

Article

Applying Machine Learning Techniques to the Audit of Antimicrobial Prophylaxis

Zhi-Yuan Shi ^{1,2,3} , Jau-Shin Hon ^{1,*}, Chen-Yang Cheng ⁴, Hsiu-Tzy Chiang ⁵ and Hui-Mei Huang ⁶

¹ Department of Industrial Engineering & Enterprise Information, Tunghai University, Taichung 407, Taiwan; zyshi@vghtc.gov.tw

² Infection Control Center, Taichung Veterans General Hospital, Taichung 407, Taiwan

³ School of Medicine, National Yang Ming Chiao Tung University, Taipei 106, Taiwan

⁴ Department of Industrial Engineering & Management, National Taipei University of Technology, Taipei 106, Taiwan; cycheng@ntut.edu.tw

⁵ Infection Control Center, Mackay Memorial Hospital, Taipei 10449, Taiwan; hideko@ms1.mmh.org.tw

⁶ Nursing Department, Taichung Veterans General Hospital, Taichung 407, Taiwan; hmhuang@vghtc.gov.tw

* Correspondence: honjs@thu.edu.tw

Abstract: High rates of inappropriate use of surgical antimicrobial prophylaxis were reported in many countries. Auditing the prophylactic antimicrobial use in enormous medical records by manual review is labor-intensive and time-consuming. The purpose of this study is to develop accurate and efficient machine learning models for auditing appropriate surgical antimicrobial prophylaxis. The supervised machine learning classifiers (Auto-WEKA, multilayer perceptron, decision tree, SimpleLogistic, Bagging, and AdaBoost) were applied to an antimicrobial prophylaxis dataset, which contained 601 instances with 26 attributes. Multilayer perceptron, SimpleLogistic selected by Auto-WEKA, and decision tree algorithms had outstanding discrimination with weighted average AUC > 0.97. The Bagging and SMOTE algorithms could improve the predictive performance of decision tree against imbalanced datasets. Although with better performance measures, multilayer perceptron and Auto-WEKA took more execution time as compared with that of other algorithms. Multilayer perceptron, SimpleLogistic, and decision tree algorithms have outstanding performance measures for identifying the appropriateness of surgical prophylaxis. The efficient models developed by machine learning can be used to assist the antimicrobial stewardship team in the audit of surgical antimicrobial prophylaxis. In future research, we still have the challenges and opportunities of enriching our datasets with more useful clinical information to improve the performance of the algorithms.

Keywords: antimicrobial prophylaxis; machine learning; Auto-WEKA; multilayer perceptron; decision tree; bagging; SMOTE



Citation: Shi, Z.-Y.; Hon, J.-S.; Cheng, C.-Y.; Chiang, H.-T.; Huang, H.-M. Applying Machine Learning Techniques to the Audit of Antimicrobial Prophylaxis. *Appl. Sci.* **2022**, *12*, 2586. <https://doi.org/10.3390/app12052586>

Academic Editor: Jerry Chun-Wei Lin

Received: 9 December 2021

Accepted: 21 February 2022

Published: 2 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The incidence of surgical site infections (SSIs) is estimated to be ~2–5% in patients undergoing surgery [1]. SSIs are associated with increased rates of morbidity and mortality. The financial impact of SSIs is the highest among all healthcare-associated infections. The annual cost of SSIs in the United States of America is estimated to be \$3.5–10 billion [2]. About 40–60% of SSIs are preventable by the interventions with evidence-based measures [1,2]. The interventions include antimicrobial prophylaxis, preoperative bathing and showering, glucose control, skin preparation, intraoperative normothermia, and wound closure [1,3–5].

Appropriate antimicrobial prophylaxis is the most important and strongly recommended measure to reduce the SSI rates [2,3,6]. To achieve these goals, appropriate practices of preoperative-administration timing, choice and dosing of antimicrobial agents, and duration of antimicrobial prophylaxis are the most important approaches [7]. However,

previous studies showed that nonadherence to the guidelines of antimicrobial prophylaxis is a worldwide problem and ranges from 20% to 50% [5,8,9].

Antimicrobial stewardship is an emerging global action plan on antimicrobial resistance. Antimicrobial stewardship implements a series of strategies and interventions to improve appropriate use of antimicrobial agents in healthcare settings [10,11]. The interventions include formulary restriction, audit and feedback, education for prescribers, clinical guidelines, and clinical decision support systems provided by information technology. Audit is the most common intervention of all antimicrobial stewardship strategies. However, the processing and analysis of enormous medical data for the manual audit is time-consuming and labor-intensive. Therefore, it is helpful to develop efficient models by machine learning to analyze the big medical data associated with antimicrobial prophylaxis.

Machine learning creates new and promising opportunities in processing enormous volume of data to provide data classification, clustering, outlier or anomaly detection, and real-time prediction [12]. Machine learning can help clinicians to diagnose infectious diseases, evaluate severity of the diseases, and decide the choice of antimicrobial agents [13,14]. In a narrative review of 60 articles of machine learning for clinical decision support systems, 37 (62%) articles focused on bacterial infections, 10 (17%) on viral infections, and 13 (22%) on other kinds of infections. Among the 60 articles, four (7%) articles addressed the prediction of antimicrobial resistance, and three (5%) addressed the selection of antimicrobial agents [13]. The popular supervised machine learning algorithms, multilayer perceptron and decision tree, were applied to the antimicrobial susceptibility dataset to predict the resistant profile of bacteria, thereby aiding the clinicians in selecting empiric antimicrobial therapy [15,16]. However, they had the limitations of features selection and imbalanced dataset. Application of machine learning to the audit of antimicrobial prophylaxis was never reported. There is strong potential for research to develop machine learning algorithms with excellent performance in this field.

The purpose of this study is to develop accurate and efficient machine learning models for auditing appropriate surgical antimicrobial prophylaxis. Auto-WEKA, multilayer perceptron, and decision tree algorithms were applied to the dataset of surveillance of healthcare-associated infections and antimicrobial use. Sampling and ensemble methods were used to enhance the performance against the imbalanced dataset. We compared the performance measures of these algorithms for the audit of antimicrobial prophylaxis.

2. Materials and Methods

This research was approved by Institutional Review Board at Taichung Veterans General Hospital, with the ethical approval no. SE13130 and SE13130#2 on 17 May 2013 and on 19 May 2014, respectively, waiving the requirement for obtaining informed consent.

2.1. Data Preprocessing

A point prevalence survey of healthcare-associated infections and antimicrobial use was conducted in 25 acute care hospitals in Taiwan. Data were collected according to the protocol version 4.3 of point prevalence survey of healthcare-associated infections and antimicrobial use in European acute care hospitals [17]. If the patient receiving at least one antimicrobial drug at the time of survey, the antimicrobial use data were recorded. For surgical patient, if any surgical prophylactic antimicrobial agent was given in the 24 h before 8:00 a.m. on the day of the survey, the antimicrobial use data were collected.

One or more antimicrobial agents were used for surgical prophylaxis in 601 of the 7377 surveyed patients. The characteristics of dataset are listed in Table 1. The dataset of 601 instances contained 26 attributes, including age (numeric), gender (binary) hospital type (nominal), patient specialty (nominal, 29 distinct values), diagnosis (nominal, 20 distinct values), central vascular catheter in place on survey date (binary), peripheral vascular catheter in place on survey date (binary), urinary catheter in place on survey date (binary), under endotracheal intubation on survey date (binary), under tracheostomy intubation on survey date, ventilator used on survey date (binary), patient has active

healthcare-associated infection (binary), blood stream infection present on survey date (binary), urinary tract infection present on survey date (binary), pneumonia present on survey date (binary), surgical site infection present on survey date (binary), antimicrobial agents used (nominal, 15 distinct values), indication for antimicrobial use (nominal, three distinct values), diagnosis sites for antimicrobial use (nominal, 16 distinct values), and finally the class label attribute (nominal, five distinct values). There were three attributes for antimicrobial agents, with their corresponding indications for antimicrobial use and diagnosis sites for antimicrobial use. The class label attribute was the type of compliance with the guidelines for antimicrobial prophylaxis.

Table 1. Characteristics of dataset.

Attribute	Type	Remarks
Hospital type	Nominal	Primary: 71 instances. Secondary: 258 instances. Tertiary: 271 instances
Age	Numeric	Mean: 56.2 years
Gender	Binary	Female: 297 instances, male: 304 instances
Patient specialty	Nominal	29 distinct values
Diagnosis	Nominal	20 distinct values
Central vascular catheter in place	Binary	Yes: 75 instances No: 526 instances
Peripheral vascular catheter in place	Binary	Yes: 477 instances No: 124 instances
Urinary catheter in place	Binary	Yes: 213 instances No: 388 instances
Under endotracheal intubation	Binary	Yes: 33 instances No: 568 instances
Under tracheostomy intubation	Binary	Yes: 9 instances No: 592 instances
Ventilator used	Binary	Yes: 34 instances No: 567 instances
Patient has active healthcare-associated infection	Binary	Yes: 6 instances No: 595 instances
Blood stream infection	Binary	Yes: 0 instances No: 601 instances
Urinary tract infection	Binary	Yes: 0 instances No: 601 instances
Pneumonia	Binary	Yes: 2 instances No: 599 instances
Surgical site infection	Binary	Yes: 3 instances No: 598 instances
Antimicrobial agents used	Nominal	15 distinct values
Indication of antimicrobial agents	Nominal	3 distinct values
Diagnosis sites for antimicrobial use	Nominal	16 distinct values
Class label attribute	Nominal	5 distinct values

The values of class label attribute were determined by infectious disease specialists, based on the common principles and procedure-specific guidelines [7,18]. For reading

comprehension, the recommendations of guidelines for antimicrobial prophylaxis were briefly stated in Table 2. For details, please refer to the articles that are more complex [7,18].

Table 2. Brief recommendations for surgical antimicrobial prophylaxis.

Wound Classification	Choice of Prophylactic Antimicrobial Agents	Duration of Prophylactic Antimicrobial Agents
Class I (Clean wound)	<ol style="list-style-type: none"> 1. Recommended agents: First generation cephalosporins 2. Alternative agents in patients with β-lactam allergy: Clindamycin or vancomycin 	A single dose or one day
Class II (Clean-contaminated wound)	<ol style="list-style-type: none"> 1. Recommended agents: First-generation cephalosporins \pm metronidazole or Second-generation cephalosporins 2. Alternative agents in patients with β-lactam allergy: Clindamycin + aminoglycosides or Metronidazole + aminoglycosides 	A single dose or one day

If the choice of prophylactic antimicrobial use was in concordance with the recommendations in the practice guidelines, then the compliance with choice of prophylactic antimicrobial agents was determined as “Yes”; otherwise, it was determined as “No”. The duration of prophylactic antimicrobial agents had the following three observed values, i.e., single dose encoded as SP1 (surgical prophylaxis 1), one day encoded as SP2, and more than one day encoded as SP3. If the duration of prophylactic antimicrobial agents was SP1 or SP2, the compliance with duration of prophylactic antimicrobial agents was determined as “Yes”. If the duration was SP3, the compliance was determined as “No”. The values of class label attribute, i.e., compliance with guidelines for antimicrobial prophylaxis, were then classified into five types (i.e., A, B, C, D, and E), according to the compliance with both choice and duration of prophylactic antimicrobial agents in Table 3.

Table 3. Description of five types of compliance with recommendations for antimicrobial prophylaxis.

Types of Compliance with Recommendations for Antimicrobial Prophylaxis Determined by Infectious Disease Specialists	Compliance with Choice of Prophylactic Antimicrobial Agents	Compliance with Duration of Prophylactic Antimicrobial Agents
A	Yes	Yes
B	Yes	No
C	No	Yes
D	No	No
E	Antimicrobial agents used for treatment of other infections rather than surgical prophylaxis.	

In this study, cefazolin was the most common first-generation cephalosporin used for surgical prophylaxis. The second-generation cephalosporins included cefoxitin mostly, but also cefuroxime. The aminoglycosides included gentamicin and amikacin. Amikacin was used for treatment of infection rather than surgical prophylaxis. The fluoroquinolones included ciprofloxacin and levofloxacin. The β -lactam/ β -lactamase combinations included amoxicillin/clavulanic acid, ampicillin/sulbactam, and piperacillin/tazobactam.

2.2. Sampling Methods

Imbalanced dataset contains instances that are distributed unequally among the different classes, i.e., there are more instances in some classes than in other classes. Classifiers

perform well with the majority class, but poorly with the minority class. Sampling approach is one of the useful methods to manage imbalanced data. The sampling techniques include two subgroups, i.e., undersampling and oversampling.

Undersampling method removes instances from the majority class to make the datasets balanced. SpreadSubsample is a supervised filter that produces a random undersampling. The relative distribution difference between the majority and the minority class can be adjusted [19].

Oversampling is a sampling method that involves replicating the instances of minority class to make the dataset balanced. SMOTE is a supervised filter that oversamples the minority class to produce synthetic instances utilizing k-nearest neighbor [20]. The oversampling percentage and the number of neighbors can be adjusted when creating synthetic instances.

Both SpreadSubsample and SMOTE algorithms were proposed to improve the predictive capability of decision tree.

2.3. Machine Learning Techniques

Among the available machine learning systems, we used WEKA as the data mining software to aid the audit of antimicrobial prophylaxis [19]. WEKA provides implementations of various machine learning algorithms that can be easily applied to a dataset. The dataset is divided into 10 approximately equal partitions (also called 10 folds). In each turn, nine-tenths were used for training and one-tenth for testing. The process was repeated 10 times. Finally, every instance would be used exactly once for testing. This is called 10-fold cross-validation. Cross-validation is widely used as a reliable approach to evaluate the performance of machine learning techniques when data are all in one set [19].

The multilayer perceptron and decision tree algorithms were applied to evaluate the appropriateness of antimicrobial prophylaxis in this study. Ensemble methods with bagging and boosting algorithms were used to improve the performance of decision tree. Auto-WEKA is a system automatically searching through the WEKA's algorithms and their respective hyperparameter settings, by a Bayesian optimization method to achieve optimal performance [21]. SimpleLogistic was the optimal classifier selected by Auto-WEKA for the whole dataset in this study. SimpleLogistic is a classifier that employs the LogitBoost algorithm to build the logistic regression functions at the nodes of a tree and uses the CART algorithm for pruning. The most relevant attributes in the dataset are selected by performing a simple regression in each iteration. The optimal number of iterations is determined by cross-validation [22]. We compared the performance metrics and execution time of all these algorithms.

Multilayer perceptron is an artificial neural network composed of multiple layers of perceptrons. The layers include one input layer, one or more hidden layers, and one output layer. Multilayer perceptron is usually trained by minimizing the squared-error loss function of network output to make a best estimate of the class probability. The weights of the connections between neurons are modified by a backpropagation algorithm, which is computed by a standard mathematical optimization algorithm, called gradient descent, to minimize the squared-error loss function [19]. The gradient descent algorithm requires derivatives of the squared-error loss function [19].

The squared-error loss function, E , is defined as:

$$E = \frac{1}{2}(y - f(x))^2 \quad (1)$$

where x is the weight of the input, $f(x)$ is the neural network's prediction function obtained from the output unit, and y is the observed value of instance's class label.

The derivative of $f(x)$ with respect to x is as the following:

$$\frac{df(x)}{dx} = f(x)(1 - f(x)) \quad (2)$$

The classification algorithm classifier J48 in WEKA provides the application of decision tree C4.5 algorithm [19]. The decision tree C4.5 algorithm is widely used because of its fast classification and high precision. Decision tree C4.5 uses gain ratio for feature selection and construction of a decision tree. Decision tree C4.5 can handle both continuous and discrete features [12,23].

The decision tree algorithm applies attribute selection measure to determine the splitting criterion to partition the instances in D into individual branches and nodes. If the instances in a partition are all of the same class, the node becomes a leaf and is labeled with that class. The attribute with the least information entropy (or the least information impurity) has the highest information gain [12].

The dataset, D , is a class-labeled training dataset. Let the class label attribute has k distinct classes, C_i .

$Info(D)$ is the information entropy of D :

$$Info(D) = - \sum_{i=1}^k p_i \log_2(p_i) \quad (3)$$

where p_i is the nonzero probability that a subset in D belongs to the corresponding class C_i .

$Info_A(D)$ is defined as the expected information entropy generated by splitting the training dataset, D , into v partitions, corresponding to the v values of attribute A :

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \quad (4)$$

$Gain(A)$ is the information gain:

$$Gain(A) = Info(D) - Info_A(D) \quad (5)$$

The highest information gain in the partitions is chosen as the splitting attribute for node, and partitioning this attribute can produce the best classification [12].

Imbalanced dataset is a common problem in machine learning algorithms. The ensemble method (for example, Bagging and Adaboost) can be used to enhance the performance measure against imbalanced data [24]. An ensemble for classification is made up of a combination of classifiers (models) to improve the prediction. If one classifier obtains more votes than others, it is taken as the correct one. Predictions will become more reliable, if they obtain more votes [19,24]. Both Bagging and boosting adopt this voting approach, but they derive the models in different ways. In Bagging, the models have equal weight, whereas in boosting, the base weak learners receive more weights, and the simple weak learners are combined into a more complex strong ensemble [25]. In this study, Bagging and AdaBoost approaches were implemented on the dataset to improve the classification performance of decision tree.

2.4. Performance of Machine Learning Techniques

2.4.1. Confusion Matrix

To assess the performance of a classifier, we considered the following results reported by WEKA: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [12]. TP and TN are correct identification by the classifier. An FP is an outcome predicted incorrectly as being positive when it is actually negative. An FN is an outcome predicted incorrectly as being negative when it is actually positive. The quality of a classifier is judged by the following formulas [12,19].

True positive rate: the number of TP is divided by the total number of actual positives, also called sensitivity:

$$\text{True positive rate} = \frac{TP}{TP + FN} \quad (6)$$

False positive rate: the number of FP is divided by the total number of actual negatives, also equal to 1—specificity:

$$\text{False positive rate} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (7)$$

Specificity: the number of TN is divided by the total number of actual negatives, also equal to 1—false positive rate:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (8)$$

Precision: the number of true positive predictions divided by the total number of positive predictions, also referred to as positive predictive value (PPV):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (9)$$

Accuracy: the probability that the model prediction is correct:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (10)$$

Area under the receiver operating characteristics curve (ROC curve): the ROC curve is created by plotting the true positive rate (also called sensitivity) against the false positive rate (also called 1-specificity) at various threshold settings. The area under the ROC curve is also called area under curve (AUC), or ROC area. The AUC can be used to evaluate the prediction performance of a classifier system. The value of AUC varies between 0 and 1. If the AUC is 0, all the predictions will be wrong. If $\text{AUC} = 1$, the classifier can predict perfectly. If $0.7 < \text{AUC} \leq 0.8$, it is defined as acceptable discrimination. If $0.8 < \text{AUC} \leq 0.9$, it is defined as excellent discrimination. If $0.9 < \text{AUC} \leq 1.0$, it is defined as outstanding discrimination [26].

2.4.2. Weighted Average of Performance Metrics

The weighted average for multiclass model is calculated by micro-average method, i.e., weighted arithmetic average of performance metrics for all classes, where the weight of each class is the prevalence of the class in the whole dataset [27].

For example, an AUC is calculated for each class in the dataset. The weighted average AUC is calculated by class-reference formulation, also called AUC_{CR} [28]. AUC_{CR} is defined as the weighted average of the class-reference AUCs, where the weight is the prevalence of the reference class.

$$\text{AUC}_{\text{CR}} = \sum_{i=1}^n \text{AUC}_{C_i} P_{C_i} \quad (11)$$

where AUC_{C_i} is the area under the class-reference ROC curve for class C_i . C_i is the corresponding class in class label attribute. CR denotes the class reference. P_{C_i} is the prevalence of each class.

2.4.3. Comparison of the Execution Time for Machine Learning Algorithms

We compared the execution time of machine learning algorithms and manual review for the dataset, by using a stopwatch to measure execution time from the moment after we had loaded the data files and started the filter and/or classifiers. The execution time included time for manual operation of the software and computation of the algorithms.

3. Results

The distributions of values of class label attribute with regard to choice and duration of prophylactic antimicrobial agents for the whole dataset are listed in Table 4. In Class A, cefazolin was the most common prophylactic antimicrobial agent and compli-

ant with recommendations of guideline for the correct choice and duration. In Class B, cefazolin, cefoxitin, clindamycin, and ciprofloxacin were correctly chosen for different surgical procedures with risks of infections. However, the duration was longer than one day and not compliant with recommended duration of antimicrobial prophylaxis. In Class C, the combination of cefazolin and gentamicin was identified as the most common type of non-compliance with choice of antimicrobial agent, because there were no special clinical indications for prophylaxis of Gram-negative bacterial infections for their surgical procedures. In Class D, the combination of cefazolin and gentamicin was identified as the most common type of non-compliance with the correct choice and duration of antimicrobial agent. In Class E, the prescriptions of ampicillin–sulbactam and third-generation cephalosporins were classified as not compliant with the recommended choice of antimicrobial agents. Their indications were identified as treatment of other infections rather than prophylaxis.

Table 4. Choice and duration of prophylactic antimicrobial agents for whole dataset.

Class	No. of Instances	Choice of Antibiotics	Duration of Prophylactic Antimicrobial Use ^a (No. of Instances)
A	Total 255		
	245	Cefazolin	SP1 (178), SP2 (67)
	5	Cefoxitin	SP1
	4	Cefoxitin	SP2
	1	Cefuroxime + metronidazole	SP2
B	Total 155		
	126	Cefazolin	SP3
		Cefazolin	
	8	+ other antibiotic (for different surgical procedures or risks)	SP3
	10	Cefoxitin	SP3
	8	Clindamycin	SP3
C	Total 30		
	25	Cefazolin + gentamicin	SP1 (13), SP2 (12)
	5	Other antibiotics	SP1 (2), SP2 (3)
D	Total 103		
	69	Cefazolin + gentamicin	SP3
	5	Cefoxitin + gentamicin	SP3
	2	Cefoxitin	SP3
	21	Oral cephalixin	SP3
	6	Others	
E	Total 57		
	17	Ampicillin–sulbactam	SP2 (4), SP3 (13)
	10	Third-generation cephalosporins	SP2 (4), SP3 (6)
	30	Others	

^a Single dose encoded as SP1, one day encoded as SP2, and more than one day encoded as SP3.

The performance metrics of machine learning techniques for the whole dataset are shown in Table 5. Classification accuracy alone can be misleading if there is an unequal number of observations in each class (type) in the dataset. Therefore, we had to look at other performance metrics in the Table 5. Multilayer perceptron had the best performance measures. SimpleLogistic was selected by Auto-WEKA as the optimal classifier for the whole dataset. However, the performance measures of SimpleLogistic were the second best among the algorithms in this study. The weighted average sensitivity, specificity, precision, and AUC of decision tree were 0.932, 0.972, 0.931, and 0.985 respectively. Bagging with decision tree, AdaBoost with decision tree, decision tree with SMOTE could increase the performance measures of decision tree, except the AUC of AdaBoost was a little less than that of decision tree. The performance measures of decision tree with SpreadSubsample were lower than those of decision tree alone.

Table 5. Weighted average of performance metrics of algorithms for whole dataset.

	Spread Subsampl- Decision Tree	Decision Tree	Bagging-Decision Tree	AdaBoost- Decision Tree	SMOTE- Decision Tree	Simple- Logistic	Multilayer Perceptron
Sensitivity	0.895	0.932	0.940	0.945	0.960	0.958	0.967
Specificity	0.970	0.972	0.968	0.980	0.990	0.990	0.992
Precision	0.908	0.931	0.941	0.944	0.961	0.959	0.967
AUC ^a	0.978	0.985	0.991	0.983	0.985	0.987	0.992

^a AUC: Area under receiver operating characteristics curve.

The performance metrics (sensitivity, specificity, precision, and AUC) of the seven algorithms for the whole dataset (601 records) were shown in Figure 1. The seven algorithms (decision tree with SpreadSubsample, decision tree, Bagging with decision tree, Adaboost with decision tree, decision tree with SMOTE, SimpleLogistic, and multilayer perceptron) are plotted in increasing order of performance with regard to sensitivity, specificity, and precision. All the algorithms presented with outstanding discrimination, i.e., AUC > 0.97.

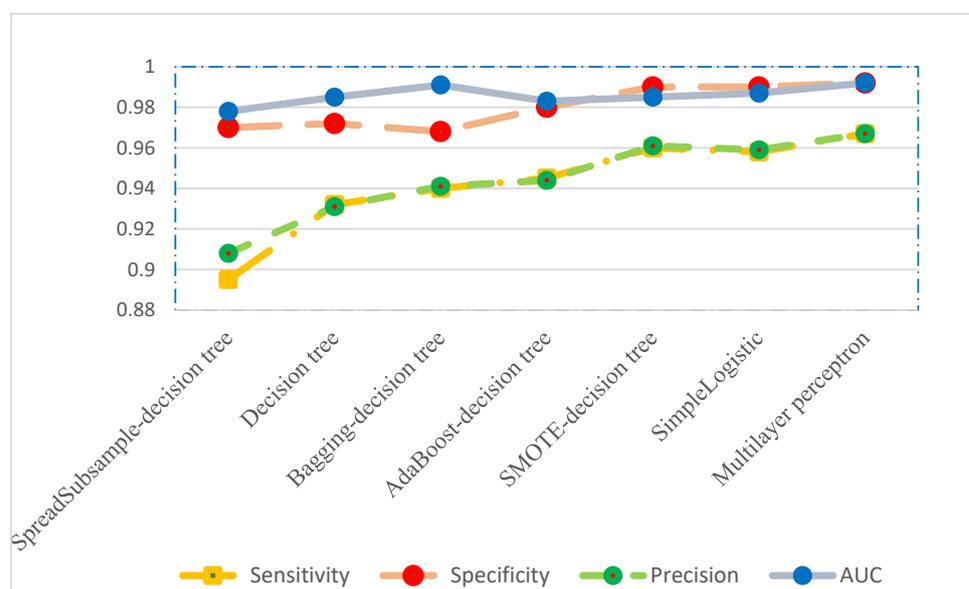


Figure 1. Performance metrics of algorithms for whole dataset (601 instances).

In this study, 544 instances were classified as surgical prophylaxis, 57 instances were classified as treatment of infections rather than surgical prophylaxis. The confusion matrix of decision tree algorithm for class-label attribute is shown in Table 6. Between-class imbalance was observed in this study, i.e., unequal number of observations in each class. Class C (5%) and E (9.5%) were much smaller subsets, as compared with that of Class A (42.6%), B (25.8%), and D (17.1%). There were also unequal sample sizes and minority instances within each labeled class, i.e., within-class imbalance. The sample sizes of minority instances were too small to be identified accurately by the machine learning tools, especially in Class E, as shown in Table 6.

For the 544 instances of antimicrobial prophylaxis (Class A, B, C, and D), 75.6% (411/544) of antimicrobial choices (Class A and B) were classified as consistent with guidelines, 52.6% (286/544) of antimicrobial prescriptions (Class A and C) were discontinued within 24 h after surgery completion.

Five specialties, orthopedics (22.5%), neurosurgery (13.8%), general surgery (14.0%), obstetrics/gynecology (13.1%), and urology (9.7%) had the greatest number of instances. The dataset was further stratified and analyzed by specialty. The compliance with the recommendations of guideline for antimicrobial prophylaxis by the five major specialties is plotted in Figure 2. The compliance with recommended choice of prophylactic antimicrobial agents for neurosurgery, obstetrics and gynecology, general surgery, orthopedics, and

urology were 64.5%, 71.6%, 80.0%, 80.5%, and 81.0%, respectively. The compliance with duration of surgical prophylaxis for neurosurgery, urology, general surgery, orthopedics, and obstetrics and gynecology were 36.8%, 46.6%, 54.7%, 67.9%, and 77.0%, respectively.

Table 6. Confusion matrix of decision tree algorithm for whole dataset.

Predicted Class					Actual Class	Number of Observations (%)
A	B	C	D	E		
255	0	1	0	0	A	256 (42.6)
0	144	0	6	5	B	155 (25.8)
5	0	25	0	0	C	30 (5.0)
1	4	0	98	0	D	103 (17.1)
8	6	0	5	38	E	57 (9.5)
						Total: 601 (100)

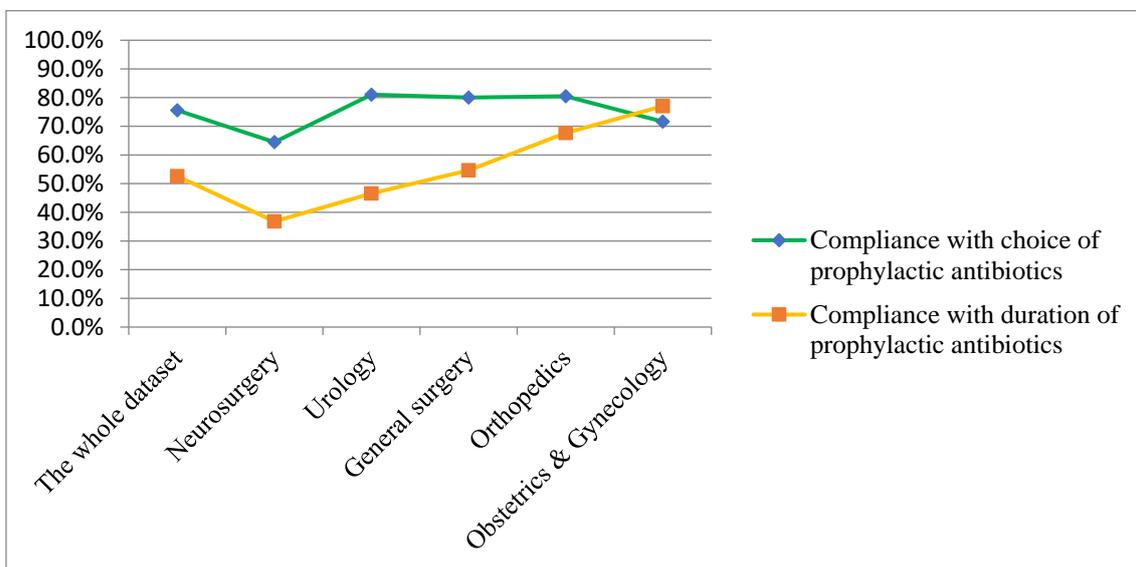


Figure 2. Compliance with recommendations of guidelines for antimicrobial prophylaxis, plotted by 5 major specialties.

The AUC of six algorithms for five major surgical specialties are plotted in Figure 3. Multilayer perceptron was the best algorithms among the five specialties. Bagging with decision tree, decision tree with SMOTE could increase the performance of decision tree for the datasets of most specialties, except for urology dataset. SimpleLogistic worked better than decision tree with sampling methods for the datasets of obstetrics and gynecology and urology, but not for neurosurgery dataset.

The hyperparameters and execution time of machine learning algorithms and manual review are shown in Table 7 (Whole dataset in Supplementary Materials). Auto-WEKA and multilayer perceptron took more execution time as compared with that of decision tree. The execution time of decision tree with SpreadSubsample and SMOTE is longer than that of decision tree alone. All the algorithms were more efficient in execution time than manual review.

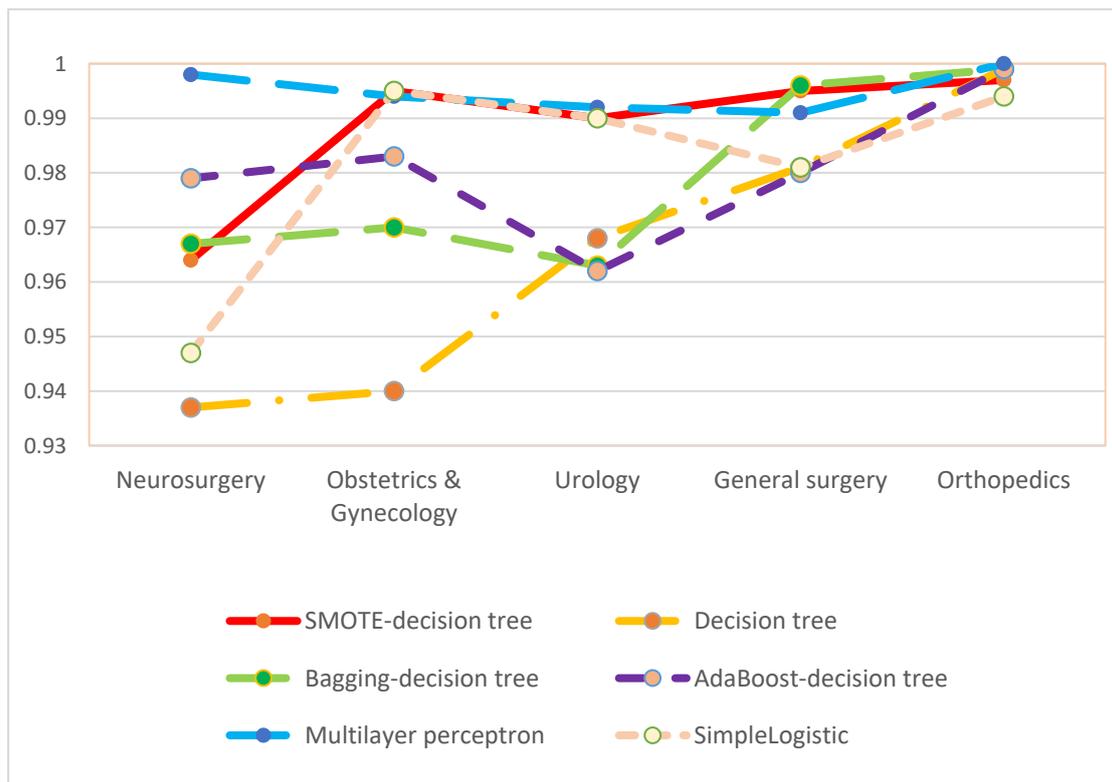


Figure 3. Area under receiver operating characteristics curve (AUC) of 7 algorithms for 5 major specialties.

Table 7. Hyperparameters and execution time for machine learning techniques and manual review.

Classifier	Hyperparameters	Sampling Method	Execution Time (s)
Decision tree	Reduced error pruning = false	No sampling	7.8
	Confidence factor = 0.2	SpreadSubsample	27.9
	distributionSpread: 2.0	SMOTE	71.8
	Percentage: frequency of minor classes adjusted to nearly the same with that of major class	Bagging	16.0
	Classifier Choose: J48	AdaBoost	16.1
SimpleLogistic	Default	No sampling	9.8
Multilayer perceptron	Learning rate = 0.3 Training Time = 500	No sampling	353.1
Auto-WEKA	Default Time = 15 min	No sampling	8586
Manual review			Estimated 24,040

4. Discussion

The National Healthcare Safety Network (NHSN), the CDC surveillance system, provides information on incidence rates for healthcare-associated infections. The NHSN cannot provide national-scale data on the antimicrobial use. Point prevalence surveys provide valuable information on healthcare-associated infections and antimicrobial use globally, and they may facilitate comparisons among different countries [29]. Although the duration of point prevalence survey in a single hospital is limited to three weeks, the criteria of compliance with duration of antimicrobial prophylaxis are less than one day or longer than one day, and the short duration of prevalence survey does not limit the analysis of data or presentation of useful information [17]. The surgical prophylaxis guideline by Bratzler et al. remains a standard and is still cited by the recent publications [7,18].

Among the 7377 patients, 1981 patients were never subjected to a surgical procedure, but only 601 instances with antimicrobial prophylaxis were collected. Because only the antimicrobial agents used on the day of survey were recorded, the dataset did not include the data of antimicrobial use for every patient who had been subjected to a surgical procedure before this survey. Some commentators may suggest that some of these patients should have been subjected to surgical prophylaxis, though they were not. If that is the case, such cases are of major importance for the purposes of this work. This issue was not described in the previous publications [8,9,30]. However, it is not a major concern in real-world practice except for human error. This is because the question, “Has antibiotic prophylaxis been given within the last 60 min before skin incision?”, is an item in the time-out surgical safety checklist proposed by WHO [8,31]. Surgeons will confirm that antibiotic prophylaxis was given within the last 60 min before surgical incision. More importantly, the appropriate use of antimicrobial prophylaxis (choice of antimicrobial agents and duration of antimicrobial use) is not on the checklist and needs to be clarified in the audit. However, we do not have data to verify this issue caused by human error, and we will take it into consideration for a perfect audit in a future study.

In this study, 544 instances were surgical prophylaxis. A total of 75.6% (411/544) of antimicrobial choices (Classes A and B) were classified as consistent with guidelines, while 52.6% (286/544) of antimicrobial prescriptions (Classes A and C) were discontinued within 24 h after surgery completion. The combination of cefazolin and gentamicin was the most common type of noncompliance with guideline. Cefazolin was widely used as an appropriate drug for many surgical procedures with its wide spectrum of susceptibility (e.g., against Gram-positive *Staphylococcus* species and Gram-negative bacilli) and adequate concentrations in the tissue. Therefore, gentamicin is used only if Gram-negative bacteria infection is a concern of postoperative infections and the patient is allergic to cefazolin. In comparison, antimicrobial agents compliant with guidelines were used for 92.6% of patients in U.S. hospitals [7] and 78.4% in Italian hospitals. [9] Prolonged surgical prophylaxis longer than 24 h was common in the worldwide surveys, including 54.2% in the European hospitals, 52.4% in the global survey, 45.7% in Australia hospitals, and 20.7% in US hospitals [8,29,30,32]. Compared with that of surgical prophylaxis used for 24 h or less, prophylactic antimicrobial agent used for more than 24 h for most surgical procedures does not reduce postoperative infections but increases the risk of antimicrobial resistance and adverse effects. In the absence of preoperative infection or postoperative complications, prolonged postoperative antimicrobial prophylaxis is not necessary [2,32]. International collaborative works are needed to improve appropriate antimicrobial use for surgical prophylaxis, such as optimal antibiotic choice, dosage, and length of prophylaxis.

WHO suggests an assessment method for antimicrobial surgical prophylaxis by reviewing the essential variables for evaluation, including indication (type of surgery), age and gender of the patients, comorbidities, antibiotics prescribed as prophylaxis, dose, time of administration, and duration of antimicrobial use [11]. Manual method is not suitable for the audit of enormous data from annual survey in a medical center, interhospital comparison, or national survey. Auditing appropriate antimicrobial use for surgical prophylaxis by manual review of enormous medical data is labor-intensive and time-consuming. It is needed to develop efficient models for the audit of appropriateness of prophylactic antibiotic use. The purpose of this study was to investigate whether machine learning techniques could be used to predict the appropriate antimicrobial use for surgical prophylaxis from a large volume of clinical data. Machine learning techniques facilitate reviewing large volumes of data to discover specific patterns and trends that would usually not be apparent to humans. As shown in the results, with the help of the machine learning techniques, we could easily identify the five types of compliance with recommendations of antimicrobial prophylaxis. In addition, multilayer perceptron, SimpleLogistic, and decision tree algorithms provided high-performance measures for identifying appropriate use (Class A) and inappropriate use (Classes B, C, and D). The appropriateness of prophylactic antibiotic

use by different surgical specialties can also be determined rapidly by the efficient models developed in this study.

Imbalanced dataset contains instances that are unevenly distributed among the different classes. Between-class imbalance means unequal numbers of instances between different classes; within-class imbalance occurs when a class consists of different subclasses with unequal numbers of instances [33]. Both between-class imbalance and within-class imbalance are observed commonly in datasets. Between-class imbalances as well as within-class imbalances can affect classification performance.

The between-class imbalance was observed in the dataset in this study, i.e., unequal distribution in each class. Classes C and E were much smaller subsets, as compared with that of Classes A and D. The performance metrics in Classes C and E were not as good as those of Classes A and D. There were also unequal sample sizes and minority instances within each labeled class, i.e., within-class imbalance. The sample sizes of minority instances were too small to be identified accurately by the machine learning tools, especially in Class E, as shown in Table 6.

Imbalanced dataset is a common problem in machine learning classification. This imbalanced data can prevent the machine learning algorithms from building accurate models for these minority classes and lead to prediction errors [25,34]. For example, SimpleLogistic worked better than decision tree with sampling methods for the datasets of obstetrics and gynecology and urology, but not for neurosurgery dataset. There are several methods to solve this problem of imbalanced data, such as resampling the datasets by under-sampling the majority class and over-sampling the minority class, modifying algorithms, and considering a different perspective, such as anomaly [24,25,34]. We used two resampling approaches (Bagging and AdaBoost) to overcome the problem of imbalanced dataset. Bagging increased the precision for decision tree algorithm in this study. AdaBoost could reduce the bias of decision tree, but could be prone to overfitting noisy, as shown in Classes C and E [25].

Overfitting is a common problem in supervised machine learning algorithms. Overfitting occurs if the model performs perfectly on the training dataset, while fitting poorly on the testing dataset. Overfitting happens because of the presence of noise in the training dataset, small size of dataset, and the complexity of classifiers. Various strategies were proposed to reduce overfitting. The first is data sampling methods, including under-sampling and oversampling [25,35]. SpreadSubsample randomly removes instances from the majority class to make the dataset balanced. However, because under-sampling method could lose useful data for classifiers, decision tree with SpreadSubsample did not achieve better performance than decision tree alone in this study. SMOTE is an oversampling method that has the advantage of no loss of useful data. SMOTE could improve the performance measures of decision tree in this study, as well as other studies [36,37]. However, SMOTE may require additional computational time if the dataset is very large, as shown in Table 7, SMOTE took more execution time as compared with decision alone or SpreadSubsample [25,37]. The other methods for reducing overfitting are cross-validation, ensemble methods, as shown in this study [25,35]. However, overfitting cannot be avoided completely. As with most applications in data science and machine learning algorithms, there is no definitive best approach that always performs excellently. Depending on the characteristics of the dataset, distribution of classes, models, and predictions, some of the above algorithms will perform better than others.

It takes about 40 s to review the data of a single instance manually. The estimated execution time of manual review for all the 601 instances will be 24,040 s (6 h, 40 min and 40 s). However, the actual execution time of manual review will exceed 24,040 s, because we cannot continuously concentrate on reviewing the data without rest. Although they have better performance measures, Auto-WEKA and multilayer perceptron require more execution time as compared with that of other algorithms. All the algorithms were more efficient in execution time than manual review.

Finally, this study demonstrated that multilayer perceptron, SimpleLogistic, and decision tree had outstanding discrimination with $AUC > 0.9$ for identifying the appropriateness of surgical prophylaxis, despite the imbalanced data. However, we still consider a different perspective that the rare instances are anomalies in the dataset. The antimicrobial use of these rare instances was inappropriate as compared to those regular practices which were according to the recommendations of antimicrobial prophylaxis guidelines. With the models determined by all the algorithms, we can easily identify the anomalies or outliers.

There are some limitations in this study. First, for the instances with active infection already present at the time of surgery, the antimicrobial agents could then be selected by clinicians according to the clinical data, such as source of infection or microbiology. These antibiotics were used for active infections rather than surgical prophylaxis. Although these instances could be identified as inappropriate use, the models that used machine learning techniques to audit antimicrobial prophylaxis do not override clinical judgment. Once the surgeons obtain the information from the audit, they can review the cases of nonadherence to guideline. Lack of documentation of the reason of antimicrobial prescription is not uncommon (about 20% in European hospitals) [32]. If the variances from standards arise from lack of documentation of pre-existing infection and treatment for infection rather surgical prophylaxis, they can improve the documentation of treatment or surgical prophylaxis on the medical records. In the future study, the instances with treatment for preexisting infection will be excluded to reduce the variances from standards. Second, the small population of outliers used in training set may not sufficiently represent the outlier patterns. The small amount of outlier samples can limit the capability of building an accurate classifier for the detection of outlier patterns, and it may also result in misclassification. Third, before administering any antibiotic prophylaxis regimen, it is necessary to define whether the patient really needs such prophylaxis. Antimicrobial prophylaxis is not recommended for most clean surgical procedures in patients without risk of postoperative infection, such as thyroidectomy or clean plastic surgery [7]. However, the risks of infection for these clean procedures were not included in the attributes. Consensus on this issue is required from the surgical society as well as adding infection risks of surgery to the attributes. Only then will the algorithms be able to provide more useful information for assessing the appropriate antimicrobial use in future studies. Fourth, some features considered in this study were not relevant. Irrelevant features can have an impact on the model performance. Regularization strategy is proposed to enhance the model performance by removing the useless features and selecting only the useful features from the model, and minimizing the weights of the features with little useful information on the classification [35]. In future work, we will focus on the database design and configuration of more useful data to improve the performance of the algorithms. Fifth, we are concerned about the issue of inappropriate surgical prophylaxis or limited antimicrobial options leading to the selection of resistant bacteria. Some strategies were proposed to prevent the emergence of resistant bacteria, such as combination regimens [38,39] and adequate dosing to maintain the serum concentration of antibiotic above the minimal inhibitory concentration of the bacteria during the treatment [39,40]. Machine learning may be used to deal with this issue, which involves alternative surgical prophylactic regimens and emergence of resistant bacteria, based on the hypothesis that the new resistant bacteria may be prevented or treated with new prophylactic regimens. Reinforcement learning is an approach that seems suitable for such a problem. However, more data on antimicrobial resistant patterns and alternative options for surgical prophylactic antimicrobial agents need to be collected before validating reinforcement learning for this issue in future studies.

5. Conclusions

Multilayer perceptron, SimpleLogistic, and decision tree algorithms have outstanding performance measures for identifying the appropriateness of surgical prophylaxis. The Bagging and SMOTE algorithms can improve the predictive power of decision tree classifiers against imbalanced datasets. The efficient models can be used to assist the antimicrobial

stewardship team in the audit of surgical antimicrobial prophylaxis. In future work, we still have the challenges and opportunities of investigating the database configuration and enriching our datasets with more useful clinical information to improve the performance of the algorithms.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app12052586/s1>.

Author Contributions: Data collection, H.-T.C. and H.-M.H.; conceptualization, analysis of data, and writing the original manuscript: Z.-Y.S.; review and editing of the manuscript: J.-S.H.; supervision: C.-Y.C. and J.-S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was granted by Taiwan Centers for Disease Control (Project no. DOH102-DC-1502 from 1 January 2013 to 31 December 2013; MOHW103-CDC-C-114-122501 from 1 January 2014 to 31 December 2014).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by Institutional Review Board of Taichung Veterans General Hospital (protocol code SE13130 on 17 May 2013 and SE13130#2 on 19 May 2014).

Informed Consent Statement: Patient consent was waived by IRB according to the following criteria: 1. The research is an assessment of a public policy, conducted either by a governmental institution or by an outsourced professional organization. 2. The research presents no more than minimal risk to the participants. The risk to the participants is not higher than non-participants. The research cannot practically be carried out without the waiver. The waiver does not adversely affect the rights and welfare of the participants.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ban, K.A.; Minei, J.P.; Laronga, C.; Harbrecht, B.G.; Jensen, E.H.; Fry, D.E.; Itani, K.M.; Dellinger, E.P.; Ko, C.Y.; Duane, T.M. American College of Surgeons and Surgical Infection Society: Surgical Site Infection Guidelines, 2016 Update. *J. Am. Coll. Surg.* **2017**, *224*, 59–74. [[CrossRef](#)] [[PubMed](#)]
2. World Health Organization. *Global Guidelines for the Prevention of Surgical Site Infection*, 2nd ed.; World Health Organization: Geneva, Switzerland, 2018. Available online: <https://www.who.int/publications/i/item/global-guidelines-for-the-prevention-of-surgical-site-infection-2nd-ed> (accessed on 6 December 2021).
3. Berrios-Torres, S.I.; Umscheid, C.A.; Bratzler, D.W.; Leas, B.; Stone, E.C.; Kelz, R.R.; Reinke, C.E.; Morgan, S.; Solomkin, J.S.; Mazuski, J.E.; et al. Centers for Disease Control and Prevention Guideline for the Prevention of Surgical Site Infection, 2017. *JAMA Surg.* **2017**, *152*, 784–791. [[CrossRef](#)] [[PubMed](#)]
4. McGee, M.F.; Kreutzer, L.; Quinn, C.M.; Yang, A.; Shan, Y.; Halverson, A.L.; Love, R.; Johnson, J.K.; Prachand, V.; Bilimoria, K.Y. Leveraging a Comprehensive Program to Implement a Colorectal Surgical Site Infection Reduction Bundle in a Statewide Quality Improvement Collaborative. *Ann. Surg.* **2019**, *270*, 701–711. [[CrossRef](#)] [[PubMed](#)]
5. Kefale, B.; Tegegne, G.T.; Degu, A.; Molla, M.; Kefale, Y. Surgical Site Infections and Prophylaxis Antibiotic Use in the Surgical Ward of Public Hospital in Western Ethiopia: A Hospital-Based Retrospective Cross-Sectional Study. *Infect Drug Resist.* **2020**, *13*, 3627–3635. [[CrossRef](#)]
6. Purba, A.K.R.; Setiawan, D.; Bathoorn, E.; Postma, M.J.; Dik, J.H.; Friedrich, A.W. Prevention of Surgical Site Infections: A Systematic Review of Cost Analyses in the Use of Prophylactic Antibiotics. *Front. Pharm.* **2018**, *9*, 776. [[CrossRef](#)]
7. Bratzler, D.W.; Dellinger, E.P.; Olsen, K.M.; Perl, T.M.; Auwaerter, P.G.; Bolon, M.K.; Fish, D.N.; Napolitano, L.M.; Sawyer, R.G.; Slain, D.; et al. Clinical practice guidelines for antimicrobial prophylaxis in surgery. *Am. J. Health Syst. Pharm.* **2013**, *70*, 195–283. [[CrossRef](#)]
8. Ierano, C.; Thursky, K.; Marshall, C.; Koning, S.; James, R.; Johnson, S.; Imam, N.; Worth, L.J.; Peel, T. Appropriateness of Surgical Antimicrobial Prophylaxis Practices in Australia. *JAMA Netw. Open* **2019**, *2*, e1915003. [[CrossRef](#)]
9. Tiri, B.; Bruzzone, P.; Priante, G.; Sensi, E.; Costantini, M.; Vernelli, C.; Martella, L.A.; Francucci, M.; Andreani, P.; Mariottini, A.; et al. Impact of Antimicrobial Stewardship Interventions on Appropriateness of Surgical Antibiotic Prophylaxis: How to Improve. *Antibiotics* **2020**, *9*, 168. [[CrossRef](#)]
10. Centers for Disease Control and Prevention. *Core Elements of Hospital Antibiotic Stewardship Programs*; US Department of Health and Human Services, CDC: Atlanta, GA, USA, 2019. Available online: <https://www.cdc.gov/antibiotic-use/healthcare/pdfs/hospital-core-elements-H.pdf> (accessed on 6 December 2021).

11. World Health Organization. *Antimicrobial Stewardship Programmes in Health-Care Facilities in Low- and Middle-Income Countries: A Practical Toolkit*; World Health Organization: Geneva, Switzerland, 2019. Available online: <https://www.who.int/publications/item/9789241515481> (accessed on 6 December 2021).
12. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Morgan Kaufmann: Burlington, MA, USA, 2012.
13. Peiffer-Smadja, N.; Rawson, T.M.; Ahmad, R.; Buchard, A.; Georgiou, P.; Lescure, F.X.; Birgand, G.; Holmes, A.H. Machine learning for clinical decision support in infectious diseases: A narrative review of current applications. *Clin. Microbiol. Infect.* **2020**, *26*, 584–595. [[CrossRef](#)]
14. Bote-Curiel, L.; Muñoz-Romero, S.; Gerrero-Curienes, A.; Rojo-Álvarez, J.L. Deep Learning and Big Data in Healthcare: A Double Review for Critical Beginners. *Appl. Sci.* **2019**, *9*, 2331. [[CrossRef](#)]
15. Feretzakis, G.; Loupelis, E.; Sakagianni, A.; Kalles, D.; Martsoukou, M.; Lada, M.; Skarmoutsou, N.; Christopoulos, C.; Valakis, K.; Velentza, A.; et al. Using Machine Learning Techniques to Aid Empirical Antibiotic Therapy Decisions in the Intensive Care Unit of a General Hospital in Greece. *Antibiotics* **2020**, *9*, 50. [[CrossRef](#)] [[PubMed](#)]
16. Martínez-Agüero, S.; Mora-Jiménez, I.; Lérída-García, J.; Álvarez-Rodríguez, J.; Soguero-Ruiz, C. Machine Learning Techniques to Identify Antimicrobial Resistance in the Intensive Care Unit. *Entropy* **2019**, *21*, 603. [[CrossRef](#)] [[PubMed](#)]
17. European Centre for Disease Prevention and Control. *Point Prevalence Survey of Healthcare-associated Infections and Antimicrobial Use in European Acute Care Hospitals—Protocol Version 4.3*; ECDC: Stockholm, Sweden, 2012. Available online: <https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/0512-TED-PPS-HAI-antimicrobial-use-protocol.pdf> (accessed on 6 December 2021).
18. Anderson, D.J.; Sexton, D.J. Antimicrobial Prophylaxis for Prevention of Surgical Site Infection in Adults. Uptodate. 2021. Available online: <https://www.uptodate.com/contents/antimicrobial-prophylaxis-for-prevention-of-surgical-site-infection-in-adults> (accessed on 6 December 2021).
19. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed.; Morgan Kaufmann: Burlington, MA, USA, 2017.
20. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
21. Kotthoff, L.; Thornton, C.; Hoos, H.H.; Hutter, F.; Leyton-Brown, K. Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
22. Landwehr, N.; Hall, M.; Frank, E. Logistic Model Trees. *Mach. Learn.* **2005**, *59*, 161–205. [[CrossRef](#)]
23. Quinlan, J.R. *C4.5. Programs for Machine Learning*; Morgan Kaufmann: San Francisco, CA, USA, 1993.
24. Rokach, L. *Ensemble Learning: Pattern Classification Using Ensemble Methods*, 2nd ed; World Scientific Publishing Company: Singapore, 2019.
25. Sánchez-Hernández, F.; Ballesteros-Herráez, J.C.; Kraiem, M.S.; Sánchez-Barba, M.; Moreno-García, M.N. Predictive Modeling of ICU Healthcare-Associated Infections from Imbalanced Data. Using Ensembles and a Clustering-Based Undersampling Approach. *Appl. Sci.* **2019**, *9*, 5287. [[CrossRef](#)]
26. Mandrekar, J.N. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *J. Thorac. Oncol.* **2010**, *5*, 1315–1316. [[CrossRef](#)]
27. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multi-Class Classification: An Overview. *arXiv* **2020**, arXiv:abs/2008.05756. Available online: <https://arxiv.org/pdf/2008.05756.pdf> (accessed on 6 December 2021).
28. Wandishin, M.S.; Mullen, S.J. Multiclass ROC Analysis. *Weather Forecast.* **2009**, *24*, 530–547. [[CrossRef](#)]
29. Magill, S.S.; O’Leary, E.; Ray, S.M.; Kainer, M.A.; Evans, C.; Bamberg, W.M.; Johnston, H.; Janelle, S.J.; Oyewumi, T.; Lynfield, R.; et al. Antimicrobial Use in US Hospitals: Comparison of Results from Emerging Infections Program Prevalence Surveys, 2015 and 2011. *Clin. Infect. Dis.* **2021**, *72*, 1784–1792. [[CrossRef](#)]
30. Plachouras, D.; Kärki, T.; Hansen, S.; Hopkins, S.; Lyytikäinen, O.; Moro, M.L.; Reilly, J.; Zarb, P.; Zingg, W.; Kinross, P.; et al. Antimicrobial use in European acute care hospitals: Results from the second point prevalence survey (PPS) of healthcare-associated infections and antimicrobial use, 2016 to 2017. *Eurosurveillance* **2018**, *23*, 1800393. [[CrossRef](#)] [[PubMed](#)]
31. World Health Organization. Surgical Safety Checklist. Available online: <https://www.who.int/teams/integrated-health-services/patient-safety/research/safe-surgery/tool-and-resources> (accessed on 6 December 2021).
32. Versporten, A.; Zarb, P.; Caniaux, I.; Gros, M.F.; Drapier, N.; Miller, M.; Jarlier, V.; Nathwani, D.; Goossens, H. Antimicrobial consumption and resistance in adult hospital inpatients in 53 countries: Results of an internet-based global point prevalence survey. *Lancet Glob. Health* **2018**, *6*, e619–e629. [[CrossRef](#)]
33. Agrawal, A.; Viktor, H.L.; Paquet, E. SCUT: Multi-class imbalanced data classification using SMOTE and cluster-based under-sampling. In Proceedings of the 2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), Lisbon, Portugal, 12–14 November 2015; pp. 226–234.
34. Kraiem, M.S.; Sánchez-Hernández, F.; Moreno-García, M.N. Selecting the Suitable Resampling Strategy for Imbalanced Data Classification Regarding Dataset Properties. An Approach Based on Association Models. *Appl. Sci.* **2021**, *11*, 8546. [[CrossRef](#)]
35. Ying, X. An Overview of Overfitting and its Solutions. *J. Phys. Conf. Ser.* **2019**, *1168*, 022022. [[CrossRef](#)]
36. Davagdorj, K.; Lee, J.S.; Pham, V.H.; Ryu, K.H. A Comparative Analysis of Machine Learning Methods for Class Imbalance in a Smoking Cessation Intervention. *Appl. Sci.* **2020**, *10*, 3307. [[CrossRef](#)]

37. Yıldırım, P. Pattern classification with imbalanced and multiclass data for the prediction of albendazole adverse event outcomes. *Procedia Comput. Sci.* **2016**, *83*, 1013–1018. [[CrossRef](#)]
38. Angst, D.C.; Tepekule, B.; Sun, L.; Bogos, B.; Bonhoeffer, S. Comparing treatment strategies to reduce antibiotic resistance in an in vitro epidemiological setting. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2023467118. [[CrossRef](#)]
39. Menz, B.D.; Charani, E.; Gordon, D.L.; Leather, A.J.M.; Moonesinghe, S.R.; Phillips, C.J. Surgical Antibiotic Prophylaxis in an Era of Antibiotic Resistance: Common Resistant Bacteria and Wider Considerations for Practice. *Infect. Drug Resist.* **2021**, *14*, 5235–5252. [[CrossRef](#)]
40. Paterson, I.K.; Hoyle, A.; Ochoa, G.; Baker-Austin, C.; Taylor, N.G.H. Optimising Antibiotic Usage to Treat Bacterial Infections. *Sci. Rep.* **2016**, *6*, 37853. [[CrossRef](#)]