

Prediction of Antibiotic Resistance in Patients with Urinary Tract Infection: Algorithm Development and Validation

Nevruz İlhanlı, Se Yoon Park, Jae-woong Kim, Jee An Ryu, Ahmet Yardımcı,
Dukyong Yoon

Submitted to: JMIR Medical Informatics
on: July 27, 2023

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	4
Supplementary Files.....	24
Figures	25
Figure 1.....	26
Figure 2.....	27
Figure 3.....	28
Figure 4.....	29
Figure 5.....	30
Multimedia Appendixes	31
Multimedia Appendix 1.....	32
Multimedia Appendix 2.....	32
Multimedia Appendix 3.....	32
Multimedia Appendix 4.....	32
Multimedia Appendix 5.....	32
Multimedia Appendix 6.....	32
Multimedia Appendix 7.....	32
Multimedia Appendix 8.....	32
Multimedia Appendix 9.....	32
Multimedia Appendix 10.....	32

Prediction of Antibiotic Resistance in Patients with Urinary Tract Infection: Algorithm Development and Validation

Nevruz ?lhanl?^{1*}; Se Yoon Park^{2*} MD, PhD; Jae-woong Kim¹ MSc; Jee An Ryu¹; Ahmet Yard?mc?³ PhD; Dukyong Yoon¹ MD, PhD

¹Department of Biomedical Systems Informatics, Yonsei University College of Medicine Yongin KR

²Department of Internal Medicine, Hanyang University College of Medicine Seoul KR

³Department of Biostatistics and Medical Informatics, Akdeniz University Antalya TR

*these authors contributed equally

Corresponding Author:

Dukyong Yoon MD, PhD

Department of Biomedical Systems Informatics, Yonsei University College of Medicine

363, Dongbaekjukjeon-daero

Yongin

KR

Abstract

Background: The early prediction of antibiotic resistance in patients with urinary tract infection is important to guide appropriate antibiotic therapy selection.

Objective: In the present study, we aimed to predict antibiotic resistance in patients with urinary tract infection. Additionally, we aimed to interpret the machine learning models we developed.

Methods: We used admission, diagnosis, prescription, and microbiology records of patients who underwent urine culture tests in Yongin Severance Hospital, South Korea. We developed 5 sub-models to classify urinary tract infection pathogens as either sensitive or resistant to cephalosporin, piperacillin/tazobactam, trimethoprim/sulfamethoxazole, fluoroquinolone, and carbapenem. To analyze how each variable contributed to the machine learning model's predictions of antibiotic resistance, we used the SHapley Additive exPlanations method. Finally, we proposed a prototype machine learning based clinical decision support system to provide clinicians the resistance probabilities for each antibiotic.

Results: The area under the curve values ranged from 0.710 to 0.826 in the training set and 0.642 to 0.812 in the test set for predicting antibiotic resistance. The administration of drugs before infection and exposure time to these drugs were found to be important variables for predicting antibiotic resistance.

Conclusions: The study results demonstrated the potential of machine learning to predict antibiotic resistance in patients with urinary tract infection. Machine learning can assist clinicians in making decisions regarding the selection of appropriate antibiotic therapy in patients with urinary tract infection.

(JMIR Preprints 27/07/2023:51326)

DOI: <https://doi.org/10.2196/preprints.51326>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Title: Prediction of Antibiotic Resistance in Patients with Urinary Tract Infection: Algorithm Development and Validation

Authors: Nevruz İlhanlı^{1,2,*}, Se Yoon Park^{1,3,4,*}, Jae-woong Kim^{1,3}, Jee An Ryu¹, Ahmet Yardımci², Dukyong Yoon^{1,4,5}

*These authors contributed equally

Affiliations:

¹Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Yongin, Gyeonggi-do, South Korea

²Department of Biostatistics and Medical Informatics, Akdeniz University, Turkey

³Department of Hospital Medicine, Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Korea

⁴Center for Digital Health, Yongin Severance Hospital, Yonsei University Health System, Yongin, South Korea

⁵Institute for Innovation in Digital Healthcare (IIDH), Severance Hospital, Seoul, Republic of Korea

Address for Correspondence:

Dukyong Yoon, MD, PhD

Department of Biomedical Systems Informatics, Yonsei University College of Medicine

363, Dongbaekjukjeon-daero, Giheung-gu, Yongin 16995, Gyeonggi-do, Republic of Korea

Tel: +82-31-5189-8450

E-mail: dukyong.yoon@yonsei.ac.kr

Title: Prediction of Antibiotic Resistance in Patients with Urinary Tract Infection: Algorithm Development

and Validation

Abstract

Background: The early prediction of antibiotic resistance in patients with urinary tract infection is important to guide appropriate antibiotic therapy selection. In the present study, we aimed to predict antibiotic resistance in patients with urinary tract infection. Additionally, we aimed to interpret the machine learning models we developed.

Methods: We used the electronic medical records of patients who were admitted to Yongin Severance Hospital, South Korea. A total of 71 features extracted from patients' admission, diagnosis, prescription, and microbiology records were used for classification. We classified urinary tract infection pathogens as either sensitive or resistant to cephalosporin, piperacillin/tazobactam, carbapenem, trimethoprim/sulfamethoxazole, and fluoroquinolone. To analyze how each variable contributed to the machine learning model's predictions of antibiotic resistance, we used the SHapley Additive exPlanations method. Finally, we proposed a prototype machine learning based clinical decision support system to provide clinicians the resistance probabilities for each antibiotic.

Results: The dataset included 3,535, 737, 708, 1,582, and 1,365 samples for cephalosporin, TZP, TMP-SMX, fluoroquinolone, and carbapenem resistance prediction models, respectively. The area under the curve values of the random forest models were 0.777 (95% CI: 0.775-0.779), 0.864 (95% CI: 0.862-0.867), 0.877 (95% CI: 0.874-0.880), 0.881 (95% CI: 0.879-0.882), 0.884 (95% CI: 0.884-0.885) in the training set and 0.638 (95% CI: 0.635-0.642), 0.630 (95% CI: 0.626-0.634), 0.665 (95% CI: 0.659-0.671), 0.670 (0.666-0.673), 0.721 (0.718-0.724) in the test set for predicting resistance to cephalosporin, TZP, carbapenem, TMP-SMX, and fluoroquinolone, respectively. The number of previous visits, first culture after admission, chronic lower respiratory diseases, administration of drugs before infection, and exposure time to these drugs were found to be important variables for predicting antibiotic resistance.

Conclusion: The study results demonstrated the potential of machine learning to predict antibiotic resistance in patients with urinary tract infection. Machine learning can assist clinicians in making decisions regarding the selection of appropriate antibiotic therapy in patients with urinary tract infection.

Keywords: Antibiotic resistance; Machine learning; Urinary tract infections

INTRODUCTION

Urinary tract infection (UTI) refers to an infection that occurs any part of the urinary system, including the kidneys, ureters, urinary bladder, urethra, and other auxiliary structures [1, 2]. Globally, UTIs are the most prevalent type of infectious disease, with around 150–250 million cases occurring each year [3]. Considerable morbidity and mortality result from these infections [4]. Typically, the most effective treatment for UTIs is the administration of antibiotics [3]. However, inappropriate use of antibiotics can permanently affect the normal microbiota of the urinary tract system and lead to antibiotic resistance [5].

The antibiotic susceptibility test is commonly used to identify antibiotic resistance, but it takes 24–48 hours to obtain test results [6, 7]. However, in the clinical workflow, clinicians need to identify antibiotic resistance quickly to provide effective treatment for patients with UTIs. For this reason, early prediction of antibiotic resistance in patients with UTIs is important to guide the selection of appropriate antibiotic therapy. Machine learning can be used to develop prediction models and clinical decision support systems (CDSS) to identify antibiotic resistance and support the selection of appropriate antibiotic therapy for patients with UTI.

Several efforts have been made to predict antibiotic resistance in patients with UTIs using data from patients' electronic medical records (EMRs), including demographics, prescriptions, comorbidities, procedures, and laboratory tests. These investigations have yielded promising results. Some of these studies were limited to specific patient groups, including patients with uncomplicated UTIs [8] and patients treated in the emergency department [9]. In other studies, researchers worked with heterogeneous data that were not limited to a specific patient group [10–12]. However, prior studies that analyzed heterogeneous data did not address the interpretation of machine learning models. The black-box nature of machine learning is a limiting factor not only in its use for antibiotic resistance prediction but also in its wider clinical usage [13, 14]. Thus, interpreting the results obtained by the machine learning model is crucial in increasing users' trust in the machine learning model [15, 16]. Furthermore, these studies did not address the development of the CDSS with the prediction models they built.

In the present study, we aimed to predict antibiotic resistance in patients with UTI. We used heterogeneous data that were not limited to a specific patient group. We classified UTI pathogens as either sensitive or

resistant to 5 commonly used antibiotics in UTI treatment: cephalosporin, piperacillin/tazobactam (TZP), carbapenem, trimethoprim/sulfamethoxazole (TMP-SMX), and fluoroquinolone. In addition, our objective was to understand and explain the inner workings of the machine learning models we developed. Eventually, we developed a prototype CDSS to provide clinicians the resistance probabilities for each antibiotic.

METHODS

Ethical approval

Ethical approval for the study was obtained from the Institutional Review Board of Yonsei University Severance Hospital on June 6, 2022 (approval number: 9-2023-0095). The informed consent was not required due to the retrospective nature of the study.

Dataset description

In the present study, we used the EMRs of patients who were admitted to Yongin Severance Hospital, South Korea between October 2012 and October 2022. To build the prediction models, we extracted admission, diagnosis, prescription, and microbiology records. The summary of the research process is presented in Figure 1.

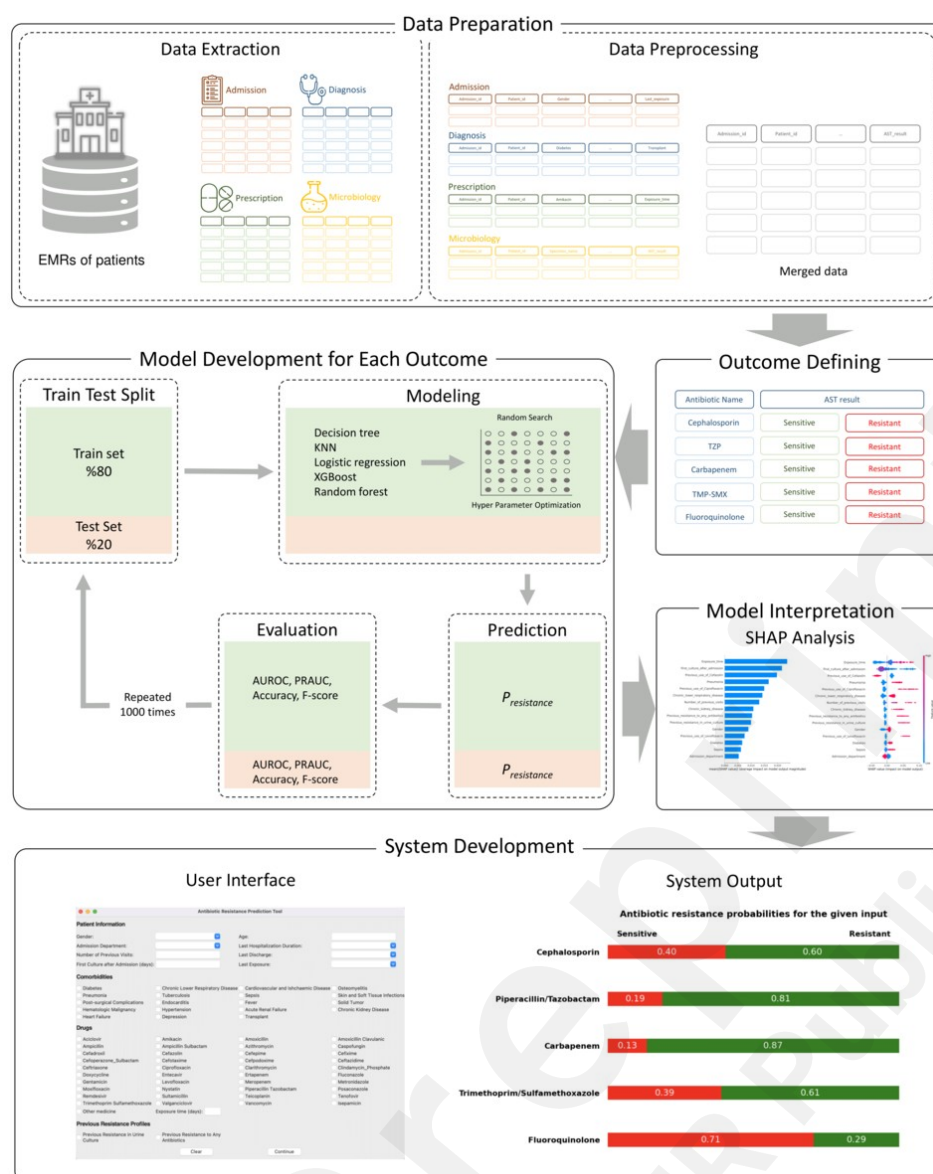


Figure 1. Summary of the research process. EMRs: electronic medical records; AST: antibiotic susceptibility test; TZP: piperacillin-tazobactam; TMP-SMX: trimethoprim-sulfamethoxazole; KNN: K-nearest neighbors; XGBoost: eXtreme Gradient Boosting; AUROC: area under the ROC curve; PRAUC: precision-recall area under the curve.

Data preprocessing

The microbiology table contained 143,114 urine cultures collected from 6,011 patients during 7,719 admissions. Since positive samples typically indicate the presence of bacteriuria, and urine culture samples were typically collected from patients with UTI symptoms, we considered these to be indicative of UTI [10]. The resistance profiles were evaluated based on the Clinical and Laboratory Standards Institute guidelines, where intermediate-level resistance was considered sensitive. To assess the resistance of UTI pathogens to

antibiotic classes, antibiotics were grouped as cephalosporin, TZP, carbapenem, TMP-SMX, and fluoroquinolone. The antibiotics included in each antibiotic class are presented in the Multimedia Appendix 1. The patients' demographic information was extracted from the admission table, their comorbidities were extracted from the diagnosis table, and their drug usage information was extracted from the prescription table. For all input variables, the time of the first culture test was considered as the endpoint, and only data collected before the first culture test were used. After preprocessing and variable extraction from the raw data, the tables were combined using the admission number as the primary key. Missing data were excluded from the study. Patients over 19 age years and under 100 years of age at admission were included in the study, and numerical variables were standardized. A total of 71 features were used to classify UTI pathogens as either sensitive or resistant to each antibiotic. The predictors for the prediction models were selected by considering related works and using clinical judgment. Additionally, the threshold values for binarization were selected according to the literature [17] and the expert assessment of a specialist in infectious diseases. Detailed information about the predictors can be found in the Multimedia Appendix 2.

Machine learning model development

We used a repeated train test split approach for modeling. The datasets were split into training and test sets using an 80:20 ratio, and the training sets were used for the development of the machine learning models. When splitting the data into training and test sets, data points from the same patient and admission were exclusively included in either the training or test dataset to prevent potential data leakage and ensure the models were evaluated on previously unseen data. At each iteration, we created different training and test datasets by changing the random seed. Decision tree, K-nearest neighbors (KNN), logistic regression, eXtreme Gradient Boosting (XGBoost), and random forest were used for modeling. The hyperparameters of the machine learning models were optimized by using the random search hyperparameter optimization method with 10-fold cross validation on the training dataset. We stored the performance of the prediction models at each iteration, and the mean of performance metrics was calculated. The procedure of splitting the data, optimizing hyperparameters, modeling, and evaluation was iteratively repeated 1000 times to classify UTI pathogens as either sensitive or resistant to cephalosporin, TZP, carbapenem, TMP-SMX, and

fluoroquinolone. The machine learning models were built using Python (version 3.10.4).

Machine learning model interpretation

To analyze the contribution of the variables to the machine learning models in predicting antibiotic resistance, we used the SHapley Additive exPlanations (SHAP) method. The SHAP values of the random forest models that showed superior performance compared to other machine learning methods were evaluated. The random forest model with the highest area under the ROC curve (AUROC) on the test set across all iterations for each antibiotic was used for SHAP analysis. Python (version 3.10.4) was used for SHAP analysis.

Clinical decision support system development

To develop the CDSS prototype, the random forest model with the highest AUROC on the test set across all iterations for each antibiotic was used. The CDSS prototype was developed using the tkinter package in Python (version 3.10.4).

Evaluation

The performance of the machine learning model for predicting antibiotic resistance was evaluated on the training (n=2,874 (81.3%) for cephalosporin, n=611 (82.9%) for TZP, n=574 (81.1%) for TMP-SMX, n=1,248 (78.9%) for fluoroquinolone, and n=1,085 (79.5%) for carbapenem) and test sets (n=661 (18.7%) for cephalosporin, n=126 (17.1%) for TZP, n=134 (18.9%) for TMP-SMX, n=334 (21.1%) for fluoroquinolone, and n=280 (20.5%) for carbapenem) using the AUROC with 95% confidence intervals (CI), precision-recall area under the curve (PRAUC), accuracy, and f-score performance metrics. Herein, the AUC value was considered the main evaluation metric. The definitions of the performance metrics we used are provided below.

AUROC: The AUROC is a widely used metric that represents a classifier's ability to discriminate between positive instances and negative instances [18].

PRAUC: PRAUC refers to the area under the precision-recall curve which plots precision as a function of recall for all the possible decision thresholds [19].

Accuracy: Accuracy is the ratio of correctly classified samples to all samples.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{False Negative} + \text{True Negative}}$$

F-score: F-score is the harmonic mean of precision and recall metrics.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$F\text{-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Python (version 3.10.4) was used to evaluate the prediction models.

RESULTS

Dataset Characteristics

The general characteristics of the dataset used in this study are presented in Table 1. The dataset included 3,535, 737, 708, 1,582, and 1,365 samples for cephalosporin, TZP, TMP-SMX, fluoroquinolone, and carbapenem resistance prediction models, respectively. *Escherichia coli* was the most frequently isolated bacterial specimen across all antibiotics.

Table 1: General characteristics of the dataset

	Cephalosporin	TZP	TMP-SMX	Fluoroquinolone	Carbapenem
n (samples)	3,535	737	708	1,582	1,365
n (admissions)	396	366	374	571	392
n (patients)	390	360	368	557	386
Resistance, n (%)	1492 (42.2)	169 (22.9)	281 (39.7)	1014 (64.1)	142 (10.4)
Age, mean $\pm$ sd	71.5 $\pm$ 14.4	71.4 $\pm$ 14.4	71.4 $\pm$ 14.4	71.9 $\pm$ 14.4	71.7 $\pm$ 14.3
Female, %	2597 (73.5)	523 (71.0)	507 (71.6)	1013 (64.0)	994 (72.8)
Most common Bacteria, n (%)	<i>Escherichia coli</i> , 1650 (46.7)	<i>Escherichia coli</i> , 312 (42.3)	<i>Escherichia coli</i> , 331 (46.7)	<i>Escherichia coli</i> , 349 (22)	<i>Escherichia coli</i> , 624 (45.7)
Second-most common bacteria, n (%)	<i>Klebsiella pneumoniae</i> , 556 (15.7)	<i>Klebsiella pneumoniae</i> , 109 (14.8)	<i>Klebsiella pneumoniae</i> , 111 (15.7)	<i>Enterococcus faecium</i> , 305 (19.3)	<i>Klebsiella pneumoniae</i> , 220 (16.1)

Third-most common bacteria, n (%)	<i>Pseudomonas aeruginosa</i> , 168 (4.7)	<i>Pseudomonas aeruginosa</i> , 69 (9.4)	<i>Citrobacter freundii</i> , 21 (3.0)	<i>Enterococcus faecalis</i> , 180 (11.4)	<i>Pseudomonas aeruginosa</i> , 83 (6.1)
TZP: piperacillin-tazobactam, TMP-SMX: trimethoprim-sulfamethoxazole; sd: standard deviation					

Model Performance

The performance analysis of the random forest models is presented in Table 2. The AUC values were 0.777 (95% CI: 0.775-0.779), 0.864 (95% CI: 0.862-0.867), 0.877 (95% CI: 0.874-0.880), 0.881 (95% CI: 0.879-0.882), 0.884 (95% CI: 0.884-0.885) in the training set and 0.638 (95% CI: 0.635-0.642), 0.630 (95% CI: 0.626-0.634), 0.665 (95% CI: 0.659-0.671), 0.670 (0.666-0.673), 0.721 (0.718-0.724) in the test set for predicting resistance to cephalosporin, TZP, carbapenem, TMP-SMX, and fluoroquinolone, respectively. The performance analysis of the other machine learning models is presented in Multimedia Appendices 3-6.

Table 2: Classification performances of the random forest models

	Training Set				Test Set			
	AUROC	PRAUC	Accuracy	F1 Score	AUC	PRAUC	Accuracy	F1 Score
Cephalosporin	0.777 (0.775 - 0.779)	0.725	0.715	0.676	0.638 (0.635 - 0.642)	0.547	0.603	0.556
TZP	0.864 (0.862 - 0.867)	0.688	0.808	0.652	0.630 (0.626 - 0.634)	0.332	0.641	0.313
Carbapenem	0.877 (0.874 - 0.880)	0.539	0.822	0.493	0.665 (0.659 - 0.671)	0.222	0.725	0.220
TMP-SMX	0.881 (0.879 -	0.829	0.822	0.781	0.670 (0.666	0.568	0.638	0.560

	0.882)				-			
					0.673)			
Fluoroquinolone	0.884	0.938	0.802	0.832	0.721	0.813	0.657	0.706
	(0.884 -				(0.718			
	0.885)				-			
					0.724)			
AUROC: area under the ROC curve; PRAUC: precision-recall area under the curve;								
TZP: piperacillin-tazobactam; TMP-SMX: trimethoprim-sulfamethoxazole								

Important features

The SHAP values of the 15 most important features in the random forest models are presented in Figure 2.

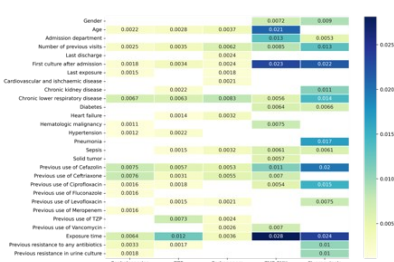


Figure 2. SHAP values of the 15 most important features in the prediction models. SHAP: SHapley Additive exPlanations; TZP: piperacillin-tazobactam; TMP-SMX: trimethoprim-sulfamethoxazole.

The SHAP feature importance bar plot (Figure 3A) and SHAP summary plot (Figure 3B) of the fluoroquinolone resistance prediction model are presented in Figure 3. The SHAP feature importance plot and SHAP summary plot of the other antibiotic prediction models are presented in Multimedia Appendices 7-10.

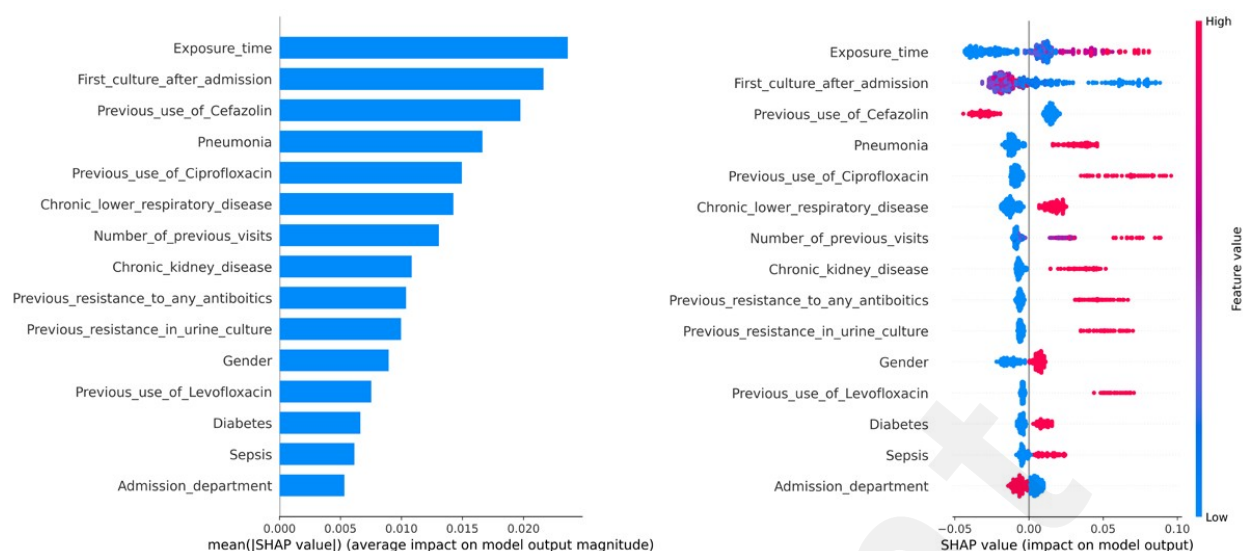


Figure 3. SHAP analysis results of fluoroquinolone resistance prediction model. (A) The feature importance bar plot. (B) The SHAP summary dot plot. SHAP: SHapley Additive exPlanations.

Clinical Decision Support System

The user interface of the CDSS is shown in Figure 4. The CDSS prototype obtains data from the user and produces antibiotic resistance probabilities for each antibiotic.

Antibiotic Resistance Prediction Tool

Patient Information

Gender: Female ☒ Age: 55

Admission Department: Outpatient ☒ Last Hospitalization Duration: Less than 5 days ☒

Number of Previous Visits: 3 Last Discharge: 30 days or less ☒

First Culture after Admission (days): 1 Last Exposure: 30 days or less ☒

Comorbidities

☒ Diabetes ☒ Chronic Lower Respiratory Disease ☐ Cardiovascular and Ischaemic Disease ☐ Osteomyelitis

☐ Pneumonia ☐ Tuberculosis ☐ Sepsis ☐ Skin and Soft Tissue Infections

☐ Post-surgical Complications ☐ Endocarditis ☐ Fever ☐ Solid Tumor

☐ Hematologic Malignancy ☐ Hypertension ☐ Acute Renal Failure ☐ Chronic Kidney Disease

☐ Heart Failure ☐ Depression ☐ Transplant

Drugs

☐ Aciclovir ☐ Amikacin ☐ Amoxicillin ☐ Amoxicillin Clavulanic

☐ Ampicillin ☐ Ampicillin Sulbactam ☐ Azithromycin ☐ Caspofungin

☐ Cefadroxil ☒ Cefazolin ☐ Cefepime ☐ Cefixime

☐ Cefoperazone_Sulbactam ☐ Cefotaxime ☐ Cefpodoxime ☐ Ceftazidime

☐ Ceftriaxone ☐ Ciprofloxacin ☐ Clarithromycin ☐ Clindamycin_Phosphate

☐ Doxycycline ☐ Entecavir ☐ Ertapenem ☐ Fluconazole

☐ Gentamicin ☐ Levofloxacin ☐ Meropenem ☐ Metronidazole

☐ Moxifloxacin ☐ Nystatin ☐ Piperacillin Tazobactam ☐ Posaconazole

☐ Remdesivir ☐ Sultamicillin ☐ Teicoplanin ☐ Tenofovir

☐ Trimethoprim Sulfamethoxazole ☐ Valganciclovir ☐ Vancomycin ☐ Isepamicin

☐ Other medicine Exposure time (days): 5

Previous Resistance Profiles

☒ Previous Resistance in Urine Culture ☐ Previous Resistance to Any Antibiotics

Clear Continue

Figure 4. The user interface of the CDSS.

We presented the CSDD prototype on a scenario. In this case, a 55-year-old female was admitted to the hospital's outpatient department. The patient previously visited the hospital three times and was re-admitted to the hospital within 30 days of her last 3-day stay. The duration between the patient's admission to the hospital and the first culture was 1-day. The patient was previously diagnosed with diabetes and chronic lower respiratory disease. Additionally, the patient had a history of cefazolin usage and resistance in urine culture. The system output for the given scenario is shown in Figure 5. The system produced resistance probabilities for each antibiotic. For the given scenario, the system produced a 71% probability of fluoroquinolone resistance, a 41% probability of cephalosporin resistance, a 39% probability of TMP-SMX resistance, a 19% probability of TZP resistance, and a 13% probability of carbapenem resistance.

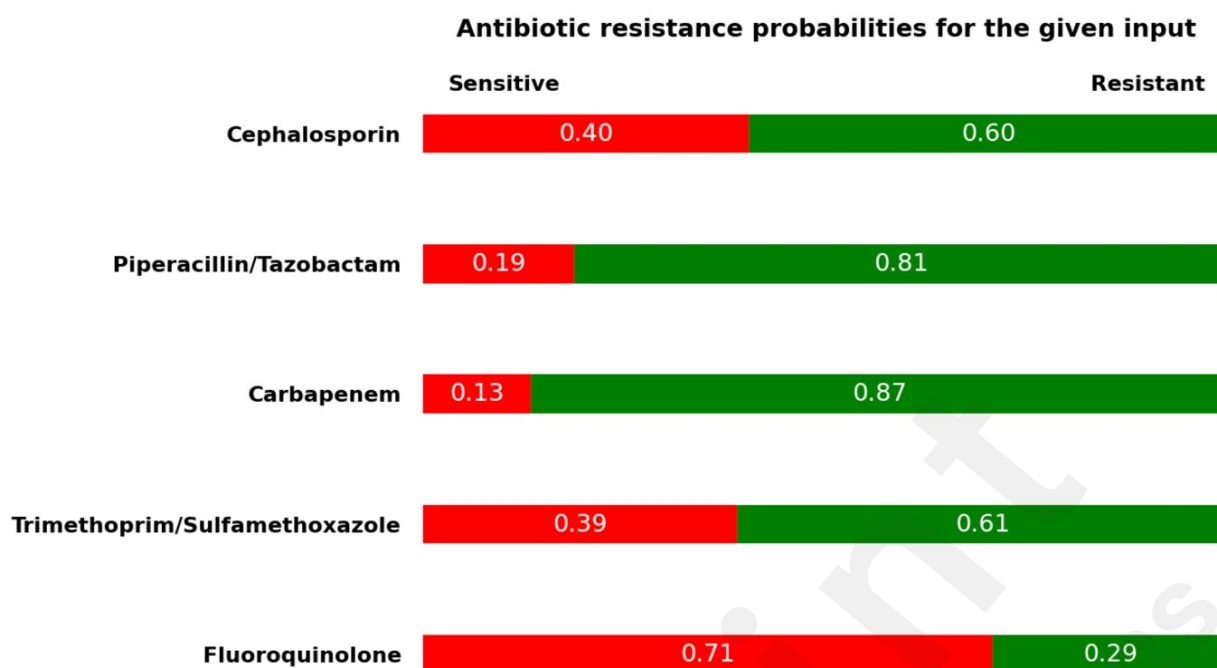


Figure 5. **The system output for the given data.**

DISCUSSION

In the present study, our main objective was to predict cephalosporin, TZP, carbapenem, TMP-SMX, and fluoroquinolone resistance in patients with UTI and develop a CDSS with the machine learning models we built. Moreover, we identified the most important features for predicting antibiotic resistance in patients with UTI using SHAP analysis.

Our prediction models achieved AUCs of 0.777 (95% CI: 0.775-0.779), 0.864 (95% CI: 0.862-0.867), 0.877 (95% CI: 0.874-0.880), 0.881 (95% CI: 0.879-0.882), 0.884 (95% CI: 0.884-0.885) in the training set and 0.638 (95% CI: 0.635-0.642), 0.630 (95% CI: 0.626-0.634), 0.665 (95% CI: 0.659-0.671), 0.670 (0.666-0.673), 0.721 (0.718-0.724) in the test set for predicting resistance to cephalosporin, TZP, carbapenem, TMP-SMX, and fluoroquinolone, respectively. The fluoroquinolone resistance prediction model showed superior performance, as confirmed by its high AUC values in both the training and test sets. On the other hand, the cephalosporin resistance prediction model showed poor performance, as confirmed by the low AUC values in both training and test sets.

Past efforts to predict antibiotic resistance in patients with UTIs have had promising results, with the lowest AUC being 0.58 for predicting TMP-SMX resistance [12] and the highest AUC being 0.83 for predicting

ciprofloxacin resistance [9]. In comparison, our prediction models demonstrated comparable performance to these prior works. Some previous studies on predicting antibiotic resistance in patients with UTIs were limited to specific patient groups, including patients with uncomplicated UTIs [8] and patients treated in the emergency department [9]. We analyzed heterogeneous data that was not limited to a specific patient group or bacteria. This approach provides a more comprehensive insight into the prediction of antibiotic resistance in patients with UTIs. Similarly, Lewin-Epstein et al. analyzed heterogeneous data and were able to achieve AUC values ranging from 0.73 to 0.79 for the prediction of ceftazidime, gentamicin, imipenem, ofloxacin, and TMP-SMX resistance [20]. Their data contained multiple culture tests, which provided a more comprehensive approach to predicting antibiotic resistance. Although urine cultures can be used to infer colonized resistance in patients, further research is needed to extend culture results beyond urine.

According to SHAP analysis, the contribution of the variables varied for each antibiotic; however, we found that the number of previous visits, first culture after admission, chronic lower respiratory diseases, administration of drugs before infection, and exposure time to these drugs were important predictors across all antibiotics. Factors such as first culture after admission, exposure time, and number of previous visits were found to affect resistance, which can be explained by the impact of healthcare-associated infections. Chronic lower respiratory and kidney diseases are also likely to be associated with frequent visits to healthcare facilities, although it is difficult to confirm the actual number of visits. However, this suggests that the characteristics of healthcare-seeking behavior in patients with specific underlying diseases may influence resistance [21]. Interestingly, the use of cefazolin had a negative impact on the development of resistance for all antibiotics. This is because cefazolin is one of the narrow-spectrum antibiotics used in less severe patients. Further research is needed to examine these results.

While our study provides insights into predicting antibiotic resistance in patients with UTIs, it has some limitations. Firstly, our study is the lack of multidrug resistance classification. The dataset we used in this study did not contain sufficient amount of multidrug resistance outcome to build a classification model for the prediction of multidrug resistance. Furthermore, our prediction models were developed using prescription records within the hospital setting. However, patients may have used antibiotics outside of the hospital setting, either during visits to other hospitals or through self-medication. The lack of information about past drug

usage could have negatively impacted the performance of our prediction models. To overcome this limitation, we intend to conduct further studies using data from the National Health Insurance Service of South Korea, which contains all past drug usage information of the patients. Thus, we will have a more comprehensive dataset. By using this approach, we may be able to develop more accurate machine learning models to predict antibiotic resistance and improve our ability to guide appropriate antibiotic therapy selection. Additionally, further development is required to address limitations of prototype CDSS, including the integration of real-time patient data and validation in larger patient cohorts. Moreover, the prototype CDSS only gives the resistance risk probability to the user. However, a more comprehensive system that can provide decision support on the selection of appropriate therapy, dosage, and duration of treatment can be developed in further studies. Such a system has the potential to reduce the duration of treatment, number of antibiotics used, cost, mortality, and morbidity [22, 23].

In conclusion, our study results demonstrated that prediction models to predict antibiotic resistance in patients with UTIs can be constructed using routinely collected EMR data alone, without requiring additional laboratory tests or specialized tests. Machine learning techniques can be used to develop systems that can guide clinicians in selecting appropriate antibiotic therapy. This has the potential to prevent the risk of inappropriate antibiotic administration, thereby reducing patients' risk of developing antibiotic resistance.

Acknowledgments

This study was supported by National Institute for International Education (NIIED) of the Government of the Republic of Korea, The Scientific and Technological Research Council of Turkey (TUBITAK) (Grant number: 2214-A), and a faculty research grant of Yonsei University College of Medicine (6-2022-0080).

Authors' Contributions

NI, AY, SYP, and DY contributed to the conceptualization of the study. JWK, JAR, and SYP were responsible for data curation. NI performed the formal analysis of the collected data. NI, AY, SYP, and DY contributed to the funding acquisition. NI, SYP, and DY contributed to the development of the study methodology. NI wrote the manuscript. SYP and DY reviewed and edited the manuscript.

Conflicts of Interest

All authors have no potential conflicts of interest.

Abbreviations

AUROC: area under the ROC curve

CDSS: clinical decision support systems

EMRs: electronic medical records

KNN: K-nearest neighbors

PRAUC: precision-recall area under the curve

SHAP: SHapley Additive exPlanations

TMP-SMX: trimethoprim/sulfamethoxazole

TZP: piperacillin/tazobactam

UTI: urinary tract infection

XGBoost: eXtreme Gradient Boosting

References

1. Tan CW, Chlebicki MP. Urinary tract infections in adults. Singapore Med J. 2016 Sep;57(9):485-90. PMID: 27662890. doi: 10.11622/smedj.2016153.
2. Belete MA, Saravanan M. A Systematic Review on Drug Resistant Urinary Tract Infection Among Pregnant Women in Developing Countries in Africa and Asia; 2005-2016. Infect Drug Resist. 2020;13:1465-77. PMID: 32547115. doi: 10.2147/Idr.S250654.
3. Santos M, Mariz M, Tiago I, Martins J, Alarico S, Ferreira P. A review on urinary tract infections diagnostic methods: Laboratory-based and point-of-care approaches. J Pharm Biomed Anal. 2022 Sep 20;219:114889. PMID: 35724611. doi: 10.1016/j.jpba.2022.114889.
4. Suskind AM, Saigal CS, Hanley JM, Lai J, Setodji CM, Clemens JQ, et al. Incidence and Management of Uncomplicated Recurrent Urinary Tract Infections in a National Sample of Women in the United States. Urology. 2016 Apr;90:50-5. PMID: 26825489. doi: 10.1016/j.urology.2015.11.051.
5. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol. 2015 May;13(5):269-84. PMID: 25853778. doi: 10.1038/nrmicro3432.

6. Boolchandani M, D'Souza AW, Dantas G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet.* 2019 Jun;20(6):356-70. PMID: 30886350. doi: 10.1038/s41576-019-0108-4.
7. Benkova M, Soukup O, Marek J. Antimicrobial susceptibility testing: currently used methods and devices and the near future in clinical practice. *J Appl Microbiol.* 2020 Oct;129(4):806-22. PMID: 32418295. doi: 10.1111/jam.14704.
8. 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 Nov 4;12(568). PMID: 33148625. doi: 10.1126/scitranslmed.aay5067.
9. Lee HG, Seo Y, Kim JH, Han SB, Im JH, Jung CY, et al. Machine learning model for predicting ciprofloxacin resistance and presence of ESBL in patients with UTI in the ED. *Sci Rep.* 2023 Feb 25;13(1):3282. PMID: 36841917. doi: 10.1038/s41598-023-30290-y.
10. 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 Jul;25(7):1143-52. PMID: 31273328. doi: 10.1038/s41591-019-0503-6.
11. 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 Jun 23;64(7). PMID: 32312778. doi: 10.1128/AAC.02236-19.
12. Rich SN, Jun I, Bian J, Boucher C, Cherabuddi K, Morris JG, Jr., et al. Development of a Prediction Model for Antibiotic-Resistant Urinary Tract Infections Using Integrated Electronic Health Records from Multiple Clinics in North-Central Florida. *Infect Dis Ther.* 2022 Oct;11(5):1869-82. PMID: 35908268. doi: 10.1007/s40121-022-00677-x.
13. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: beyond the black box. *BMJ.* 2019 Mar 12;364:l886. PMID: 30862612. doi: 10.1136/bmj.l886.

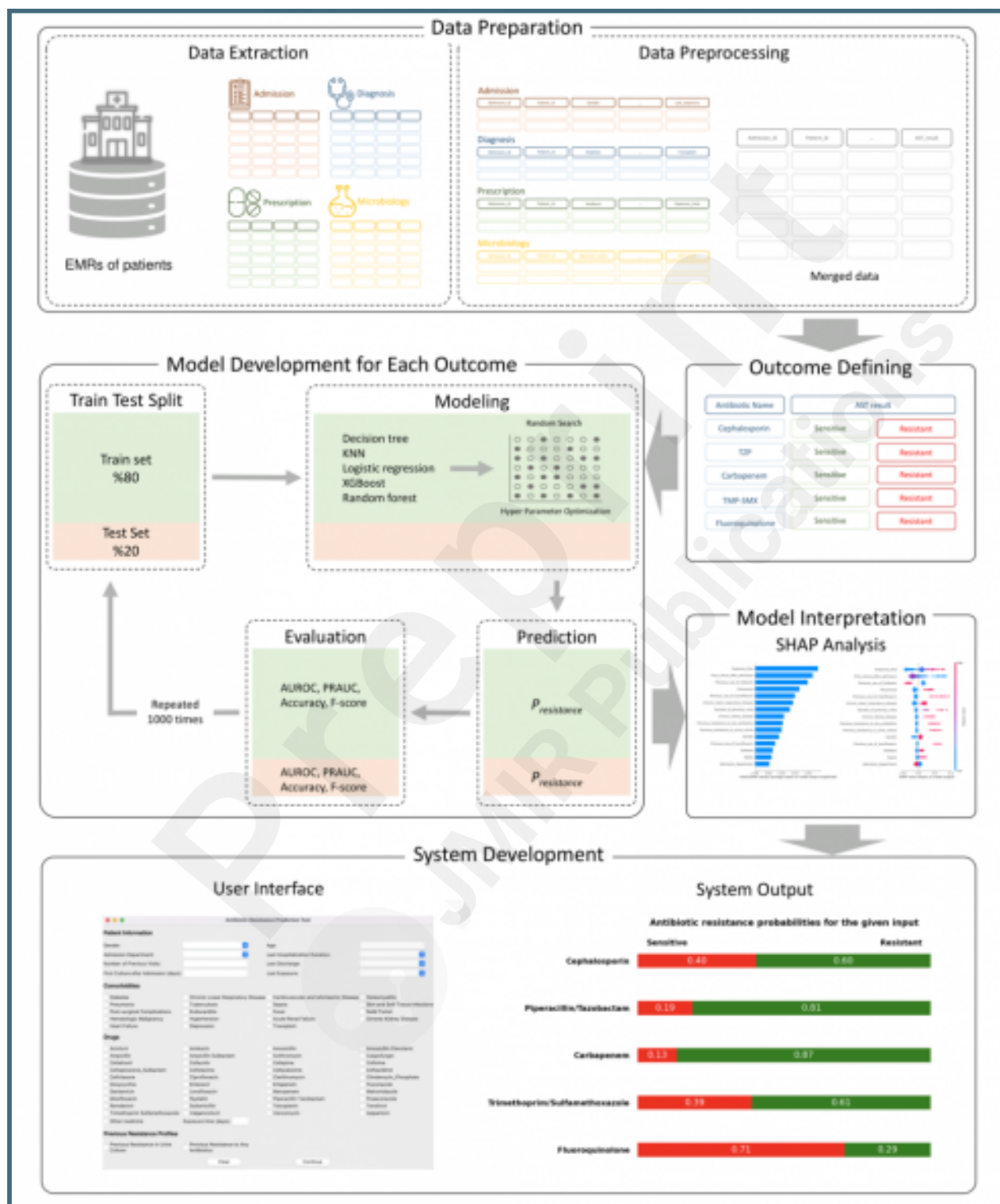
14. Macesic N, Polubriaginof F, Tatonetti NP. Machine learning: novel bioinformatics approaches for combating antimicrobial resistance. *Curr Opin Infect Dis*. 2017 Dec;30(6):511-7. PMID: 28914640. doi: 10.1097/QCO.0000000000000406.
15. Tucci V, Saary J, Doyle TE. Factors influencing trust in medical artificial intelligence for healthcare professionals: a narrative review. *Journal of Medical Artificial Intelligence*. 2021;5. doi: 10.21037/jmai-21-25.
16. Elshaw R, Al-Mallah MH, Sakr S. On the interpretability of machine learning-based model for predicting hypertension. *BMC Med Inform Decis Mak*. 2019 Jul 29;19(1):146. PMID: 31357998. doi: 10.1186/s12911-019-0874-0.
17. Tacconelli E. New strategies to identify patients harbouring antibiotic-resistant bacteria at hospital admission. *Clinical Microbiology and Infection*. 2006 2006/02/01;12(2):102-9. PMID: 16441446. doi: 10.1111/j.1469-0691.2005.01326.x.
18. Janssens A, Martens FK. Reflection on modern methods: Revisiting the area under the ROC Curve. *Int J Epidemiol*. 2020 Aug 1;49(4):1397-403. PMID: 31967640. doi: 10.1093/ije/dyz274.
19. Cook J, Ramadas V. When to consult precision-recall curves. *Stata J*. 2020 Mar;20(1):131-48. doi: 10.1177/1536867x20909693.
20. 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*. 2021 Jun 1;72(11):e848-e55. PMID: 33070171. doi: 10.1093/cid/ciaa1576.
21. Park H, Son MJ, Jung DW, Lee H, Lee JY. National Trends in Hospitalization for Ambulatory Care Sensitive Conditions among Korean Adults between 2008 and 2019. *Yonsei Medical Journal*. 2022;63(10):948-55. PMID: 36168248. doi: 10.3349/ymj.2022.0110.
22. Curtis CE, Al Bahar F, Marriott JF. The effectiveness of computerised decision support on antibiotic use in hospitals: A systematic review. *PLoS One*. 2017;12(8):e0183062. PMID: 28837665. doi: 10.1371/journal.pone.0183062.
23. Laka M, Milazzo A, Merlin T. Can evidence-based decision support tools transform antibiotic management? A systematic review and meta-analyses. *Journal of Antimicrobial Chemotherapy*. 2020;75(5):1099-111. PMID: 31960021. doi: 10.1093/jac/dkz543.



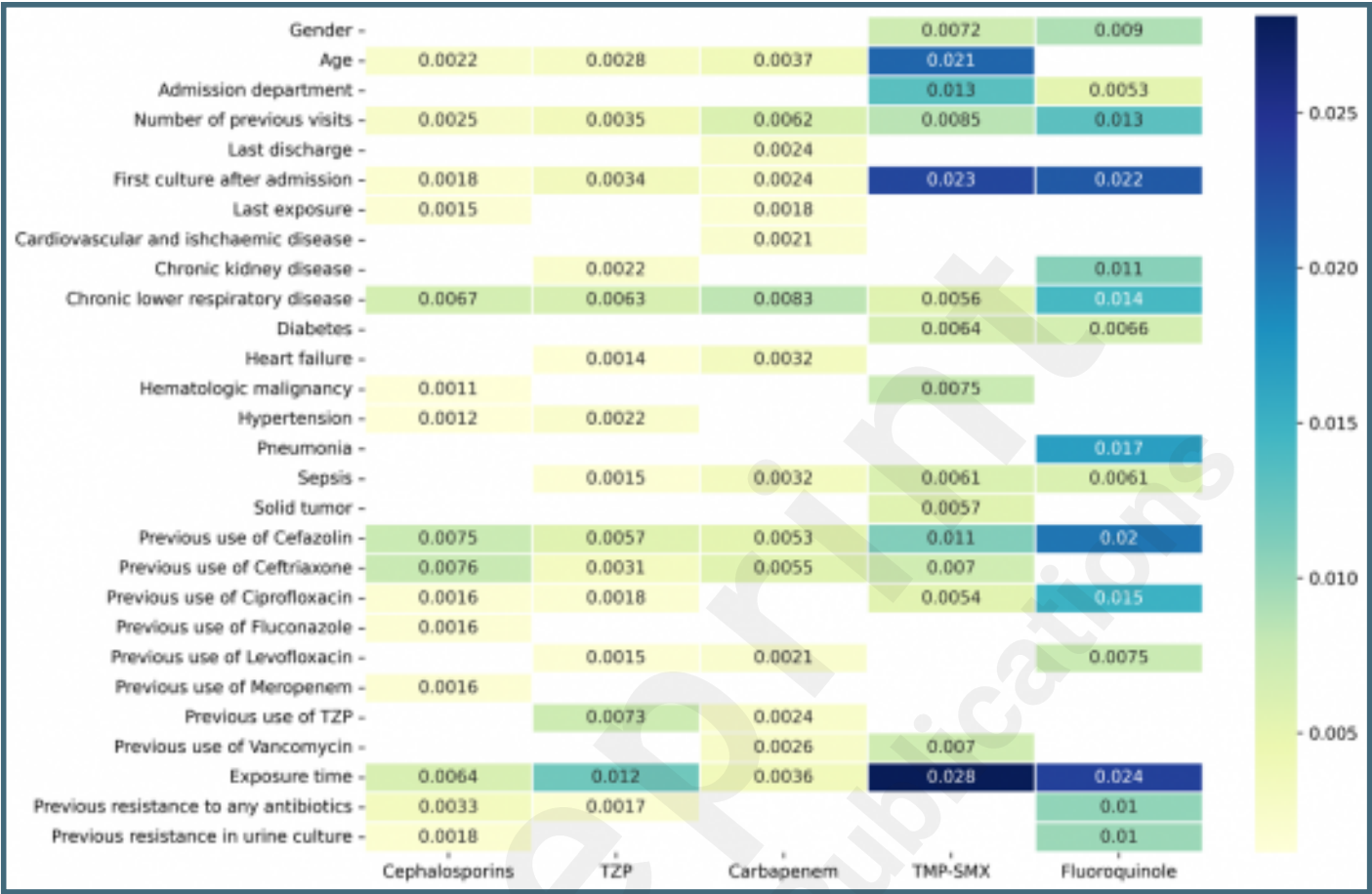
Supplementary Files

Figures

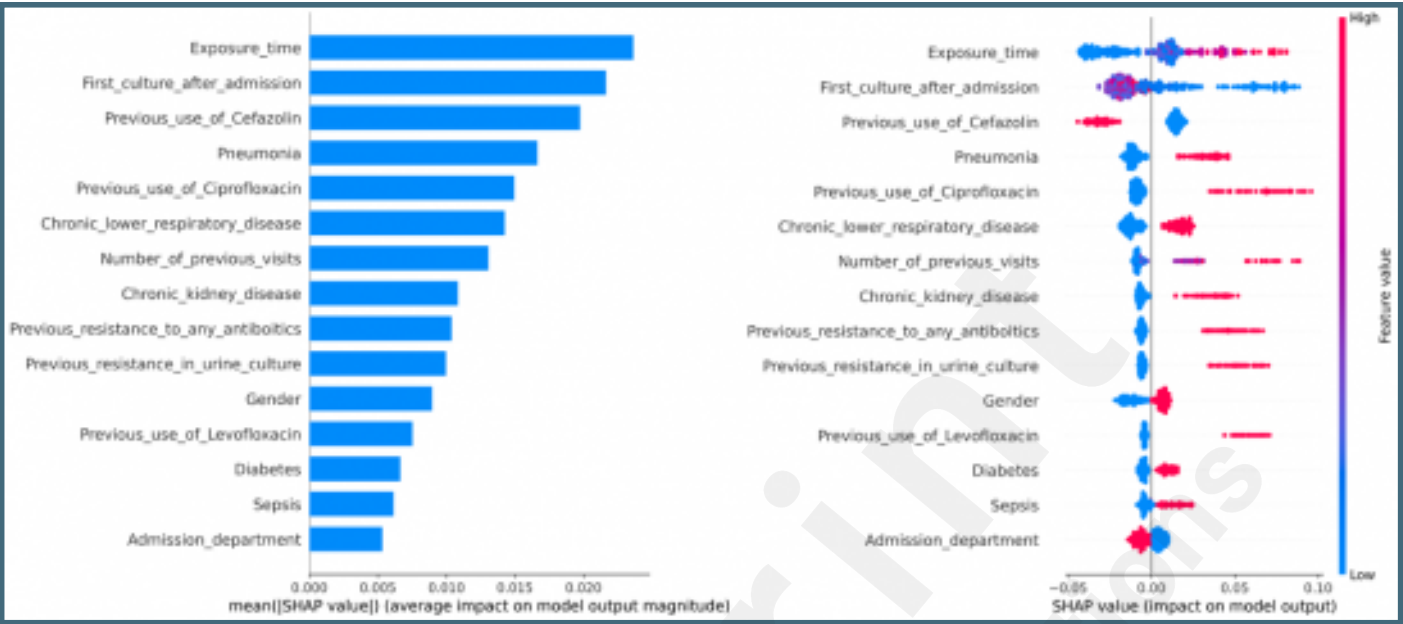
Summary of the research process. EMRs: electronic medical records; AST: antibiotic susceptibility test; TZP: piperacillin-tazobactam; TMP-SMX: trimethoprim-sulfamethoxazole; KNN: K-nearest neighbors; XGBoost: eXtreme Gradient Boosting; AUROC: area under the ROC curve; PRAUC: precision-recall area under the curve.



SHAP values of the 15 most important features in the prediction models. SHAP: SHapley Additive exPlanations; TZP: piperacillin-tazobactam; TMP-SMX: trimethoprim-sulfamethoxazole.



SHAP analysis results of fluoroquinolone resistance prediction model. (A) The feature importance bar plot. (B) The SHAP summary dot plot. SHAP: SHapley Additive exPlanations.



The user interface of the CDSS.

Antibiotic Resistance Prediction Tool

Patient Information

Gender: Age:

Admission Department: Last Hospitalization Duration:

Number of Previous Visits: Last Discharge:

First Culture after Admission (days): Last Exposure:

Comorbidities

☒ Diabetes ☒ Chronic Lower Respiratory Disease ☐ Cardiovascular and Ischaemic Disease ☐ Osteomyelitis

☐ Pneumonia ☐ Tuberculosis ☐ Sepsis ☐ Skin and Soft Tissue Infections

☐ Post-surgical Complications ☐ Endocarditis ☐ Fever ☐ Solid Tumor

☐ Hematologic Malignancy ☐ Hypertension ☐ Acute Renal Failure ☐ Chronic Kidney Disease

☐ Heart Failure ☐ Depression ☐ Transplant

Drugs

☐ Aciclovir ☐ Amikacin ☐ Amoxicillin ☐ Amoxicillin Clavulanic

☐ Ampicillin ☐ Ampicillin Sulbactam ☐ Azithromycin ☐ Caspofungin

☐ Cefadroxil ☒ Cefazolin ☐ Cefepime ☐ Cefixime

☐ Cefoperazone_Sulbactam ☐ Cefotaxime ☐ Cefpodoxime ☐ Ceftazidime

☐ Ceftriaxone ☐ Ciprofloxacin ☐ Clarithromycin ☐ Clindamycin_Phosphate

☐ Doxycycline ☐ Entecavir ☐ Ertapenem ☐ Fluconazole

☐ Gentamicin ☐ Levofloxacin ☐ Meropenem ☐ Metronidazole

☐ Moxifloxacin ☐ Nystatin ☐ Piperacillin Tazobactam ☐ Posaconazole

☐ Remdesivir ☐ Sulfamethoxazole ☐ Teicoplanin ☐ Tenofovir

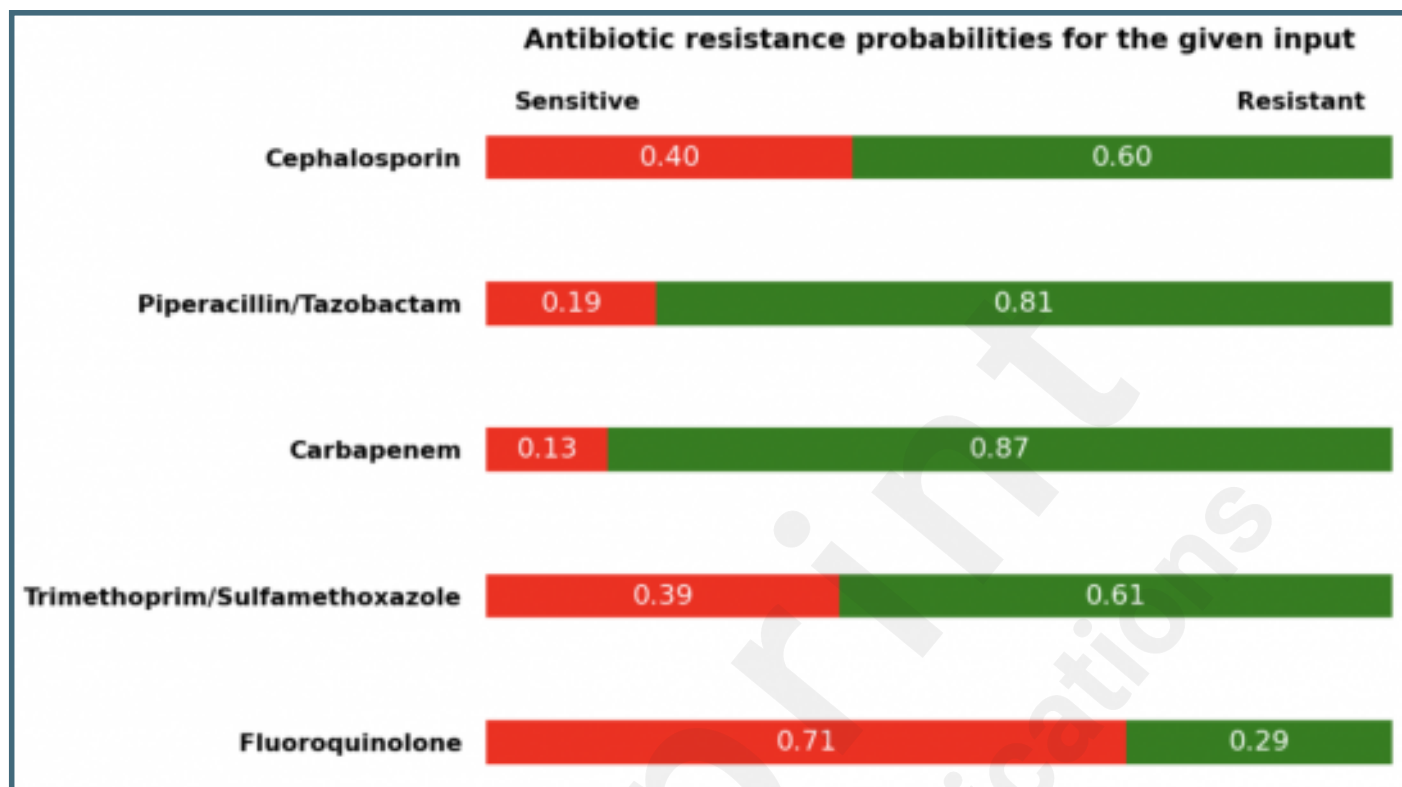
☐ Trimethoprim Sulfamethoxazole ☐ Valganciclovir ☐ Vancomycin ☐ Isepamicin

☐ Other medicine Exposure time (days):

Previous Resistance Profiles

☒ Previous Resistance in Urine Culture ☐ Previous Resistance to Any Antibiotics

The system output for the given data.



Multimedia Appendixes

The antibiotics included in each antibiotic class.

URL: <http://asset.jmir.pub/assets/00ef413d978501856ce75cfa0683059e.docx>

Description of input variables.

URL: <http://asset.jmir.pub/assets/fb6c0ffaba0dcec7579f5b7005e52391.docx>

Classification performances of the decision tree models.

URL: <http://asset.jmir.pub/assets/150839c3d0779e1816a22df489951521.docx>

Classification performances of the KNN models.

URL: <http://asset.jmir.pub/assets/c0a8a62ca08c14972b470d138d41d79b.docx>

Classification performances of the logistic regression models.

URL: <http://asset.jmir.pub/assets/bef1645943d28e0445f927ac6c8af16b.docx>

Classification performances of the XGBoost models.

URL: <http://asset.jmir.pub/assets/7f04044a1b7829c71c864bb5d79d291e.docx>

The SHAP feature importance bar plot (A) and SHAP summary plot (B) of the cephalosporin resistance prediction model.

URL: <http://asset.jmir.pub/assets/5f180fd8f096d5aa8afc331832e96032.png>

The SHAP feature importance bar plot (A) and SHAP summary plot (B) of the TZP resistance prediction model.

URL: <http://asset.jmir.pub/assets/19b3cd5a7e5f36a5277d039a26d76bc3.png>

The SHAP feature importance bar plot (A) and SHAP summary plot (B) of the carbapenem resistance prediction model.

URL: <http://asset.jmir.pub/assets/57739a0f3090395da593817eb7c69c57.png>

The SHAP feature importance bar plot (A) and SHAP summary plot (B) of the TMP-SMX resistance prediction model.

URL: <http://asset.jmir.pub/assets/ea6389ad051429144ee2f0ae2fcfdad.png>