International Journal of Antimicrobial Agents 62 (2023) 106966

Contents lists available at ScienceDirect

International Journal of Antimicrobial Agents

journal homepage: www.elsevier.com/locate/ijantimicag

Translation of Machine Learning-Based Prediction Algorithms to Personalised Empiric Antibiotic Selection: A Population-Based Cohort Study

Chungsoo Kim[a], Young Hwa Choi[b], Jung Yoon Choi[c], Hee Jung Choi[d,e], Rae Woong Park[a,f,**], Sandy Jeong Rhie[c,g,*]

[a] Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Gyeonggi-do, Republic of Korea
[b] Department of Infectious Diseases, Ajou University School of Medicine, Suwon, Republic of Korea
[c] Graduate School of Pharmaceutical Sciences, Ewha Womans University, Seoul, Republic of Korea
[d] College of Medicine, Ewha Womans University, Seoul, Republic of Korea
[e] Department of Internal Medicine, Ewha Womans University Mokdong Hospital, Seoul, Republic of Korea
[f] Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Republic of Korea
[g] College of Pharmacy, Ewha Womans University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 1 December 2022
Accepted 3 September 2023

Editor: Dr Y-W Lin

Keywords:
Machine-learning
Prediction models
Non-susceptibility
Empiric antibiotics
Urinary tract infection

ABSTRACT

Background: Prediction of antibiotic non-susceptibility based on patient characteristics and clinical status may support selection of empiric antibiotics for suspected hospital-acquired urinary tract infections (HA-UTIs).

Methods: Prediction models were developed to predict non-susceptible results of eight antibiotic susceptibility tests ordered for suspected HA-UTI. Eligible patients were those with urine culture and susceptibility test results after 48 hours of admission between 2010–2021. Patient demographics, diagnosis, prescriptions, exposure to multidrug-resistant organisms, transfer history, and a daily calculated antibiogram were used as predictors. Lasso logistic regression (LLR), extreme gradient boosting (XGB), random forest, and stacked ensemble methods were used for development. Parsimonious models were also developed for clinical utility. Discrimination was assessed using the area under the receiver operating characteristic curve (AUROC).

Results: In 10 474 suspected HA-UTI cases, the mean age was 62.1 ± 16.2 years and 48.1% were male. Non-susceptibility prediction for ampicillin/sulbactam, cefepime, ciprofloxacin, imipenem, piperacillin/tazobactam, and trimethoprim/sulfamethoxazole performed best using the stacked ensemble (AUROC 76.9, 76.1, 77.0, 80.6, 76.1, and 76.5, respectively). The model for ampicillin performed best with LLR (AUROC 73.4). Extreme gradient boosting only performed best for gentamicin (AUROC 66.9). In the parsimonious models, the LLR yielded the highest AUROC for ampicillin, ampicillin/sulbactam, cefepime, gentamicin, and trimethoprim/sulfamethoxazole (AUROC 70.6, 71.8, 73.0, 65.9, and 73.0, respectively). The model for ciprofloxacin performed best with XGB (AUROC 70.3), while the model for imipenem performed best in the stacked ensemble (AUROC 71.3). A personalised application using the parsimonious models was publicly released.

Conclusions: Prediction models for antibiotic non-susceptibility were developed to support empiric antibiotic selection for HA-UTI.

Introduction

Hospital-acquired urinary tract infections (HA-UTIs) account for 40% of hospital-acquired infections [1,2]. Serious morbidity from HA-UTI remains high, and HA-UTI may often lead to urosepsis and septic shock, which can be fatal [3]. Antibiotic therapy is imperative for the treatment of HA-UTIs; therefore, early initiation of appropriate antibiotic treatment is given on an empiric ba-

* Corresponding author: College of Pharmacy, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Republic of Korea. Tel: +82-2-3277-3023; (Sandy Jeong Rhie).
** Co-corresponding author: Department of Biomedical Informatics, Ajou University School of Medicine, Suwon 16499, Gyeonggi-do, Republic of Korea. Tel: +82-31-219-4470; (Rae Woong Park).
E-mail addresses: veritas@ajou.ac.kr (R.W. Park), sandy.rhie@ewha.ac.kr (S.J. Rhie).

https://doi.org/10.1016/j.ijantimicag.2023.106966
0924-8579/

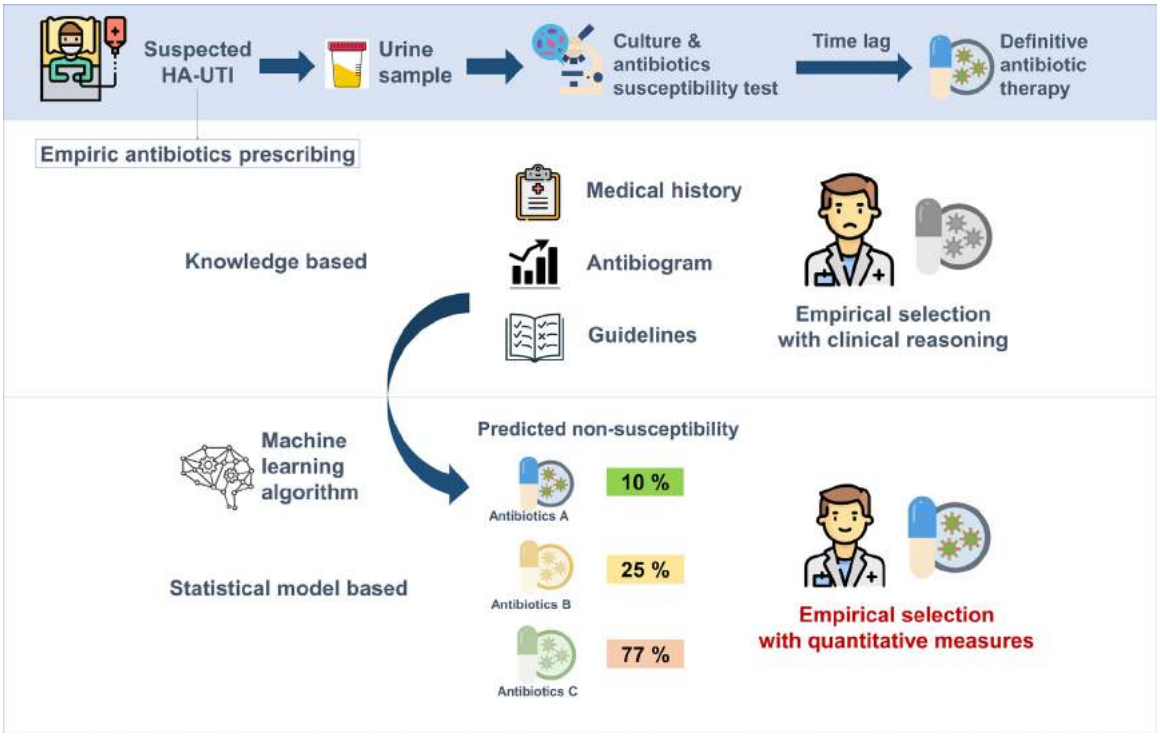

Figure 1. Decision support scheme for empirical antibiotics selection based on machine-learning model.

sis [4]. Broad-spectrum antibiotics—such as ceftriaxone, cefepime (CFP), piperacillin-tazobactam (PPT), meropenem, imipenem (IPM), ciprofloxacin, and levofloxacin—are the most commonly recommended empiric antimicrobial selections for treating HA-UTI [5]. Decisions over which empiric antibiotic to prescribe remain challenging, given the lack of head-to-head trials, especially in the contexts of multiple patient-specific factors and continuous changes in institutional antimicrobial resistance patterns [6]. Which empiric antibiotic to use relies on clinical decisions, with limited clinical evidence only focused on the average treatment effect. This may result in suboptimal and unfavourable practices like a worse early clinical response, longer hospital stay, and eventual increasing resistance [7,8].

To resolve such challenges in infectious diseases several studies have reported on the development of prediction algorithms for personalised antimicrobial treatment [9–12]. However, these studies had their own limitations, for example: some studies were conducted for all infectious diseases regardless of their classification [12] or showed insufficient performance [9–11]. When developing a clinical prediction model, it is important to set up the prediction problem using available variables in a circumstance that resembles an actual situation [13]. However, most previous studies have developed prediction models using key variables deemed to be clinically relevant, and thus have not fully considered all predictors recommended by clinical guidelines – just real-world decisions. Local antibiogram information is easily overlooked [14,15]; consequently, it was impossible to find a model to be applied in the settings of HA-UTI patients, and the translation of these prediction models into clinical settings has been weak.

This study aimed to develop prediction models to estimate the probability of antibiotic non-susceptibility in patients with suspected HA-UTIs. Further, the models were translated to clinical applications via explainable features to propose patient-specific empirical antibiotics (Figure 1).

Methods

Study Setting and Definition

This study was conducted with a retrospective population-based analysis using an electronic health record database from Ajou University School of Medicine, South Korea, which is formatted into the Observational Medical Outcomes Partnership (OMOP) common data model version 5.3.1 [16].

The study population included patients who were aged ≥ 18 years and thought to have HA-UTIs managed between 2010–2021 at Ajou University Medical Center. Patients with suspected HA-UTI were defined as those who had first urine culture and antibiotic susceptibility (C&S) tests ordered at least 48 hours after being admitted (index date). Each qualified hospitalisation was counted as a separate index admission. Patients who had urine C&S tests within 30 days before the index date were excluded to ensure that the first case was tested in the hospital.

Outcomes were non-susceptibility (including resistant and intermediate) to eight antibiotics among bacterial colonies grown from the urine culture samples. Cases in which the colony did not grow in the sample or had no test result (including missing) were excluded. These target antibiotics were ampicillin (AMP), ampicillin/sulbactam (AMS), CFP, ciprofloxacin (CIP), gentamicin (GEN), IPM, PPT, and sulfamethoxazole/trimethoprim (SXT). These antibiotics are commonly tested for susceptibility at this medical centre when HA-UTI is suspected.

Model Development and Validation

The models were developed using machine learning algorithms, including Lasso logistic regression (LLR), extreme gradient boosting (XGB), random forest (RF), and stacked ensemble method [17,18]. The stacked model was developed using two-level architecture consisting of level 0 and level 1 models. The LLR, XGB, and RF

models were used as level 0 models, and XGB was used for the level 1 model. All hyperparameters were optimised by the grid searching method.

The data were split at a ratio of 75:25 for the training and test sets, respectively. Three-fold cross-validation was used in the training process. The performance of the models was evaluated based on the accuracy, discrimination, and calibration of the final models using the test set [19]. The accuracy of the model was calculated as the percentage of the total correctly classified (the sum of true positive and true negative per total case). Model discrimination was evaluated by calculating the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC). Model calibration was evaluated using the slopes of the calibration plots according to the Hosmer-Lemeshow method [20]. The final models for each antibiotic were decided according to their discrimination based on the AUROC.

Model Specification

Candidate predictors were demographics (age, sex), diagnostic codes, medications, observations (obesity, smoking, etc.), procedure, and visit records. In particular, previous antibiotic exposure, transfer records (whether transferred from other tertiary healthcare systems or skilled nursing facilities), previous exposure to multidrug-resistant organisms (MDROs) and daily-calculated antibiotic non-susceptibility rates were included (Supplementary Table E1). All predictors, except age and antibiotic non-susceptibility rates, were dichotomised as binary variables. The diagnoses were differentiated according to clinical aspects (e.g. diagnostic codes were characterised with different time periods). Comorbid diagnoses (e.g. hypertension, diabetes mellitus, hyperlipidaemia) were extracted from between the index date and 1 year before; however, the current conditions (e.g. pneumonia, sepsis) were defined based on diagnostic codes at the index date only. Previous exposure to antibiotics and corticosteroids was included, and antibiotics were differentiated according to their classifications, formulations, and time periods.

To avoid inaccurate predictions due to variations in the distribution of historical antibiotic resistance rates, the concept of daily-calculated non-susceptibility rates was introduced and incorporated into these models. Similar to the traditional antibiogram, non-susceptibility rates were defined as the incidence rates of resistance (R) and intermediate susceptibility (I) within 90 days before the index date. Non-susceptibility rates were determined for 11 bacterial species that commonly cause UTIs: *Escherichia coli, Klebsiella pneumoniae, Enterobacter aerogenes, Enterobacter cloacae, Serratia marcescens, Acinetobacter baumannii, Pseudomonas aeruginosa, Enterococcus faecalis, Enterococcus faecium, Staphylococcus aureus*, and coagulase-negative staphylococci (Supplementary Table E2). All calculations were conducted on a daily basis at a hospital level. Only covariates with missing values $< 10\%$ were included as predictors, and missing values were imputed by the exponential weighted moving average method [21].

Models using two feature selection strategies were developed: full models and parsimonious models. The full models included all available features; however, this approach has less applicability and utility for clinical settings so parsimonious models with selected features were also developed. Feature selections for the parsimonious models were based on the variable importance (β coefficient) of the Lasso shrinkage logistic regression model. Features with strong correlations (Pearson's coefficient > 0.7) were removed to resolve multicollinearities and validated with the variance inflation factor < 10.0. The additional step for the feature selection was conducted by experts with clinical knowledge for improving biological plausibility. If a strong correlation was found between variables, the variable with the more comprehensive information was selected. The variable importance was calculated by the Shapley Additive exPlanations (SHAP) values [22], which is a model agnostic method.

Development of Personalised Insusceptibility Estimator

The Personalised Insusceptibility Estimator (PIE) application was developed for clinical translation, based on parsimonious models in the form of an interactive R shiny. This application took baseline covariates as inputs and returned the estimated probability of non-susceptibilities for each antibiotic agent. For the feasibility test of application, it was intended to further investigate the probability of non-susceptibility by cumulatively adding one variable to the baseline demographic to see how the probability changed with the accumulation of predictors in four clinical scenarios. The non-susceptibility changes per variable added on or off were also determined (like an ablation test).

All analyses were conducted using R, version 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria) and its open-source statistical packages, including the Health Analytics Data-to-Evidence Suites packages of Observational Health Data Sciences and Informatics.

Results

Population Characteristics

There were 10 474 suspected HA-UTI cases; the mean patient age was 62.1 ± 16.2 years and 48.1% were male. The median and interquartile range of length of hospital stay was 12.0 (18.0) days. The summary information for each antibiotic susceptibility test is presented in Table 1. The numbers of urine samples for susceptibility tests to each antibiotic were as follows: 5025 for AMP, 4701 for AMS, 3771 for CFP, 5929 for CIP, 6084 for GEN, 7572 for IMP, 3762 for PPT, and 5402 for STX. Ampicillin (48.2%) yielded the highest non-susceptible rate, and *Escherichia coli* (44.2%) was the most non-susceptible species to AMP.

Model Specification

The full models were developed using ca. 20 000 predictors, including demographics, diagnosis, medications, etc. The detailed number of predictors is presented in Supplementary Table E3. After the feature selection process for the development of the parsimonious models, 140 variables were finally included as model predictors.

The final predictors included demographics (sex and age), diagnosis, drugs, procedures, visit records, and calculated non-susceptibility rate. The detailed final predictors of the parsimonious models are presented in Supplementary Table E1. The calculated non-susceptibility rates were also included in the models. Seventy-six non-susceptibility rates (one for each antibiotic-microorganism pair) were included (Supplementary Table E2). Variable importance, according to the SHAP value of each predictor, was plotted for each model (Figure 2, Supplementary Figure E1). Age, sex, and previous exposure to antibiotics were commonly positioned within the top 15 predictors of non-susceptibility in the CIP and PPT models. The calculated non-susceptibility rates were mainly selected as important predictors in the LLR models. The overall figures for all parsimonious models are presented in Supplementary Figure E1.

Model Performance

Table 2 presents the discrimination performance for the full and parsimonious models for each antibiotic susceptibility test.

Table 1
Baseline patient characteristics and urine culture results.

Variables	AMP (n = 5025)	AMS (n = 4701)	CFP (n = 3771)	CIP (n = 5929)	GEN (n = 6084)	IMP (n = 7572)	PPT (n = 3762)	SXT (n = 5402)
Age, mean ± SD	64.6 ± 15.6	64.4 ± 15.8	63.8 ± 16.1	63.5 ± 16.2	62.9 ± 16.2	62.5 ± 16.1	63.8 ± 16.1	63.1 ± 16.3
Female, *n* (%)	3017 (60.0)	2658 (56.5)	1648 (56.3)	3065 (51.7)	3103 (51.0)	4176 (55.2)	2118 (56.3)	2769 (51.3)
Length of stay, days, median (IQR)	11.0 (15.0)	12.0 (16.0)	13.0 (17.0)	12.0 (16.0)	12.0 (17.0)	13.0 (19.0)	13.0 (18.0)	12.0 (17.0)
Cancer, *n* (%)	1937 (38.5)	1737 (36.9)	1173 (31.1)	2186 (36.9)	2145 (35.3)	2612 (34.5)	1305 (34.7)	1930 (35.7)
Type 2 diabetes mellitus, *n* (%)	1211 (24.1)	1092 (23.2)	795 (21.1)	1376 (23.2)	1436 (23.6)	1655 (21.9)	887 (23.6)	1276 (23.6)
Pneumonia, *n* (%)	689 (13.7)	715 (15.2)	489 (13.0)	944 (15.9)	970 (15.9)	1193 (15.8)	592 (15.7)	872 (16.1)
Previous antibiotic use, *n* (%)	4496 (89.5)	4242 (90.2)	2981 (90.4)	5336 (90.0)	5522 (90.1)	6705 (88.5)	3420 (90.9)	4861 (90.0)
Transferred, *n* (%)	6 (0.1)	6 (0.1)	4 (0.1)	5 (0.1)	6 (0.1)	8 (0.1)	5 (0.1)	4 (0.1)
Previous MDRO exposure, *n* (%)								
CRE	38 (0.8)	41 (0.9)	33 (1.0)	40 (0.7)	42 (0.7)	52 (0.7)	40 (1.1)	38 (0.7)
ESBL	306 (6.1)	313 (6.7)	257 (7.8)	341 (5.8)	354 (5.8)	487 (6.4)	285 (7.6)	317 (5.9)
MRSA	367 (7.3)	421 (9.0)	329 (10.0)	579 (9.8)	610 (10.0)	871 (11.5)	383 (10.2)	558 (10.3)
VRE	145 (2.9)	136 (2.9)	93 (2.8)	145 (2.4)	143 (2.4)	254 (3.4)	108 (2.9)	127 (2.4)
Non-susceptibility, *n* (%)	2421 (48.2)	758 (16.1)	878 (23.3)	1745 (29.4)	998 (16.4)	842 (11.1)	486 (12.9)	1493 (27.6)
Prevalent pathogen								
In all samples, (%)	*E.coli* (23.0)	*E.coli* (24.6)	*E.coli* (33.4)	*E.coli* (18.6)	*E.coli* (19.0)	*E.coli* (30.2)	*E.coli* (30.7)	*E.coli* (20.4)
In insusceptible samples, (%)	*E.coli* (44.2)	*E. faecium* (56.9)	*E.coli* (60.6)	*E.coli* (36.4)	*E.coli* (46.5)	*E. faecium* (55.2)	*E.coli* (29.6)	*E.coli* (36.6)

Abbreviations: AMP, ampicillin; AMS, ampicillin and sulbactam; CFP, cefepime; CIP, ciprofloxacin; GEN, gentamicin; PPT, piperacillin and tazobactam; SXT, trimethoprim and sulfamethoxazole; SD, standard deviation; IQR, interquartile range; MDRO, multidrug-resistant organisms; CRE, carbapenem-resistant Enterobacteriaceae; *E. coli, Escherichia coli; E. faecium, Enterococcus faecium*; ESBL, extended-spectrum beta-lactamase–producing Enterobacteriaceae; MRSA, methicillin-resistant *Staphylococcus aureus*; VRE, vancomycin-resistant enterococci.
Cancer, pneumonia, and type 2 diabetes mellitus were determined within 365 days before the index date.
Transferred means that the patient was transferred from a tertiary healthcare system or nursing facility.

Table 2
Discrimination performances of non-susceptibility prediction models according to the full covariate and parsimonious covariate strategies.

	AMP (n = 5025)	AMS (n = 4701)	CFP (n = 3771)	CIP (n = 5929)	GEN (n = 6084)	IMP (n = 7572)	PPT (n = 3762)	SXT (n = 5402)
Outcome rate (%)	48.2	16.1	23.3	29.4	16.4	11.1	12.9	27.6
Full model, AUROC (%)								
LLR	73.4‡	71.0	66.0	71.7	62.0	61.8	69.5	73.9
GBM	72.4	74.8	70.7	70.4	66.9‡	73.0	65.0	72.1
RF	72.3	72.2	67.3	70.7	65.9	75.2	68.8	75.7
Stacked ensemble	68.8	76.9‡	76.1‡	77.0‡	64.0	80.6‡	76.1‡	76.5‡
Parsimonious model, AUROC (%)								
LLR	70.6§	71.8§	73.0§	70.0	65.9§	70.8	64.4	73.0§
GBM	70.2	68.6	70.4	70.3§	62.3	70.7	67.2§	71.0
RF	69.1	67.0	70.9	69.8	59.4	68.3	63.9	68.5
Stacked ensemble	63.7	55.9	57.4	62.5	54.0	71.3§	56.0	54.5

Abbreviations: AMP, ampicillin; AMS, ampicillin and sulbactam; CFP, cefepime; CIP, ciprofloxacin; GEN, gentamicin; PPT, piperacillin and tazobactam; SXT, trimethoprim and sulfamethoxazole; AUC, area under the receiver operating characteristics curve; LLR, Lasso logistic regression; GBM, gradient boosting machine; RF, random forest; Stacked, stacked ensemble.
‡ Highest performance among different algorithms applied to the full models.
§ Highest performance among different algorithms applied to parsimonious models.

Among the full models, the models applied to the stacked ensemble method for six antibiotics (AMS, CFP, CIP, IMP, PPT, and SXT) had the highest performance (AUROC in test set 76.9%, 76.1%, 77.0%, 80.6%, 76.1%, and 76.5%, respectively). The model to which LLR was applied for AMP (AUROC 73.4%) and the model to which GBM was applied for GEN (AUROC 66.9%) had the highest AUROC values. The detailed performance results—including AUROC, AUPRC, and calibration slope—are presented in Supplementary Table E3. In the parsimonious models, the model using the LLR algorithm had the highest AUROCs in the five susceptibility tests to AMP, AMS, CFP, GEN, and SXT: 70.6%, 71.8%, 73.0%, 65.9%, and 73.0%, respectively. In the models for CIP and PPT, GBM yielded the highest AUROCs (70.3% and 67.2%, respectively) compared with other algorithms. The stacked ensemble method showed the highest AUROC associated with IMP testing (71.2%). The detailed results are presented in Supplementary Table E4.

Personal Insusceptibility Estimates

An R-based shiny application for non-susceptibility calculation, PIE, was developed using the finally selected parsimonious models based on highest AUROC (Figure 3). The personal estimates of antibiotic non-susceptibility were calculated using PIE, and this application was applied to simulations of various clinical scenarios. The PIE application has been released online (https://cskim-abmi.shinyapps.io/PIEapp/).

Figure 4 shows the probability of non-susceptibility to each antibiotic with four example cases: (a) a 40-year-old woman hospitalised on 25 December 2018; (b) a 40-year-old man; (c) a 70-year-old woman; and (d) a 40-year-old woman hospitalised on 7 February 2011. The patients' clinical variables were accumulated one by one from left to right and used to calculate non-susceptibility rates per algorithm. The estimates varied according

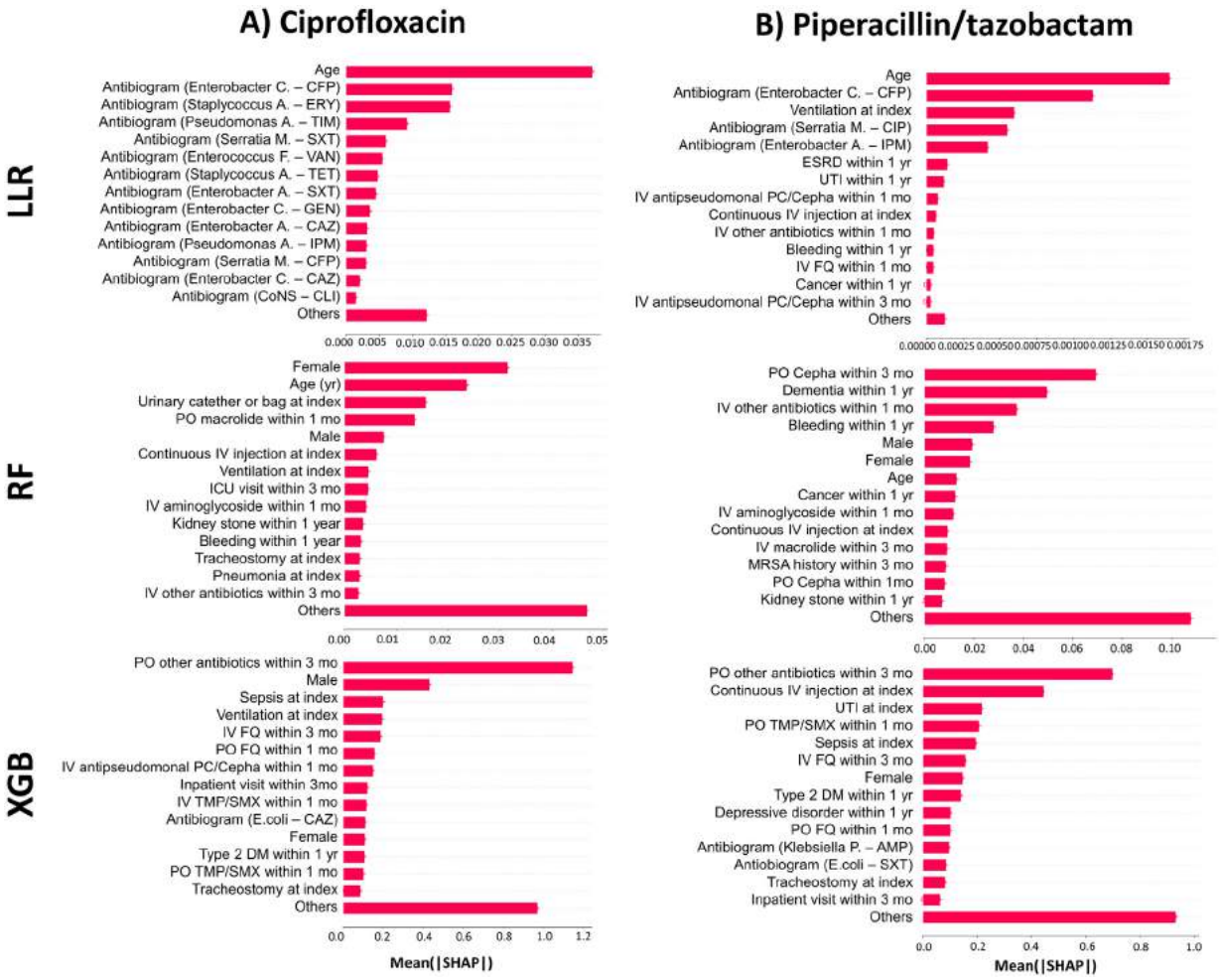

Figure 2. Predicted non-susceptibility rate for various patients' scenarios. From left to right, the characteristics corresponding to the column are accumulated. (A) Baseline case with 40-year-old female patient hospitalised on 25 December 2018; (B) 40-year-old male patient hospitalised on the same date; (C) 70-year-old female patient hospitalised on the same date; (D) 40-year-old female patient hospitalised on 7 February 2011.
Abbreviations: MDRO, multidrug-resistant organism; abx, antibiotics; IV, intravenous; PO, per os; AMP, ampicillin; AMS, ampicillin/sulbactam; CFP, cefepime; CIP, ciprofloxacin; GEN, gentamicin; PPT, piperacillin/tazobactam; SXT, sulfamethoxazole/trimethoprim; IPM, imipenem; HT, hypertension; DM, diabetes mellitus; CKD, chronic kidney disease; ICU, intensive care unit; VRE, vancomycin-resistance enterococci; MRSA, methicillin-resistant *Staphylococcus aureus*; CRE, carbapenem-resistant Enterobacteriaceae; ESBL, extended spectrum beta-lactamase producing Enterobacteriaceae; PC, penicillin; Cefa, cephalosporin; FQ, fluoroquinolone.

to demographic characteristics in all cases. Most estimates represented higher non-susceptibility for women than men and older than young individuals. Results also differed between patients with the same demographic characteristics but different hospitalisation dates. This calculator was also used to compare estimates while changing specific predictors one by one (see Supplementary Table E5). Predictors related to previous antibiotic exposure were associated with increased non-susceptible probabilities.

Discussion

This study developed machine learning-based models to predict patient-specific non-susceptible probabilities to broad-spectrum antibiotics among hospitalised patients with clinically suspected HA-UTIs. Further, parsimonious prediction models were developed for the purpose of clinical translation to adopt as a decision-supporting tool for the selection of personalised empiric antibiotics. A series of prediction algorithms of LLR, XGB, RF, and stacked ensemble was developed to obtain better discrimination and calibration. The values of AUROC and AUPRC, and calibration plot slopes of the 64 models indicated high performance in the task of predicting non-susceptibility outcomes, and the highest AUROC values of the best performance models among four algorithms remained across the various antibiotics, ranging 65.4–80.6%; these were superior to those determined by previous similar studies that involved predicting such resistance rates [10,11].

In addition to performance of the prediction models, there were several meaningful advantages to be noted. In practice, antibiotic resistance data derived from an institutional antibiogram ('local antibiogram') are important components of the treatment success associated with the initial selection of empiric antibiotics, and their use in clinical practice is strongly recommended [14,23–25]. However, many frontline clinicians are faced with outdated accessibility to the use of resistance rates due to scheduled release (e.g. 6 months), as well as unknown antibiotic resistance profiles with limited MDRO prevalence data [24]. In the current study, the models embedded a daily calculation function of non-susceptibility rate of antibiotics to common causative pathogens, which was designed to include both resistance and intermediate susceptible data based on local C&S results within the 90 days before the index date. It increased the applicability of the models by providing timely changes of non-susceptibility results of the institution over the cohort observation window periods; it is believed that this led to more accurate predictability for the treatment of appropriate em-

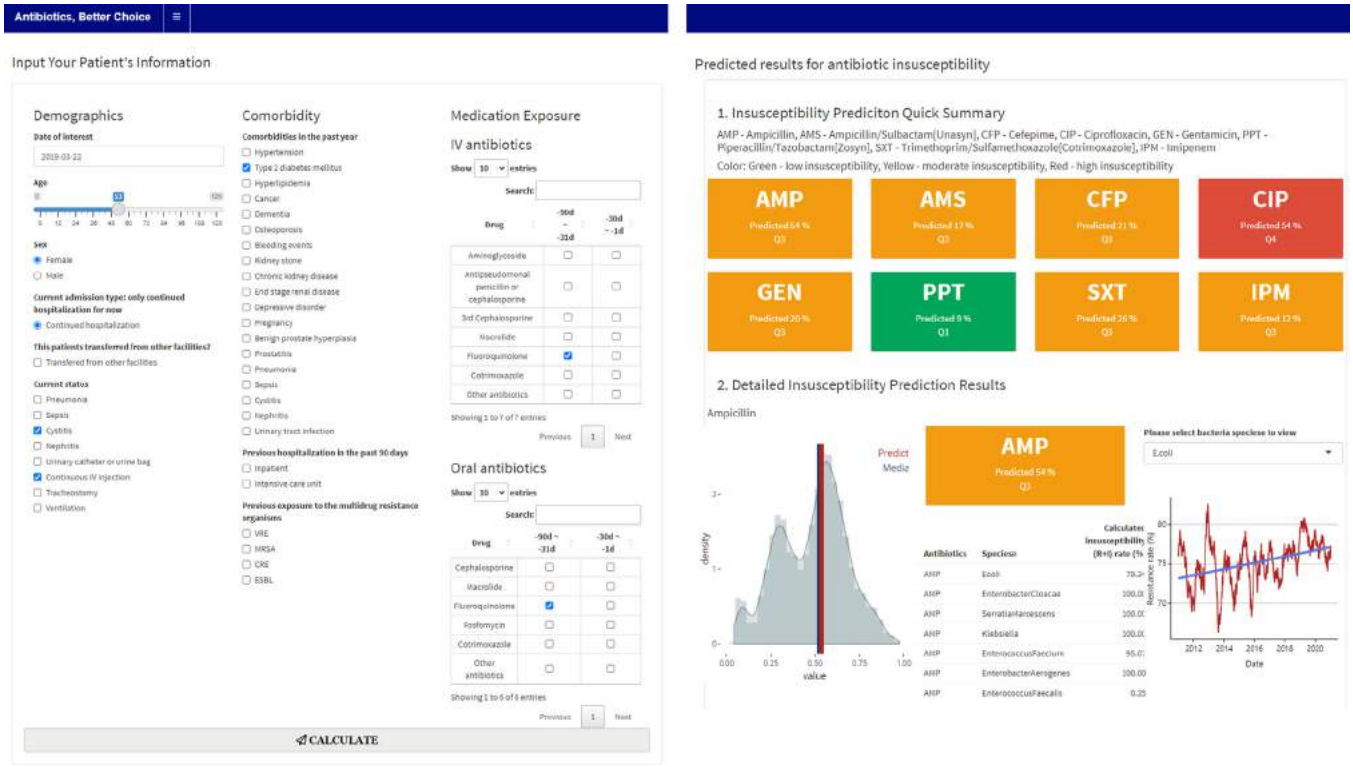

Figure 3. Variable importance based on the absolute Shapley additive explanations (SHAP) of two different prediction models (the ciprofloxacin and piperacillin/tazobactam models). SHAP is a model-agnostic tool that assesses importance through a predictor's contribution to a given prediction compared with the average prediction.
Abbreviations: LLR, Lasso logistic regression; RF, random forest; XGB, extreme gradient boosting machine; MDRO, multidrug-resistant organism, IV, intravenous; PO, per os; AMP, ampicillin; AMS, ampicillin/sulbactam; CFP, cefepime; CIP, ciprofloxacin; GEN, gentamicin; PPT, piperacillin/tazobactam; SXT, sulfamethoxazole/trimethoprim; IPM, imipenem; HT, hypertension; DM, diabetes mellitus; CKD, chronic kidney disease; ICU, intensive care unit; VRE, vancomycin-resistance enterococci; MRSA, methicillin-resistant *Staphylococcus aureus*; CRE, carbapenem-resistant Enterobacteriaceae; ESBL, extended-spectrum beta-lactamase producing Enterobacteriaceae; PC, penicillin; Cefa or Cepha, cephalosporin; FQ, fluoroquinolone.

piric antibiotics. Of note, an intermediate susceptibility category was included as non-susceptibility to follow the clinical decision consideration. A previous study used susceptibility data in the prediction models that relied on expert consensus to fill in unreported susceptibility values, which constituted almost half of the data, and this gave rise to concern about subjective bias and frequent updates of the models [10]. Another pivotal point of the current models was that they addressed the problem of a traditional antibiogram not reflecting patient characteristics, by incorporating patient demographics and infection-related information.

Directly motivated by the need for clinical translation to these findings, a predictor shrinkage in the full models was proposed to induce parsimonious models. The goal beyond developing the prediction models was clinical implementation. Therefore, the full models were derived; however, as with other prediction algorithms, the full prediction models contained many (> 20 000 predictors) independent features, which limited the capacity for clinical interpretation and translation. Thus, these features were mapped into subgroup clusters, and the variables of major subgroup clusters were used to build parsimonious prediction models for implementing patient care. The AUROCs of the parsimonious models reflected similar discrimination and calibration to the performance of the full prediction models, and highly observed features were consistently matched with previously well-reported or permutation-based factors. It is believed that this process helped disentangle the complex relationships among the variables, while providing reasonable values for the number of predictors. If the variables that are highly ranked as important for each algorithm are examined, it can be seen that they are different from the Lasso logistic regression and tree-based XGB and RF. Lasso logistic regression had more continuous variables, while XGB and RF did not; this is likely due to the different prediction algorithms that handle continuous and categorical variables [26].

The feasibility of the parsimonious models was further assessed by ablation-like experiments and patient case scenarios. As expected, the parsimonious algorithms revealed distinct differences in non-susceptibility as a function of patient characteristics, clinical observations, disease severity, history of MDRO infection, and previous antibiotic exposure. For example, the models predicted a higher non-susceptible probability in the 70-year-old woman than the 40-year-old woman, and higher non-susceptible probability with previous antibiotics exposures was also able to be confirmed.

To maximise the potential feasibility of the decision-supporting function, a PIE application was designed where the listed options consisted of features with final parsimonious models. The PIE application allowed clinicians to easily visualise the patient-specific non-susceptible probability to eight broad-spectrum antibiotics. The PIE provided further detailed information on comparative resistance distribution, indicating the interquartile location of a patient's resistance relative to the median value of the cohort population, the resistance rates of pathogens to other antibiotics, and the resistance trend in previous years. Thus, clinicians can make informed antibiotic selection decisions based on pertinent data regarding patient demographics, clinical aspects, and institutional environmental components. Moreover, it was developed based on the OMOP formatted database, so it is applicable to other institutions and can be updated in line with the OMOP common data model conversion cycle.

This study had some limitations. First, the eligible criteria for the study population did not include HA-UTI diagnosis codes be-

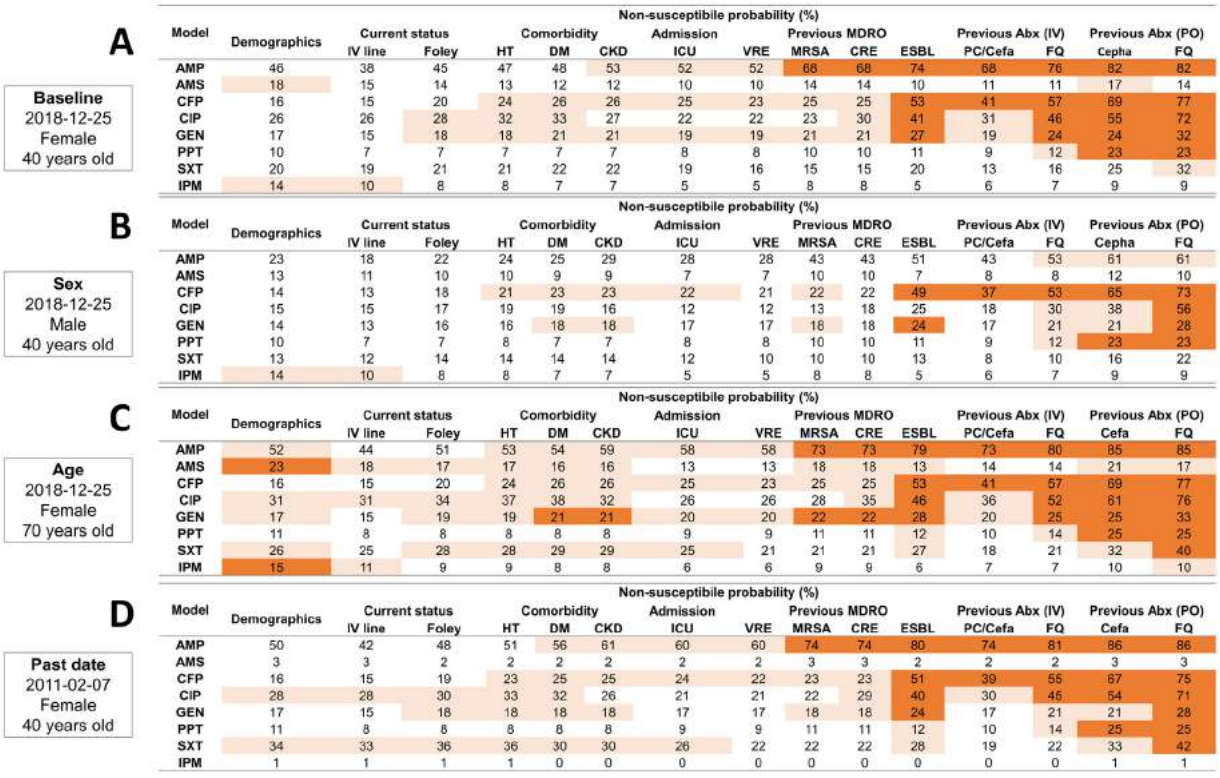

Figure 4. Shiny application including patient clinical characteristics input and calculated results for probabilities of predicted non-susceptibility. The results of individual antibiotic susceptibility tests were predicted through clinical information.
Abbreviations: MDRO, multidrug-resistant organism; IV, intravenous; PO, per os; AMP, ampicillin; AMS, ampicillin/sulbactam; CFP, cefepime; CIP, ciprofloxacin; GEN, gentamicin; PPT, piperacillin/tazobactam; SXT, sulfamethoxazole/trimethoprim; IPM, imipenem; HT, hypertension; DM, diabetes mellitus; CKD, chronic kidney disease; ICU, intensive care unit; VRE, vancomycin-resistance enterococci; MRSA, methicillin-resistant *Staphylococcus aureus*; CRE, carbapenem-resistant Enterobacteriaceae; ESBL, extended-spectrum beta-lactamase producing Enterobacteriaceae; PC, penicillin; Cefa, cephalosporin; FQ, fluoroquinolone.

cause those specifically representing hospital-acquired or nosocomial infections were presumably not often recorded by clinicians in the electronic health records, and there was concern about underutilisation of actual HA-UTI patient data. Moreover, the algorithms were intended to be built based on the times of urine C&S test orders as a surrogate of the clinicians' suspicion of UTI. Second, the average performance of the parsimonious models was ca. AUROC 70%, and further effort is required to improve performance despite high discrimination (which does not directly mean high predictability) in clinical settings. The limited performance could be partly due to unique characteristics of acute illness of infection, representing sudden onset, rapid progression, urgent treatment, multiple therapies, and finally a short course. This may be because it is difficult for the model to learn consistent tendencies of the predictors. Nonetheless, the models showed moderate discrimination and excellent calibration performance. Continuous efforts to improve performance should be maintained for better prediction of pathogen non-susceptibility to antibiotics for treating HA-UTI. Third, it is important to note that the feasibility of the aforementioned does not allow causal inferences to be made about the predictors utilised in the models [27,28]. Since the primary objective of this study was to develop a good performance predictive model, not inferring predictors' direction and magnitude of association, the importance of the predictors cannot be linked to causal inference. Lastly, the applicability of the prediction models in patient care needs to be validated through prospective clinical trials. Retrospective data were used to develop the model and validate its performance. In the future, well-designed prospective clinical trials will need to be conducted to assess whether this model can actually aid in empirical antibiotic prescribing. Despite the limitations, the results appear to be clinically applicable in predicting antibiotic non-susceptibility and facilitate the selection of empiric antibiotics.

Conclusion

Machine learning models were developed for predicting antibiotic resistance to facilitate the selection of empiric antibiotics to treat HA-UTI. An application tool was developed for personalised antibiotic therapy selection.

Declarations

Funding: This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea Government Ministry of Science and ICT [2020R1A2C1009224] and by Basic Science Research Program funded by the Ministry of Education [NRF-2020R1A6A1A03043528]. This work was also supported by the NRF grant funded by the Ministry of Education [2017R1D1A1B03033389] and grant of the project for Infectious Disease Medical Safety, funded by the Ministry of Health & Welfare, Republic of Korea [grant number: HG22C0024].

Competing Interests: The authors declare no conflicts of interest.

Ethical Approval: This study was conducted after approval by the institutional review board of Ajou University Medical Center (AJIRB-MED-MBD-21-489). The study was conducted according to the guidelines of the Declaration of Helsinki for human studies from the World Medical Association. The requirement for written

informed consent was waived because the analyses used deidentified data.

Author Contributions: Conceptualisation: S.J.R., R.W.P., and C.K.; methodology: C.K. and S.J.R.; software: C.K. and R.W.P.; validation: C.K. and S.J.R.; formal analysis: C.K.; investigation: Y.W.C., S.J.R., and R.W.P.; resources: C.K. and R.W.P.; data curation: C.K.; writing—original draft preparation: C.K. and S.J.R.; writing—review and editing: C.K., S.J.R., Y.W.C., and R.W.P.; visualization: C.K.; supervision: R.W.P. and S.J.R.; project administration: S.J.R.; funding acquisition: S.J.R. and R.W.P. All authors have read and agreed to the submitted version of the manuscript.

Patient and public involvement: Patients and/or the public were not involved in the design, or conduct, reporting, or dissemination plans of this research.

Data availability statement: All data relevant to the study can be provided upon reasonable requests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijantimicag.2023.106966.

References

[1] Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol 2015;13(5):269–84.
[2] Nicolle LE. The prevention of hospital-acquired urinary tract infection. Clin Infect Dis 2008;46(2):251–3.
[3] Sabih A, Leslie SW. Complicated urinary tract infections. StatPearls [Internet], Treasure Island (FL): StatPearls Publishing; Jan 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK436013/ [accessed 18 Jan 2023]..
[4] Gupta K, Hooton TM, Naber KG, Wullt B, Colgan R, Miller LG, et al. International Clinical Practice Guidelines for the Treatment of Acute Uncomplicated Cystitis and Pyelonephritis in Women: A 2010 Update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. Clin Infect Dis 2011;52(5):e103–20.
[5] Gupta K. Acute complicated urinary tract infection (including pyelonephritis) in adults. UpToDate 2022.
[6] World Health Organization. Antimicrobial stewardship programmes in healthcare facilities in low-and middle-income countries: a WHO practical toolkit. Available at: https://www.who.int/publications/i/item/9789241515481.
[7] Lee SS, Kim Y, Chung DR. Impact of discordant empirical therapy on outcome of community-acquired bacteremic acute pyelonephritis. J Infection 2011;62(2):159–64.
[8] Brusselaers N, Vogelaers D, Blot S. The rising problem of antimicrobial resistance in the intensive care unit. Ann Intensive Care 2011;1(1):47.
[9] Yelin I, Snitser O, Novich G, Katz R, Tal O, Parizade M, et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. Nat Med 2019;25(7):1143–52.
[10] Hebert C, Gao Y, Rahman P, Dewart C, Lustberg M, Pancholi P, et al. Prediction of antibiotic susceptibility for urinary tract infection in a hospital setting. Antimicrob Agents Chemother 2020;64(7):e02236 –19.
[11] Kanjilal S, Oberst M, Boominathan S, Zhou H, Hooper DC, Sontag D. A decision algorithm to promote outpatient antimicrobial stewardship for uncomplicated urinary tract infection. Sci Transl Med 2020;12(568):eaay5067.
[12] Lewin-Epstein O, Baruch S, Hadany L, Stein GY, Obolski U. Predicting antibiotic resistance in hospitalized patients by applying machine learning to electronic medical records. Clin Infect Dis 2020;72(11):e848–ee55.
[13] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J 2014;35(29):1925–31.
[14] Pakyz AL. The utility of hospital antibiograms as tools for guiding empiric therapy and tracking resistance: insights from the Society of Infectious Diseases Pharmacists. Pharmacotherapy 2007;27(9):1306–12.
[15] Truong WR, Hidayat L, Bolaris MA, Nguyen L, Yamaki J. The antibiogram: key considerations for its development and utilization. JAC-Antimicrob Resist 2021;3(2).
[16] Yoon D, Ahn EK, Park MY, et al. Conversion and data quality assessment of electronic health record data at a Korean tertiary teaching hospital to a common data model for distributed network research. Healthc Inform Res 2016;22(1):54–8.
[17] Wolpert DH. Stacked generalization. Neural Networks 1992;5(2):241–59.
[18] Kim C, You SC, Reps JM, Cheong JY, Park RW. Machine-learning model to predict the cause of death using a stacking ensemble method for observational data. J Am Med Inform Assoc 2021;28(6):1098–107.
[19] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology Jan 2010;21(1):128–38.
[20] Hosmer DW, Lemesbow S. Goodness of fit tests for the multiple logistic regression model. Comm Stat Theory Methods 1980;9(10):1043–69.
[21] Hause SMSGEWR. Time Series Missing Value Imputation. Available at: https://github.com/SteffenMoritz/imputeTS [accessed 23 Novermber 2022].
[22] Lee SMLS-I. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17); 2017.
[23] Morales A, Campos M, Juarez JM, Canovas-Segura B, Palacios F, Marin R. A decision support system for antibiotic prescription based on local cumulative antibiograms. J Biomed Inform 2018;84:114–22.
[24] Auzin A, Spits M, Tacconelli E, Rodriguez-Bano J, Hulscher M, Adang E, et al. What is the evidence base of used aggregated antibiotic resistance percentages to change empirical antibiotic treatment? A scoping review. Clin Microbiol Infect 2022;28(7):928–35.
[25] Cook A, Sharland M, Yau Y, Group* P, Bielicki J. Improving empiric antibiotic prescribing in pediatric bloodstream infections: a potential application of weighted-incidence syndromic combination antibiograms (WISCA). Expert Rev Anti-Infect Ther 2022;20(3):445–56.
[26] Song YY, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry 2015;27(2):130–5.
[27] Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. CHANCE 2019;32:1 42–9.
[28] Tredennick AT, Hooker G, Ellner SP, Adler PB. A practical guide to selecting models for exploration, inference, and prediction in ecology. Ecology 2021;102(6):e03336.