https://doi.org/10.31925/farmacia.2026.5.22 **ORIGINAL ARTICLE**

MACHINE LEARNING-BASED PREDICTION OF ANTIBIOTIC SUSCEPTIBILITY TO SUPPORT PERSONALISED PHARMACOTHERAPY IN HOSPITALIZED PATIENTS: EVIDENCE FROM AN INDONESIAN REFERRAL HOSPITAL

IKHWAN YUDA KUSUMA [1]*, MUHAMMAD SYAIFUL ALIIM [2], FIQIH NURKHOLIS [1], MUH.AKBAR BAHAR [3]

[1]*Pharmacy Study Program, Faculty of Health, Universitas Harapan Bangsa, 53182 Purwokerto, Indonesia*
[2]*Department of Electrical Engineering, Faculty of Engineering, Universitas Jenderal Soedirman, 53371 Purbalingga, Indonesia*
[3]*Department of Pharmacy, Faculty of Pharmacy, Universitas Hasanuddin, 90245 Makassar, Indonesia*

**corresponding author: ikhwanyudakusuma@uhb.ac.id*

Manuscript received: April 2026

Abstract

Antimicrobial resistance (AMR) is a major global health threat, particularly in resource-limited settings such as Indonesia, where empirical antibiotic selection is often unguided due to delayed diagnostics. This study developed and evaluated a multi-label, multi-class machine learning model to predict antibiotic susceptibility using routine clinical and microbiological data. A retrospective dataset of hospitalized patients in 2024 receiving systemic antibiotics classified under Anatomical Therapeutic Chemical (ATC) code J01 in an Indonesian secondary referral hospital was analyzed. Predictors included age, sex, International Classification of Diseases, Tenth Revision (ICD-10) diagnosis, hospital unit, specimen type, and microorganism. A multi-output Random Forest model was trained using an 80:20 stratified split and evaluated using accuracy, precision, recall, F1-score, subset accuracy, and area under the receiver operating characteristic curve (AUC-ROC). Across the evaluated antibiotics, model performance varied substantially when assessed using macro-averaged metrics and majority-class baselines. Meropenem showed high overall accuracy (0.94), but its macro-F1 score was low (0.32), and the accuracy lift above the no-information rate was minimal (0.005), indicating that the predominance of susceptible isolates largely drove accuracy. The summary AUC-ROC for meropenem was 0.70, whereas the resistant-class one-*vs*-rest AUC was 0.81. Therefore, model performance was interpreted primarily using macro-F1, class-wise metrics, AUC-ROC, no-information rate, and accuracy lift rather than accuracy alone. Misclassification mainly occurred between susceptible and resistant β-lactam classes, while intermediate categories showed lower AUC due to imbalance. Key predictors included age, microorganism, and diagnosis. The model shows potential to support empirical prescribing, although prospective validation is needed.

Rezumat

Studiul a dezvoltat și evaluat un model de învățare automată multi-etichetă și multi-clasă, destinat predicției sensibilității la antibiotice pe baza datelor clinice și microbiologice de rutină. A fost analizat un set retrospectiv de date privind pacienți spitalizați în anul 2024, care au primit antibiotice sistemice (ATC J01) într-un spital de referință secundar din Indonezia. Pentru antibioticele evaluate, performanța modelului a variat considerabil atunci când a fost analizată prin metrici mediate la nivel de macro și prin raportare la valorile de referință ale clasei majoritare. Meropenemul a înregistrat o acuratețe globală ridicată (0,94), însă scorul macro-F1 a fost scăzut (0,32), iar creșterea acurateței față de rata de neinformare a fost minimă (0,005), ceea ce indică faptul că predominanța izolatelor sensibile a determinat, în mare măsură, acuratețea obținută. Valoarea AUC-ROC pentru meropenem a fost 0,70, în timp ce AUC de tip „unu-versus-restul" pentru clasa rezistentă a fost 0,81. Prin urmare, performanța modelului a fost interpretată în principal pe baza scorului macro-F1, a metricilor pe clase, a AUC-ROC, a ratei de neinformare și a creșterii acurateței, și nu exclusiv pe baza acurateței. Erorile de clasificare au apărut preponderent între clasele sensibile și rezistente de β-lactamine, în timp ce categoriile intermediare au prezentat valori AUC mai reduse, ca urmare a dezechilibrului dintre clase. Modelul și-a demonstrat potențialul de a sprijini prescrierea empirică a antibioticelor, deși este necesară validarea prospectivă.

Keywords: anti-bacterial agents, drug resistance, electronic health records, hospital, machine learning

Introduction

Antimicrobial resistance (AMR) is one of the major global health threats of the 21st century, which the World Health Organization (WHO) has described as a "silent pandemic" that may reverse the achievements of modern medicine (1). In 2019, AMR-associated infections were directly attributed to 1.27 million deaths worldwide, while the burden in larger low- and middle-income countries (LMICs) was much

higher, especially in Southeast Asia (2). In this regard, Indonesia is among the countries with the highest rates of antimicrobial use and rising resistance due to inappropriate prescribing practices, the availability of antibiotics without a prescription, and underdeveloped laboratory facilities (3). In hospitals, where AMR-related infections are common, it is vital to select appropriate empiric therapy (4).

Empirical antibiotic use, however, is often fraught with errors owing to the inevitable delays in antimicrobial susceptibility testing (AST) results (5). There is a substantial mismatch between the antibiotics prescribed and the susceptibility profile of the infecting organisms. This leads to failure of treatment, high morbidity rates, and resistance (6). These limitations become even more pronounced in LMIC hospitals, where rapid diagnosis and timely clinical decision-making tools are limited (7). Therefore, there is an urgent need for new solutions that enable timely predictions based on available data to guide empirical antibiotic use (8). Thus, there is an urgent need for innovative approaches that can provide timely, data-driven predictions to guide empiric antibiotic selection (9). In the setting of an LMIC hospital, poor laboratory capabilities and AST results that are either delayed or expensive prevent timely antibiotic prescriptions. Thus, the need for cheap and scalable decision-making tools becomes paramount (10-12).

The most recent developments in the field of artificial intelligence (AI) and machine learning (ML) provide opportunities to overcome these issues (13). ML models have been applied to predict AMR by analyzing different datasets related to clinical and microbiological settings and have achieved outstanding performance. Created preliminary ML models that analyzed pre-laboratory parameters to accurately predict AST results (AUC-ROC > 0.90) (14, 15) proved that multi-label models utilizing matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry could predict resistance to several antibiotics at once with high interpretability. Moreover, a study based in the United Kingdom (UK) (16) and one based in the United States of America (USA) (17) have used a gradient-boosting approach and electronic health records (EHRs) to predict AMR and guide antimicrobial stewardship programs.

Though there have been positive strides in this direction, the vast majority of studies have been conducted in high-income nations and have relied upon carefully curated research databases or have focused on single-output classification methods (18). Few studies have explored multi-output classification approaches for the simultaneous prediction of antibiotic susceptibility in routine clinical practice (19). Even fewer have tested these methods in real-life situations such as those encountered in LMIC hospitals (20). In LMIC hospitals, ML is especially appealing as it uses data already available in hospitals and does not need any extra laboratory investment. Trained ML algorithms can be integrated into the hospital information system to provide real-time predictions at low cost (21). This is especially relevant to a country like Indonesia, where multiple infections and antibiotic use need to be predicted simultaneously due to the country's diversity (21).

This study aimed to design and assess a multi-label machine learning classifier to predict systemic antibiotic resistance using clinical and microbiological data collected at a secondary referral hospital in Indonesia, to improve empirical antibiotic prescribing.

Materials and Methods

Study design and setting

This retrospective observational study utilized antimicrobial susceptibility and patient-level clinical data obtained from Cilacap Teaching Hospital, a 400-bed secondary referral hospital in Central Java, Indonesia. The hospital serves as a regional referral center for over two million residents. Data were extracted from the hospital's electronic medical record platform, Pilar Data Esensi, which is integrated into the Hospital Information System (SIM-RS). The study included all hospitalized patients who received systemic antibiotics classified under the Anatomical Therapeutic Chemical (ATC) code J01 during 2024. The initial dataset was extracted from the hospital electronic medical record system and subsequently screened according to the predefined eligibility criteria. Antibiotics with fewer than 30 valid AST results were excluded from model development to improve the stability of performance estimates. Ethical considerations regarding the privacy of information were maintained, and all information was anonymized before performing any data analyses. The modeling process in general is depicted in Figure 1.

Data collection and variables

The dataset comprised structured clinical and microbiological records. Predictor variables included patient sex, age, ICD-10 diagnosis code, hospital unit (location), specimen type (*e.g.*, blood, sputum, urine), and identified microorganism. For admissions with multiple cultures, only the earliest eligible AST *per* admission was retained for analysis. The target variables were antimicrobial susceptibility outcomes reported for a panel of systemic antibiotics derived from the earliest eligible AST *per* admission, with categorical labels indicating Susceptible (S), Intermediate (I), or Resistant (R). Susceptibility outcomes were initially exported as numeric codes for software interoperability (S = 0, R = 1, I = 2). These numeric codes were used only during data import and were mapped back to the corresponding clinical string labels (S, I, and R) before model evaluation, performance calculation, and visualization.

Consequently, all reported confusion matrices, class-wise metrics, and ROC curves are presented using the clinical labels rather than the underlying numeric codes. All class-wise metrics, confusion matrices, and ROC analyses were generated using the fixed clinical label order S, I, and R to avoid ambiguity in downstream interpretation. No ordinal assumption was applied to the susceptibility categories. A binary formulation (S *vs*. non-S, where non-S = I + R) was not implemented in the current analysis and is reserved for prospective benchmarking. Antibiotics with fewer than 30 valid susceptibility test results were excluded from the analysis to improve the statistical stability of estimates, as studies have shown that using fewer than 30 isolates leads to high error rates and unreliable antibiogram results (22). Only systemic antibiotics (ATC J01) were retained; antifungals, antivirals, antiparasitic, antituberculosis agents, and topical antibacterials were excluded.

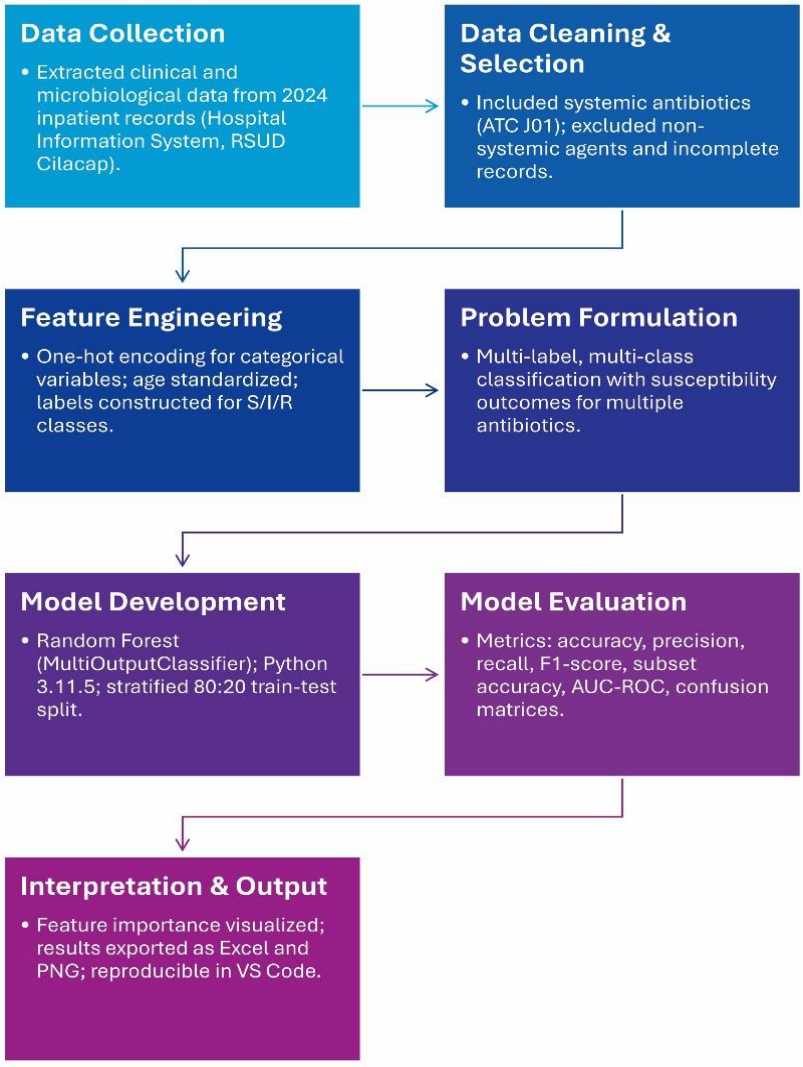


Figure 1.
Workflow of the multi-label machine learning model

Problem formulation and label construction

The predictive modeling task was framed as a multi-output, multi-class classification problem, in which a single isolate was simultaneously assigned susceptibility predictions for multiple antibiotics. In the implemented MultiOutputClassifier framework, each antibiotic was modeled independently without explicitly learning label dependencies. Each sample could be associated with multiple antibiotic labels, and each label contained three possible outcome classes. Categorical input features, such as sex, specimen, ICD-10 code, location, and organism, were one-hot encoded, while the numerical variable (age) was standardized using z-score normalization. These preprocessing steps were fitted on the training dataset and subsequently applied to the testing dataset to avoid information leakage. Records with missing

or invalid values in either predictors or labels were excluded to ensure the integrity of the training dataset.

Model development

Machine learning models were developed using Python (v3.11.5) and scikit-learn (v1.2.2). The dataset was randomly divided into training (80%) and testing (20%) subsets before model development. Because the prediction task involved multiple output labels, the split was performed without simultaneous multi-label stratification. To prevent information leakage, all preprocessing steps, including one-hot encoding and standardization, were fitted using the training data and subsequently applied to the testing data (23). A multi-label classification framework was implemented using the Multi-Output Classifier wrapper with Random Forest Classifier as the base estimator (24). Each antibiotic was modeled as an independent classification task, with predictions aggregated across all antibiotics for each patient sample. Random Forest models were configured with 100 decision trees and default parameters unless otherwise specified. Random Forest was chosen because it can accommodate high-dimensional categorical variables and provides good model interpretability. The algorithm is not inherently robust to severe class imbalance; model performance was interpreted using complementary metrics beyond accuracy.

Model evaluation

Model performance was evaluated independently for each antibiotic label using complementary metrics, with macro-averaged F1-score treated as the main summary measure because the susceptibility classes were imbalanced. Accuracy was reported only as a secondary descriptive metric because it may be inflated when a single susceptibility category dominates. Macro-averaged precision, macro-averaged recall, and macro-averaged F1-score were calculated to provide a more balanced assessment across the susceptible, intermediate, and resistant classes. (25). For each antibiotic, we also calculated the no-information rate, defined as the proportion of the majority class in the test set. This value represents the performance expected from a classifier that always predicts the most frequent susceptibility category. Accuracy lift was calculated as the difference between model accuracy and the no-information rate. To enable direct assessment of performance across the susceptible, intermediate, and resistant categories, the principal antibiotic-specific performance measures are presented in Table I and the corresponding figures in the main text.

Subset accuracy was calculated as the proportion of samples in which all predicted antibiotic outcomes were correctly classified. Confusion matrices were generated to compare observed and predicted susceptibility classes. ROC curves were constructed using a one-*vs*-rest approach only for antibiotics with sufficient representation across all three susceptibility classes (S, I, and R). For each ROC analysis, the positive class was defined as the target susceptibility category: AUC-S represented susceptible *versus* intermediate plus resistant (S *vs*. I + R), AUC-I represented intermediate *versus* susceptible plus resistant (I *vs*. S + R), and AUC-R represented resistant *versus* susceptible plus intermediate (R *vs*. S + I). Antibiotics with one or more insufficiently represented or absent classes were excluded from AUC-ROC analysis because reliable multiclass AUC estimation was not possible (27). These curves were visualized and saved as high-resolution plots to complement the tabulated metrics (27,28).

Because susceptibility classes were imbalanced, the macro-averaged F1-score was prioritized as the main summary metric. Accuracy was retained as a secondary descriptive metric because it may be inflated when one susceptibility class dominates. For each antibiotic, the majority class was identified as the most frequent susceptibility category, and the no-information rate was calculated as the proportion of this majority class. Accuracy lift was calculated as the difference between model accuracy and the no-information rate to estimate the improvement over a simple majority-class classifier.

Feature importance analysis

To improve the interpretability, the feature importance was measured based on the Gini index values of the Random Forest classifiers (29). The values were calculated for each antibiotic classifier and averaged to assess the average feature importance across all the antibiotics. The contribution of features was illustrated through bar graphs and heat maps, allowing the comparison of feature influence on the output of the model.

Data visualization

All evaluation outputs, including measures of performance, confusion matrices, ROC curves, and feature importance plots, were saved in standardized formats. Output in tabular form was saved as Excel spreadsheets, while figures were saved as high-resolution Portable Network Graphics (PNG) files ready for publishing. Modeling and evaluation took place using Visual Studio Code in a Python virtual environment.

Results and Discussion

The following subsection outlines the results from the multi-label and multi-class classification model for predicting antimicrobial susceptibility using clinical and microbiological attributes. The result evaluation involves performance measurement for each of the antibiotics individually, class-wise classification, AUC-ROC analysis, and model interpretability through feature importance.

This section presents the outcomes of the multi-label, multi-class classification model for predicting antimicrobial susceptibility based on clinical and microbiological features. The evaluation includes individual performance metrics *per* antibiotic, class-wise classification results, AUC-ROC analyses, and model interpretability through feature importance.

Model Performance Across Antibiotics

The performance of the multi-output Random Forest classifier varied across the selected antibiotics, as shown in Table I. To ensure consistency across the Abstract, Results, and Discussion, headline interpretation was restricted to antibiotics included in the revised main results table. Among these antibiotics, macro-F1 ranged from 0.32 to 0.53, indicating a limited to moderate balance in three-class classification. No antibiotic in the main table was interpreted as having high balanced performance based on macro-F1 alone.

Table I

Performance of antibiotic-specific classifiers relative to the majority-class baseline

Antibiotic	Samples	Majority class	No-information rate	Macro-F1	Macro-precision	Macro-recall	Accuracy	Accuracy lift	AUC-ROC
MEM	681	S	0.935	0.32	0.32	0.33	0.94	0.005	0.70
CAZ	679	R	0.474	0.41	0.42	0.40	0.52	0.046	0.62
AMC	628	R	0.479	0.40	0.40	0.40	0.51	0.031	0.61
CRO	552	R	0.585	0.40	0.39	0.40	0.59	0.005	0.65
CHL	493	S	0.517	0.36	0.35	0.37	0.54	0.023	0.55
CIP	736	S	0.595	0.36	0.35	0.37	0.56	-0.035	-
CTX	612	R	0.688	0.33	0.35	0.34	0.59	-0.098	0.54
LVX	542	S	0.616	0.32	0.31	0.33	0.50	-0.116	0.48
AMP	165	R	0.958	0.49	0.48	0.50	0.97	0.012	-
SAM	666	S	0.634	0.53	0.56	0.51	0.66	0.026	-

Note: The majority class denotes the most frequent susceptibility category for each antibiotic. The no-information rate represents the accuracy obtained by always predicting the majority class and was used as the baseline for comparison. Accuracy lift was calculated as the difference between model accuracy and the no-information rate. AUC-ROC was calculated using a one-vs-rest approach only when all three susceptibility classes (S, I, and R) were sufficiently represented. "–" indicates that reliable AUC-ROC estimation was not possible because one or more susceptibility classes were absent or insufficiently represented. Antibiotic abbreviations: MEM, meropenem; CAZ, ceftazidime; AMC, amoxicillin-clavulanic acid; CRO, ceftriaxone; CHL, chloramphenicol; CIP, ciprofloxacin; CTX, cefotaxime; LVX, levofloxacin; AMP, ampicillin; SAM, ampicillin-sulbactam. S, susceptible; I, intermediate; R, resistant; AUC-ROC, area under the receiver operating characteristic curve.

For meropenem, the model achieved high overall accuracy (0.94), but the majority class was susceptible, with a no-information rate of 0.935. The resulting accuracy lift was only 0.005, while macro-F1 was 0.32. This indicates that the high accuracy largely reflected the predominance of susceptible isolates rather than balanced discrimination across susceptible, intermediate, and resistant classes.

Accuracy lift was also limited for several other antibiotics, including ceftriaxone (0.005), chloramphenicol (0.023), ampicillin (0.012), and sulbactam/ampicillin (0.026). Negative accuracy lift was observed for ciprofloxacin, cefotaxime, and levofloxacin, indicating that their model accuracy did not exceed the majority-class baseline. These findings highlight the importance of interpreting accuracy in relation to class distribution.

In Table I, the summary AUC-ROC for meropenem was 0.70. The value of 0.81 refers specifically to the resistant-class one-*vs*-rest AUC shown in Figure 4 and is now described only as a class-specific ROC result. This distinction was added to avoid conflating antibiotic-level summary AUC-ROC with class-specific one-*vs*-rest AUC values.

Evaluation of class-wise prediction performance

To further assess classification capability, confusion matrices were generated for selected antibiotics (Figure 3). The model correctly classified nearly all susceptible cases of MEM, with minor misclassification of intermediate or resistant instances. In contrast, the confusion matrices for AMC, CAZ, and CRO revealed greater dispersion, particularly among S and R classes, suggesting overlap in feature space for those categories.

Receiver operating characteristic analysis

Figure 4 presents the multiclass ROC curves for the same four antibiotics, plotted using the one-*vs*-rest strategy for the three susceptibility classes. MEM showed robust AUC values, particularly for the Resistant class (0.81), while CRO and AMC achieved moderately strong AUCs for Susceptible and Resistant classes. The intermediate class generally had lower AUC scores across antibiotics, reflecting class imbalance and greater difficulty in prediction.

Contribution of clinical features to model predictions

Feature importance analysis based on mean Gini importance values from the random forest models revealed that age and organism type were the most influential predictors across all antibiotic models.

These were followed by ICD-10 code and hospital location, while sex and specimen type had the least influence. The aggregate contributions are displayed in Figure 5.



Figure 2.

Heatmap showing accuracy, precision, recall, F1-score, and AUC-ROC for each antibiotic-specific classifier. Color intensity reflects metric performance. Antibiotic abbreviations: AMC, amoxicillin-clavulanic acid; AMK, amikacin; AMP, ampicillin; ATM, aztreonam; CAZ, ceftazidime; CHL, chloramphenicol; CIP, ciprofloxacin; CLI, clindamycin; CRO, ceftriaxone; CTX, cefotaxime; CZO, cefazolin; ERY, erythromycin; FEP, cefepime; FOX, cefoxitin; IPM, imipenem; LNZ, linezolid; LVX, levofloxacin; MEM, meropenem; MXF, moxifloxacin; OXA, oxacillin; RIF, rifampicin; SAM, ampicillin-sulbactam; TCY, tetracycline; TZP, piperacillin-tazobactam; VAN, vancomycin; COL, colistin; MHL, Madurahydroxylactone; NFT, nitrofurantoin; PCG, penicillin G; PIP, piperacillin; TEC, teicoplanin; TMP, trimethoprim; TXS, trimethoprim-sulfamethoxazole. AUC-ROC, area under the receiver operating characteristic curve

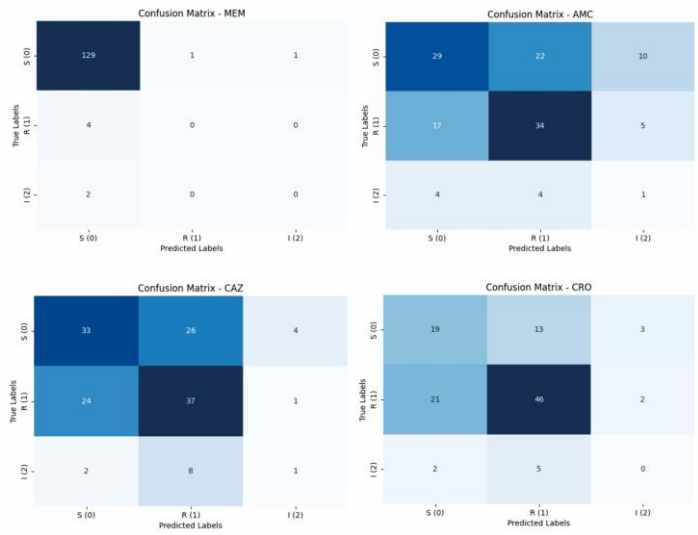


Figure 3.

Confusion matrices for selected antibiotics. True labels are shown on the y-axis and predicted labels on the x-axis using the fixed clinical label order S, I, and R, after mapping from numeric codes: (a) meropenem (MEM); (b) amoxicillin-clavulanic acid (AMC); (c) ceftazidime (CAZ); and (d) ceftriaxone (CRO). S, susceptible; I, intermediate; R, resistant

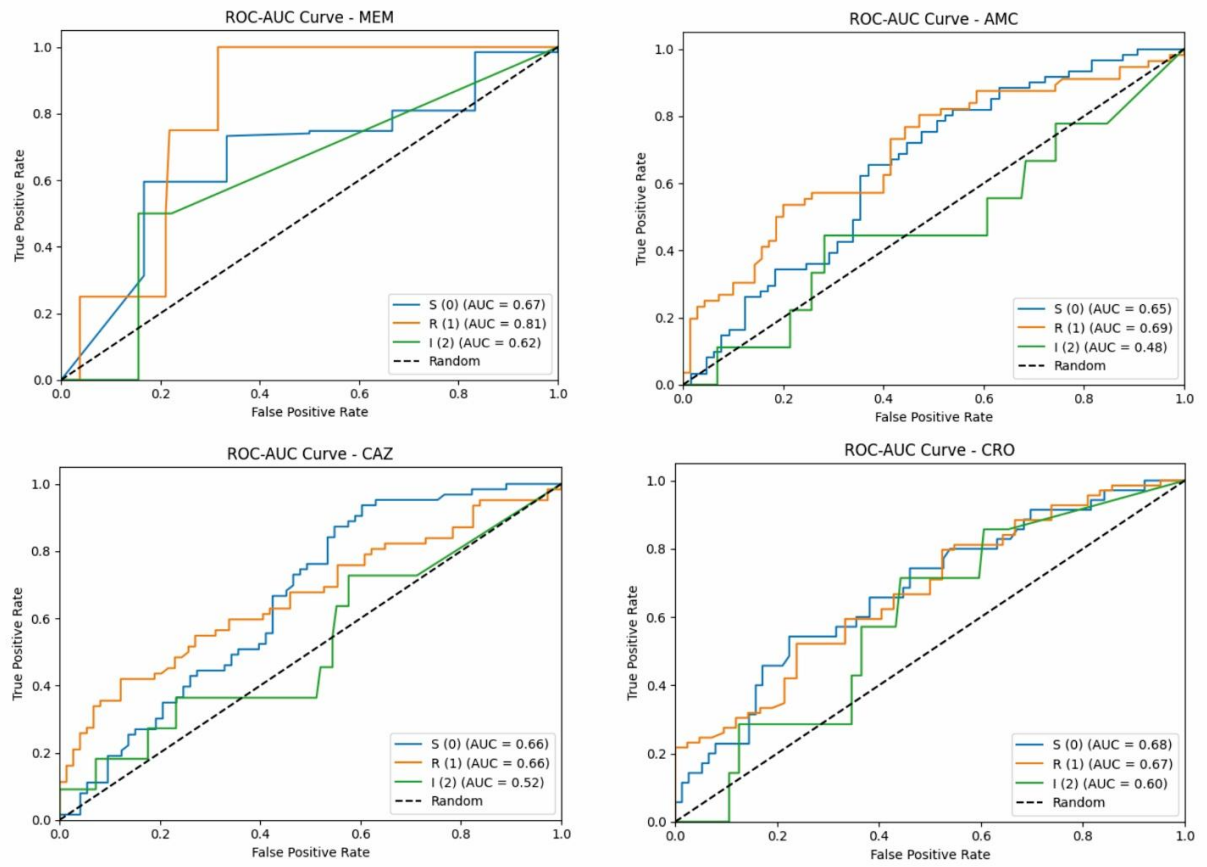


Figure 4.

Receiver operating characteristic (ROC) curves for one-vs-rest classification of the three susceptibility classes (S, I, and R). (a) MEM: AUC-S = 0.67, AUC-R = 0.81, AUC-I = 0.62; (b) AMC: AUC-S = 0.65, AUC-R = 0.69, AUC-I = 0.48; (c) CAZ: AUC-S = 0.66, AUC-R = 0.66, AUC-I = 0.52; and (d) CRO: AUC-S = 0.68, AUC-R = 0.67, AUC-I = 0.60. MEM, meropenem; AMC, amoxicillin-clavulanic acid; CAZ, ceftazidime; CRO, ceftriaxone; AUC, area under the curve; S, susceptible; I, intermediate; R, resistant.

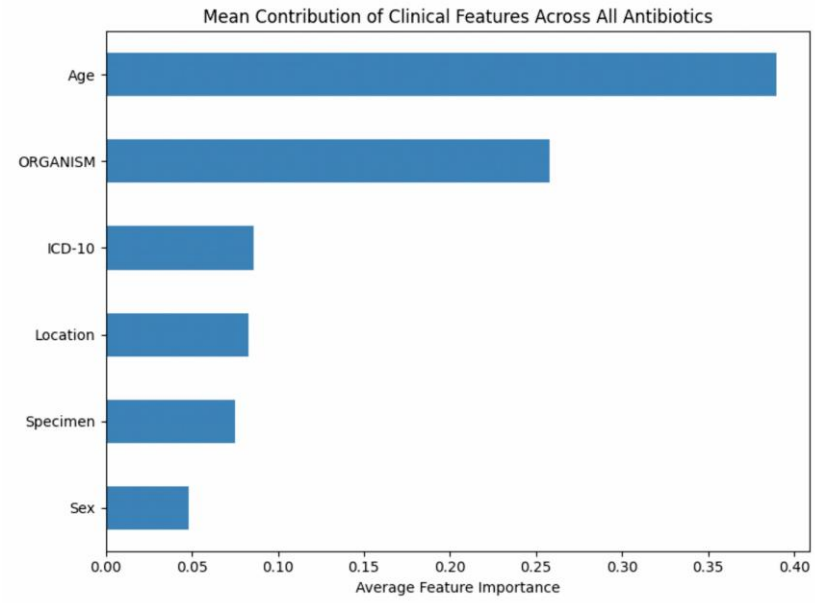


Figure 5.

Mean contribution of clinical features across all antibiotic-specific models. Values represent average feature importance scores.

To the best of our knowledge, this is the first study to develop and evaluate a multi-output, multi-class machine learning model capable of simultaneously predicting susceptibility outcomes (S, R, and I) across a broad panel of systemic antibiotics in a referral hospital in Indonesia. Unlike prior AMR prediction studies, which have predominantly used single-label binary models in high-income settings, our approach simultaneously predicts susceptibility outcomes for multiple antibiotics while explicitly incorporating the Intermediate category. However, the implemented MultiOutputClassifier models each antibiotic independently and do not explicitly capture co-resistance relationships.

By incorporating predictors such as age, sex, ICD-10 diagnosis, isolated organism, specimen type, and hospital location, the model produced heterogeneous performance across antibiotic-specific tasks. The revised analysis indicates that high accuracy should not be interpreted as evidence of strong model performance when class imbalance is substantial. This was most evident for meropenem, where accuracy was high (0.94), but macro-F1 was low (0.32), and the accuracy lift above the no-information rate was minimal (0.005). Therefore, meropenem was not interpreted as a high-performing classifier based on accuracy alone (30). Antibiotics with limited sample sizes (*e.g.*, trimethoprim (TMP), nitrofurantoin (NFT), and colistin (COL)) were not included in the main performance interpretation because their estimates may be unstable and require validation in larger datasets (32). Amoxicillin-clavulanate (AMC) and chloramphenicol (CHL) were less effective, probably because of the heterogeneity of their prescription practices and numerous intermediate findings blurring the definition of their classes (33, 34). Drugs in the cephalosporin class, such as ceftriaxone (CRO) and ceftazidime (CAZ), exhibited intermediate accuracy, with substantial classification confusion between susceptible and resistant categories, likely due to extended-spectrum β-lactamase (ESBL)/AmpC-mediated resistance in the region (35, 36). This approach demonstrates the feasibility of supporting early antimicrobial susceptibility prediction before laboratory confirmation. However, prospective multicenter validation and clinical impact assessment are required before implementation in antimicrobial stewardship or routine clinical decision-making.

The implementation of the entire three-class model (S/I/R) instead of the reduced two-class model represents an important step forward from previous ML-AMR research (30). The use of the "Intermediate" class in our model allows us to more precisely reflect the practice of clinical microbiology and prevent the simplification of complex situations of drug resistance (31). The accuracy of classification for the Intermediate class across most antibiotics is reduced due to the complexity of this group's differentiation, which overlaps phenotypically with other categories (32). This fact indicates that the results in this class can be caused by biological diversity, differences in minimum inhibitory concentrations (MICs), or even clinical practice (33).

Interpretability of the model remains an essential criterion for its use in clinical practice (34). It was shown during our analysis of feature importance that age and organism have the greatest predictive power regarding the patient's susceptibility, followed by the ICD-10 diagnosis and hospitalization site. Age is a known predictor of infection risk, pharmacokinetics, and immune response, whereas the knowledge of the organism helps to determine the susceptibility to different antibiotic classes (35). The association between age and antibiotic resistance differs among antibiotic classes and bacterial species and can be manifested by the resistance increasing with age in some cases or by constant resistance regardless of the patient's age (35,36). The reduced importance of the sex and sample type shows that these criteria have less discriminatory power, even though the data is collected systematically in clinical practice (37). The determination of the pathogen type as one of the most important factors implies the inclusion of preliminary culture-based information (Gram stain, rapid ID) into the algorithm of treatment decisions (38). From the perspective of clinical pharmacy practice, the application of such algorithms would facilitate a more accurate selection of an empirical antibiotic, particularly in circumstances where resistance is unknown or delayed (39). These patient-specific antimicrobial susceptibility predictions may provide complementary information to support empirical antibiotic selection when integrated with microbiological findings and clinical judgment (40). Their potential contribution to antimicrobial stewardship should be confirmed through prospective implementation studies (41). An integration with electronic health records would allow making susceptibility predictions during the encounter, thus reducing broad-spectrum coverage and preventing further emergence of resistance (42). These algorithms would facilitate a compromise between the need to have broad coverage and the precision of antimicrobial administration, which could improve antibiotic stewardship by using less broad-spectrum antibiotics and prevent further emergence of resistance (41). These findings also have practical implications for the work of the antimicrobial stewardship program, regarding identifying patients who might need earlier de-escalation or escalation of therapy based on the prediction of patient- and pathogen-level factors (43).

Despite these promising results, several limitations should be acknowledged. First, the study used retro-

spective data from a single hospital and did not include external validation from other centers, which may limit its generalizability. Second, class imbalance, particularly for the Intermediate and Resistant categories, likely contributed to the lower balanced classification performance. No class weighting, resampling, threshold optimization, or probability calibration was applied in the present analysis; therefore, the model may retain bias toward the majority susceptibility class. Future studies should evaluate imbalance-aware learning strategies, including class-weighted Random Forest, resampling techniques, calibration methods, and threshold optimization to improve minority-class prediction. Third, the models did not incorporate temporal variables, prior antibiotic exposure, or microbiological MIC values, key laboratory values (*e.g.*, complete blood count [CBC], C-reactive protein [CRP], renal function), comorbidity index (*e.g.*, Charlson Index), or hospital level of care (Intensive Care Unit *vs.* general ward). Where multiple cultures occurred within an admission, we analyzed the earliest available AST; however, the dataset was split at the record level rather than the patient or admission level, and we did not explicitly verify that individual patients were confined to a single data partition. Consequently, residual within-patient correlation or potential information leakage cannot be completely excluded. Future studies should perform patient- or admission-level splitting to ensure strict independence between training and testing datasets. Finally, although we report high model performance for selected antibiotics, results for others with low sample counts (*e.g.*, TMP, NFT, or COL) should be interpreted cautiously due to potential overfitting. AUC-ROC analyses were restricted to antibiotics with sufficient representation across all susceptibility classes. Antibiotics with inadequate class representation were excluded because reliable multiclass AUC-ROC estimation was not feasible. Methodologically, we evaluated a single algorithm (Random Forest) without head-to-head benchmarking against alternative classifiers or a binary formulation (S *vs*. non-S); the latter is specified for prospective validation. We also did not assess usability, workflow integration, or clinical impact, so the real-world effectiveness of this decision support remains to be demonstrated. Future work should focus on external and temporal validation across sites, the incorporation of additional patient-level and microbiological features, and the evaluation of real-world clinical impact through prospective implementation studies.

Class imbalance remains a major limitation of this study. For several antibiotics, particularly meropenem and ampicillin, one susceptibility category dominated the dataset, which inflated overall accuracy and reduced the clinical interpretability of accuracy as a stand-alone metric. Although we added majority-class baselines and accuracy lift to contextualize model performance, no resampling, cost-sensitive learning, or threshold optimization was applied in the present analysis. In addition, model performance was evaluated using a single train/test split, and confidence intervals were not estimated. Consequently, the stability and uncertainty of the reported performance metrics could not be fully assessed. The current modeling framework does not explicitly learn dependencies among antibiotic susceptibility labels. Future studies should evaluate methods such as classifier chains, label powerset approaches, or shared-representation models to better capture co-resistance patterns across antibiotics. Future work should also incorporate repeated cross-validation or bootstrap resampling to estimate confidence intervals and provide more robust performance evaluation.

Conclusions

This study demonstrates the feasibility of developing a multi-output, multi-class machine learning model to simultaneously predict antimicrobial susceptibility outcomes (S/I/R) across a broad panel of systemic antibiotics in a secondary referral hospital in Indonesia. By leveraging routinely available clinical and microbiological data, the model demonstrated the feasibility of predicting antimicrobial susceptibility across multiple systemic antibiotics. However, given the retrospective single-center design and the moderately balanced classification performance, external validation and prospective clinical evaluation are required before routine clinical implementation. Feature importance analysis underscored the central role of pathogen identity, patient age, and diagnostic context in shaping resistance risk.

Acknowledgement

The authors gratefully acknowledge the support of Universitas Hasanuddin, Universitas Harapan Bangsa, and Universitas Jenderal Soedirman. We also thank the team at Cilacap Teaching Hospital for their assistance in data provision and administrative coordination.

Funding

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate

Not applicable.

AI use disclosure

During the preparation of this manuscript, the authors used ChatGPT (OpenAI, GPT-5.5) for language editing and writing assistance to improve grammar, spelling, word choice, readability, and clarity of the manuscript. The authors carefully reviewed, verified, and substantially revised all outputs generated by the tool and take full responsibility for the final content of the manuscript. In addition, the revised manuscript was evaluated using an AI-detection tool, yielding an AI-generated text similarity score of approximately 2%, which is below the 20% threshold required by the journal. No artificial intelligence tool was used to generate research ideas, formulate hypotheses, design the study, analyze or interpret data, or draw scientific conclusions.

Conflict of interest

The authors declare no conflict of interest.

References

1. Giri P, Aikat A. The impending global challenge of antimicrobial resistance: A call for urgent action. Indian J Community Health. 2024;36(5):624-628.
2. Bertagnolio S, Suthar AB, Tosas O, Van Weezenbeek K. Antimicrobial resistance: Strengthening surveillance for public health action. PLoS Med. 2023;20(7):e1004265.
3. Limato R, Nelwan EJ, Mudia M, Brabander JD, Guterres H, Enty E, et al. A multicentre point prevalence survey of patterns and quality of antibiotic prescribing in Indonesian hospitals. JAC Antimicrob Resist. 2021;3(2):dlab047.
4. Zhang P, Hao J, Zhang Y, Su J, Sun G, Xie J, et al. Understanding the clinical and molecular epidemiological characteristics of carbapenem-resistant *Acinetobacter baumannii* infections within intensive care units of three teaching hospitals. Ann Clin Microbiol Antimicrob. 2025;24(1):2.
5. Edmondson R, Saeed K, Green S, O'Dwyer M. Improving turnaround times for routine antimicrobial sensitivity testing following European Committee on Antimicrobial Susceptibility Testing methodology in patients with bacteraemia. Antibiotics (Basel). 2024;13(11):1094.
6. Tallman GB, Vilches-Tran RA, Elman MR, Bearden DT, Taylor JE, Gorman PN, et al. Empiric antibiotic prescribing decisions among medical residents: The role of the antibiogram. Infect Control Hosp Epidemiol. 2018;39(5):578-583.
7. Ozturk T, Weiss DS. Heteroresistance is a cause of discrepant antibiotic susceptibility testing results. Lancet Microbe. 2024;5(4):e312.
8. Alexander BD, Irish WD, Rosato AE, Eisenstein BI, Fragala MS, Goldberg SE, et al. Is pathogen molecular testing reshaping outpatient antibiotic prescribing? Am J Med Qual. 2025;40(1):21-23.
9. MacFadden DR, Daneman N. Can decision support tools improve empiric antibiotic prescribing? NEJM Evid. 2023;3(1):EVIDtt2300234.
10. Feretzakis G, Sakagianni A, Loupelis E, Kalles D, Skarmoutsou N, Martsoukou M, et al. Machine learning for antibiotic resistance prediction: A prototype using off-the-shelf techniques and entry-level data to guide empiric antimicrobial therapy. Healthc Inform Res. 2021;27(3):214-221.
11. Iskandar K, Molinier L, Hallit S, Sartelli M, Hardcastle TC, Haque M, et al. Surveillance of antimicrobial resistance in low- and middle-income countries: A scattered picture. Antimicrob Resist Infect Control. 2021;10(1):63.
12. Dunachie SJ, Day NP, Dolecek C. The challenges of estimating the human global burden of disease of antimicrobial resistant bacteria. Curr Opin Microbiol. 2020;57:95-101.
13. Khandagale P. Emerging trends in drug discovery: Harnessing artificial intelligence and machine learning for drug development. IP-Planet. 2024;12(3):56-61.
14. Hur R, Golik S, She Y. Leveraging large data, statistics, and machine learning to predict the emergence of resistant *Escherichia coli* infections. Pharmacy (Basel). 2024;12(2):53.
15. Astudillo CA, López-Cortés XA, Ocque E, Manríquez-Troncoso JM. Multi-label classification to predict antibiotic resistance from raw clinical MALDI-TOF mass spectrometry data. Sci Rep. 2024;14(1):31283.
16. 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;72(11):e848-e855.
17. Ardila CM, González-Arroyave D, Tobón S. Machine learning for predicting antimicrobial resistance in critical and high-priority pathogens: A systematic review considering antimicrobial susceptibility tests in real-world healthcare settings. PLoS One. 2025;20(2):e0319460.
18. Restrepo D, Quion J, Vásquez-Venegas C, Villanueva C, Celi LA, Nakayama LF. A scoping review of the landscape of health-related open datasets in Latin America. PLoS Digit Health. 2023;2(10):e0000368.
19. Domingues CPF, Rebelo JS, Dionisio F, Nogueira T. Multi-drug resistance in bacterial genomes: A comprehensive bioinformatic analysis. Int J Mol Sci. 2023;24(14):11438.
20. Nakahara S, Hoang BH, Mayxay M, Pattanarattanamolee R, Jayatilleke AU, Ichikawa M, et al. Development of an emergency medical system model for resource-constrained settings. Trop Med Int Health. 2019;24(10):1140-1150.
21. Syabrina R, Wang G. Predictive analysis of malaria cases in Indonesia using machine learning. JSL. 2024;9(9):4849-4859.
22. Tran C, Hargy J, Hess B, Pettengill MA. Estimated impact of low isolate numbers on the reliability of cumulative antibiogram data. Microbiol Spectr. 2023;11(1):e03939-22.
23. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One. 2019;14(11):e0224365.
24. Wu X, Gao Y, Jiao D. Multi-label classification based on random forest algorithm for non-intrusive load monitoring system. Processes. 2019;7(6):337.

25. Hinojosa Lee MC, Braet J, Springael J. Performance metrics for multilabel emotion classification: Comparing micro, macro, and weighted F1-scores. Appl Sci (Basel). 2024;14(21):9863.
26. Verbakel JY, Steyerberg EW, Uno H, De Cock B, Wynants L, Collins GS, et al. ROC curves for clinical prediction models part 1: ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. J Clin Epidemiol. 2020;126:207-216.
27. Cantarino I, Carrion MA, Goerlich F, Martinez Ibañez V. A ROC analysis-based classification method for landslide susceptibility maps. Landslides. 2019;16(3):265-282.
28. Nembrini S, König I, Wright MN. The revival of the Gini importance? Bioinformatics. 2018;34(21):3711-3718.
29. Lee ALH, To CCK, Lee ALS, Chan RCK, Wong JSH, Wong CW, et al. Deep learning model for prediction of extended-spectrum beta-lactamase (ESBL) production in community-onset Enterobacteriaceae bacteraemia from a high ESBL prevalence multi-centre cohort. Eur J Clin Microbiol Infect Dis. 2021;40(5):1049-1061.
30. Mendes RE, Critchley IA, Cotroneo N, Streit JM, Sader HS, Castanheira M, et al. Tebipenem *in vitro* activity against a collection of pathogens responsible for urinary tract infections in the US. Open Forum Infect Dis. 2021;8(Suppl 1):S620-S621.
31. Riccobene T, Ai C, Yu KC, Gregory S, Kim B, Debabov D, et al. Real-world *in vitro* activity of newer antibiotics against Enterobacterales and *Pseudomonas aeruginosa*, including carbapenem-non-susceptible and multidrug-resistant isolates: A multicenter analysis. Microbiol Spectr. 2023;11(6):e0312923.
32. Tang R, Luo R, Tang S, Song H, Chen X. Machine learning in predicting antimicrobial resistance: A systematic review and meta-analysis. Int J Antimicrob Agents. 2022;60(5-6):106684.
33. He Q, Chaillet JK, Labbé F. Antigenic strain diversity predicts different biogeographic patterns of maintenance and decline of antimalarial drug resistance. eLife. 2024;12:RP90888.
34. Theissler A, Thomas M, Burch M, Gerschner F. ConfusionVis: Comparative evaluation and selection of multi-class classifiers based on confusion matrices. Knowl Based Syst. 2022;247:108651.
35. Maddock KJ, Gefroh SJ, Burbick CR. β-Lactam resistance in veterinary β-hemolytic Streptococcus species: Are we experiencing a public health or test method crisis? J Am Vet Med Assoc. 2023;261(9):1403-1406.
36. Anahtar MN, Yang JH, Kanjilal S. Applications of machine learning to the problem of antimicrobial resistance: An emerging model for translational research. J Clin Microbiol. 2021;59(7):e0126020.
37. Yusuf E, Zeitlinger M, Meylan S. A narrative review of the intermediate category of the antimicrobial susceptibility test: Relation with dosing and possible impact on antimicrobial stewardship. J Antimicrob Chemother. 2023;78(2):338-345.
38. Hu K, Meyer F, Deng ZL, Asgari E, Kuo TH, Münch PC, et al. Assessing computational predictions of antimicrobial resistance phenotypes from microbial genomes. Brief Bioinform. 2024;25(3):bbae206.
39. Li X, Liu X, Mao J, Liu D, Jiao Z. Evaluation of population pharmacokinetic models of micafungin: Implications for dosing regimen optimization in critically ill patients. Pharmaceutics. 2024;16(9):1145.
40. Mesinovic M, Watkinson P, Zhu T. Explainable AI for clinical risk prediction: a survey of concepts, methods, and modalities. arXiv [Preprint]. 2023 [cited 2026 March 7]. Available from: https://arxiv.org/abs/2308.08407
41. Binsted LE, McNally L. Diverse relationships between antibiotic resistance and host age: a meta-analysis across antibiotic classes and bacterial genera. medRxiv [Preprint]. 2024 [cited 2026 March 7]. Available from: https://www.medrxiv.org/content/10.1101/2024.02.25.24303263v1
42. Garcia A, Delorme T, Nasr P. Patient age as a factor of antibiotic resistance in methicillin-resistant *Staphylococcus aureus*. J Med Microbiol. 2017;66(12):1782-1789.
43. Munson E, Moore J, Krueger T, Zapp A, Lavey SC, Munson KL, et al. Mycoplasmoides genitalium nucleic acid semi-quantitation and molecular macrolide resistance detection *via* automated assays: Gender and specimen source considerations. J Clin Microbiol. 2024;62(6):e0048524.
44. Nafea AM, Wang Y, Wang D, Salama AM, Aziz MA, Xu S, et al. Application of next-generation sequencing to identify different pathogens. Front Microbiol. 2024;14:1329330.
45. Kim C, Choi YH, Choi JY, Choi HJ, Park RW, Rhie SJ. Translation of machine learning-based prediction algorithms to personalised empiric antibiotic selection: A population-based cohort study. Int J Antimicrob Agents. 2023;62(5):106966.
46. MacFadden DR, Coburn B, Shah N, Robicsek A, Savage R, Elligsen M, et al. Decision-support models for empiric antibiotic selection in Gram-negative bloodstream infections. Clin Microbiol Infect. 2019;25(1):108.e1-108.e7.
47. Elligsen M, Pinto R, Leis JA, Walker SAN, Daneman N, MacFadden DR. Improving decision making in empiric antibiotic selection (IDEAS) for Gram-negative bacteremia: A prospective clinical implementation study. Clin Infect Dis. 2021;73(2):e417-e425.
48. Sullivan T, Ichikawa O, Dudley J, Li L, Aberg J. The rapid prediction of carbapenem resistance in patients with *Klebsiella pneumoniae* bacteremia using electronic medical record data. Open Forum Infect Dis. 2018;5(5):ofy091.
49. Adimadhyam S, Barreto EF, Cocoros NM, Toh S, Brown JS, Maro JC, et al. Leveraging the capabilities of the FDA's Sentinel System to improve kidney care. J Am Soc Nephrol. 2020;31(11):2506-2516.