Communications in Mathematics and Applications
Vol. 16, No. 3, pp. 929–948, 2025
ISSN 0975-8607 (online); 0976-5905 (print)
Published by RGN Publications
DOI: 10.26713/cma.v16i3.3603

http://www.rgnpublications.com

Research Article

Optimizing Surgical Site Infection Prediction Performance Through a Combined Approach of SMOTE-ENN, SMOTE-Tomek, and Wrapper Feature Selection

Salha Al-Ahmari

Department of Informatics and Computer Systems, Faculty of Computer Science, King Khalid University, Abha, Saudi Arabia
ssalahmry@kku.edu.sa

Received: September 7, 2024 **Revised:** November 15, 2024 **Accepted:** December 20, 2024

Abstract. Risk stratification can be enhanced by assessing preoperative risk factors, which are essential for guiding surgical decision-making. *Machine learning* (ML) and AI-based expert systems can predict, detect, and monitor *surgical site infections* (SSIs) using data from *electronic healthcare records* (EHRs). The predictive capability of classification algorithms is impaired by class imbalance and depends heavily on the quality of features in a dataset, which may include irrelevant or redundant information. The primary goal of feature selection is to remove such features to improve classification accuracy. This study utilized a dataset of 64,793 surgical records, each featuring 25 variables, to evaluate six machine learning classification methods: *logistic regression* (LR), *K-Nearest Neighbors* (KNN), *decision tree classifier* (DTC), *support vector machine* (SVM), *Gaussian Naive Bayes* (GNB), and *artificial neural network* (ANN). These techniques, intended to enhance classification accuracy, frequently prioritize the majority class in an imbalanced dataset, hence distorting the accuracy metric. To address the imbalanced classification problem in SSI prediction. The study applied the hybrid sampling techniques SMOTEMOTE-ENN and SMOTE-Tomek, combined with wrapper feature selection, to improve the model performance. The wrapper feature selection model optimizes the feature set by reducing the number of features while simultaneously enhancing the classification accuracy. Findings represented that the combined sampling of *Edited Nearest Neighbors and Synthetic Minority Oversampling Technique* (SMOTE-ENN) with wrapper feature selection outperformed the SMOTE-Tomek in all performance metrics (KNN; AUC: 98, recall: 61, precision: 80, F1-score: 69, and accuracy: 98) and (DTC; AUC: 93, recall: 60, precision: 76, F1-score: 67, and accuracy: 98), particularly for minority class. When paired with SMOTE-ENN sampling and optimized through wrapper feature selection, the SSI prediction accuracy and AUC were significantly enhanced. The proposed method effectively mitigates the overfitting and underfitting issues, though the wrapper method is computationally intensive, which results in longer training times.

Keywords. Imbalanced classification, Combined sampling, Surgical site infection, Machine learning, Wrapper feature selection

Mathematics Subject Classification (2020). 68T09, 62P10, 92C50

1. Introduction

Surgical site infections (SSIs) are a significant and severe postoperative complication, with a frequency ranging from 0.2% to 16.1%. (Petrosyan *et al*. [35]). SSIs can cause catastrophic consequences like prolonged hospitalisation, osteomyelitis, instrumentation failure, increasing patient suffering, readmissions, and increased hospital costs (Xiong *et al*. [47]). Therefore, risk stratification can be improved by assessing preoperative risk factors, which help guide surgical decision-making. ML and expert systems based on AI can predict, detect, and monitor SSIs based on the data stored in *electronic healthcare records* (EHRs) (Samareh *et al*. [36]). In the recent era, artificial intelligence has played a significant role in medicine for disease prediction, image diagnosis, and confidentiality management (Xiong *et al*. [47]).

In data mining, classification aims to extract knowledge from huge datasets where the method of complete classification is applicable for balanced datasets, whereas partial classification phenomena are applied to imbalanced datasets. A fundamental problem in ML is the imbalanced classification problem, which is resolved at three proposed levels: the hybrid, the data, and the algorithm (Mahani and Ali [28]). In combination with large data, ML is applied to predict SSIs where category 'imbalance' influences data structures (Xiong *et al*. [47]). Nonetheless, imbalanced datasets, missing data, and interpretability issues may influence the usability of the ML predictive models. The feature selection and sampling processes can be useful for dealing with imbalanced classification problems as these techniques efficiently select subsets of vital features, improving predictive accuracy and precision (Al Mamlook *et al*. [2]). The second approach is feature selection, which involves algorithms categorized primarily into filtering, embedded, and wrapper algorithms. The filtering algorithms depend on the internal architecture of the features to select relevant features. Typically, methods that measure information gain are employed for categorical feature filtering, whereas methods that assess correlation are used for numerical feature filtering, as noted in reference (Chandrashekar and Sahin *et al*. [5]). Although computationally efficient, interactions between features can compromise feature selection effectiveness. The embedded feature selection method is integrated into the model-building process and is highly efficient and capable of managing interactions between various features. LASSO-based algorithms are frequently used in linear combination models, whereas tree-based algorithms are applied in nonlinear combination models (Khalid *et al*. [18]). Wrapper algorithms implement an iterative strategy to assess the suitability of a feature subset to enhance model performance on a specified dataset. The process was continued until optimal performance was achieved (Visalakshi and Radha [43]).

Conventional machine learning methods often fall short when there is a significant imbalance because they were primarily designed under the assumption that classes have roughly equal numbers (Haixiang *et al*. [15], and Nieto-del-Amor *et al*. [30]). When sufficient data is available, machine learning algorithms may be able to learn from skewed distributions. Comparable conditional distributions among classes have been shown to significantly affect model performance, as noted in the study by Luque *et al*. [26].

Xiong *et al*. [47] highlighted that the *synthetic minority oversampling technique* (SMOTE) is a well-established method for addressing imbalanced data to optimize *machine learning* (ML) predictive models. Clinicians believe that high-risk patients with SSI can be identified with prediction models developed by combining ML models with SMOTE for preoperative management and optimized patient selection. Consequently, early preventive intervention can be vital for reducing serious factors causing the prevalence of SSI (Xiong *et al*. [48]). Furthermore, Xu *et al*. [48] noted that the imbalanced classification problem with traditional classification algorithms can be resolved when the patient samples are used in large numbers. Besides, a commonly used oversampling method that enhances random oversampling is called SMOTE, which recommended for data resampling under various ML techniques, preferably a decision tree and random forest classifier (Xu *et al*. [48]). However, while replicating the minority cases in oversampling, data may result in overfitting (Haixiang *et al*. [15], and Yap *et al*. [51]), some earlier researchers have proposed that random undersampling is preferable to random oversampling (Junsomboon and Phienthrakul [17]). The SMOTE can be utilised to prevent overfitting for oversampling (Xiong *et al*. [47], and Yap *et al*. [51]). SMOTE creates a new instance of the minority label at an arbitrary spot along the line connecting existing examples and their nearest neighbor for each instance of the minority class (Junsomboon and Phienthrakul [17]).

According to previous research, the imbalanced classification problem in prediction of SSIs has been resolved by implying under sampling and oversampling techniques (Xiong *et al*. [47], and Xu *et al*. [48]). However, a gap has been observed in SSI prediction using combined or hybrid sampling techniques. The accuracy of the classification model is strongly dependent on the quality of the features in the dataset, which may contain unnecessary or redundant information. The primary goal of feature selection is to remove such features to improve classification accuracy. The wrapper feature selection model optimizes the feature set by reducing the number of features while simultaneously enhancing classification accuracy.

Therefore, the current research aims to tackle the imbalanced classification problem in SSI prediction by integrating SMOTE-ENN and SMOTE-Tomek sampling techniques, optimized through wrapper feature selection. To evaluate model performance, six machine learning classifiers were employed: DTC, GNB, KNN, LR, ANN, and SVM.

The remaining parts of this work are structured as follows. Section 2 highlights the related work. Section 3 delineates the methodology employed in this paper, including the data preprocessing procedures, modeling approach, and implementation hyperparameters. Section 4 presents the evaluation metrics used to report the findings. Section 5 presents the results of the SSI prediction models, focusing on the impact of employing combined resampling techniques and model-based feature selection. Section 6 discusses the results, and Section 7 concludes the study.

2. Related Work

According to Bartz-Kurycki *et al*. [4], resampling approaches are essential for improving the effectiveness of models in high-dimensional data analysis. The selection of sampling method can profoundly influence model performance, particularly when the variable count is extensive.These sampling techniques are mainly categorized as oversampling and under sampling, used for improving the model performance involving imbalanced data. For example, in the SMOTE oversampling approach, new samples are generated from the minority class to enhance imbalance data classification and choose a predicted cut point probability that optimizes the AUC curve. ML may clarify complicated relationships for huge datasets and shed light on crucial aspects (Bartz-Kurycki *et al*. [4], and Colborn *et al*. [8]).

Furthermore, in a recent study conducted by Lu *et al*. [25], machine learning methods were employed for the purposes of predicting (SSIs) post-posterior cervical surgery. The study assesses the performance of eight distinct models with the assessment criteria encompassing AUC, sensitivity, specificity, accuracy, true positive and true negative predictive value. It also became clear that both Random Forest and Gradient Boosting models had satisfactory non-SSI predictive capabilities and accuracies above the rest of the classifiers (Lu *et al*. [25]). Yang *et* *al*. [50] contributed to the research by providing a hybrid sampling technique that integrates SMOTE and ENN. ENN was employed to undersample the oversampled data in the missed abortion and diabetes datasets. Subsequently, RF was applied, and the results indicated that RF attained superior performance on both datasets (Yang *et al*. [50]). Lu *et al*. [25] also concluded that despite exhibiting a higher false positive rate, the KNN algorithm holds potential for early SSI identification. Notably, the study emphasized the pivotal role of serum albumin levels as a crucial indicator; it underscored the importance of preoperative nutrition support among patients with low albumin levels. Moreover, changes in BMI were not very effective for predicting SSI and it was found out that subcutaneous fat thickness was more effective (Lu *et al*. [25]).

A machine learning predictive model can aid in identifying patients at elevated risk for *surgical site infections* (SSI). Ying *et al*. [52] conducted a study involving 351 patients who underwent *open reduction and internal fixation* (ORIF) between January 2018 and October 2022. The patients' characteristics, surgery information, and laboratory values were included. The research paper utilized ten different machine learning algorithms to create the prediction model, indicating that ET, LR, and RF are the top performers. The performance of the models illustrated in the figure indicates that the proposed ET model is valid, attaining the highest AUC scores of 0.85 and 0.86. The decision curve and calibration curve illustrated the model's clinical efficacy. The study ultimately validated that all scores are strongly associated with patient conditions, and the multivariate prediction model demonstrates sufficient accuracy regarding its clinical predictive capability (Ying *et al*. [52]).

Few studies using national databases revealed two intriguing points. First, the results proposed by previous studies that compared ML models indicated the best performance achieved by *Random Forest* (RF) and AdaBoost classifiers (Petrosyan *et al*. [34]) and (Xiong *et al*. [47]) with 89% and 87% AUC, respectively. However, they have limited clinical implications because of the low predictability with PPV (23% and 34%). Second, sampling methods like SMOTE were applied to balance the dataset (Xiong *et al*. [47]), whereas other studies did not deal with imbalanced classification problems that affected those prediction models' performance. The number of features generated from national databases is usually high; some are irrelevant. Therefore, the feature selection step should ensure the key and important features from the original features set in the recent research. For instance, Petrosyan *et al*. [34] used stepwise feature selection before applying the RF model. Whereas other studies did not use feature selection methods except (Bartz-Kurycki *et al*. [4]) used previous literature and clinical experts to feed the models with only the set of important features.

A different study introduces a novel predictive model for SSI utilizing temporal data. The model utilizes maximum likelihood with L1 norm regularization in the penalized logistic regression analysis to detect SSIs from patients' blood test data prior to the operation date. The strategy integrates prior knowledge of blood test predictors, penalty factors commensurate with test costs, and a specific parameter for early termination designed to limit the amount of features. The fundamental C-reactive protein classifier yielded a mean AUC of 0.80, while the optimal full LASSO model, limited to 20 features, achieved a mean AUC of 0.96. The study evidenced that these findings imply the potential of the proposed models in enhancing the accuracy of identifying high-risk patients for SSI, thereby assisting the domain specialists in decision-making towards establishing novel guidelines for preoperative SSI prevention and monitoring (Kocbek *et al*. [20]).

Moreover, a recent study concluded that predictive model not only help surgeons to identify the potential factors for developing SSI but also help patients to be objectively aware of those factors and try to avoid the complications (Gutierrez-Naranjo *et al*. [13]) Additionally, various important metrics were calculated in the validation set to compare the result with previous literature, such as precision, accuracy, recall, F-score, and AUC (Al Mamlook *et al*. [2], Bartz-Kurycki *et al*. [4], Chen *et al*. [6], Song *et al*. [37], Xiong *et al*. [47]). The results from past researchers showed that accuracy rates are up to 89% by employing ML models to predict SSIs (Petrosyan *et al*. [35]). In 2024, Gutierrez-Naranjo *et al*. *et al*. [13] evaluated a number of clinical and demographic factors using four medical models: neural networks, boosted generalized linear models, naïve Bayes, and penalized discriminant analysis. The method attained a maximum score of 77% for AUC, exceeding Youdon's index of 62.5% and producing a Brier score between 5.1% and 5.6%. Nonetheless, it is also noteworthy that these ML models displayed low predictability, offering limited clinical implementations. Low predictability means the positive predictive value was 34%, and the high imbalance between the two classes may lead to an accuracy paradox (Luque *et al*. [26]). A high accuracy number does not always indicate a high-quality model since the model is biased towards the majority class and may obscure the findings (Kim *et al*. [19]). This occurrence is known as the accuracy paradox.

Further investigation is required to consider the diverse and conflicting results for predicting SSI using ML models, the imbalanced data situation, and the many features specified in the national dataset. The main goal of this paper is to accurately predict the possibility of SSI after different types of surgical procedures. Therefore, the current research, Confusion Matrix, Accuracy, Precision, Recall, F-score, and AUC evaluation measures are compared in the findings to discuss the following objectives:

(1) SMOTE-EEN improves model robustness by balancing class distribution and reducing noise, leading to more accurate predictions on imbalanced datasets.

(2) The wrapper feature selection method enhances model performance by retaining only the most relevant features, improving accuracy and efficiency.

(3) Combined Benefits: Using SMOTE-EEN and wrapper feature selection together optimizes dataset quality and feature relevance, resulting in high model performance and more reliable predictions specifically with KNN.

3. Method

The current study predicted SSIs after various surgical procedures. The researcher used a large dataset and six ML models to achieve the objective. Initially, the researchers cleaned the data, handled missing values, removed unnecessary features and outlier values, and implemented two sample approaches for combined sampling to improve the process. The study used a wrapper feature selection approach to feed the ML models with the most important feature set. This study intended to address the issue of imbalanced datasets by employing wrapper feature selection and a hybrid sampling strategy, hence improving model predictability via several assessment criteria. The structure displayed in Figure 1 directed the training as well as evaluation of the prediction models.

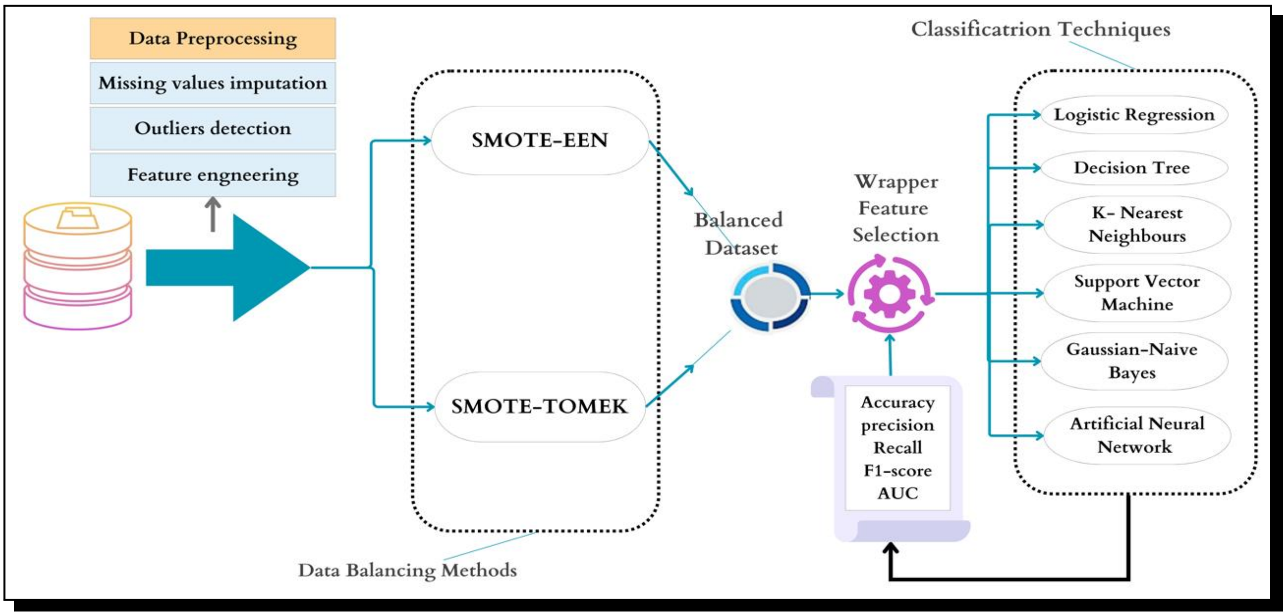

Figure 1. Flow diagram of the proposed methodology

3.1 Data Source

The dataset was acquired from the General Administration for Infection Control for Infection Control in Health Facilities at the Ministry of Health in the Kingdom of Saudi Arabia. The dataset was labelled with SSI and non-SSI cases. In this experiment, I collected 64,793 individual data points.

3.2 Data Pre-Processing

This study uses a raw dataset that contains many missing values and errors, necessitating cleaning and preprocessing before use. The data cleaning stage includes data transformation, outlier removal, and the management of missing values. I treated datasets differently to ensure they didn't miss the NaN values. For instance, I dropped rows or columns that had more than half of their values as NaN; otherwise, I filled in the mean, mode, or median, or used a constant.

Three concepts are involved in data transformation: feature engineering, aggregation, and normalization. According to Sun *et al*. [38], data normalization sets up the data such that each field and record is comparable. Consequently, data transformation in this study improved the coherence of entry types, leading to data cleaning, lead generation, segmentation, and quality improvement; for instance, 'CORE_TEMPERATURE', 'WEIGHT', and 'HEIGHT' normalized to float. Feature engineering constructs new features from existing ones, including 'cancer_fever_diabetics' from the pre_procedure_diagnosis feature, 'BMI' from 'height' and 'weight', and 'adult_child_infants' also from the 'pre_procedure_diagnosis' column. Data aggregation identifies common traits. For example, the World Health Organization has classified BMI into four groups (underweight, normal weight, pre-obesity, and obesity) (Weir and Jan [45]).

Finally, the data preprocessed in this study included 25 features, and records from 64,793 patients were considered. However, I assigned a value of '0' to patients without SSI and '1' to those with SSI. The dataset included 97.5% of patients classified as 'non-SSI' and 2.5% of patients classified as 'SSI'. I divided the dataset into a 70% training set and a test set with the remaining 30%.

3.3 Class Balancing Using Combined Method

The combined sampling method leverages the strengths of different sampling techniques to achieve better balance and improve the performance of classifiers (Kraiem *et al*. [22]). The combined method involves integrating two or more sampling methods, such as oversampling and undersampling, in a sequential or parallel manner (Tarimo *et al*. [39]). The sequential combination involves first applying one sampling method and then applying another method to the modified dataset. In contrast, parallel combinations apply multiple sampling methods simultaneously and combine the resulting datasets (Kraiem *et al*. [22]).

3.3.1 SMOTE-ENN

SMOTE-ENN is a sampling technique that integrates over-sampling and under-sampling through SMOTE and *edited nearest neighbors* (ENN); it enhances the minority class representation via SMOTE and subsequently employs ENN to eliminate noisy instances. This methodology mitigates the issue of imbalanced datasets by furnishing a more equitable and dependable training set for classifiers (Ghorbani and Ghousi [12]). In instances of class distribution imbalance, ML algorithms may demonstrate bias toward the majority class, rendering this approach beneficial (Kraiem *et al*. [22]). SMOTE-ENN initially use interpolation to augment the minority class, thereafter applying the ENN method to remove redundant samples. When integrated with machine learning algorithms, it produces class-balanced data that ultimately achieves the desired outcomes (Ghorbani and Ghousi [12]).

3.3.2 SMOTE-Tomek Links

Another combined method to address data imbalance in classification issues. The SMOTE-Tomek Link method integrates the Tomek Link and SMOTE techniques (Hairani *et al*. [14]). SMOTE generates synthetic instances to augment the minority class. Conversely, Tomek Link removes noise or samples that are proximate to each other within both the minority and majority classes (Tarimo *et al*. [39]). SMOTE-Tomek Link seeks to mitigate data imbalance by reducing noisy samples and augmenting the quantity of synthetic samples in the minority class.Upon finishing the class balancing technique, the resampled data was examined using a feature selection method to ascertain the most pertinent features and enhance the classifier's efficacy, hence refining the prediction of the SSI outcome. The integrated sampling method utilizes the advantages of many sample techniques to enhance balance and optimize classifier performance (Kraiem *et al*. [22]). The combined approach entails the integration of two or more sampling techniques, such as oversampling and under sampling, executed either sequentially or in tandem (Tarimo *et al*. [39]). The sequential combination involves first implementing one sampling technique, then applying another approach to the modified dataset. Conversely, parallel combinations utilize various sampling strategies concurrently and amalgamate the resultant datasets (Kraiem *et al*. [22]).

3.4 Wrapper Feature Selection

When ML chooses a limited number of features as inputs for the model, the data may not be sufficient for making predictions. The curse of dimensionality causes an increase in execution time and diminishes generalizability due to the presence of many features (Visalakshi and Radha [43]). Consequently, to ensure accurate forecasts, it is essential to select the most significant components that influence the outcomes (Chandrashekar and Sahin [5]). Wrapper feature selection is a method that determines the most pertinent features for a machine learning model

Table 1. Surgical site infection class distribution

Dataset	Class	Number of datapoints
Original distribution	Non-SSI SSI	0: 64,793 1: 1632
Training set	50103	0: 49040 1:1063
Test set	19655	0: 19086 1: 569
SMOT-EEN resampled	Non-SSI SSI	0: 50528 1: 47187
SMOTE-Tomek resampled	Non-SSI SSI	0: 44298 1: 8674

by assessing the performance of several feature subsets (Visalakshi and Radha [43]). In wrapper approaches, a search strategy incrementally incorporates or excludes information from a dataset. The predominant search strategies are forward, backward, and recursive feature removal. *Recursive feature elimination* (RFE) commences with all features in the training dataset and systematically eliminates features until the specified quantity is achieved, as illustrated in Figure 3. This research employed 5-folds cross-validation to determine the optimal amount of features. The results were shown on an elbow curve to make sure the performance metric was accurate and to avoid overfitting to the training data. The elbow point signifies the juncture at which metric enhancement begins to plateau, indicating that the addition of further features results in diminishing returns, as depicted in Figure 4. The elbow plot shows that

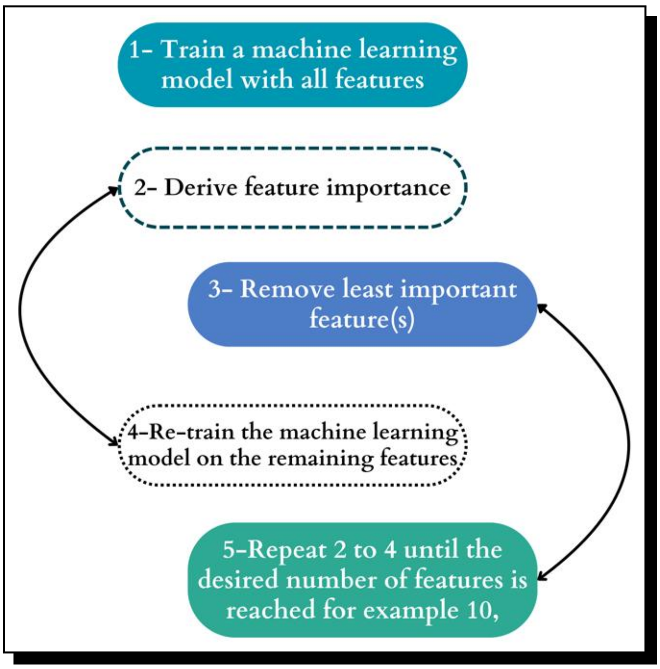

Figure 2. The process of *Recursive Feature Elimination* (RFE) [43]

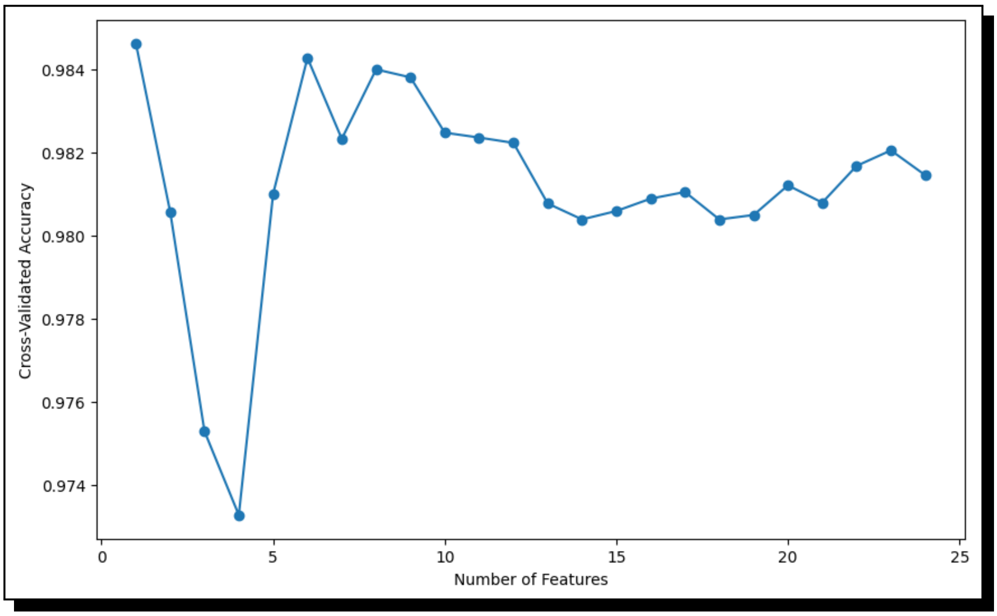

Figure 3. Elbow curve with optimal number of features

the curve typically levels off markedly after a pronounced initial ascent, marking the optimal amount of features. The performance improved to around five features. The performance seems to reach its zenith at around 10 features. Following the inclusion of 10 features, the recall measure experiences a minor decline and subsequently stabilizes at around 15 features. Ultimately, ten features emerge as the optimal selection according to this elbow plot, as performance attains its zenith before a subsequent fall.

3.5 Classification Algorithms and Hyperparameters

The aim of the current research was to examine the effectivity of various ML algorithms in the prediction of SSIs. LR, DTC, SVM, GNB, ANN, and KNN were used to examine the aim of the research and develop predictive models. However, the models' convergence was enhanced by using considerable sampling techniques for data preprocessing and balancing. The main combined sampling techniques and related techniques for resolving the imbalanced classification problem are presented in this paper. A wrapper feature selection technique was used by the researchers to determine the best combination that yielded the most effective prediction models. In this study, 30% and 70% of the testing and training datasets were selected. The PYTHON version 3.9.12 package was used to predict SSI (Pedregosa *et al*. [33]). For combined sampling, seven classifiers were used. The model was evaluated using the selected hyperparameters (Table 2).

3.5.1 Logistic Regression (LR)

This approach uses a supervised ML algorithm to address classification issues (Bartz-Kurycki *et al*. [4]). LR is adequately applied when there are categorical values in the target class to categorise and predict the disease (Ahmed *et al*. [1]). The statistical method is appropriate for analysing the relationship between a dichotomous or binary outcome (0/1) using a set of independent predictors (Das [9]).

Table 2. Classifiers and hyperparameters implemented for combined sampling in SSI prediction

Classifier	Hyperparameters
Logistic Regression	random_state=1, C=1.00, solver='lbfgs'
Decision Tree Classifier	random_state=1, min_samples_split=2, splitter='best'
K-Neighbors Classifier	K=[3,6,9]; weight options [uniform, distance], and metric (Minkowski)
Gaussian Naive Bayes	Default
Random Forest Classifier	n_estimators=[5000,7500,10000,12500,15000], split measured criteria=['gini','entropy']
Support Vector Classifier	kernel='linear', gamma=scale, and tolerance=0.001
Stochastic Gradient Descent Classifier	loss='log_loss'

3.5.2 Decision Tree Classifier (DTC)

This technique categorises instances according to feature values. Notably, in a DT, each node indicates a feature in an instance that needs to be categorised, and every branch provides a possible value for the node (Song *et al*. [37]). A best-suited variable is attributed at the root dividing the algorithm into parts to unmix the dataset. An iterative splitting occurs until the data is grouped into homogenous partitions (Ahmed *et al*. [1]).

3.5.3 K-Nearest Neighbours (KNN)

KNN is another well-liked classification method for its easy interpretation and quick computation (Nasteski [29]). It works well with the discrete target classes by calculating the distances of the point of a query for each of the instances to find the K minimum distances. KNN is an effortless algorithm in which data is grouped into coherent subsets or clusters, which categories the data depending on its similarity with the trained dataset assigning the input variable to the nearest neighbors (Taunk *et al*. [40]).

3.5.4 Support Vector Machine (SVM)

A supervised ML approach, SVM, aims to find an n-dimensional repeatable hyperplane that maximises the distance among the support vectors of two different class labels (Al Mamlook *et al*. [2], and Song *et al*. [37]). It is a classification-regression method developed for multiple classification problems applied simultaneously with feature selection (Lengua and Quiroz [23]).

3.5.5 Gaussian-Naive Bayes (GNB)

It is a classification algorithm that assigns a label to a class, increasing the posterior probability of an individual sample. It is applicable with the assumption of voxel contributions which conditionally are independent following the rules of Gaussian distribution. The decision rule of GNB is written by obeying the discriminant function for every class (Ontivero-Ortega *et al*. [31]). GNB assumes that each variable predicts the outcome attribute independently, such that the final prediction categorises the dependent variable into each group considering the highest probability of the respective set of items (Nasteski [29]).

3.5.6 Artificial Neural Networks (ANNs)

These advanced machine learning algorithms, known as ANNs, are inspired by the intricate architecture and operation of living neural networks in the human brain (Lydia and Francis [27]).

These networks consist of interconnected nodes, referred to as artificial neurons, organized in layers (Lu *et al*. [25]). The depth of an artificial neural network is determined by the quantity of hidden layers it possesses. Essentially, basic neural networks and deep learning models often comprise three or more layers. The depth of ANN is defined by the number of hidden layers it contains. Fundamentally, basic neural networks and deep learning models, with the latter often consisting of three or more layers (Lu *et al*. [25]).

4. Model Evaluation

The confusion matrix illustrates the predictive models' ability in properly categorizing data into actual classes. The confusion matrix, also known as the error matrix, is a specific table layout that facilitates the presentation of the algorithm's predictive accuracy. Figure 4 delineates the classification of the confusion matrix into four components. *True Positive* (TP) indicates that the instance has been accurately classified as affirmative (1); *False Positive* (FP) signifies that the instance has been erroneously classified as positive; *True Negative* (TN) denotes that the instance has been correctly classified as negative (0); and *False Negative* (FN) indicates that the positive instance has been incorrectly classified as negative. The performance attribute of accuracy, particularly indicates the generated ML model's comprehensive capability of prediction, is the first performance metric determined by the confusion matrix (Al Mamlook *et al*. [2], and Yang and Berdine [49]).

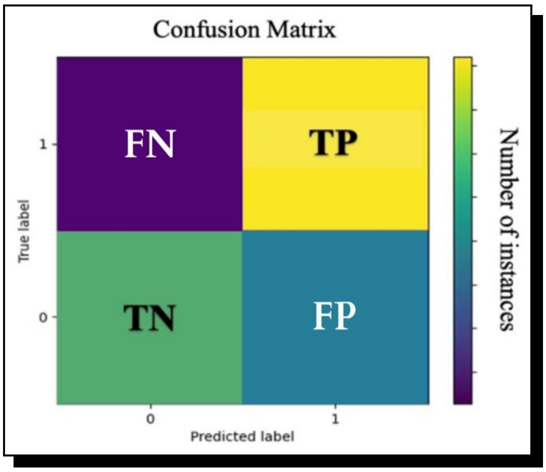

Figure 4. Confusion matrix

The Area under Receiving Operator Characteristic (AUC), a performance indicator for ML classifiers, is well-recognized. The crucial AUC is calculated using the ROC, a probability curve showing sensitivity, i.e., TP rate on the vertical and specificity axes, whereas TN rate on the horizontal axis. The predicted value of the AUC ranges from 0.5 (the lowest predictability) to 1 (the highest predictability) (Al Mamlook *et al*. [2], and Yang and Berdine [49]).

Furthermore, to compare the results of all ML models, several significant metrics were also computed in the validation defined below:

Accuracy: It is the measure of analyzing the proportion of the number of results which are measured correctly (Chowdhury and Schoen [7]),

$$\text{Acc} = (\text{TP}) + (\text{TN})/(\text{TP}) + (\text{TN}) + (\text{FP}) + (\text{FN}).$$

Precision: It determines the ratio of correct predicted outcomes and the total number of positive predictions (Badia *et al*. [3], Chowdhury and Schoen [7], Luque *et al*. [26]),

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}).$$

Recall: It identifies the data outcomes that are positive and predicted correctly (Chowdhury and Schoen [7]),

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}).$$

F1-score: It combines precision and recall calculating the harmonic mean between 0 and 1 (Chowdhury and Schoen [7], Luque *et al*. [26], and Umscheid *et al*. [42]).

$$\text{F1} - \text{score} = 2 \times [(\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})].$$

5. Results

I evaluated the performance of six machine learning models—LR, DTC, KNN, GNB, SVM, and ANN—using a combined sampling approach and wrapper feature selection. The data were pre-processed and balanced using the SMOTE-ENN and SMOTE-Tomek techniques, and feature selection was performed using Wrapper. The models were then constructed, trained, and validated.

From Table 3, using the SMOTE-ENN combined sampling approach, the accuracy was highest for both the KNN and DTC classifiers, i.e., 99%. The precision of DTC was highest for the majority class, i.e., 99%, whereas KNN demonstrated the highest precision for the minority class, i.e., 80%. The highest recall was obtained for the majority class, i.e., 99% for the DTC model; however, the KNN model demonstrated the highest recall for the minority class, i.e., 61Among the six models, the F1-score for the majority class was high, and the KNN classifier achieved the highest F1-score (69% for the minority class).

Table 3. Performance metrics of ML algorithms by SMOTE-EEN and SMOT-Tomek with wrapper feature selection

ML	Combined method	Accuracy %	Precision %		Recall %		F1-score %	
			0	1	0	1	0	1
KNN	SMOTE-ENN	98	99	80	98	61	99	69
DTC	SMOTE-ENN	98	99	77	99	60	99	68
ANN	SMOTE-ENN	94	99	26	95	60	97	36
LR	SMOTE-ENN	89	99	15	90	60	94	24
GNB	SMOTE-ENN	87	99	13	88	60	93	21
SVM	SMOTE-ENN	74	98	7	74	60	85	13
KNN	SMOTE-Tomek	99	99	87	99	60	99	74
DTC	SMOTE-Tomek	99	99	89	99	60	99	72
LR	SMOTE-Tomek	88	99	14	89	60	93	23
GNB	SMOTE-Tomek	86	99	12	86	60	92	20
ANN	SMOTE-Tomek	85	99	11	86	60	92	19
SVM	SMOTE-Tomek	73	98	6	74	60	84	12

When using SMOTE-Tomek combined sampling, the lowest accuracy results were 73%, with precision values of 98% and 6% for the majority and minority classes, recall values of 74% for the majority class and 60% for the minority class, and F1-scores of 84% for the majority class and 12% for the minority class for the SVM model.

The performance metrics evaluation can be used to explain why the KNN and DTC classifiers utilizing both sampling approaches demonstrated the best levels of precision, recall, accuracy, and F1-score, demonstrating their value for reducing FP instances. The results further demonstrate that the proposed KNN can capture increasing TP occurrences and exhibits a meaningful and authentic balance between recall and precision. In conclusion, KNN is the most accurate predictor of both positive and negative outcomes of SSIs.

Confusion Matrix

Figures 4 and 5 show confusion matrices that link the model's predicted result to the actual class. As shown in Figure 4, the confusion matrices for all six ML models were split into four quadrants for SMOTE-ENN combined sampling and wrapper feature selection. It should be noted that the highest TP was obtained using the KNN model, i.e., 343 indicating the best performance in detecting infected cases. The lowest FP (i.e., 44) was observed for DTC compared with the other ML models under SMOTE-ENN sampling, explaining that DTC showed the lowest number of non-infected cases as infected cases of SSIs. The lower FN (i.e., 226) compared to the other ML models explains why KNN showed the lowest number of infected cases as non-infected SSI cases. It was predicted that the DTC model would have the highest TN, i.e., 19042. Thus, it denoted that the DTC model showed the maximum number of correct predictions for non-infected SSI cases.

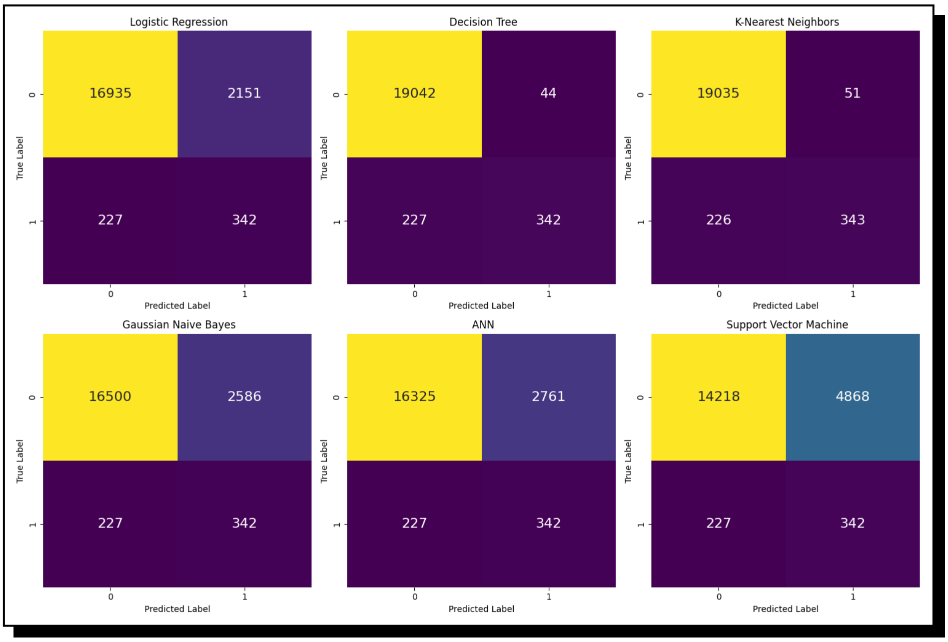

Figure 5. Confusion matrices of models with combined sampling method (SMOTE-ENN)

Figure 5 shows all confusion matrices for ML models with SMOTE Tomek using wrapper feature selection. The results show the highest TP value from the KNN classifier, i.e., 345; the lowest FN value was noted for KNN, i.e., 224; the highest FP value was observed from the SVM classifier's confusion matrix, i.e., 5019; and the highest TN value was noted for KNN, i.e., 18,998.

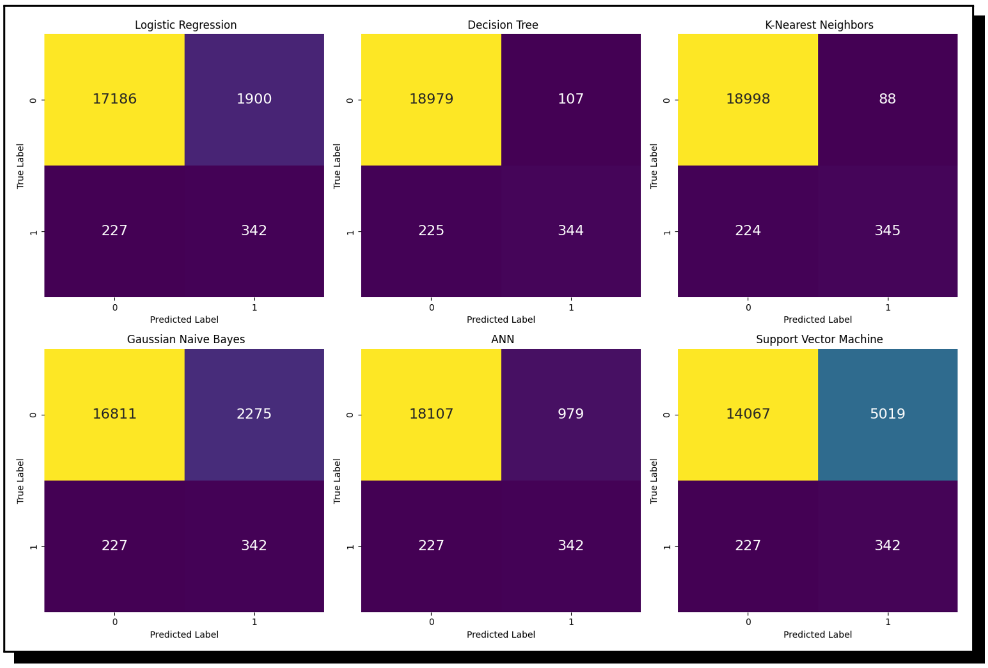

Figure 6. Confusion matrices of models with combined sampling method (SMOTE Tomek)

The ROC curves for the Six ML models employing SMOTE-ENN and SMOTE Tomek combined sampling are shown in Figures 6 and 7. When the AUC value exceeds 0.5, the ML model can successfully distinguish between negative and positive instances. KNN validated the best performance compared to the other models since it applied both SMOTE-ENN and SMOTE Tomek combined sampling strategies and displayed the highest AUCROC (i.e., 0.90 (90%) and 0.89 (89%)). Also, DTC, when used with SMOTE-ENN, also showed the highest AUCROC of 90%, as shown in Figure 7.

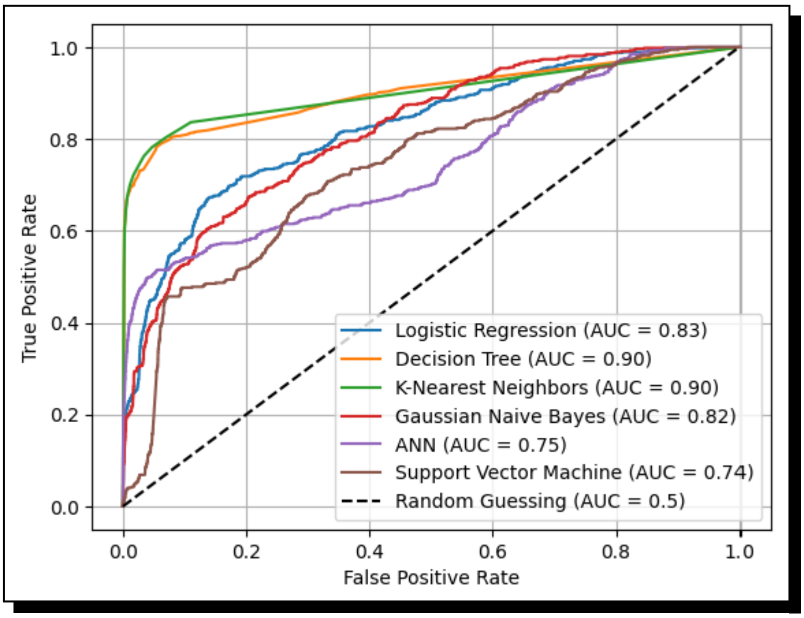

Figure 7. Models performance under ROC CURVE using combined sampling method (SMOTE-ENN)

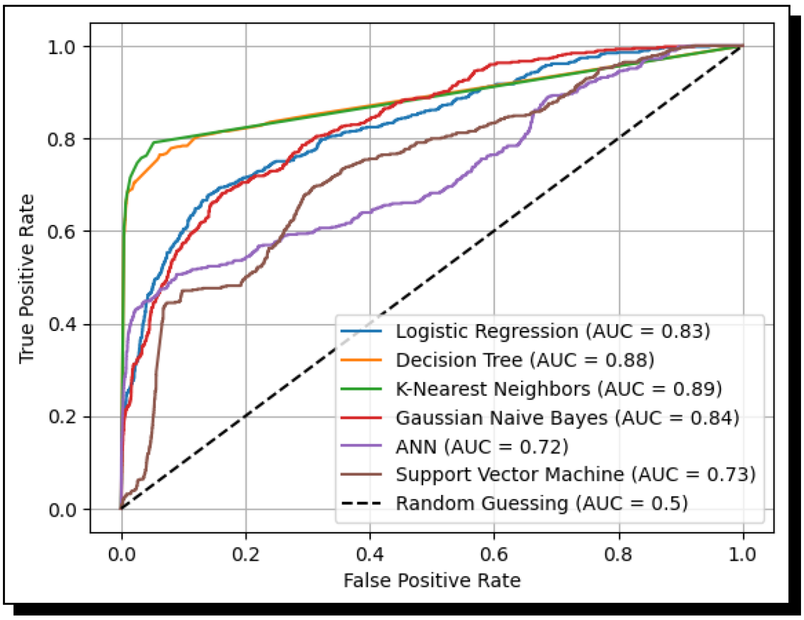

Figure 8. Models performance under ROC CURVE using combined sampling method (SMOTE Tomek)

6. Discussion

In the current study, the application of the KNN algorithm resulted in the highest AUC of 0.90 when the combined sampling approach of SMOTE-ENN was used. This finding is supported by previous research that used oversampling. For example, Xiong *et al*. [47] developed a prediction model based on SMOTE and ML for the successful prediction of the risk of SSI for early prediction. The proposed model employs four ML algorithms such as LR, SGD, RF, and AdaBoost Classification Trees. Results revealed that the AUC of the AdaBoost Classification Trees classifier with SMOTE sampling was the highest (0.90), and it correctly identified 15 out of 16 patients as having a higher risk of developing SSI (Xiong *et al*. [47]). Another study by Fletcher *et al*. [10] used LR and SVM and showed that LR had an AUC accuracy of 96.5%, while SVM performed best with image data with an accuracy of 99.5%. It was concluded that in combination, they both performed well (Fletcher *et al*. [10]). However, a previous study that also used six ML classifiers revealed that the DTC and KNN models had the highest AUC of 0.90 under the combined sampling of SMOTE-ENN. The results demonstrate that by selecting an appropriate ML model, combined sampling under Wrapper feature selection is a highly suitable approach to obtain higher accuracy than oversampling techniques.

Some heuristic algorithms of under sampling, like Tomek and ENN links, are significant for solving the classification problem of imbalanced datasets based on the *instance hardness threshold* (IHT) (Xu *et al*. [48]). Wang and Liu [44]) also introduced the SMOTE oversampling technique. The author applied SMOTE sampling rates and concluded that the SMOTE and SVM models demonstrated significant improvements in accuracy and assurance for tackling imbalanced data classification issues (Wang and Liu [44]). Similarly, the outcomes of the current research showed credible findings because KNN using both the SMOTE-ENN and SMOTE Tomek combined sampling techniques demonstrated the greatest AUC (i.e., 0.90 (90%) and 0.89 (89%) compared to the other models; it verified the best performance makes accurate predictions for unbalanced datasets.

7. Conclusion

Adopting machine learning algorithms to decipher complex associations is crucial for identifying key variables in large datasets. These techniques significantly enhance the effectiveness of prediction models, leading to improved clinical outcomes. The current research also hypothesized the integration of a KNN classifier as the model with the highest accuracy and AUC under SMOTE-ENN and SMOTE-Tomek sampling techniques in SSI prediction when certain risk factors were selected using wrapper feature selection. The issues of data overfitting and underfitting are also resolved when analyzed using ML algorithms, which limit the challenges faced in imbalanced datasets. By selecting only the most relevant features, the proposed wrapper method improves model accuracy and effectiveness. This focused approach reduces noise and irrelevant information, resulting in better predictions and more reliable results. However, wrapper methods can be computationally intensive, especially with SVM and other complex models because they require multiple iterations to evaluate different feature subsets. This can lead to long training times and increased resource consumption.

Competing Interests

The author declares that she has no competing interests.

Authors' Contributions

The author wrote, read and approved the final manuscript.

References

[1] K. Ahmed, T. R. Shahidi, S. M. I. Alam and S. Momen, Rice leaf disease detection using machine learning techniques, in: *Proceedings of the 2019 International Conference on Sustainable Technologies for Industry 4.0* (STI, Dhaka, Bangladesh, 2019), pp. 1 – 5, (2019), DOI: 10.1109/sti47673.2019.9068096.

[2] R. E. Al Mamlook, L. J. Wells and R. Sawyer, Machine-learning models for predicting surgical site infections using patient pre-operative risk and surgical procedure factors, *American Journal of Infection Control* **51**(5) (2023), 544 – 550, DOI: 10.1016/j.ajic.2022.08.013.

[3] J. M. Badia, A. L. Casey, N. Petrosillo, P. M. Hudson, S. A. Mitchell, C. Crosby, Impact of surgical site infection on healthcare costs and patient outcomes: a systematic review in six European countries, *Journal of Hospital Infection* **96**(1) (2017), 1 – 15, DOI: 10.1016/j.jhin.2017.03.004.

[4] M. A. Bartz-Kurycki, C. Green, K. T. Anderson, A. C. Alder, B. T. Bucher, R. A. Cina, R. Jamshidi, R. T. Russell, R. F. Williams and K. Tsao, Enhanced neonatal surgical site infection prediction model utilizing statistically and clinically significant variables in combination with a machine learning algorithm, *The American Journal of Surgery* **216**(4) (2018), 764 – 777, DOI: 10.1016/j.amjsurg.2018.07.041.

[5] G. Chandrashekar and F. Sahin, A survey on feature selection methods, *Computers & Electrical Engineering* **40**(1) (2014), 16 – 28, DOI: 10.1016/j.compeleceng.2013.11.024.

[6] K. A. Chen, C. U. Joisa, J. Stem, J. G. Guillem, S. M. Gomez and M. R. Kapadia, Improved prediction of surgical-site infection after colorectal surgery using machine learning, *Diseases of the Colon & Rectum* **66**(3) (2023), 458 – 466, DOI: 10.1097/dcr.0000000000002559.

[7] S. Chowdhury and M. P. Schoen, Research paper classification using supervised machine learning techniques, in: *Proceedings of the 2020 Intermountain Engineering, Technology and Computing* (IETC2020, Orem, UT, USA, 2020), pp. 1 – 6, (2020), DOI: 10.1109/ietc47856.2020.9249211.

[8] K. L. Colborn, M. Bronsert, E. Amioka, K. Hammermeister, W. G. Henderson and R. Meguid, Identification of surgical site infections using electronic health record data, *American Journal of Infection Control* **46**(11) (2018), 1230 – 1235, DOI: 10.1016/j.ajic.2018.05.011.

[9] A. Das, Logistic regression, in *Encyclopedia of Quality of Life and Well-Being Research*, Springer, Cham., pp. 1 – 2 (2021), DOI: 10.1007/978-3-319-69909-7_1689-2.

[10] R. R. Fletcher, G. Schneider, B. Hedt-Gauthier, T. Nkurunziza, B. Alayande, R. Riviello and F. Kateera, Use of convolutional neural nets and transfer learning for prediction of surgical site infection from color images, in: *Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (EMBC2021, Mexico, 2021), pp. 5047 – 5050, (2021), DOI: 10.1109/embc46164.2021.9630430.

[11] R. R. Fletcher, O. Olubeko, H. Sonthalia, F. Kateera, T. Nkurunziza, J. L. Ashby, R. Riviello and B. Hedt-Gauthier, Application of machine learning to prediction of surgical site infection, in: *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC2019, Berlin, Germany, 2019), pp. 2234 – 2237, (2019), DOI: 10.1109/embc.2019.8857942.

[12] R. Ghorbani and R. Ghousi, *Comparing different resampling methods in predicting students' performance using machine learning techniques*, *IEEE Access* **8** (2020), 67899 – 67911, DOI: 10.1109/access.2020.2986809.

[13] J. M. Gutierrez-Naranjo, A. Moreira, E. Valero-Moreno, T. S. Bullock, L. A. Ogden and B. A. Zelle, A machine learning model to predict surgical site infection after surgery of lower extremity fractures, *International Orthopaedics* **48** (2024), 1887 – 1896, DOI: 10.1007/s00264-024-06194-5.

[14] H. Hairani, A. Anggrawan and D. Priyanto, Improvement performance of the random forest method on unbalanced diabetes data classification using Smote-Tomek Link, *International Journal on Informatics Visualization* **7**(1) (2023), 258 – 264, DOI: 10.30630/joiv.7.1.1069.

[15] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue and G. Bing, Learning from class-imbalanced data: Review of methods and applications, *Expert Systems With Applications* **73** (2017), 220 – 239, DOI: 10.1016/j.eswa.2016.12.035.

[16] T. Humayun, A. M. Alkhamis, G. M. Saleh, M. A. Al Qahtani, M. AlSaedi, M. M. El Dalatony and K. H. Alanezi, Designing and implementing national program of Health Electronic Surveillance Network (HESN); Infection control module in Saudi Arabia, *American Journal of Infectious Diseases and Microbiology* **9**(2) (2021), 61 – 70, URL: https://pubs.sciepub.com/ajidm/9/2/6/index.html.

[17] N. Junsomboon and T. Phienthrakul, Combining over-sampling and under-sampling techniques for imbalance dataset, in: *Proceedings of the 9th International Conference on Machine Learning and Computing* (ICMLC'17), Association for Computing Machinery, New York, USA, pp. 243 – 247, (2017), DOI: 10.1145/3055635.3056643.

[18] S. Khalid, T. Khalil and S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, in: *Proceedings of the 2014 Science and Information Conference* (SIC 2014, London, UK, 2014), pp. 372 – 378, (2014) DOI: 10.1109/sai.2014.6918213.

[19] J. H. Kim, J.-K. Shin, H. Lee, D. H. Lee, J.-H. Kang, K. H. Cho, Y.-G. Lee, K. Chon, S.-S. Baek and Y. Park, Improving the performance of machine learning models for early warning of harmful algal blooms using an adaptive synthetic sampling method, *Water Research* **207** (2021), 117821, DOI: 10.1016/j.watres.2021.117821.

[20] P. Kocbek, N. Fijacko, C. Soguero-Ruiz, K. Ø. Mikalsen, U. Maver, P. Povalej Brzan, A. Stozer, R. Jenssen, S. O. Skrøvseth and G. Stiglic, Maximizing interpretability and cost-effectiveness of Surgical Site Infection (SSI) predictive models using feature-specific regularized logistic regression on preoperative temporal data, *Computational and Mathematical Methods in Medicine* **2019**(1) (2019), 2059851, DOI: 10.1155/2019/2059851.

[21] M. H. Kotb and R. Ming, Comparing SMOTE family techniques in predicting insurance premium defaulting using machine learning models, *International Journal of Advanced Computer Science and Applications* **12**(9) (2021), 621 – 629, DOI: 10.14569/ijacsa.2021.0120970.

[22] M. S. Kraiem, F. Sánchez-Hernández and M. N. Moreno-García, Selecting the suitable resampling strategy for imbalanced data classification regarding dataset properties. An approach based on association model, *Applied Sciences* **11**(18) (2021), 8546, DOI: 10.3390/app11188546.

[23] M. A. C. Lengua, E. A. P. Quiroz, A systematic literature review on support vector machines applied to classification, in: *Proceedings of the 2020 IEEE Engineering International Research Conference* (EIRCON 2020, Lima, Peru, 2020), pp. 1 – 4, (2020), DOI: 10.1109/eircon51178.2020.9254028.

[24] W.-C. Liu, H. Ying, W.-J. Liao, M.-P. Li, Y. Zhang, K. Luo, B.-L. Sun, Z.-L. Liu and J.-M. Liu, Using preoperative and intraoperative factors to predict the risk of surgical site infections after lumbar spinal surgery: A machine learning–based study, *World Neurosurgery* **162** (2022), e553 – e560, DOI: 10.1016/j.wneu.2022.03.060.

[25] K. Lu, Y. Tu, S. Su, J. Ding, X. Hou, C. Dong, H. Jin and W. Gao, Machine learning application for prediction of surgical site infection after posterior cervical surgery, *International Wound Journal* **21**(4) (2024), e14607, DOI: 10.1111/iwj.14607.

[26] A. Luque, A. Carrasco, A. Martín and A. de Las Heras, The impact of class imbalance in classification performance metrics based on the binary confusion matrix, *Pattern Recognition* **91** (2019), 216 – 231, DOI: 10.1016/j.patcog.2019.02.023.

[27] A. A. Lydia and S. Francis, Adagrad—An optimizer for stochastic gradient descent, *International Journal of Information and Computing Science* **6** (2019), 566 – 568.

[28] A. Mahani and A. R. B. Ali, Classification problem in imbalanced datasets, in: *Recent Trends in Computational Intelligence*, A. Sadollah and T. S. Sinha (editors), *IntechOpen*, London, (2020), DOI: 10.5772/intechopen.78839.

[29] V. Nasteski, An overview of the supervised machine learning methods, *Horizons Series B* **4** (2017), 51 – 62.

[30] F. Nieto-del-Amor, G. Prats-Boluda, J. Garcia-Casado, A. Diaz-Martinez, V. J. Diago-Almela, R. Monfort-Ortiz, D. Hao and Y. Ye-Lin, Combination of feature selection and resampling methods to predict preterm birth based on electrohysterographic signals from imbalance data, *Sensors* **22**(14) (2022), 5098, DOI: 10.3390/s22145098.

[31] M. Ontivero-Ortega, A. Lage-Castellanos, G. Valente, R. Goebel and M. Valdes-Sosa, Fast Gaussian Naïve Bayes for searchlight classification analysis, *NeuroImage* **163** (2017), 471 – 479, DOI: 10.1016/j.neuroimage.2017.09.001.

[32] A. Parmar, R. Katariya and V. Patel, A review on random forest: An ensemble classifier, in: *Proceedings of the International Conference on Intelligent Data Communication Technologies and Internet of Things* (ICICI 2018), Lecture Notes on Data Engineering and Communications Technologies, Vol. 26, Springer, Cham., (2019), DOI: 10.1007/978-3-030-03146-6_86.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss and V. Dubourg, *Scikit-learn: Machine learning in Python*, *Journal of Machine Learning Research* **12** (2011), 2825 – 2830, URL: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

[34] Y. Petrosyan, K. Thavorn, G. Smith, M. Maclure, R. Preston, C. van Walravan and A. J. Forster, Predicting postoperative surgical site infection with administrative data: a random forests algorithm, *BMC Medical Research Methodology* **21** (2021), article number 179, DOI: 10.1186/s12874-021-01369-9.

[35] Y. Petrosyan, K. Thavorn, M. Maclure, G. Smith, D. McIsaac, D. Schramm, H. Moloo, P. Roanne and A. J. Alan, Long-term health outcomes and health system costs associated with surgical site infections: A retrospective cohort study, *Annals of Surgery* **273**(5) (2021), 917 – 923, DOI: 10.1097/SLA.0000000000003285.

[36] A. Samareh, X. Chang, W. B. Lober, H. L. Evans, Z. Wang, X. Qian and S. Huang, Artificial intelligence methods for surgical site infection: Impacts on detection, monitoring, and decision making, *Surgical Infections* **20**(7) (2019), 546 – 554, DOI: 10.1089/sur.2019.150.

[37] J. Song, E. Sanabria-Buenaventura, B. Cohen, J. Liu, D. Yao and E. Larson, Predictive Models for Surgical Site Infection (SSI) in Patients with a Permanent Pacemaker (PPM) using machine learning methods, *Authorea Preprints* **2020** (2020), DOI: 10.22541/au.159188536.65812462.

[38] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang and G. Wang, Data processing and text mining technologies on electronic medical records: A review, *Journal of Healthcare Engineering* **2018**(1) (2018), 4302425, DOI: 10.1155/2018/4302425.

[39] C. S. Tarimo, S. S. Bhuyan, Q. Li, W. Ren, M. J. Mahande and J. Wu, Combining resampling strategies and ensemble machine learning methods to enhance prediction of neonates with a low Apgar score after induction of labor in northern Tanzania, *Risk Management and Healthcare Policy* **14** (2021), 3711 – 3720, DOI: 10.2147/rmhp.s331077.

[40] K. Taunk, S. De, S. Verma and A. Swetapadma, A brief review of nearest neighbor algorithm for learning and classification, in: *2019 International Conference on Intelligent Computing and Control Systems* (ICCS 2019, Madurai, India, 2019), pp. 1255 – 1260, DOI: 10.1109/iccs45141.2019.9065747.

[41] Z. Ullah, F. Saleem, M. Jamjoom, B. Fakieh, F. Kateb, A. M. Ali and B. Shah, Detecting high-risk factors and early diagnosis of diabetes using machine learning methods, *Computational Intelligence and Neuroscience* **2022**(1) (2022), 2557795, DOI: 10.1155/2022/2557795.

[42] C. A. Umscheid, M. D. Mitchell, J. A. Doshi, R. Agarwal, K. Williams and P. Brennan, Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs, *Infection Control & Hospital Epidemiology* **32**(2) (2011), 101 – 114, DOI: 10.1086/657912.

[43] S. Visalakshi and V. Radha, A literature review of feature selection techniques and applications: Review of feature selection in data mining, in: *2014 IEEE International Conference on Computational Intelligence and Computing Research* (ICCICR 2014, Coimbatore, India, 2014), pp. 1 – 6, (2014), DOI: 10.1109/iccic.2014.7238499.

[44] Z. Wang and Q. Liu, Imbalanced data classification method based on LSSASMOTE, *IEEE Access* **11** (2023), 32252 – 32260, DOI: 10.1109/access.2023.3262460.

[45] C. B. Weir and A. Jan, *BMI Classification Percentile and Cut off Points*, StatPearls Publishing, Treasure Island, (2021).

[46] H. Xiaoli and S. Qiang, A hybrid FRFS-CSRF model for surgical site infection prediction, in: *2019 16th International Conference on Service Systems and Service Management* (ICSSSM 2019, Shenzhen, China, 2019), pp. 1 – 4, (2019), DOI: 10.1109/icsssm.2019.8887701.

[47] C. Xiong, R. Zhao, J. Xu, H. Liang, C. Zhang, Z. Zhao, T. Huang and X. Luo, Construct and validate a predictive model for surgical site infection after posterior lumbar interbody fusion based on machine learning algorithm, *Computational and Mathematical Methods in Medicine* **2022**(1) (2022), 2697841, DOI: 10.1155/2022/2697841.

[48] Z. Xu, D. Shen, T. Nie and Y. Kou, A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data, *Journal of Biomedical Informatics* **107** (2020), 103465, DOI: 10.1016/j.jbi.2020.103465.

[49] S. Yang and G. Berdine, The receiver operating characteristic (ROC) curve, *The Southwest Respiratory and Critical Care Chronicles* **5**(19) (2017), 34 – 36, URL: https://doi.org/10.12746/swrccc.v5i19.391.

[50] F. Yang, K. Wang, L. Sun, M. Zhai, J. Song and H. Wang, A hybrid sampling algorithm combining synthetic minority over-sampling technique and edited nearest neighbor for missed abortion diagnosis, *BMC Medical Informatics and Decision Making* **22**(1) (2022), article number 344, DOI: 10.1186/s12911-022-02075-2.

[51] B. W. Yap, K. A. Rani, H. A. A. Rahman, S. Fong, Z. Khairudin and N. N. Abdullah, An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets, in: *Proceedings of the First International Conference on Advanced Data and Information Engineering* (DaEng-2013), T. Herawan, M. Deris and J. Abawajy (editors), Lecture Notes in Electrical Engineering, Vol. 285, Springer, Singapore, (2014), DOI: 10.1007/978-981-4585-18-7_2.

[52] H. Ying, B.-H. Guo, H.-J. Wu, R.-P. Zhu, W.-C. Liu and H.-F. Zhong, Using multiple indicators to predict the risk of surgical site infection after ORIF of tibia fractures: A machine learning based study, *Frontiers in Cellular and Infection Microbiology* **13** (2023), 1206393, DOI: 10.3389/fcimb.2023.1206393.