset - SMOTE

Variables	Non-PI (*n*=2,252)	PI (*n*=2,252)	*p* value
Sex, *n* (%)	1,329 (59.0%)	821 (36.5%)	< 0.001
Age, years	53 [0-98]	74 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	329 (14.6%)	778 (34.5%)	< 0.001
Hyperlipidemia, *n* (%)	11 (0.5%)	0 (0%)	0.011
Diabetes Mellitus (DM), *n* (%)	91 (4.0%)	39 (1.7%)	< 0.001
Cardiovascular Disease (CVD), *n* (%)	4 (0.2%)	0 (0%)	0.153
History of Malignant Tumor, *n* (%)	127 (5.6%)	22 (1.0%)	< 0.001
Smoking Status, *n* (%)	230 (10.2%)	32 (1.4%)	< 0.001
Alcohol Consumption Status, *n* (%)	2 (0.1%)	0 (0%)	0.515
Body Temperature, °C	36.6 [35.3-40.7]	36.5 [36.0-40.3]	0.247
Pulse Rate (PR), bpm	88 [18-198]	83 [52-165]	< 0.001
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	< 0.001
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	81 [56-134]	< 0.001
Systolic Blood Pressure (SBP), mmHg	124 [53-277]	135 [50-195]	0.604
Body Mass Index (BMI), kg/m²	21.6 [7.2-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.8 [16.5-53.9]	33.2 [18.4-45.7]	0.886
Operating Time, h	2.89 [0.15-6.84]	2.76 [0.25-10.02]	< 0.001
Intraoperative Blood Transfusion, *n* (%)	222 (9.9%)	389 (17.3%)	< 0.001
Intraoperative Hypotension (IH), *n* (%)	198 (8.8%)	38 (1.7%)	0.017
Surgical Position	0.8 [0-3]	0.4 [0-3]	< 0.001
Surgical Dressing, *n* (%)	1,888 (83.8%)	1,315 (58.4%)	< 0.001
Dressing Site, *n* (%)	364 (16.2%)	159 (7.1%)	< 0.001
Anesthesia Grade	2.8 [1-5]	3.2 [2-5]	< 0.001
Anesthesia Method, *n* (%)	2,145 (95.2%)	1,519 (67.5%)	< 0.001
Oxygen Saturation (SpO2), %	97 [76-100]	97 [65-100]	0.003
Self-Care Ability Grade	2.9 [1-3]	2.9 [2-3]	< 0.001
Blood Glucose (BG)	7.9 [1.4-28.0]	7.9 [3.8-13.4]	0.575

Table 3. Supplementary data-enhanced dataset - ADASYN

Variables	Non-PI (*n*=2,252)	PI (*n*=2,252)	*p* value
Sex, *n* (%)	1,311 (58.2%)	733 (32.5%)	< 0.001
Age, years	54 [0-98]	73 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	346 (15.3%)	64 (2.8%)	< 0.001
Hyperlipidemia, *n* (%)	9 (0.4%)	0 (0%)	0.101
Diabetes Mellitus (DM), *n* (%)	96 (4.3%)	36 (1.6%)	0.109
Cardiovascular Disease (CVD), *n* (%)	3 (0.1%)	0 (0%)	0.785
History of Malignant Tumor, *n* (%)	138 (6.1%)	34 (1.5%)	< 0.001
Smoking Status, *n* (%)	223 (9.9%)	33 (1.5%)	< 0.001
Alcohol Consumption Status, *n* (%)	1 (0.1%)	0 (0%)	0.305
Body Temperature, °C	36.6 [35.9-40.7]	36.6 [35.2-40.3]	0.079
Pulse Rate (PR), bpm	88 [18-198]	84 [52-165]	< 0.001
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	< 0.001
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	81 [54-134]	0.001
Systolic Blood Pressure (SBP), mmHg	124 [47-277]	134 [50-195]	0.987
Body Mass Index (BMI), kg/m²	21.7 [5.4-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.8 [16.5-59.0]	33.0 [18.4-45.7]	0.119
Operating Time, h	2.89 [0.17-6.84]	2.80 [0.25-10.02]	< 0.001
Intraoperative Blood Transfusion, *n* (%)	206 (9.1%)	41 (18.2%)	< 0.001
Intraoperative Hypotension (IH), *n* (%)	209 (9.3%)	2 (0.1%)	< 0.001
Surgical Position	0.8 [0-3]	0.4 [0-3]	< 0.001
Surgical Dressing, *n* (%)	1,876 (83.3%)	1,433 (63.6%)	< 0.001
Dressing Site, *n* (%)	376 (16.7%)	95 (4.2%)	< 0.001
Anesthesia Grade	2.8 [1-5]	3.2 [2-5]	< 0.001
Anesthesia Method, *n* (%)	2,143 (95.2%)	1,660 (73.7%)	< 0.001
Oxygen Saturation (SpO2), %	97 [71-100]	97 [65-100]	0.002
Self-Care Ability Grade	2.9 [1-3]	2.9 [2-3]	< 0.001
Blood Glucose (BG)	7.9 [1.4-23.0]	8.0 [3.8-17.6]	0.007

Table 4. Supplementary data-enhanced dataset - GAN

Variables	Non-PI (*n*=2,252)	PI (*n*=2,252)	*p* value
Sex, *n* (%)	1,290 (59.0%)	821 (36.5%)	0.715
Age, years	55 [0-98]	74 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	329 (14.6%)	778 (34.5%)	0.077
Hyperlipidemia, *n* (%)	11 (0.5%)	0 (0%)	0.396
Diabetes Mellitus (DM), *n* (%)	91 (4.0%)	39 (1.7%)	0.023
Cardiovascular Disease (CVD), *n* (%)	4 (0.2%)	0 (0%)	0.500
History of Malignant Tumor, *n* (%)	127 (5.6%)	22 (1.0%)	0.302
Smoking Status, *n* (%)	230 (10.2%)	32 (1.4%)	0.173
Alcohol Consumption Status, *n* (%)	2 (0.1%)	0 (0%)	0.166
Body Temperature, °C	36.6 [35.3-40.7]	36.5 [36.0-40.3]	0.687
Pulse Rate (PR), bpm	88 [18-198]	83 [52-165]	< 0.001
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	0.199
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	81 [56-134]	0.412
Systolic Blood Pressure (SBP), mmHg	124 [53-277]	135 [50-195]	0.322
Body Mass Index (BMI), kg/m²	21.6 [7.2-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.8 [16.5-53.9]	33.2 [18.4-45.7]	0.649
Operating Time, h	2.89 [0.15-6.84]	2.76 [0.25-10.02]	0.026
Intraoperative Blood Transfusion, *n* (%)	222 (9.9%)	389 (17.3%)	0.919
Intraoperative Hypotension (IH), *n* (%)	198 (8.8%)	38 (1.7%)	0.371
Surgical Position	0.8 [0-3]	0.4 [0-3]	0.121
Surgical Dressing, *n* (%)	1,888 (83.8%)	1,315 (58.4%)	0.187
Dressing Site, *n* (%)	364 (16.2%)	159 (7.1%)	0.056
Anesthesia Grade	2.8 [1-5]	3.2 [2-5]	< 0.001
Anesthesia Method, *n* (%)	2,145 (95.2%)	1,519 (67.5%)	< 0.001
Oxygen Saturation (SpO2), %	97 [76-100]	97 [65-100]	< 0.001
Self-Care Ability Grade	2.9 [1-3]	2.9 [2-3]	0.002
Blood Glucose (BG)	7.9 [1.4-28.0]	7.9 [3.8-13.4]	0.447

Table 5. Supplementary data-enhanced dataset - SMOTE + ENN

Variables	Non-PI (*n*=2,252)	PI (*n*=2,252)	*p* value
Sex, *n* (%)	1,329 (59.0%)	809 (35.9%)	< 0.001
Age, years	53 [0-98]	74 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	329 (14.6%)	89 (4.0%)	< 0.001
Hyperlipidemia, *n* (%)	11 (0.5%)	0 (0%)	0.489
Diabetes Mellitus (DM), *n* (%)	91 (4.0%)	30 (1.3%)	0.103
Cardiovascular Disease (CVD), *n* (%)	4 (0.2%)	0 (0%)	0.360
History of Malignant Tumor, *n* (%)	127 (5.6%)	31 (1.4%)	< 0.001
Smoking Status, *n* (%)	230 (10.2%)	32 (1.4%)	< 0.001
Alcohol Consumption Status, *n* (%)	2 (0.1%)	0 (0%)	0.267
Body Temperature, °C	36.6 [35.3-40.7]	36.5 [36.0-40.3]	0.586
Pulse Rate (PR), bpm	88 [18-198]	83 [52-165]	< 0.001
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	< 0.001
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	81 [56-134]	0.003
Systolic Blood Pressure (SBP), mmHg	124 [53-277]	134 [50-195]	0.979
Body Mass Index (BMI), kg/m²	21.6 [7.2-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.8 [16.5-53.9]	33.2 [18.4-45.7]	0.451
Operating Time, h	2.89 [0.15-6.84]	2.76 [0.25-10.02]	< 0.001
Intraoperative Blood Transfusion, *n* (%)	222 (9.9%)	16 (0.7%)	< 0.001
Intraoperative Hypotension (IH), *n* (%)	198 (8.8%)	1 (0.1%)	< 0.001
Surgical Position	0.8 [0-3]	0.3 [0-3]	< 0.001
Surgical Dressing, *n* (%)	1,888 (83.8%)	1,335 (59.3%)	< 0.001
Dressing Site, *n* (%)	364 (16.2%)	150 (6.7%)	< 0.001
Anesthesia Grade	2.8 [1-5]	3.2 [2-5]	< 0.001
Anesthesia Method, *n* (%)	2,145 (95.2%)	1,555 (69.0%)	< 0.001
Oxygen Saturation (SpO2), %	97 [76-100]	97 [65-100]	< 0.001
Self-Care Ability Grade	2.9 [1-3]	2.9 [2-3]	< 0.001
Blood Glucose (BG)	7.9 [1.4-28.0]	7.9 [3.8-13.4]	0.845

Table 6. Supplementary data-enhanced dataset - Borderline-SMOTE

Variables	Non-PI (*n*=2,252)	PI (*n*=2,252)	*p* value
Sex, *n* (%)	1,329 (59.0%)	701 (31.1%)	< 0.001
Age, years	53 [0-98]	79 [0-94]	< 0.001
Hypertension (HTN), *n* (%)	329 (14.6%)	773 (34.3%)	< 0.001
Hyperlipidemia, *n* (%)	11 (0.5%)	0 (0%)	0.014
Diabetes Mellitus (DM), *n* (%)	91 (4.0%)	53 (2.4%)	0.001
Cardiovascular Disease (CVD), *n* (%)	4 (0.2%)	0 (0%)	0.061
History of Malignant Tumor, *n* (%)	127 (5.6%)	11 (0.5%)	< 0.001
Smoking Status, *n* (%)	230 (10.2%)	12 (0.5%)	< 0.001
Alcohol Consumption Status, *n* (%)	2 (0.1%)	0 (0%)	0.985
Body Temperature, °C	36.6 [35.3-40.7]	36.6 [36.0-40.3]	0.718
Pulse Rate (PR), bpm	88 [18-198]	84 [52-165]	< 0.001
Respiratory Rate (RR), bpm	20 [10-78]	18 [12-33]	< 0.001
Diastolic Blood Pressure (DBP), mmHg	75 [24-152]	83 [56-134]	< 0.001
Systolic Blood Pressure (SBP), mmHg	124 [53-277]	138 [50-195]	0.005
Body Mass Index (BMI), kg/m²	21.6 [7.2-49]	22.7 [12.8-41.6]	< 0.001
Serum Albumin, g/L	36.8 [16.5-53.9]	33.2 [18.4-45.7]	0.034
Operating Time, h	2.89 [0.15-6.84]	2.49 [0.25-10.02]	0.137
Intraoperative Blood Transfusion, *n* (%)	222 (9.9%)	372 (16.5%)	< 0.001
Intraoperative Hypotension (IH), *n* (%)	198 (8.8%)	39 (1.7%)	0.468
Surgical Position	0.8 [0-3]	0.3 [0-3]	< 0.001
Surgical Dressing, *n* (%)	1,888 (83.8%)	1,168 (51.9%)	< 0.001
Dressing Site, *n* (%)	364 (16.2%)	172 (7.6%)	< 0.001
Anesthesia Grade	2.8 [1-5]	3.2 [2-5]	< 0.001
Anesthesia Method, *n* (%)	2,145 (95.2%)	1,466 (65.1%)	< 0.001
Oxygen Saturation (SpO2), %	97 [76-100]	97 [65-100]	0.151
Self-Care Ability Grade	2.9 [1-3]	2.9 [2-3]	0.001
Blood Glucose (BG)	7.9 [1.4-28.0]	7.9 [3.8-13.4]	0.969

Table 7. Confusion matrix for each model (TN/FP/FN/TP)

Model	RAW Data	SMOTE	Borderline-SMOTE
SVM	2814/0/158/0	1477/775/461/1791	1683/569/250/2002
LR	2802/12/151/7	1938/314/367/1885	1995/257/292/1960
RF	2812/2/157/1	2184/68/57/2195	2203/49/80/2172
ET	2813/1/157/1	2182/70/56/2196	2199/53/76/2176
KNN	2787/27/153/5	1767/485/28/2224	1909/343/38/2214
DT	2655/159/123/35	2059/193/114/2138	2067/185/113/2139
XGBoost	2780/34/147/11	2195/57/83/2169	2189/63/86/2166
Model	SMOTE+ENN	ADASYN	GAN
SVM	1477/775/461/1791	1415/837/421/1831	2252/0/341/1911
LR	1938/314/367/1885	1916/336/369/1883	2237/15/302/1950
RF	2184/68/57/2195	2168/84/76/2176	2242/10/83/2169
ET	2182/70/56/2196	2147/105/89/2163	2238/14/87/2165
KNN	1767/485/28/2224	1734/518/21/2231	2218/34/75/2177
DT	2059/193/114/2138	2019/233/149/2103	2102/150/124/2128
XGBoost	2195/57/83/2169	2181/71/84/2168	2192/60/89/2163

all ML models improvedsignificantly after using SMOTE and its variants (*e.g.*, SMOTE+ENN, Borderline-SMOTE). For example, XGBoost improved its AUC value from 0.800 to 0.996 after using SMOTE, showing that methods of data enhancement play an important role in improving the model's ability to recognize a small number of classes.

Figure 4 shows the ROC analyses of the ML models under different conditions: (A) based on raw data, (B) based on SMOTE, (C) based on SMOTE+ENN, (D) based on Borderline-SMOTE, (E) based on GAN, and (F) based on ADASYN. These curves demonstrate the models' performance under various thresholds, with higher AUC values indicating better classification performance. The results clearly indicate that the models using SMOTE and its variants hadsignificantly higher AUC values compared to those using raw data, highlighting the effectiveness of these techniques in addressing a data imbalance.

When dealing with imbalanced data, SMOTE generates minority class samples by interpolation, which effectively increases sample diversity but may introduce

Table 8. Model performance comparison

Models	Precision	Recall	F1-score	Accuracy	AUC
Raw data					
SVM	0.000	0.000	0.000	0.947	0.565
LR	0.350	0.025	0.046	0.947	0.781
RF	0.200	0.012	0.023	0.946	0.795
ET	0.100	0.006	0.012	0.947	0.786
KNN	0.065	0.026	0.037	0.939	0.649
DT	0.151	0.165	0.152	0.898	0.573
XGBoost	0.290	0.082	0.124	0.938	0.800
SMOTE					
SVM	0.695	0.794	0.741	0.723	0.805
LR	0.846	0.816	0.829	0.834	0.919
RF	0.975	0.982	0.977	0.978	0.998
ET	0.974	0.974	0.973	0.974	0.996
KNN	0.832	0.991	0.904	0.895	0.961
DT	0.918	0.95	0.932	0.932	0.930
XGBoost	0.975	0.964	0.966	0.969	0.996
SMOTE+ENN					
SVM	0.701	0.786	0.741	0.725	0.802
LR	0.858	0.835	0.844	0.849	0.925
RF	0.974	0.978	0.975	0.976	0.997
ET	0.972	0.974	0.972	0.973	0.996
KNN	0.826	0.986	0.899	0.889	0.957
DT	0.917	0.946	0.930	0.930	0.932
XGBoost	0.982	0.961	0.967	0.971	0.993
Borderline-SMOTE					
SVM	0.782	0.887	0.831	0.820	0.879
LR	0.884	0.871	0.875	0.879	0.949
RF	0.978	0.964	0.968	0.971	0.992
ET	0.984	0.971	0.975	0.977	0.992
KNN	0.872	0.985	0.925	0.921	0.969
DT	0.927	0.955	0.938	0.940	0.938
XGBoost	0.974	0.960	0.963	0.968	0.989
ADASYN					
SVM	0.690	0.783	0.729	0.715	0.791
LR	0.834	0.818	0.824	0.827	0.930
RF	0.957	0.924	0.939	0.941	0.995
ET	0.949	0.907	0.927	0.930	0.993
KNN	0.813	0.934	0.869	0.859	0.957
DT	0.881	0.876	0.878	0.879	0.921
XGBoost	0.958	0.944	0.949	0.951	0.994
GAN					
SVM	1.000	0.930	0.958	0.965	0.933
LR	0.993	0.930	0.954	0.962	0.932
RF	0.995	0.930	0.955	0.963	0.939
ET	0.994	0.930	0.955	0.962	0.939
KNN	0.985	0.930	0.950	0.957	0.957
DT	0.933	0.930	0.922	0.928	0.927
XGBoost	0.975	0.930	0.945	0.952	0.951

noise near boundary samples. Nevertheless, SMOTE performs well in most models, and especially in XGBoost and Random Forest (RF), with an AUC value close to 1, indicating strong classification ability.SMOTE+ENN combines the undersampling techniques of SMOTE and ENN, which aim to remove noisy samples and further optimize the quality of the minority class samples. Although its performance is slightly inferior to SMOTE in most models, its performance is close in some models (*e.g.*, KNN), suggesting that it is effective in removing noise but may have sacrificed some of the sample diversity.Borderline-SMOTE focuses on generating samples near the category boundaries, which helps to improve the model's ability to discriminate between the boundary regions, but has limited performance improvement in most models and with high-dimensional data, the definition of boundary samples may not be clear enough, limiting its effectiveness.ADASYN is similar to SMOTE, but focuses more on the hard-to-learn regions of the minority class samples and improves the model performance through adaptive sampling.ADASYN performs well in some models (*e.g.*, XGBoost), but the overall performance is slightly lower than that of SMOTE, probably because the way it generates samples relies more on the local distribution of the minority class samples.GAN, as a state-of-the-art generative adversarial

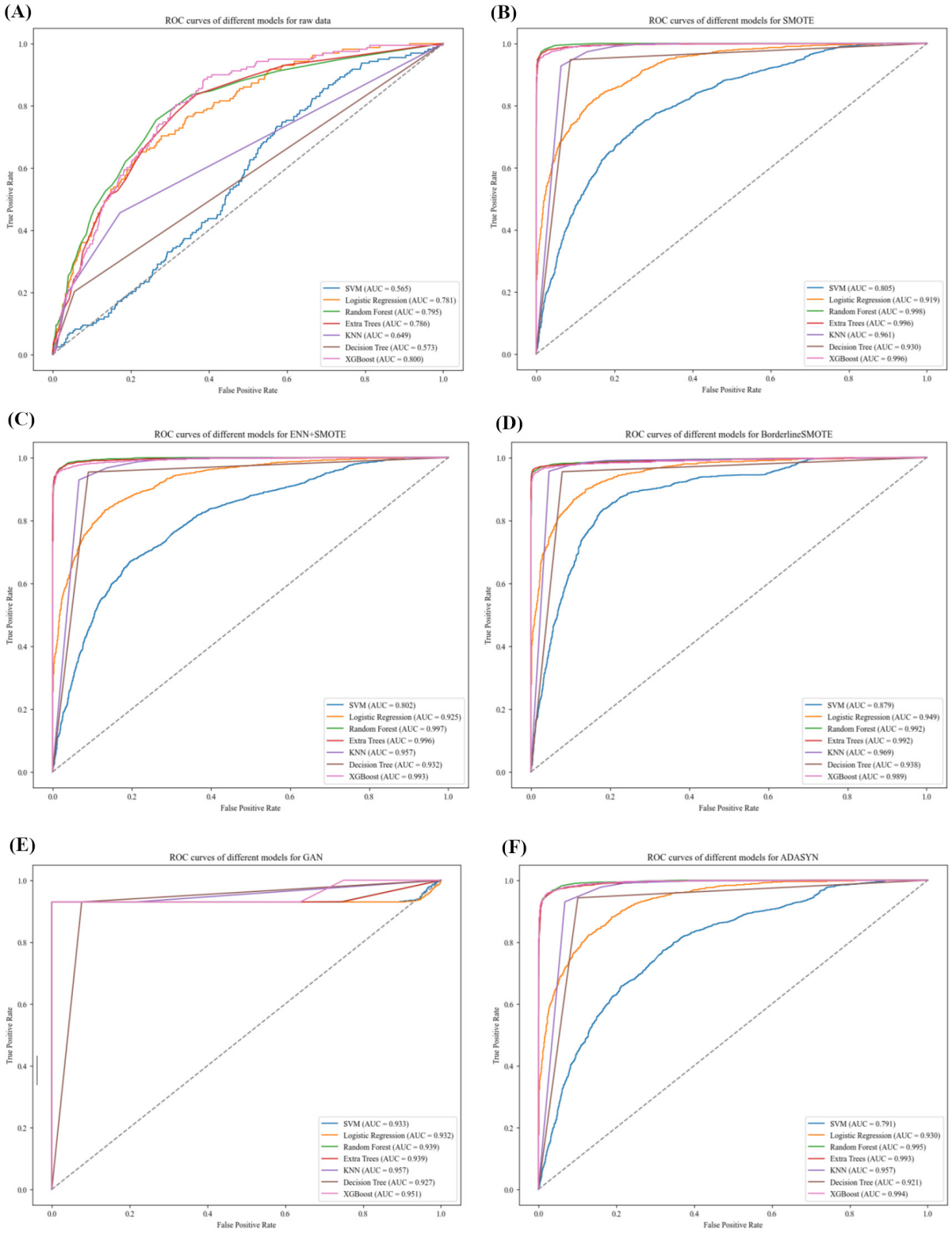

Figure 4. ROC analyses of applied machine learning models.(A) based on raw data, **(B)** based on SMOTE, **(C)** based on SMOTE+ENN, **(D)** based on Borderline-SMOTE, **(E)** based on GAN, **(F)** based on ADASYN.

network, generates high-quality minority class samples through the adversarial training of generators and discriminators. It performs well in some models, but its computational cost is high and it may face the problem of unstable training with high-dimensional data. In addition, the samples generated by GANs may be too close to the original samples, increasing the risk of overfitting.

In conclusion, methods of data enhancement, and especially SMOTE and its variants, have significant effects on improving the performance of models. In practical use, the most appropriate methods of data enhancement can be selected dependingto the specific

problems and models. These methods effectively improve the performance of the model withimbalanced datasets by increasing the number and diversity of samples, which improves the precision, recall, and overall classification performance of the model.

3.3. Validation and interpretability

Table 9 shows the five-fold cross-validation of SMOTE-based processed data.After comparing the performance of the model in 5-fold cross-validation and the original dataset, RF and XGBoost displayed the great stability and consistency in both methods of evaluation, with an AUC value close to 1, indicating its excellent generalization ability across different datasets.

In order to prevent possible overfitting after SMOTE processing and to test the generalization ability of the model, external validation of the constructed modelwas

Table 9. Fifty-fold cross-validation of the models

Models	Precision	Recall	F1-score	Accuracy	AUC
SVM	0.695 (0.664 - 0.721)	0.794 (0.772 - 0.823)	0.741 (0.729 - 0.763)	0.723 (0.700 - 0.744)	0.805 (0.788 - 0.831)
LR	0.846 (0.777 - 0.854)	0.816 (0.584 - 0.894)	0.829 (0.666 - 0.872)	0.834 (0.709 - 0.856)	0.919 (0.802 - 0.952)
RF	0.975 (0.925 - 0.990)	0.982 (0.857 - 1.000)	0.977 (0.958 - 0.987)	0.978 (0.922 - 0.992)	0.998 (0.997 - 1.000)
ET	0.974 (0.929 - 0.990)	0.974 (0.798 - 1.000)	0.973 (0.879 - 0.990)	0.974 (0.892 - 0.990)	0.996 (0.968 - 1.000)
KNN	0.832 (0.724 - 0.888)	0.991 (0.860 - 1.000)	0.904 (0.729 - 0.940)	0.895 (0.819 - 0.914)	0.961 (0.932 - 0.971)
DT	0.918 (0.868 - 0.949)	0.950 (0.766 - 0.970)	0.932 (0.848 - 0.949)	0.932 (0.863 - 0.949)	0.932 (0.863 - 0.949)
XGBoost	0.975 (0.938 - 0.985)	0.964 (0.660 - 1.000)	0.966 (0.790 - 0.987)	0.969 (0.825 - 0.995)	0.997 (0.975 - 1.000)

Table 10. External validation dataset distribution

Variables	Non-PI (*n*=277)	PI (*n*=277)	*p* value
Sex, *n* (%)	160 (57.7%)	192 (68.6%)	0.004
Age, years	53 [0-91]	73 [2-94]	< 0.001
Hypertension (HTN), *n* (%)	244 (88.1%)	173 (62.4%)	< 0.001
Hyperlipidemia, *n* (%)	276 (99.6%)	277 (100.0%)	0.751
Diabetes Mellitus (DM), *n* (%)	13 (4.7%)	8 (2.9%)	0.037
Cardiovascular Disease (CVD), *n* (%)	1 (0.4%)	0 (0%)	0.266
History of Malignant Tumor, *n* (%)	11 (4.0%)	2 (0.7%)	0.084
Smoking Status, *n* (%)	28 (10.1%)	4 (1.4%)	0.067
Alcohol Consumption Status, *n* (%)	1 (0.4%)	0 (0%)	0.481
Body Temperature, °C	36.6 [35.9-38.4]	36.5 [36.0-39.7]	0.599
Pulse Rate (PR), bpm	88 [38-170]	82 [53-150]	0.044
Respiratory Rate (RR), bpm	20 [11-70]	18 [12-32]	0.058
Diastolic Blood Pressure (DBP), mmHg	75 [30-113]	81 [56-120]	0.012
Systolic Blood Pressure (SBP), mmHg	123 [57-190]	133 [86-194]	0.133
Body Mass Index (BMI), kg/m²	21.5 [11.1-49]	22.7 [13.9-29.9]	0.090
Serum Albumin, g/L	37.1 [18.6-34.3]	33.1 [20.0-42.8]	0.401
Operating Time, h	2.89 [0.17-6.43]	2.78 [0.37-9.71]	< 0.001
Intraoperative Blood Transfusion, *n* (%)	256 (92.4%)	230 (83.0%)	0.004
Intraoperative Hypotension (IH), *n* (%)	248 (89.5%)	270 (97.5%)	0.423
Surgical Position	0.8 [0-3]	0.3 [0-3]	< 0.001
Surgical Dressing, *n* (%)	233 (84.1%)	156 (56.3%)	< 0.001
Dressing Site, *n* (%)	44 (15.9%)	15 (5.4%)	< 0.001
Anesthesia Grade	2.8 [1-5]	3.1 [2-5]	0.541
Anesthesia Method, *n* (%)	263 (94.9%)	195 (70.4%)	0.001
Oxygen Saturation (SpO2), %	97 [76-100]	96 [67-100]	0.442
Self-Care Ability Grade	2.9 [2-3]	2.9 [2-3]	0.206
Blood Glucose (BG)	7.9 [2.6-23.0]	7.9 [3.8-12.9]	0.655

Table 11. Performance of each model under external validation

Models	Precision	Recall	F1-score	Accuracy	AUC
SVM	0.731	0.733	0.731	0.731	0.731
LR	0.825	0.829	0.825	0.824	0.825
RF	0.939	0.945	0.939	0.938	0.939
ET	0.940	0.947	0.940	0.940	0.940
KNN	0.872	0.872	0.872	0.872	0.872
DT	0.942	0.948	0.942	0.942	0.942
XGBoost	0.977	0.978	0.977	0.977	0.977

attempted, but since the original amount of data was very small and had already been internally validated, cutting the data inside for external validation would have affected the performance of the original model, so the missing values that had been excluded from the original set of 9-14 data strips were used on a 1:1 basis. Positive and negative sampleswere selected in the order of missing values, and the missing values were added according to the data filling method in the previous section to serve as the external validation set. The data distribution of the external validation set is shown in Table 10, and the external validation results are shown in Table 11.

Based onthe external validation data, the precision, recall, F1 score, accuracy and AUC values of all models are very close to each other, indicatinga high degree of consistency in the performance of the models on the new dataset. The XGBoost model performed best in external validation, with a precision, recall, F1 score, accuracy and AUC value of 0.977, which is close to perfect, indicating excellent generalizability and prediction performance.

Combining the results of internal cross-validation and external validation, XGBoost, Random Forest (RF), and Extra Trees (ET) performed the best in terms of performance and stability. These models not only displayed low variability and high stability in internal cross-validation but also exhibited extremely high accuracy and AUC values in external validation, indicating their excellent generalizability. Especially, XGBoost, with its near-perfect external validation results, is the best choice among all models.

In response to the pervasive black-box problem of ML, SHAP (SHapley Additive exPlanations)has been introduced to increase the interpretability of the model. The scatterplot of SHAP values reveals the extent to which different features contribute to the predicted results of a ML model. Each point in the graph represents the SHAP value of a sample, which measures the contribution of a particular feature to the model output. The color gradient goes from blue to red, representing low to high feature values, respectively. Figure 5 shows that Anesthesia Grade has a significant effect on the model output. A high Anesthesia Grade (red points) is generally located on the right side of the graph, which indicates that it tends to increase the predictive value of the model when the Anesthesia Grade is high. Conversely, low anesthesia levels (blue points) tend to decrease the predicted value of the model, and most of these points are located on the left side of the graph. Age is also a key factor that has a broad impact on model predictions. Older people (red dots) tend to have

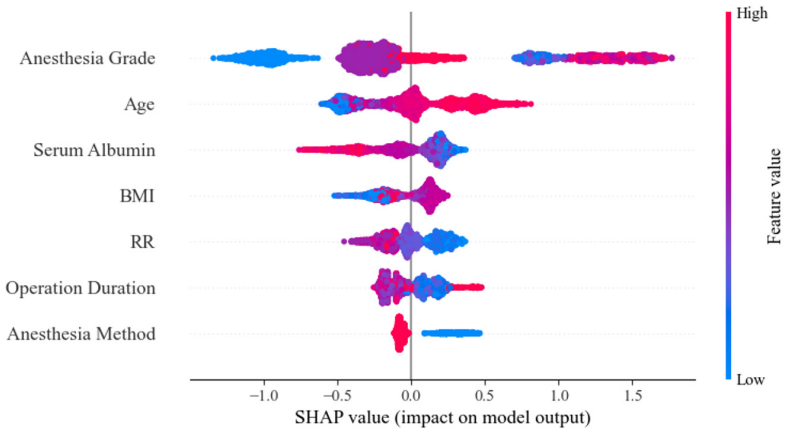

Figure 5. SHAP Summary plot of key factors.

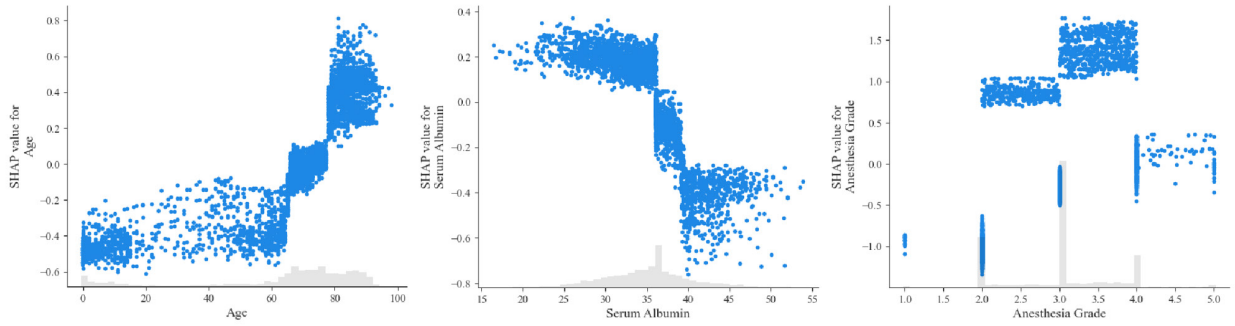

Figure 6. SHAP dependency plot.

positive SHAP values, implying that an increase in age may improve the model's predictions. In contrast, SHAP values for younger people (blue dots) tend to be negative, suggesting that a younger age may decrease model predictions. Serum albumin levels also had a significant impact on model predictions. Samples with high serum albumin levels (red dots) tend to have negative SHAP values in the graph, which could mean that higher serum albumin levels are associated with lower predictions in the model. Low serum albumin levels (blue dots), in contrast, are associated with positive SHAP values, suggesting that lower serum albumin levels may improve the predictive value of the model. The corresponding SHAP values when the variable of interest is a particular value are shown in Figure 6.

4. Discussion

The SMOTE algorithm yielded significant results whendealing with a data imbalance, but there are some potential limitations and risks. First, SMOTE may introduce noise, and especially when noise or outliers are present in a few class samples, and the synthesized samples may also contain that noise, affecting the model performance. Second, SMOTE is sensitive to the choice of parameter k (number of nearest neighbors), and improper values for k may lead to overfitting or the introduction of excessively noisy data. In addition, SMOTE is computationally expensive, and especially when dealing with large-scale datasets, and calculating the k nearest neighbors can be very time-consuming. More importantly, SMOTE increases the number of samples in a few classes by synthesizing samples that may be too close to the original samples, increasing the risk of overfitting and reducing the generalizability of the model. Finally, SMOTE may introduce noise or produce unrealistic data points when generating new samples near the boundary samples, affecting the classification effectiveness of the model.

SMOTEhas been widely used to address the issue of imbalanced class distribution in various ML applications. Sáez *et al.* introduced SMOTE-IPF, a re-sampling method with filtering, to tackle the problem of noisy and borderline examples in imbalanced classification(*21*). Rastogi *et al.* focused on implementing SMOTE in a distributed environment under Spark, highlighting the importance of applyingSMOTE to big data classification(*22*). Bao *et al.*integrated SMOTE with KNN and long short-term memory networks (LSTMs) to detect anomalies in high-dimensional and imbalanced data(*23*). Hemalatha *et al.* proposed FG-SMOTE, a fuzzy-based Gaussian synthetic minority oversampling algorithm, to handle imbalanced data and improve classifier performance(*24*). However, that study identified limitations such as the need to apply FG-SMOTE to multiclass imbalanced datasets and to evaluate theproblem of imbalancein a distributed environment. Mukherjee *et al.* introduced SMOTE-ENC, a novel SMOTE-based method for generating synthetic data with both nominal and continuous features(*25*). That study found that SMOTE-ENC outperformed SMOTE-NC in datasets with a substantial number of nominal features and associations between categorical features and the target class. Xia *et al.* proposed GBSMOTE, a sampling method based on granular-ball computing and SMOTE, to address the limitations of SMOTE such as noisy generated samples and boundary blurring(*26*). In the context of specific applications, Ismail *etal.* combined oversampling and undersampling techniques in SMOTE-RUS to classify imbalanced autism spectrum disorder datasets effectively(*27*). Nazarudin *et al.*usedsynthetic data generation techniques, including SMOTE and GAN-SMOTE, to train ML models to predictTenaga Nasional Berhad stock price movements(*28*). Overall, SMOTE has been a valuable tool in addressing a class imbalance, but studies have identified its limitations such as noisy samples, boundary blurring, and challenges in handling multiclass datasets and distributed environments. Future research may focus on enhancing SMOTE algorithms to overcome these limitations and improve their effectiveness in various applications.

To address these limitations and risks, future work can explore several directions. First, improved versions of SMOTE, such as Borderline-SMOTE or ADASYN, can be investigated and developed to improve the performance and stability of the algorithm through different strategies ofselecting the original samples used for generating new samples or adjusting the way in which new samples are generated. Second, the SMOTE algorithm can be used in conjunction with other techniques (*e.g.*, undersampling and integrated learning) to further improve the performance of the model. For example, the SMOTE algorithm can be used to oversample a small number of classes first, and then integrated learning methods can be used to train multiple models and obtain the final prediction results by voting or averaging. In addition, suitable evaluation metrics need to be used to assess the performance of the models, and especially withimbalanced datasets, where metrics such as recall and F1 scores often reflect the actual performance of the models better than accuracy. New learning algorithms designed specifically for imbalanced data can also be developedto improve the recognition of minority classes by adjusting sample weights or other mechanisms without increasing the number of samples. Finally, with the advent of the big data era, the useof SMOTEin big data environments can beexploredto address the challenges posed by the expanded size of data, such as computational efficiency and storage issues, is also an important direction for future work. Through these efforts, we can address the problem of a data imbalance more effectively and improve the predictive performance and generalization ability of the model.

This study had several limitations.First,the total

number of samples is still somewhat small relative to ML, so the model performance after SMOTE is bound to have overfitting to a certain extent. With the subsequent supplementation of the external validation set, there is also a certain amount of contamination of the training data. Second, based on data from only one hospital, the population is affected by the geographic area and may not necessarily be generalizable to other geographic areas.Further research will be conducted based on these issues in conjunction with multiple hospitals.

5. Conclusion

This study underscores the significance of usingML models to address the challenge of data imbalances in the prediction of perioperative PIs. The integration of synthetic minority oversampling techniques, and particularly SMOTE, with ML algorithms has been found to markedly enhance predictive accuracy, and especially in scenarios with few positive samples. The useof SMOTE and its variants, such as SMOTE+ENN and Borderline-SMOTE, has been shown to bolster the model's capacity to recognize minority classes, leading to more nuanced predictive modeling for PIs in emergency patient populations.

Among the seven ML models assessed, the combination of XGBoost with SMOTE emerged as the most effective, withan internally validated AUC of 0.996 and an externally validated AUC of 0.977. This result underscores the superior discriminative power of the XGBoost model when combined with SMOTE, outperforming other models across various metrics including precision, recall, F1 score, and accuracy. This study not only highlights the clinical utility of ML models augmented with SMOTE technology in predicting PIs but also underscores the importance of controlling a data imbalance toenhance the predictive value of ML models in healthcare settings. The findings suggest that the synergy of SMOTE with ML algorithms presents a viable strategy for mitigating the limitations of conventional risk assessment tools and dealing with the inherent data imbalances present in healthcare data. Future research is warranted to refine SMOTE techniques, explore their integration with other methodologies, and develop novel algorithms tailored for imbalanced datasets, thereby improving the reliability and accuracy of ML models in healthcare.

Acknowledgements

The authorswould like to sincerely thank the hospital's clinical research center for their invaluable support and assistance in database maintenance and data collection.

Funding: This work was financially supported by Shanghai Science and Technology Innovation Action Plan (23410761400), Xinhua Hospital CRU (21XHDB03), the Ministry of Education of China Humanities and Social Sciences Youth Fund Project (22YJC790189), Shanghai Key Laboratory of Urban Design and Urban Science Open Topic Grants of NYU Shanghai (Grant No.2023YWZhou_LOUD), and the Cultivation Project of School of Intelligent Emergency Management of University of Shanghai for Science and Technology.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Bajer D, Zonć B, Dudjak M, Martinović G. Performance analysis of SMOTE-based oversampling techniques when dealing with data imbalance. 2019 International Conference on Systems, Signals and Image Processing (IWSSIP). 2019; 265-271. https://doi.org/10.1109/IWSSIP.2019.8787306.
2. Kumar V, Singh Lalotra G, Sasikala P, Singh Rajput D, Kaluri R, Lakshmanna K, Shorfuzzaman M, Alsufyani A, Uddin M. Addressing binary classification over class imbalanced clinical datasets using computationally intelligent techniques. Healthcare (Basel). 2022; 10:1293.
3. Ishaq A, Sadiq S, Umer M, Ullah S, Mirjalili S, Rupapara V, Nappi M. Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE Access. 2021; 9:39707-39716.
4. van den Goorbergh R, van Smeden M, Timmerman D, Van Calster B. The harm of class imbalance corrections for risk prediction models: Illustration and simulation using logistic regression. J Am Med Inform Assoc. 2022; 29:1525-1534.
5. Ridwan AM, Setyawan GD. Perbandingan berbagai model machine learning untuk mendeteksi diabetes. Teknokon. 2023; 6:127-122.
6. Javid SA, Feghhi MM. Early diagnosis of Alzheimer's disease from MRI images with deep learning model. 2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP). 2024; 1-5. https://ieeexplore.ieee.org/document/10475240.
7. Kottner J, Cuddigan J, Carville K, Balzer K, Berlowitz D, Law S, Litchford M, Mitchell P, Moore Z, Pittman J, Sigaudo-Roussel D, Yee CY, Haesler E. Prevention and treatment of pressure ulcers/injuries: The protocol for the second update of the international Clinical Practice Guideline 2019. J Tissue Viability. 2019; 28:51-58.
8. Anderson C, Bekele Z, Qiu Y, Tschannen D, Dinov ID. Modeling and prediction of pressure injury in hospitalized patients using artificial intelligence. Med Inform Decis Mak. 2021; 21:253.
9. Vera-Salmerón E, Domínguez-Nogueira C, Sáez JA, Romero-Béjar JL, Mota-Romero E. Differentiating pressure ulcer risk levels through interpretable classification models based on readily measurable indicators. Healthcare (Basel). 2024; 12:913.
10. Chumachenko D, Yakovlev S. Artificial intelligence algorithms for healthcare. Algorithms. 2024; 17:105.
11. Chawla LV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Art Intel Res. 2002; 16:321-357.
12. He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive

synthetic sampling approach for imbalanced learning. IEEE International Joint Conference on Neural Networks. 2008; 1322-1328. https://ieeexplore.ieee.org/document/4633969.
13. Nizam-Ozogur H, Orman Z. A heuristic-based hybrid sampling method using a combination of SMOTE and ENN for imbalanced health data. Expert Systems. 2024; e13596.
14. Ma H, Song M, Zhu Y. Improved Borderline-SMOTE oversampling method for boundary classification. J Nanjing Univ (Natural Science). 2023; 59:1003-1012. (in Chinese)
15. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. AdvanNeural Info Process Syst. 2014; 2:2672-2680.
16. Araveeporn A, Wanitjirattikal W. Comparison of machine learning methods for binary classification of multicollinearity data. 2024 International Conference on Machine Learning and Cybernetics (ICMLC). IEEE. 2024; 44-49. https://doi.org/10.1145/3686592.3686600.
17. Kyriazos T, Poga M. Application of machine learning models in social sciences: Managing nonlinear relationships. Encyclopedia. 2024; 4:1790-1805.
18. Mujahid M, Kına E, Rustam F, Gracia Villar M, Silva Alvarado E, De La Torre Diez I, Ashraf I. Data oversampling and imbalanced datasets: An investigation of performance for machine learning and feature engineering. J Big Data. 2024; 11:87.
19. Halder RK, Uddin MN, Uddin MA, Aryal S, Khraisat A. Enhancing K-nearest neighbor algorithm: A comprehensive review and performance analysis of modifications. J Big Data. 2024; 11:113.
20. Azevedo BF, Rocha AMAC, Pereira AI. Hybrid approaches to optimization and machine learning methods: a systematic literature review. Mach Learn. 2024; 113:4055-4097.
21. Sáez JA, Luengo J, Stefanowski J, Herrera F. SMOTE-IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering. InfoSci. 2015; 291:184-203.
22. Rastogi, A. K., Narang, N., & Siddiqui, Z. A. Imbalanced big data classification: A distributed implementation of SMOTE. Proceedings of the Workshop Program of the 19th International Conference on Distributed Computing and Networking. 2018; 14:1-6. https://doi.org/10.1145/3170521.3170535.
23. Bao F, Wu Y, Li Z, Li Y, Liu L, Chen G. Effect improved for high-dimensional and imbalanced data anomaly detection model based on KNN-SMOTE-LSTM. Complexity. 2020; 2020:1-17.
24. Hemalatha P, Amalanathan GM. FG-SMOTE: Fuzzy-based Gaussian synthetic minority oversampling with deep belief networks classifier for skewed class distribution. Interntl J Intel Computing Cybernet. 2021; 14:270-287.
25. Mukherjee M, Khushi M. SMOTE-ENC: A novel SMOTE-based method to generate synthetic data for nominal and continuous features. Appl Syst Innov. 2021; 4:18.
26. Xia T, Shao Y, Xia S, Xiong Y, Lian X, Ling W. GBSMOTE: A robust sampling method based on Granular-ball Computing and SMOTE for class imbalance. Proceedings of the 2023 8th International Conference on Pattern Recognition and Artificial Intelligence. 2023: 19-24. https://dl.acm.org/doi/10.1145/3594300.3594304.
27. Ismail E, Gad WK, Hashem M. SMOTE-RUS: Combined oversampling and undersampling technique to classify the imbalanced Autism Spectrum Disorder dataset. Interntl J Intel Computing Cybernet. 2023; 23:83-94.
28. Mohd Nazarudin NAS, Mohd Ariffin NH, Maskat R. Leveraging on synthetic data generation techniques to train machine learning models for Tenaga Nasional Berhad stock price movement prediction. Internatl Arab J Info Tech. 2024; 21:483-494.

Received January 11, 2025; Revised February 25, 2025; Accepted March 26, 2025.

*Address correspondence to:
Liping Jiang and Ping Liu, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, No. 1665 Kongjiang Road, Yangpu District, Shanghai 200092, China.
E-mail: jiangliping@xinhuamed.com.cn (LJ), 13868311990@163.com (PL)

Released online in J-STAGE as advance publication April 15, 2025.