Investigating Infection and Clinical Data to Predict Effective Antibiotic Therapy and Monitor Resistance Trends

Marco Benedetto , Giuseppe Piccinni , Cristina De Leo , Corrado Cirielli , Francesco Facchiano , Angelo Facchiano , Annarita Panebianco , Stefano Tagliaferri , Michele La Rocca , Paolo Marchetti , Roberto Tagliaferri , Antonio Facchiano

PII: S1201-9712(26)00477-7
DOI: https://doi.org/10.1016/j.ijid.2026.108842
Reference: IJID 108842

To appear in: *International Journal of Infectious Diseases*

Received date: 28 November 2025
Revised date: 17 April 2026
Accepted date: 26 May 2026

Please cite this article as: Marco Benedetto , Giuseppe Piccinni , Cristina De Leo , Corrado Cirielli , Francesco Facchiano , Angelo Facchiano , Annarita Panebianco , Stefano Tagliaferri , Michele La Rocca , Paolo Marchetti , Roberto Tagliaferri , Antonio Facchiano , Investigating Infection and Clinical Data to Predict Effective Antibiotic Therapy and Monitor Resistance Trends, *International Journal of Infectious Diseases* (2026), doi: https://doi.org/10.1016/j.ijid.2026.108842

Highlights

(Benedetto et al.)

- Delays in antibiotic selection worsen patient outcomes and promote resistance.
- Machine learning helps rapid identification of optimal antibiotic treatments, based on personal clinical history.
- XGBoost – based calculations achieved extremely relevant AUROC accurateness scores in predicting antibiotic susceptibility.
- Routine clinical and microbiological data can be used to effectively predict antibiotic resistance profiles.

Revision 2

Investigating Infection and Clinical Data to Predict Effective Antibiotic Therapy and Monitor Resistance Trends

Marco Benedetto[1,5], Giuseppe Piccinni[2], Cristina De Leo[2], Corrado Cirielli[2], Francesco Facchiano[3], Angelo Facchiano[4], Annarita Panebianco[2], Stefano Tagliaferri[5], Michele La Rocca[6], Paolo Marchetti[2], Roberto Tagliaferri[1], Antonio Facchiano[2,7]

[1] Department of Management and Innovation Systems, University of Salerno, Fisciano (SA) 84084, Italy

[2] Istituto Dermopatico dell'Immacolata, IDI-IRCCS, Rome, Italy

[3] Department of Oncology and Molecular Medicine, Istituto Superiore di Sanità, Rome, Italy

[4] Laboratory of Bioinformatics and Computational Biology, Institute of Food Science, CNR, via Roma 52 A/C, 83100 Avellino, Italy

[5] Kelyon S.r.l., via Benedetto Brin, 59 C5/C6, 80100 Naples, Italy

[6] Department of Economics and Statistics, University of Salerno, Fisciano (SA) 84084, Italy

[7] Department of Life Sciences, Health and Health Professions, Link University, 00165 Rome, Italy

Corresponding author: Antonio Facchiano

Email of the corresponding author: a.facchiano@idi.it; a.facchiano@unilink.it

Abstract word count: 249 words

Manuscript word count: 4874 words (headings included; abstract, highlights, tables, figure captions, references, and acknowledgments excluded)

Abstract

Background: Data-driven approaches to effectively select antibiotics are crucial to improving patient outcomes and reducing antibiotic resistance. This study aimed to determine whether routinely collected clinical and microbiological data can be used to train machine learning (ML) models to predict antibiotic resistance in patients with bacterial infections.

Methods: We conducted a retrospective observational study on clinical data from two separate Italian hospitals, analyzing 15,581 bacterial isolates collected from 9,966 patients between 2018 and 2024 at a multi-department hospital located in Rome and secondary clinic located in Capranica, about 70 Km away from Rome. Multiple ML models were trained, using both unbalanced and SMOTE-balanced datasets. Performance was assessed using cross-validation and independent test sets, comprising 2023-2024 isolates. Performance was measured using the Area Under the Receiver Operating Characteristic curve (AUROC), F1 score, accuracy, precision, and recall.

Results: XGBoost consistently outperformed other trained models, achieving AUROCs of 0.882 and 0.878 for Gram-positive and Gram-negative datasets, respectively. Species-specific models further improved discrimination, reaching AUROC scores up to 0.946 for *P. aeruginosa*, 0.941 for *K. pneumoniae*, 0.938 for *E. faecalis*, 0.919 for *E. coli*, 0.894 for *P. mirabilis*, and 0.891 for *S. aureus*.

Conclusions: These results demonstrate the utility of ML models in accurately predicting antibiotic susceptibility from routine clinical data, thus facilitating rapid initiation of targeted therapy. Such an approach can potentially reduce treatment delays by up to 48 hours compared to traditional diagnostic methods, presenting a useful tool to manage patients in critical conditions and combat antibiotic resistance in clinical practice.

Keywords: Antimicrobial resistance; machine learning; precision medicine; antimicrobial stewardship; antibiotic susceptibility prediction.

Glossary

ACC: Accuracy

AM: Ampicillin

AMR: Antimicrobial resistance

AMX: Amoxicillin

AN: Amikacin

AST: Antibiotic susceptibility testing

AUROC/AUC: Area under the receiver operating characteristic curve

CC: Clindamycin

CFM: Cefixime

CIP: Ciprofloxacin

CN: Cefotetan

CST: Colistin

DAP: Daptomycin

ER: Erythromycin

ESBL: Extended-spectrum beta-lactamase

ETP: Ertapenem

FA: Fusidic acid

FEP: Cefepime

FOS: Fosfomycin

FTN: Ceftriaxone

GM: Gentamicin

ICU: Intensive care unit

IMI: Imipenem

LASSO: Least absolute shrinkage and selection operator

LEV: Levofloxacin

LNZ: Linezolid

LR: Logistic regression

MEM: Meropenem

ML: Machine learning

MXF: Moxifloxacin

OXS: Oxacillin

PENG: Penicillin G

PIP: Piperacillin

PREC: Precision

RF: Random Forest

ROXH: Roxithromycin

SENS: Sensibility

SMOTE: Synthetic Minority Oversampling Technique

SPEC: Specificity

SXT: Trimethoprim/Sulfamethoxazole

TAX: Cefotaxime

TET: Tetracycline

TMP: Trimethoprim

TOB: Tobramycin

TPN: Teicoplanin

TZP: Piperacillin/Tazobactam

TZPN: Piperacillin/Tazobactam

VA: Vancomycin

WGS: Whole Genome Sequencing

WHO: World Health Organization

XGBoost: Extreme gradient boosting

Introduction

Infections represent a constant threat to global health. Particularly, bacterial infections are a major cause of morbidity and mortality worldwide. Investigating the way infections take place and develop allows us to understand better how pathogens invade the human body, how they spread, and the underlying pathogenic mechanisms, leading to the discovery of new and effective treatments, vaccines, and preventive strategies. Infections affect individuals of all ages, especially more fragile individuals such as infants, the elderly, and immune-compromised patients [1,2]. Investigating personal data and other clinical features of infected patients may therefore help develop strategies to prevent the spread of disease in communities, hospitals, and healthcare facilities [3–9].

We are currently facing increasing antibiotic resistance, with a progressive increase in strains resistant to currently available antibiotics [10,11]. Such a condition is leading to an increasingly serious emergence related to: i) increasing infection-associated mortality; ii) reducing treatment options with an increase in time and costs for the treatment and iii) the rapid global spread of resistant strains.

Therefore, developing innovative approaches to improve the clinical management of infections and to reduce the use of ineffective antibiotics is a global clinical need. In fact, one of the most critical aspects of current health management is the emergence and spread of antibiotic resistance, exacerbated by the overuse and inappropriate use of antibiotics, both in humans and in animal farming. Multidrug-resistant bacteria represent a global threat because infections easily treatable in the past may now become potentially fatal. For example, *Escherichia coli*, *Klebsiella pneumoniae*, *Staphylococcus aureus,* and *Pseudomonas aerouginosa* infections are often more difficult to treat due to antibiotic resistance. In some cases, effective treatment options are lacking. We have recently shown the use of a particular tobramycin formulation to treat chronic multidrug-resistant *P. aerouginosa* infected ulcers, otherwise incurable [12].

An additional issue is the role of early diagnosis of serious infections. Under such clinically critical conditions, it is essential to start as soon as possible an effective antibiotic treatment, to increase the chances of successful treatment and reduce the toxicity of ineffective antibiotics [13,14]. Early

diagnosis and effective therapy also help limit the spread of infections and prevent the overuse of antibiotics, as targeted and timely treatment reduces the need for empirical or prolonged drug administration. Previous research demonstrated that machine learning (ML) models can achieve moderate to high predictive performance in various clinical settings. However, previous studies typically focus on a single specimen type, specific bacterial species, or a single antibiotic, limiting their broader application [5–9]. For instance, Mintz et al. developed an ensemble of several ML algorithms, including LASSO, random forest, gradient-boosted trees, and neural networks, to predict ciprofloxacin resistance in hospitalized patients, achieving an area under the receiver operating characteristic curve (AUROC) of 0.837 when incorporating information on the infecting bacterial species [9]. Similarly, Feretzakis et al. focused on ICU-specific data and Gram stain classification to predict resistance profiles for several antibiotics, yielding an AUROC of 0.726, but their approach was restricted to a specific clinical setting [8]. Further studies focused on urine tract infections as a single specimen type, reporting AUROCs of 0.74 [7], 0.77 [5], and 0.83 [6], respectively. Integrating whole-genome sequencing (WGS) with high-throughput sequencing data is an alternative approach that has been increasingly explored in the literature. Van Camp et al. developed XGBoost-based models to predict drug resistance in five nosocomial Gram-negative organisms using WGS data as input, achieving AUROC values ranging from 0.80 for cefotaxime to 0.97 for tobramycin [15]. Other bacterium-specific studies have also reported strong performance in predicting antibiotic resistance by integrating WGS with antibiotic susceptibility test (AST) data, with AUROC values up to 0.96 for *Escherichia coli* [16] and mean accuracies ≥0.98 for *Pseudomonas* aeruginosa [17]. Despite their promising results, compared to traditional methods, WGS remains relatively expensive, and its need for substantial storage capacity and high-performance computing restricts its availability in clinical practice, thereby limiting the applicability of these approaches [18].

In contrast, the current study integrates multiple specimen type, bacterial species, and a broad panel of 34 antibiotics simultaneously, leveraging routinely collected data from 9,966 patients at IDI-IRCCS (Istituto Dermopatico dell'Immacolata - Istituto di Ricovero e Cura a Carattere Scientifico) hospital from 2018 to 2024. During 10-fold cross-validation, ML model trained on the

Gram-positive dataset achieved an AUROC of 0.882, while the model trained on the Gram-negative dataset achieved 0.878. For models trained on a subset of isolates positive to a known bacterial species, AUROC scores of 0.919 for *E. coli*, 0.938 for *E. faecalis*, 0.941 for *K. pneumoniae*, 0.946 for *P. aeruginosa*, 0.894 for *P. mirabilis*, and 0.891 for *S. aureus* were achieved.

Methods

Patient enrollment

Data for this clinical study were collected over seven years (2018-2024) from two sites owned by the Fondazione Luigi Maria Monti, i.e. from IDI-IRCCS located in Rome and the hospital located in Capranica, both in the Lazio region of Italy. The study was approved by the IDI-IRCCS Ethic Committee (CE IDI, n. 695.1, June 7th, 2022). Patients were enrolled retrospectively according to the protocol approved by the Ethics Committee, i.e. if they were ≥ 18 years old and if they were positive for any infection test at the IDI-IRCCS laboratory in the 2018–2024 time span. The dataset consists of 15,581 isolates collected from 9,966 inpatient and outpatient individuals. It is worthwhile to notice that during the COVID-19 emergency in Italy (2020-2021), routine microbiological testing was reduced due to the reallocation of healthcare resources, the suspension of outpatient services, and reduced inpatient admission for non-COVID conditions.

Bacterial susceptibility data

In the current study, bacterial cultures demonstrating susceptible-increased exposure or resistant results were grouped as non-susceptible to address the class imbalance in the dataset, given that the number of susceptible results exceeded the other outcomes. To further handle the class imbalance and create a more balanced training set, we employed the synthetic minority oversampling technique (SMOTE), an oversampling technique that creates new synthetic samples of the minority class.

The dataset contained the following variables of interest: age (numerical), sex (binary), specimen type (blood, urine, sputum, swab, microbiology, miscellaneous; categorical), unit type (medical or surgical; binary), admission type (inpatient or outpatient care; binary), Gram stain (positive or negative; binary), bacterial species (categorical), 43 antibiotics (categorical), and their corresponding antibiotic susceptibility results (susceptible or non-susceptible; categorical). Additional features related to previous infections and antibiotic resistance were engineered from the dataset. After data cleaning, the number of antibiotics used for model training was reduced to 34.

Identification of microorganisms and antibiotic susceptibility test

To identify microorganisms in the human samples under investigation, and to test the susceptibility to antibiotics Antimicrobial Susceptibility Testing (AST), the Becton-Dickinson technology BD-Phoenix was used, according to the diagnostics standards, at the Diagnostics Laboratory at IDI-IRCCS in Rome. The BD-Phoenix 50 instrument was used along with the Epicenter software, and the diagnostics procedures were followed, according to the user manual, as briefly reported below, and following the EUCAST recommendations.

Phoenix technology uses a range of chromogenic and fluorogenic tests to identify the organism via both growth and enzymatic reactions. The tests are based on the microbial degradation of specific substrates detected by various indicator systems, such as phenol red indicator or other chromogenic or fluorogenic producing a fluorescent coumarin derivative or a resazurin-based indicator. In addition, other tests detect an organism's ability to hydrolyze, degrade, reduce, or otherwise utilize a substrate.

Testing for susceptibility to antimicrobial agents was performed by determining bacterial growth in Phoenix microwell panels containing various concentrations of the antibiotic and a growth indicator. Incubation and reading of these panels were done by a continuous process in the instrument. This technique was used according to diagnostic standards in AST predetermined panels, to measure the susceptibility or resistance of any given microorganism to various antibiotics or antimicrobial agents. For antibiogram and antibiotic resistance measures, Gram+ bacteria were tested with 23

antibiotics while Gram- were tested with 20 antibiotics, using 1:2 serial dilutions. The AST Phoenix method is a broth microdilution test, using an oxidation-reduction indicator to detect the growth of organisms in the presence of any given antibiotic [19]. Continuous measurements of changes in both indicator and bacterial turbidity were used to determine bacterial growth. All diagnostic tests carried out were done according to the international EUCAST guidelines annually updated and certified.

Feature Engineering and Data Preprocessing

The initial dataset contained multiple columns corresponding to susceptibility test results for 43 different antibiotics, with each antibiotic represented as a separate column indicating whether the isolate was susceptible (S), susceptible-increased exposure (I), or resistant (R). After data cleaning, we included in the dataset only those antibiotics included in at least 800 susceptibility test results from more than 600 distinct patients. As a result, the total number of antibiotics considered for further analysis decreased from 43 to 34. We transformed the dataset from a wide to a long format to streamline the modeling process and enable the development of a predictive model with a single output. This transformation involved two steps:

1. The multiple antibiotic-specific columns were combined into a single column, representing the susceptibility outcome (susceptible (S) or non-susceptible (NS), which includes susceptible with increased exposure (I) and resistant (R))
2. A new categorical column was introduced to indicate the name of the antibiotic corresponding to each susceptibility result.

For each patient, new binary features were added to capture historical non-susceptibility identified in previous antibiotic susceptibility test results. Accordingly, prior non-susceptibility to any antibiotic included in this study was defined as the presence of at least one historical test result for that antibiotic classified as resistant (R) or susceptible with increased exposure (I). This feature reflects a previous infection caused by a bacterial strain exhibiting a reduced susceptibility phenotype. This

restructuring allowed us to handle antibiotics as a categorical feature within the model and to predict susceptibility outcomes across different antibiotics using a unified modeling framework. By converting the dataset into a long format, we increased the effective sample size for each subset and enabled the model to learn patterns applicable to various antibiotics, potentially enhancing predictive performance.

During preprocessing, categorical variables were encoded using one-hot encoding, whereas numerical variables, such as age, were standardized.

Statistical Analysis and Machine Learning Algorithms

A χ^2 test was used to assess the association between bacterial distribution and the type of admission. For pairwise comparisons of proportions, two-proportion z-tests were conducted for the most detected bacteria. A Bonferroni correction was applied to account for multiple comparisons. The significance level was set at $P<.05$.

We performed a comparative analysis of several ML algorithms: logistic regression [20], random forest [22], eXtreme Gradient Boosting (XGBoost) [23], and TabNet [24]. Data collected between 2018 and 2022 (11,213 isolates collected from 7,395 patients) were used for model training and hyperparameter fine-tuning, employing a 10-fold cross-validation to minimize overfitting. To address class imbalance and evaluate its effect on model performance, training was performed both with and without SMOTE. Since Gram staining results are available largely before full bacterial identification, models were trained using both partial (i.e., only Gram staining) and complete information (i.e., Gram staining plus full bacterial identification).

To minimize temporal bias, the models' performance in predicting antibiotic susceptibility results (susceptible or non-susceptible) was assessed on the data collected from 2023-2024, ensuring evaluation that approximate real-world clinical deployment conditions. To mitigate patient-level data leakage, isolates from 648 patients included in the training set were later excluded from the test set, thereby generating an additional test set. While this prevents inflated performance estimates, it

may also result in underestimation, since this test set contains only patients without prior infection or resistance history.

Models were trained and tested on subsets of Gram-positive, Gram-negative, *E. coli*, *E. faecalis*, *K. pneumoniae*, *P. aeruginosa*, *P. mirabilis*, and *S. aureus* infected isolates. The pathogens were selected based on their prevalence within the IDI-IRCCS institution. Performance statistical metrics included the area under the receiver operating characteristic (ROC) curve (AUROC), F1 score, accuracy, precision, and recall. As part of our analysis, we evaluated feature importance by computing the normalized total reduction in the impurity criterion contributed by each feature, known as the Gini importance [25]. To further improve model interpretability beyond aggregate feature importance rankings, we applied SHapley Additive exPlanations (SHAP) to the XGBoost model trained on E.coli isolates. SHAP values provide both global and local explanations of model predictions, quantifying the contribution of each feature to single outcomes. Local SHAP plots were generated for a representative set of patient profiles to illustrate how clinical features interact in specific prediction scenarios.

Analyses were performed with R version 4.5.1 and Python version 3.12.3.

Results

Table 1 presents the summary statistics of clinical and microbiological data for the study population, which included 15,581 isolates from 9,966 patients with a mean age of 61.9 ± 21.4 years. The majority of the population was female (59.5%), with most patients receiving care in medical units (98.0%) and primarily as outpatients (78.9%). The most frequently analyzed specimens were urine (41.7%), followed by miscellaneous biological materials (23.9%), and swabs from ulcers (21.9%). Gram-negative bacteria were predominant, accounting for 63.3% of isolates, with *E. coli* being the most commonly identified species (28.6%). In the training set, 73.1% of isolates were susceptible to the tested antibiotics, whereas 70.0% of isolates in the test set were susceptible.

The distribution of bacterial isolates differed significantly between inpatient and outpatient settings over the seven-year study period, as shown in Figure 1. In inpatient care (Figure 1a), *Staphylococcus* was the most frequently identified genus, accounting for 30.8% of isolates, followed by *Escherichia* (20.5%) and *Pseudomonas* (17.5%). In contrast, outpatient settings (Figure 1b) were dominated by *Escherichia* (28.4%), followed by *Staphylococcus* (24.2%) and *Pseudomonas* (9.2%). Statistical analysis using a χ^2 test confirmed a significant association between bacterial distribution and type of care (P<.05). A two-proportion z-test indicated that *Staphylococcus* was significantly more prevalent in inpatients than outpatients (P<.05), while *Escherichia* was more frequent in outpatients compared to inpatients (P<.05). For *Pseudomonas* and *Proteus*, the proportion was significantly higher in inpatients than outpatients (P< .05). However, the prevalence of *Klebsiella* and *Enterococcus* showed no significant difference between inpatients (7.7% Klebsiella-positive isolates, 6.4% Enterococcus-positive isolates) and outpatients (8.7% Klebsiella-positive isolates, 5.7% Enterococcus-positive isolates; P>.05).

Using prevalence data in the different hospital units led us to the observation of microbial "fingerprints" that may discriminate the different units within a given hospital. Figure 2 illustrates bacterial distributions across different hospital units. In the dermatology unit (Figure 2a), *Staphylococcus* was the most frequently identified genus (38.6%), followed by *Escherichia* (22.2%) and *Pseudomonas* (17.4%). The oncology unit (Figure 2b) had *Escherichia* as the dominant genus (27.0%), while *Staphylococcus* and *Klebsiella* were present at 18.5% and 16.7%, respectively. In the surgery unit (Figure 2c), *Staphylococcus* accounted for 23.9% of isolates, with *Pseudomonas* (20.4%) and *Enterococcus* (15.0%) also frequently detected. In the IDI clinic (Figure 2d), *Escherichia* was most prevalent over the five years (30.2%), followed by *Staphylococcus* (25.1%) and *Pseudomonas* (9.6%).

Figure 3 presents the distribution of the three mostly present bacterial species across different hospital departments. The dermatology unit consistently reported a high prevalence of *S. aureus*, peaking at 46.9% in 2024, with the highest absolute number of cases occurring in 2019 (152 cases, 35.6%). *E. coli* remained the dominant pathogen in the outpatient IDI clinic, maintaining a

stable presence at around 30% across all years. In the oncology unit, *E. coli* reached its highest proportion in 2020 (40.5%), while the operating theatre exhibited fluctuating trends, with *P. aeruginosa* prevalence peaking in 2018 (35.5%) and 2023 (40.0%).

Heatmaps in Figure 4 depict the increasing antibiotic resistance of Gram-positive and Gram-negative isolates over time. Among Gram-positive bacteria (Figure 4a), the most resistant antibiotics included penicillin G, ampicillin, ciprofloxacin, erythromycin, with ciprofloxacin resistance exceeding 75% after 2021. For Gram-negative bacteria (Figure 4b), ampicillin, roxithromycin, ciprofloxacin, amoxicillin, and cefepime displayed the highest resistance rates. Since mid-2021, roxithromycin resistance in Gram-negative isolates surpassed 60%.

Figure 5 highlights resistance rates in different care settings: the rising resistance of *K. pneumoniae* to cefepime and amoxicillin in inpatient care, with near-complete resistance to ampicillin observed in both inpatient and outpatient settings throughout the study period. *S. aureus* demonstrated high resistance to gentamicin, ciprofloxacin, erythromycin and penicillin G, while *P. aeruginosa* showed increasing resistance to ciprofloxacin and cefepime reaching a peak in 2022 among inpatients and outpatients.

Predicting antibiotic susceptibility

Table 2 summarizes the performance of machine learning models used to predict antibiotic susceptibility, while the corresponding ROC curves are shown in Figure 6. Model performance was evaluated on both unbalanced and SMOTE-balanced datasets using 10-fold cross-validation and independent test sets. Among all models, XGBoost demonstrated the highest discriminative ability. On the unbalanced Gram-positive dataset, it achieved an AUROC of 0.882, an F1-score of 0.827, and an accuracy of 0.834. The Gram-negative dataset exhibited a similarly strong results, with an AUROC of 0.878, an F1-score of 0.834, and an accuracy of 0.841. Although XGBoost generally outperformed other algorithms, random forest surpassed it in AUROC when trained on SMOTE-balanced datasets and evaluated on the 2023-2024 datasets, achieving a 0.857 and 0.782 for the

Gram-positive and Gram-negatives, respectively. Species-specific results further highlighted XGBoost's superior performance, with AUROC scores of 0.919 for *E. coli*, 0.938 for *E. faecalis*, 0.941 for *K. pneumoniae*, 0.946 for *P. aeruginosa*, 0.894 for *P. mirabilis*, and 0.891 for *S. aureus*. Logistic regression, random forest, and TabNet performed worse than XGBoost during cross-validation; however, random forest surpassed XGBoost when trained on SMOTE-balanced datasets and evaluated on the independent test sets.

Figure 7 presents feature importance rankings in XGBoost models. Among Gram-positive isolates (Figure 7a), moxifloxacin and roxithromycin had the highest importance, followed by specimen type, ertapenem, and vancomycin. For Gram-negative isolates (Figure 7b), colistin, ceftriaxone, roxythromycin were strong predictors. Historical non-susceptibility patterns ranked among the top features for XGBoost trained on both partial (i.e., only Gram staining) and complete information (i.e., bacterial species).

To illustrate case-level interpretability, we focused on the E. coli model, as it represents the most prevalent pathogen in our dataset. Representative SHAP waterfall plots (Figure 8) highlight distinct decision patterns across cases with higher versus lower predicted probability of non-susceptibility. Panel A of Figure 8 shows the SHAP waterfall plot for a female inpatient aged 61 with an extensive documented prior resistance history, including previous non-susceptibility to gentamicin, carbapenems, and ciprofloxacin. Despite the tested antibiotic (amoxicillin) exerting a protective contribution (SHAP = −0.26), the cumulative weight of historical resistance patterns drove the model toward a non-susceptibility prediction, demonstrating its ability to integrate prior microbiological data in a manner consistent with clinical reasoning. Similarly, panel B of Figure 8 illustrates how prior non-susceptibility to ciprofloxacin and ampicillin, combined with demographic risk factors, contributes to a non-susceptibility prediction in an 85-year-old male inpatient. Conversely, panel C of Figure 8 shows that, in a female inpatient aged 72 with no prior clinical history, the model predicted non-susceptibility based solely on contextual features, including age, admission type, and a winter sampling month, demonstrating its applicability even in newly admitted patients with unknown resistance history. Overall, these examples indicate that the model

integrates clinically meaningful predictors in a transparent and clinically plausible manner across diverse scenarios.

Discussions

Investigating responses to antibiotics and bacterial resistance is essential for improving global health management, given the continuous emergence of new antibiotic resistances [26]. The findings of the present study may contribute to optimizing antibiotic selection, improving patient outcomes, and reducing ineffective prescriptions. In fact, the ability to quickly and accurately diagnose infections and determine effective treatments is crucial for limiting the spread of resistant strains and minimizing the risks associated with inappropriate antibiotic use [27].

The proposed predictive models demonstrated strong discriminative performance, with XGBoost achieving AUROC values of 0.882 and 0.878 for Gram-positive and Gram-negative bacteria, respectively, when the Gram-positivity or negativity was the only bacterium-type information available. Performances further improved when XGBoost was trained on bacterium-specific datasets (Table 2). These findings suggest that machine learning tools can be successfully applied in clinical settings to improve decision-making in antibiotic therapy using data routinely collected during clinical practice.

One of the most pressing issues in infection management is the delay in obtaining susceptibility test results, often forcing clinicians to rely on broad-spectrum antibiotics. The approach tested in the present study responds to the need for faster bacterial identification and characterization methods [28] and its integration could significantly reduce the time required to initiate effective treatment, thereby improving patient outcomes.

The economic burden of antimicrobial resistance (AMR) is substantial. Nelson et al. (2021) estimate that multidrug-resistant infections cost the US healthcare system $1.9 billion annually and contribute to over 11,000 deaths among elderly patients [29]. Furthermore, the World Health Organization (WHO) reports that the European region experiences over 670,000 AMR-related

infections annually, leading to approximately 33,000 fatalities [30-31]. The predictive models presented in this study have the potential to optimize resource utilization and reduce hospital admissions and patient stays, ultimately leading to significant cost savings for healthcare systems.

Interestingly, Figures 1-3 revealed distinct bacterial profiles across different units within the same hospital. It is reasonable to expect that such patterns would likely differ across diverse geographical contexts as well as across differently specialized hospitals. In fact, the infections-patterns may differ according to the clinical and geographical characteristics of the patient population, underscoring the importance of applying locally tailored analytical and predictive approaches. By capturing these unit-specific patterns, the proposed approach concretely aligns with precision and evidence-based medicine principles, supporting data-driven clinical decision-making adapted to specific local epidemiological conditions.

Figure 6 and Table 2 confirm that the XGBoost model consistently outperformed other classifiers and that prior non-susceptibilities to antibiotics are key predictors. Reliance on prior non-susceptibilities reflects an important system feature: it mirrors the clinical reasoning process by which physicians integrate prior medical history into therapeutic decisions. When such information is available, incorporating it into predictive models strongly enhances its clinical value. We acknowledge that applicability may be reduced in scenarios involving newly admitted patients with no medical history; however, the proposed framework was designed to support “what-if” simulations in cases of incomplete records, allowing clinicians to explore alternative scenarios (e.g., assuming prior non-susceptibility to cephalosporins) and assess their potential impact on treatment decisions. SHAP-based interpretability analysis (Figure 8) confirmed that the XGBoost model's predictions are clinically plausible across diverse patient profiles. At the individual level, prior non-susceptibility to antibiotics emerged as the dominant driver of resistance predictions in patients with documented infection history, consistent with the known co-resistance mechanisms associated with extended-spectrum beta-lactamase (ESBL) activity and multidrug-resistant *E. coli* strains. In patients without prior microbiological data, the model appropriately relied on contextual variable. These findings support the clinical translational potential of the proposed framework and

address concerns regarding the interpretability of black-box machine learning models in clinical settings.

This dual utility underscores the potential of the model both as a decision-support tool in routine practice and as a means of guiding therapy when historical data are partial or absent. The observed drop in performance on the independent test sets may be explained by XGBoost's reliance on prior non-susceptibility results. Indeed, as expected, models achieved higher performances when tested on patients whose previous microbiological history was available compared with patients lacking previous history. To prevent patient-level data leakage and overestimated performance, isolates from patients admitted between 2018 and 2022 were excluded in the test set. As a result, the additional test set comprised only patients without previous clinical histories, potentially leading to an underestimation of the model's real-world predictive performance. The introduction of the SMOTE oversampling method to balance the training sets did not lead to substantial improvements in model performance during cross-validation. Notably, random forest models trained on SMOTE-balanced data surpassed XGBoost when evaluated on the independent test sets, suggesting that oversampling may enhance model generalization in certain conditions, where no previous clinical history is known.

The study builds upon previous machine learning research, which demonstrated that predictive algorithms could accurately determine antibiotic susceptibility. However, these earlier efforts were narrowly focused on specific infections, clinical setting, bacterial species, or antibiotics, reducing their generalizability [5–9]. In contrast, the present study proposes models that extend beyond those narrow scopes by training on bacterial-positive isolates from multiple specimen types and clinical contexts, while simultaneously incorporating antibiotic susceptibility data for 34 different antibiotics. This integrative approach not only strengthens predictive performance but also improves its applicability in real-world clinical settings. The broad coverage across bacterial species, specimen types, and antibiotics overcomes the limitations of earlier narrow-focused studies, establishing the proposed method as a more generalizable and clinically relevant tool for guiding antibiotic therapy.

Limitations

The primary limitations of this study are the dataset specificity and generalizability. Since the models were trained on data from IDI-IRCCS clinics in a single Italian region, they may reflect local infection patterns rather than broader global trends. This geographic limitation could affect the models' applicability to healthcare settings with different bacterial distributions and antibiotic resistance profiles. Additionally, the dataset is largely derived from dermatology-focused units, which may limit its relevance to other medical specialties, such as cardiology or gastroenterology, where infection profiles and resistance dynamics could differ significantly. Nonetheless, this should be viewed less as a limitation and more as a feature of the system. Complex biological phenomena, such as bacterial infections, are naturally influenced by patient characteristics, geographical factors, and disease-specific contexts. This reflects the context-dependent nature of infectious disease dynamics rather than a shortcoming of the proposed modelling approach.

Another limitation is the study's timeframe, which spans pre-pandemic, pandemic, and post-pandemic periods. The COVID-19 pandemic led to disruptions in routine microbiological testing and a reallocation of healthcare resources, skewing the dataset toward more severe infections and potentially introducing biases in antibiotic resistance trends. Such limitations can be overcome by periodic re-training for each hospital based on the most recent four or five years of microbiological history.

Novelty

The novelty of the present study lies primarily in the methodological and evaluation framework, rather than in the use of a single modeling technique. Compared with previous studies, we propose:

i) an integrated modeling approach that jointly considers multiple specimen types and a broader panel of antibiotics; ii) a two-stage prediction framework that explicitly accounts for the temporal availability of bacterial species identification, reflecting real diagnostic workflows; iii) robust strategies to mitigate temporal bias and patient-level data leakage, which are frequently overlooked in prior studies. These design choices are intended to ensure a fair and clinically realistic evaluation of model performance, Specific attention was paid in approximating real-world deployment conditions rather than simple retrospective assessments.

A recent study by Valavarasu S. et al. [32] (published simultaneously to the submission of our study), reports a multiple machine learning models trained on the ATLAS dataset, achieving high AUC values with XGBoost. While the reported XGBoost performance is comparable or slightly higher than that obtained in our study, a key difference lies in the evaluation methodology. In the work of Valavarasu S. et al, model performance was assessed using only a random 80:20 split, without evaluating the model's ability to generalize across subsequent years.

In contrast to this common approach in the machine learning literature, we explicitly evaluate temporal generalization by testing model performance on data from later years and by constructing a test set composed exclusively of patients with no prior clinical history. This approach represents a deliberately challenging and clinically relevant "worst-case" scenario, which we believe provides more realistic insight into the performance expected at deployment time.

An additional novelty of our study compared to Valavarasu et al. consists of the type and number of features used to perform the antibiogram prediction. In our study the patient-related features used in the training set are more numerous, pointing toward a more personalized medicine-oriented approach.

For these reasons, we believe our study offers meaningful novelty by emphasizing: robust evaluation, performance assessment over time, prevention of data leakage (i.e. critical aspects for the translation of predictive models of antibiotic resistance into clinical practice), real-life data acquisition, use of patients-related features offering a more personalized approach analysis.

Future work will address these limitations by expanding the dataset to include multiple centers in different countries, or multiple hospitals from the same location, incorporating data from diverse clinical units and settings. Such efforts will be crucial to enhancing the robustness and clinical utility of ML-driven antibiotic resistance prediction tools. We are aware that training the system with data obtained from multiple centers is necessary to give general clinical applicability to the current findings.

Conclusions

This study demonstrates the high effectiveness of ML models in helping physicians predict antibiotic susceptibility, significantly reducing the time required to determine the most appropriate treatment. The timely selection of the right antibiotic can improve patient outcomes, shorten hospital stays, and ultimately reduce healthcare costs. While the predictions are statistically based, the approach is personalized for each patient, incorporating historical infection data, while the physician's role remains crucial to integrate data-driven predictions with real-life experience. Although this requirement for additional clinical data may seem a limitation, it enables personalized medicine and capitalizes on hospital data digitization, which is increasingly available in modern healthcare facilities.

Future research should focus on validating these models in multi-center studies and integrating genetic resistance markers to enhance accuracy, particularly for complex resistance mechanisms such as carbapenemase activity. By combining AI-driven predictive analytics with real-time microbiology, healthcare providers can advance precision antimicrobial therapy, optimize infection control strategies, and ultimately curb the growing threat of antimicrobial resistance.

Declarations

Ethics approval and consent to participate

The study was approved by the IDI-IRCCS Ethic Committee (CE IDI, n. 695.1, June 7th, 2022).

The study is carried out in accordance with the Declaration of Helsinki.

Consent to participate is collected in accordance to the Authorization no. 9/2016 from the Italian Data Protection Authority.

Consent for publication

Not applicable.

Availability of data and materials

The data included in the study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that there is no conflict of interest.

Funding

Antonio Facchiano reports financial support was provided by the Government of Italy Ministry of Health (RC2024;3.3). Other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contributions

Marco Benedetto: Writing – original draft, Software, Conceptualization, Formal analysis, Data Curation, Visualization. **Giuseppe Piccinni**: Data Curation, Conceptualization. **Cristina De Leo**: Data Curation. **Corrado Cirielli**: Data Curation. **Francesco Facchiano**: Writing – review and editing, Conceptualization. **Angelo Facchiano**: Conceptualization. **Annarita Panebianco**: Conceptualization. **Stefano Tagliaferri**: Conceptualization. **Michele La Rocca**: Methodology, Conceptualization. **Paolo Marchetti**: Conceptualization. **Roberto Tagliaferri**: Writing – review and

editing, Supervision, Conceptualization, Methodology. **Antonio Facchiano**: Writing – review and editing, Supervision, Conceptualization, Formal analysis, Visualization.

Acknowledgments

The authors would like to thank Luciano Sobrino and the Information System Department of IDI-IRCCS for their continuous support.

M.B., G.P., and A.F. had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

References

[1] Feng Q, Zhou J, Zhang L, Fu Y, Yang L. Insights into the molecular basis of c-di-GMP signalling in Pseudomonas aeruginosa. *Crit Rev Microbiol*. 2024;50(1):20-38. doi:10.1080/1040841X.2022.2154140

[2] d'Enfert C. Hidden killers: persistence of opportunistic fungal pathogens in the human host. *Curr Opin Microbiol*. 2009;12(4):358-364. doi:10.1016/j.mib.2009.05.008

[3] Abukhalil A, Barakat S, Mansour A, Al-Shami N, Naseef H. ESKAPE Pathogens: Antimicrobial Resistance Patterns, Risk Factors, and Outcomes a Retrospective Cross-Sectional Study of Hospitalized Patients in Palestine. *Infect Drug Resist*. 2024;Volume 17:3813-3823. doi:10.2147/IDR.S471645

[4] Bereanu AS, Bereanu R, Mohor C, Vintilă BI, Codru IR, Olteanu C et al. Prevalence of Infections and Antimicrobial Resistance of ESKAPE Group Bacteria Isolated from Patients Admitted to the Intensive Care Unit of a County Emergency Hospital in Romania. *Antibiotics*. 2024;13(5):400. doi:10.3390/antibiotics13050400

[5] Tzelves L, Lazarou L, Feretzakis G, Kalles D, Mourmouris P, Loupelis E et al. Using machine learning techniques to predict antimicrobial resistance in stone disease patients. *World J Urol*. 2022;40(7):1731-1736. doi:10.1007/s00345-022-04043-x

[6] Yelin I, Snitser O, Novich G, Kats R, Tal O, Parizade M et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. *Nat Med*. 2019;25(7):1143-1152. doi:10.1038/s41591-019-0503-6

[7] Corbin CK, Sung L, Chattopadhyay A, Noshad M, Chang A, Deresinski S et al. Personalized antibiograms for machine learning driven antibiotic selection. *Communications Medicine*. 2022;2(1):38. doi:10.1038/s43856-022-00094-8

[8] Feretzakis G, Loupelis E, Sakagianni A, Kalles D, Martsoukou M, Lada M et al. Using Machine Learning Techniques to Aid Empirical Antibiotic Therapy Decisions in the Intensive Care Unit of a General Hospital in Greece. *Antibiotics*. 2020;9(2):50. doi:10.3390/antibiotics9020050

[9] Mintz I, Chowers M, Obolski U. Prediction of ciprofloxacin resistance in hospitalized patients using machine learning. *Communications Medicine*. 2023;3(1):43. doi:10.1038/s43856-023-00275-z

[10] Miller WR, Arias CA. ESKAPE pathogens: antimicrobial resistance, epidemiology, clinical impact and therapeutics. *Nat Rev Microbiol*. 2024;22(10):598-616. doi:10.1038/s41579-024-01054-w

[11] Luo Q, Lu P, Chen Y, Shen P, Zheng B, Ji J et al. ESKAPE in China: epidemiology and characteristics of antibiotic resistance. *Emerg Microbes Infect*. 2024;13(1). doi:10.1080/22221751.2024.2317915

[12] Cirielli C, Facchiano F, Picardo M, Benedetto M, Tagliaferri R, Tagliaferri S et al. Topical use of tobramycin: A possible innovative treatment of severe skin chronic ulcers. *Exp Dermatol*. 2024;33(7). doi:10.1111/exd.15131

[13] Plata-Menchaca EP, Ruiz-Rodríguez JC, Ferrer R. Early Diagnosis of Sepsis: The Role of Biomarkers and Rapid Microbiological Tests. *Semin Respir Crit Care Med*. 2024;45(04):479-490. doi:10.1055/s-0044-1787270

[14] Wei W, Zhao Y, Zhang Y, Shou S, Jin H. The early diagnosis and pathogenic mechanisms of sepsis-related acute kidney injury. *Open Life Sci*. 2023;18(1). doi:10.1515/biol-2022-0700

[15] Van Camp PJ, Haslam DB, Porollo A. Prediction of Antimicrobial Resistance in Gram-Negative Bacteria From Whole-Genome Sequencing Data. *Front Microbiol*. 2020;11. doi:10.3389/fmicb.2020.01013

[16] Ren Y, Chakraborty T, Doijad S, Falgenhauer L, Falgenhauer J, Goesmann A et al. Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning. *Bioinformatics*. 2022;38(2):325-334. doi:10.1093/bioinformatics/btab681

[17] Noman SM, Zeeshan M, Arshad J, Amentie MD, Shafiq M, Yuan Y et al. Machine Learning Techniques for Antimicrobial Resistance Prediction of Pseudomonas Aeruginosa from Whole Genome Sequence Data. *Comput Intell Neurosci*. 2023;2023(1). doi:10.1155/2023/5236168

[18] Bagger FO, Borgwardt L, Jespersen AS, Hansen AR, Bertelsen B, Kodoma M et al. Whole genome sequencing in clinical practice. *BMC Med Genomics*. 2024;17(1):39. doi:10.1186/s12920-024-01795-w

[19] Lancaster MW, Fields RD. *Antibiotic and Cytotoxic Drug Susceptibility Assays Using Resazurin and Poising Agents*. Vol 485.; 1994.

[20] Cox DR. The Regression Analysis of Binary Sequences. *J R Stat Soc Series B Stat Methodol*. 1958;20(2):215-232. doi:10.1111/j.2517-6161.1958.tb00292.x

[21] Krzywinski M, Altman N. Classification and regression trees. *Nat Methods*. 2017;14(8):757-758. doi:10.1038/nmeth.4370

[22] Tin Kam Ho. Random decision forests. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. IEEE Comput. Soc. Press; 278-282. doi:10.1109/ICDAR.1995.598994

[23] Chen T, Guestrin C. XGBoost. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM; 2016:785-794. doi:10.1145/2939672.2939785

[24] Arik SÖ, Pfister T. TabNet: Attentive Interpretable Tabular Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021;35(8):6679-6687. doi:10.1609/aaai.v35i8.16826

[25] Dunn J, Mingardi L, Zhuo YD. Comparing interpretability and explainability for feature selection. Published online May 11, 2021.

[26] Van Goethem MW, Marasco R, Hong P, Daffonchio D. The antibiotic crisis: On the search for novel antibiotics and resistance mechanisms. *Microb Biotechnol*. 2024;17(3). doi:10.1111/1751-7915.14430

[27] Monk EJM, Jones TPW, Bongomin F, Kibone W, Nsubuga Y, Ssewante N et al. Antimicrobial resistance in bacterial wound, skin, soft tissue and surgical site infections in Central, Eastern, Southern and Western Africa: A systematic review and meta-analysis. *PLOS Global Public Health*. 2024;4(4):e0003077. doi:10.1371/journal.pgph.0003077

[28] Szabó S, Feier B, Capatina D, Tertis M, Cristea C, Popa A. An Overview of Healthcare Associated Infections and Their Detection Methods Caused by Pathogen Bacteria in Romania and Europe. *J Clin Med*. 2022;11(11):3204. doi:10.3390/jcm11113204

[29] Nelson RE, Hyun D, Jezek A, Samore MH. Mortality, Length of Stay, and Healthcare Costs Associated With Multidrug-Resistant Bacterial Infections Among Elderly Hospitalized Patients in the United States. *Clin* Infect Dis. 2022;74(6):1070-1080. doi:10.1093/cid/ciab696

[30] Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. Lancet Infect Dis. 2019;19(1):56-66. doi:10.1016/S1473-3099(18)30605-4

[31] Mills JP, Marchaim D. Multidrug-Resistant Gram-Negative Bacteria. Infect Dis Clin North Am. 2021;35(4):969-994. doi:10.1016/j.idc.2021.08.001

[32] Valavarasu S, Sangu Y, Mahapatra T. Prediction of antibiotic resistance from antibiotic susceptibility testing results from surveillance data using machine learning. Sci Rep. 2025 Aug 20;15(1):30509. doi: 10.1038/s41598-025-14078-w.

Figures legend

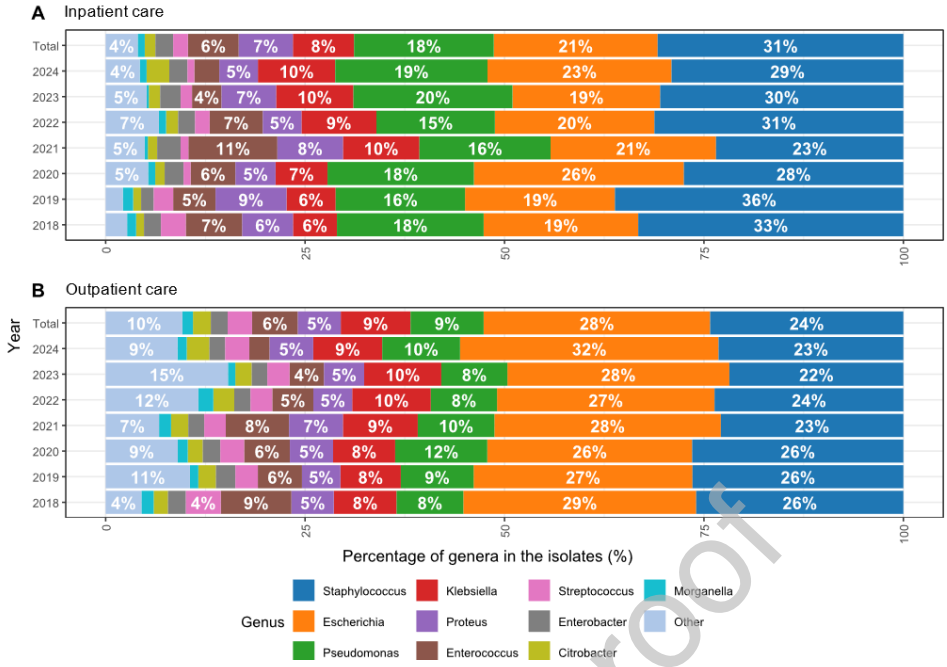


Figure 1 Distribution of bacterial genera in different care settings. Temporal distribution of bacterial genera isolated over the seven years (2018-2024) in inpatient (A) and outpatient (B) care settings. Outpatient care setting includes IDI clinic, Dermatology Day Hospital, Day Hospital 5th unit, Dermatology Day Surgery, and Villa Paola clinic.

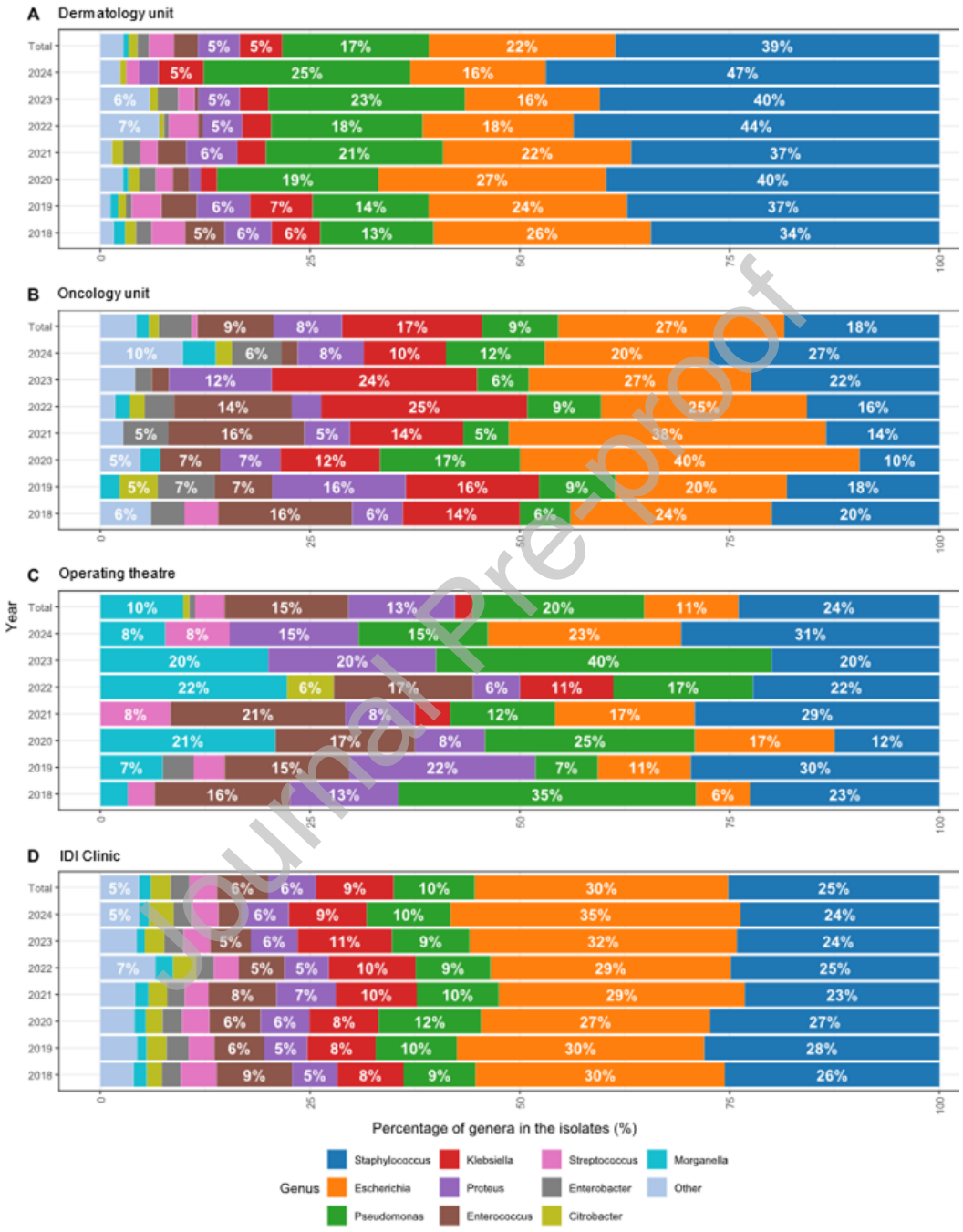


Figure 2. **Distribution of bacterial genera isolated from selected units.** Temporal distributions of bacterial genera isolated over the seven years (2018-2024), for the dermatology (a) and oncology (b) units, the operating theatre (c), and IDI clinic (d). Panels display temporal trends in isolation frequency within the operating theatre (n = 142), oncology unit (n = 330), IDI clinic (n = 11,368), and dermatology unit (n = 1,610).

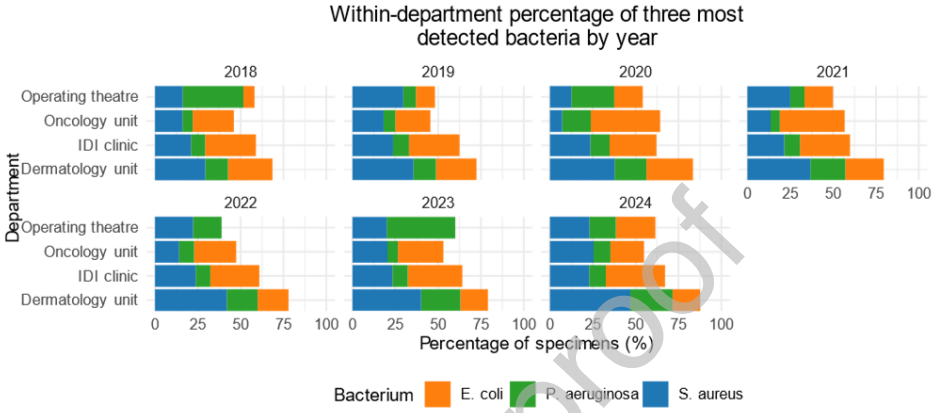


Figure 3. **Comparison of the three most frequently isolated bacterial species across selected units**. Comparison of distributions of the three most detected bacteria (i.e., *S. aureus*, *E. coli*, and *P. aeruginosa*) across four different units from 2018 to 2022. Panels display temporal trends in isolation frequency for each of the three bacterial species within the operating theatre (n = 142), oncology unit (n = 330), IDI clinic (n = 11,368), and dermatology unit (n = 1,610).

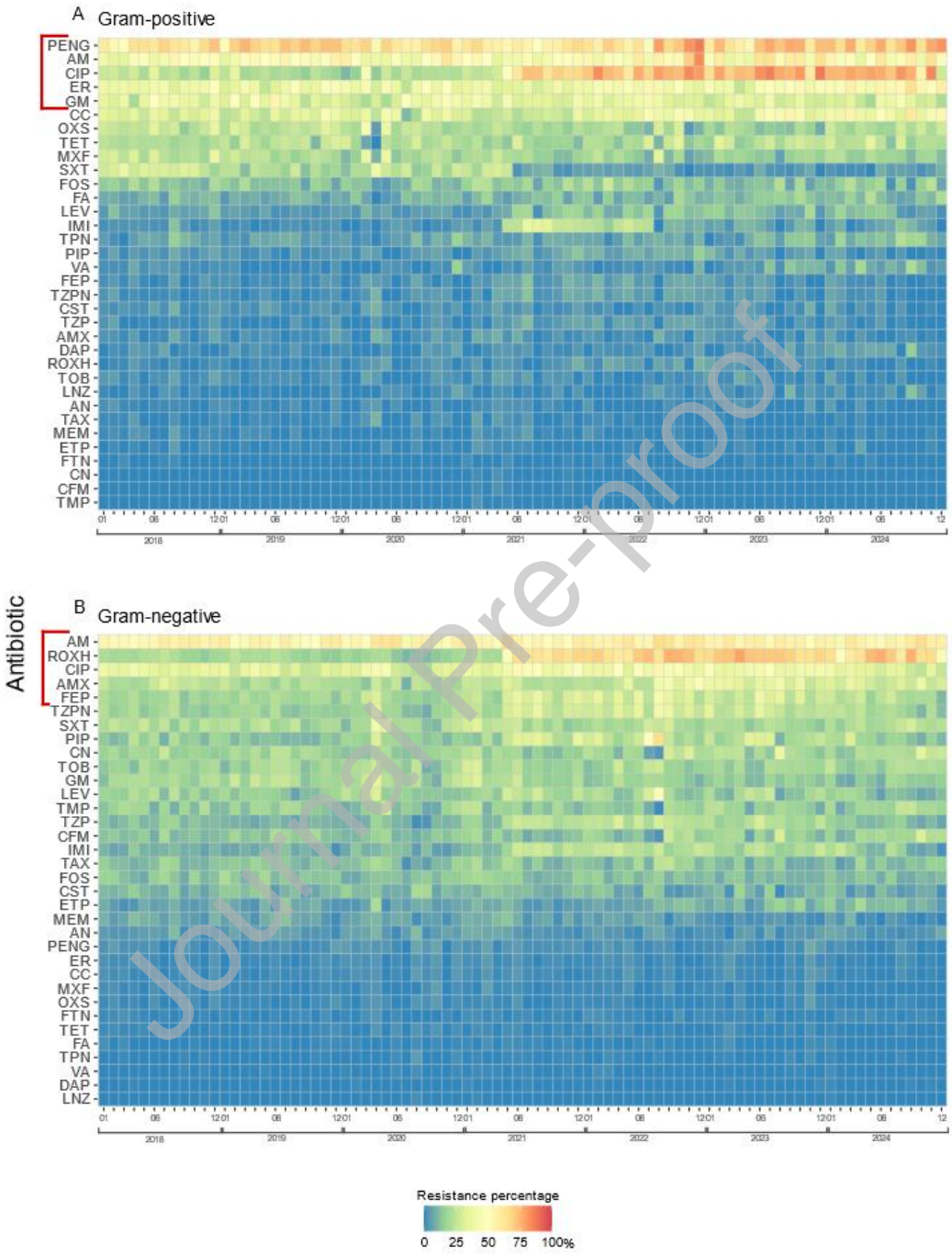


Figure 4. **Heatmaps of the antibiotic resistance in Gram-positive and Gram-negative isolates.** The heatmaps illustrate the percentage of resistance to the most commonly tested antibiotics over the study period. Each cell represents the percentage of resistant isolates for a specific antibiotic (horizontal axis) and month (vertical axis). The color scale ranges from blue (0% resistance) to red (100% resistance).

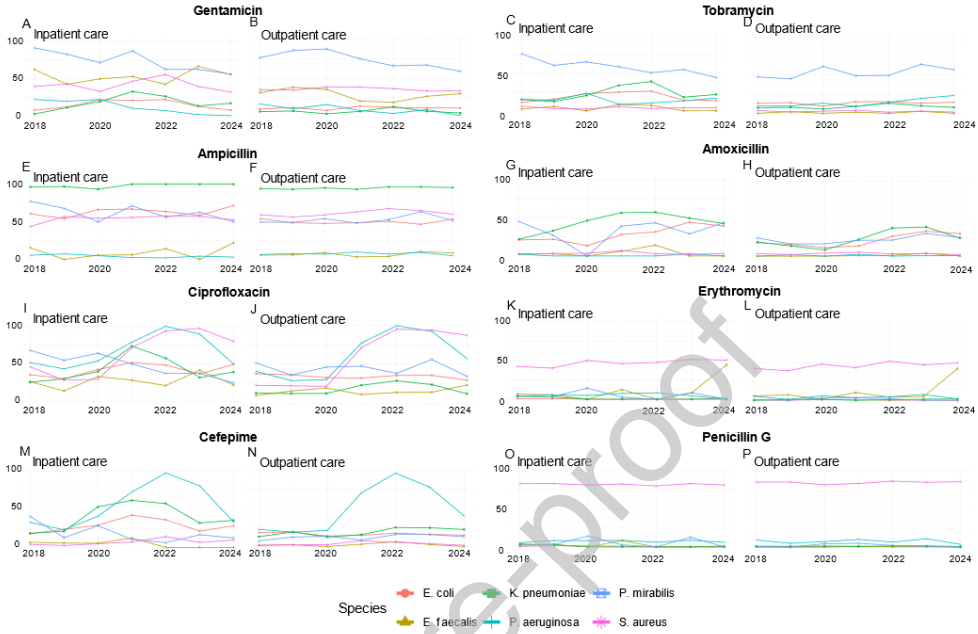


Figure 5. **Comparison of antibiotic resistance rates in different care settings.** Percentage of resistant strains in inpatient (left-hand panels) and outpatient (right-hand panels) care settings, for some representative antibiotics, exhibiting notable trends (gentamicin [a, b], tobramycin [c, d], ampicillin [e, f], amoxicillin [g, h], ciprofloxacin [i, j], erythromycin [k, l], cefepime [m, n], penicillin G [o, p]).

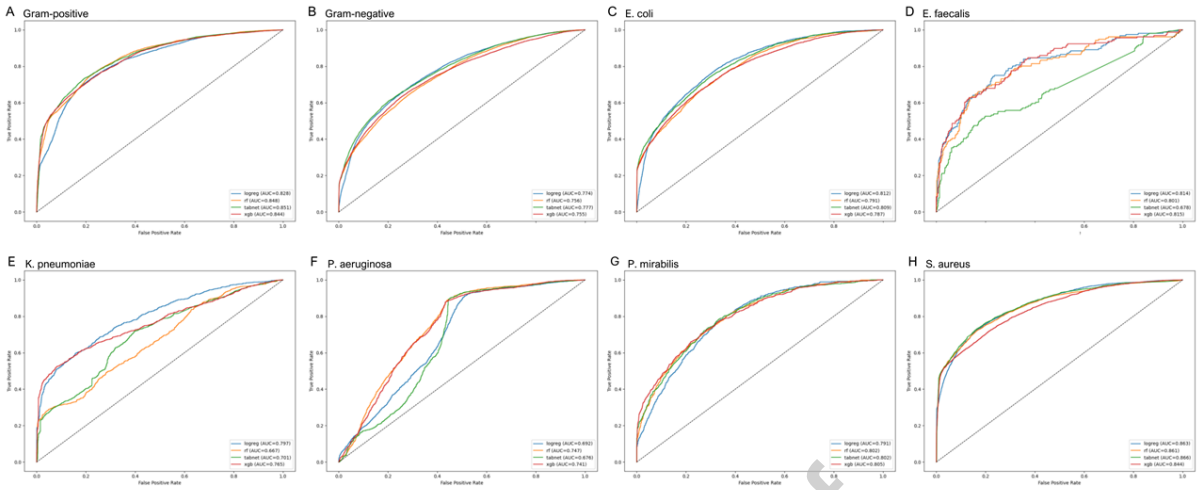


Figure 6. **Comparison of ROC curves of ML models for antibiotic resistance prediction.** ROC curves of logistic regression (logreg), random forest (rf), XGBoost (xgb), and TabNet for the Gram+ (a), Gram- (b), *E. coli* (c), *E. faecalis* (d), *K. pneumoniae* (e), *P. aeruginosa* (f), *P. mirabilis* (g), *S. aureus* (h) datasets. The colors represent different algorithms. AUC: Area under the receiver operating characteristic curve.

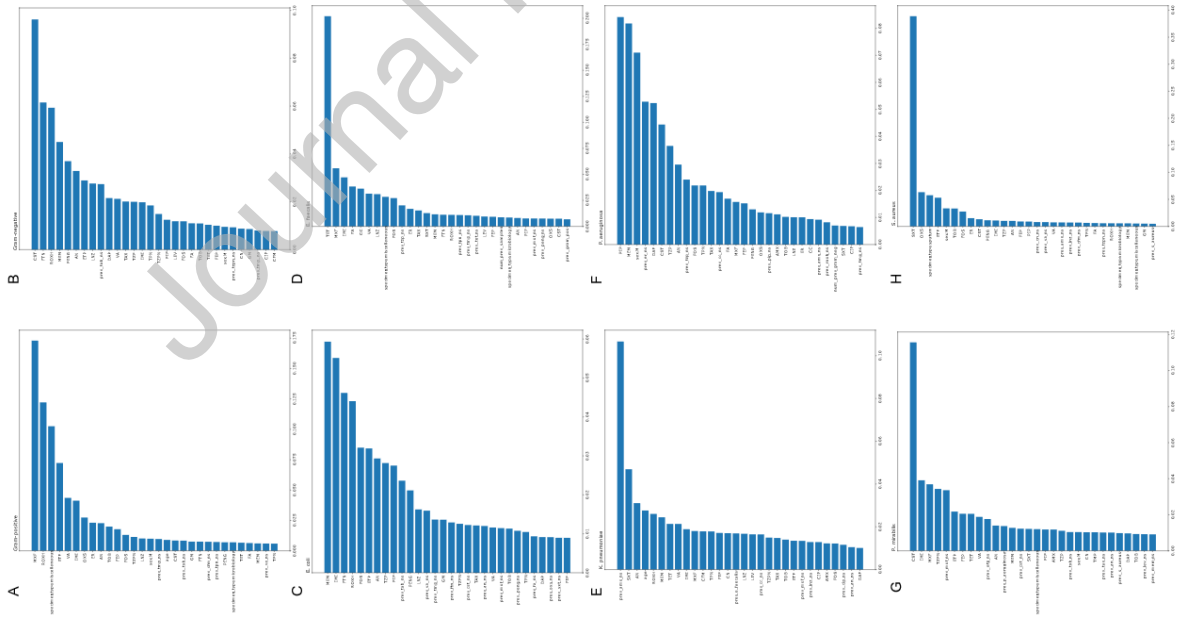
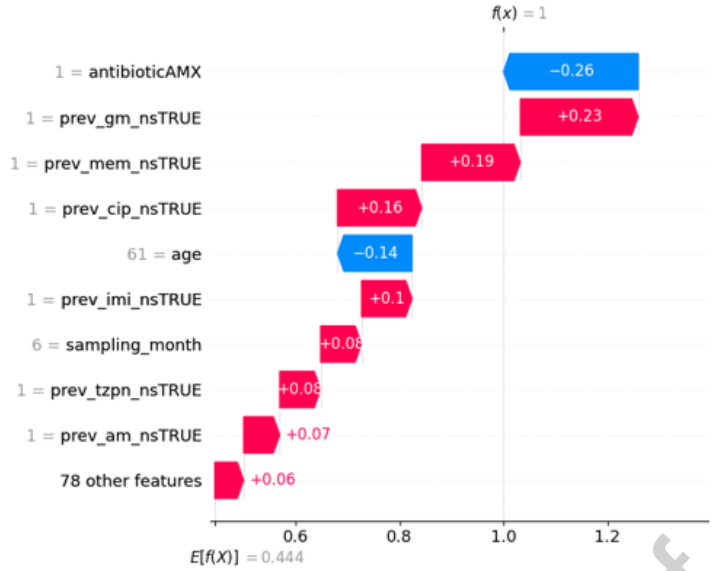
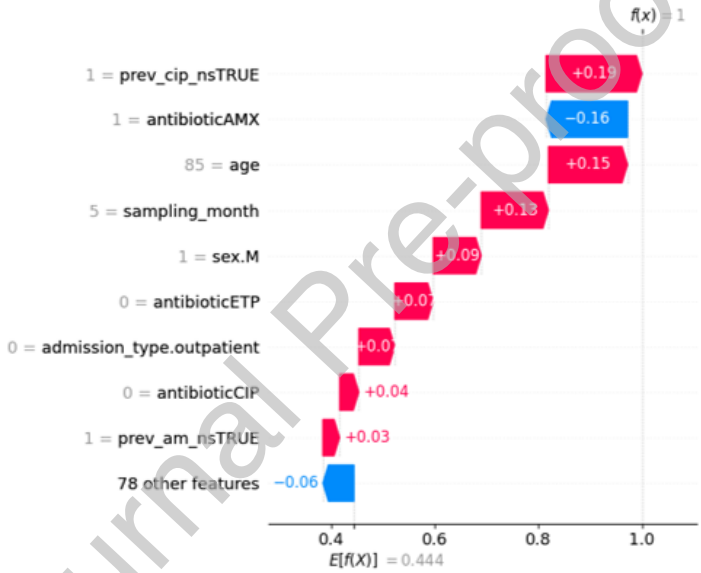


Figure 7. **Comparison of feature importance plots of XGBoost models.** Feature importance plots of XGBoost for the Gram+ (a), Gram- (b), E. coli (c), E. faecalis (d), K. pneumoniae (e), P. aeruginosa (f), P. mirabilis (g), S. aureus (h) datasets.

A



B



C

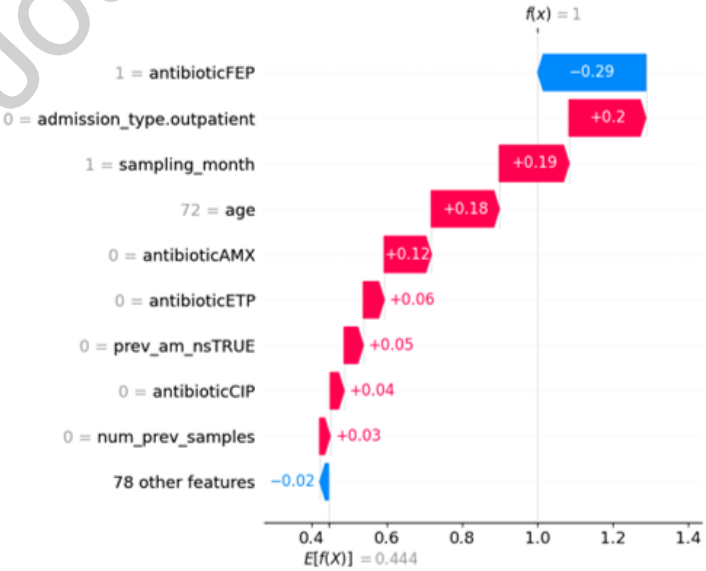


Figure 8. **SHAP-based local interpretability analysis of individual XGBoost model predictions for *E. coli* isolates.** Waterfall plots illustrate the contribution of individual features to the predicted probability of non-susceptibility relative to the baseline expected value (E[f(X)] = 0.444). Red bars indicate features increasing the predicted probability of non-susceptibility; blue bars indicate features decreasing it. Panels show representative cases for: a 61-year-old female inpatient predicted non-susceptible to amoxicillin (A); an 85-year-old male inpatient predicted non-susceptible to amoxicillin (B); and a 72-year-old female inpatient predicted non-susceptible to cefepime (C).

Table 1. **Summary of clinical and microbiological data**.

Feature	Raw dataset (n=15,581)	Training set (n=11,213)	Test set (n=3,328)
Age, mean (SD), years	61.9 (21.4)	61.8 (21.0)	59.5 (23.3)
Sex, No. (%)			
Male	6,317 (40.5%)	4,660 (41.6%)	1,343 (40.4%)
Female	9,263 (59.5%)	6,553 (58.4%)	1945 (59.4%)
Specimen analyzed, No. (%)			
Blood	221 (1.4%)	131 (1.2%)	87 (2.6%)
Sputum	165 (1.1%)	123 (1.1%)	34 (1.0%)
Miscellaneous	3,719 (23.9%)	2,810 (25.1%)	844 (25.4%)
Microbiology	1,679 (10.8%)	1,200 (10.7%)	423 (12.7%)
Swab	3,406 (21.9%)	2,593 (23.1%)	578 (17.4%)
Urine	6,391 (41.0%)	4,356 (38.8%)	1,362 (40.9%)
Unit type, No. (%)			
Medical	15,263 (98.0%)	11,001 (98.1%)	3,246 (97.5%)
Surgical	318 (2.0%)	212 (1.9%)	82 (2.5%)
Admission type, No. (%)			
Inpatient care	3,285 (21.1%)	2,583 (23.0%)	593 (17.8%)
Outpatient care	12,296 (78.9%)	8,630 (77.0%)	2,735 (82.2%)
Department, No. (%)			
Vascular surgery unit	25 (0.2%)	13 (0.1%)	9 (0.3%)
Dermatology unit	1610 (10.4%)	1,275 (11.4%)	256 (7.8%)
IDI Clinic	11,368 (76.0%)	8,104 (76.1%)	2,376 (72.5%)
General surgery unit	14 (0.1%)	12 (0.1%)	2 (0.1%)
Covid19 unit	140 (0.9%)	140 (1.3%)	0 (0%)
Plastic surgery unit	24 (0.2%)	21 (0.2%)	3 (0.1%)
Dermatology Day Hospital	115 (0.8%)	87 (0.8%)	21 (0.6%)
Day Hospital 5th unit	15 (0.1%)	15 (0.1%)	0 (0%)
Dermatology Day Surgery	3 (<0.1%)	3 (<0.1%)	0 (0%)
General Medicine unit	321 (2.2%)	125 (1.2%)	181 (7.4%)
Oncology unit	330 (2.2%)	230 (2.2%)	96 (2.9%)
Post-surgical care unit	40 (0.3%)	30 (0.3%)	9 (0.3%)
Operating theatre	142 (0.9%)	124 (1.2%)	15 (0.5%)
Clinical surveillance	1 (0.1%)	1 (0.1%)	0 (0%)
Intensive care unit	59 (0.4%)	57 (0.5%)	2 (0.1%)
Self-paying patients unit	12 (0.1%)	6 (0.1%)	6 (0.2%)
Villa Paola clinic	727 (4.9%)	411 (3.9%)	296 (72.5%)
Other services	615 (3.9%)	559 (4.9%)	56 (1.7%)
Gram stain, No. (%)			

Feature	Raw dataset (n=15,581)	Training set (n=11,213)	Test set (n=3,328)
Negative	9,198 (63.3%)	7,168 (61.8%)	2,231 (62.8%)
Positive	5,343 (36.7%)	4,045 (38.2%)	1,097 (37.2%)
Bacterial species, No. (%)			
Achromobacter spp.	10 (0.1%)	9 (0.1%)	0 (0%)
A. baumanii	24 (0.2%)	20 (0.2%)	4 (0.1%)
A. baumanii/calcoaceticus complex	25 (0.2%)	22 (0.2%)	3 (0.1%)
A. lwoffii/haemolyticus	1 (<0.1%)	1 (<0.1%)	0 (0%)
A. viridians	1 (<0.1%)	1 (<0.1%)	0 (0%)
C. davisae	1 (<0.1%)	0 (0%)	1 (<0.1%)
C. lapagei	1 (<0.1%)	1 (<0.1%)	0 (0%)
C. amalonaticus	1 (<0.1%)	1 (<0.1%)	0 (0%)
C. braakii	7 (<0.1%)	4 (<0.1%)	3 (0.1%)
C. farmeri	5 (<0.1%)	2 (<0.1%)	3 (0.1%)
C. freundii	63 (0.4%)	43 (0.4%)	17 (0.6%)
C. koseri	229 (1.6%)	157 (1.5%)	50 (1.7%)
C. youngae	1 (<0.1%)	0 (0%)	0 (0%)
Citrobacter spp.	14 (0.1%)	8 (0.1%)	3 (0.1%)
E. aerogenes	153 (1.1%)	116 (1.1%)	28 (0.9%)
E. cloacae	298 (2.0%)	221 (2.1%)	65 (2.2%)
Enterobacter spp.	35 (0.2%)	23 (0.2%)	11 (0.4%)
E. faecalis	850 (5.8%)	708 (6.7%)	104 (3.5%)
E. faecium	43 (0.3%)	36 (0.3%)	6 (0.2%)
E. raffinosus	1 (<0.1%)	1 (<0.1%)	0 (0%)
Enterococcus spp.	21 (0.2%)	18 (0.2%)	3 (0.1%)
E. coli	4,162 (28.6%)	2918 (27.5%)	831 (28.0%)
Escherichia spp.	4 (<0.1%)	3 (<0.1%)	1 (<0.1%)
H. alvei	2 (<0.1%)	2 (<0.1%)	0 (0%)
K. oxytoca	197 (1.4%)	142 (1.3%)	38 (1.3%)
K. pneumoniae spp ozaenae	19 (0.1%)	12 (0.1%)	5 (0.2%)
K. pneumoniae spp pneumoniae	934 (6.4%)	635 (6.0%)	189 (6.4%)
Klebsiella spp.	24 (0.2%)	17 (0.2%)	4 (0.1%)
K. ascorbata	1 (<0.1%)	1 (<0.1%)	0 (0%)
Moraxella species	1 (<0.1%)	1 (<0.1%)	0 (0%)
M. morganii	193 (1.3%)	151 (1.4%)	33 (1.1%)
P. mirabilis	845 (5.8%)	623 (5.9%)	158 (5.3%)
P. vulgaris	24 (0.2%)	17 (0.2%)	5 (0.2%)
P. vulgaris/penneri	6 (<0.1%)	5 (<0.1%)	1 (<0.1%)
Proteus spp.	9 (0.1%)	6 (0.1%)	1 (<0.1%)
P. alcalifaciens	1 (<0.1%)	1 (<0.1%)	0 (0%)
P. rettgeri	36 (0.2%)	21 (0.2%)	10 (0.3%)
P. stuartii	27 (0.2%)	22 (0.2%)	4 (0.1%)
Providencia spp.	1 (<0.1%)	1 (<0.1%)	0 (0%)
P. aeruginosa	1,637 (11.2%)	1,184 (11.2%)	354 (11.9%)
P. fluorescens	5 (<0.1%)	5 (<0.1%)	0 (0%)
P. luteola	1 (<0.1%)	1 (<0.1%)	0 (0%)
P. putida	8 (<0.1%)	6 (0.1%)	0 (0%)
Pseudomonas spp.	51 (0.3%)	40 (0.4%)	9 (0.3%)
S. liquefaciens	2 (<0.1%)	1 (<0.1%)	1 (<0.1%)

Feature	Raw dataset (n=15,581)	Training set (n=11,213)	Test set (n=3,328)
S. marcescens	104 (0.7%)	86 (0.8%)	14 (0.5%)
S. plymuthica	5 (<0.1%)	1 (<0.1%)	4 (0.1%)
Serratia spp.	7 (<0.1%)	4 (<0.1%)	2 (0.1%)
S. boydii	1 (<0.1%)	1 (<0.1%)	0 (0%)
S. aureus	3,638 (25.0%)	2,642 (24.9%)	851 (28.7%)
S. capitis	1 (<0.1%)	1 (<0.1%)	0 (0%)
S. caprae	1 (<0.1%)	1 (<0.1%)	0 (0%)
S. cohnii spp urelyticum	1 (<0.1%)	1 (<0.1%)	0 (0%)
S. epidermidis	20 (0.1%)	14 (0.1%)	5 (0.2%)
S. haemolyticus	283 (1.9%)	255 (2.4%)	21 (0.7%)
S. hominis	36 (0.2%)	34 (0.3%)	1 (<0.1%)
S. intermedius	1 (<0.1%)	0 (0%)	1 (<0.1%)
S. lentus	1 (<0.1%)	1 (<0.1%)	0 (0%)
S. lugdunensis	2 (<0.1%)	2 (<0.1%)	0 (0%)
S. pasteuri	2 (<0.1%)	2 (<0.1%)	0 (0%)
S. saprophyticus	1 (<0.1%)	1 (<0.1%)	0 (0%)
S. sciuri	4 (<0.1%)	4 (<0.1%)	0 (0%)
S. simulans	1 (<0.1%)	0 (0%)	1 (<0.1%)
S. warneri	1 (<0.1%)	1 (<0.1%)	0 (0%)
Staphylococcus spp.	3 (<0.1%)	3 (<0.1%)	0 (0%)
S. maltophilia	30 (0.2%)	24 (0.2%)	5 (0.2%)
S. agalactiae	323 (2.2%)	238 (2.2%)	79 (2.7%)
S. anginosus	4 (<0.1%)	4 (<0.1%)	0 (0%)
S. dysgalactiae spp equisimilis	1 (<0.1%)	0 (0%)	1 (<0.1%)
S. group A	9 (0.1%)	4 (<0.1%)	5 (0.1%)
S. group B	5 (<0.1%)	5 (<0.1%)	0 (0%)
S. group C	8 (0.1%)	8 (0.1%)	0 (0%)
S. group C/G	1 (<0.1%)	1(<0.1%)	0 (0%)
S. group G	7 (<0.1%)	7 (0.1%)	0 (0%)
S. pyogenes	82 (0.6%)	60 (0.6%)	22 (0.7%)

The test set presented in this table corresponds to the additional test set obtained after excluding isolates from 648 patients

Table 2. Performance metrics of machine learning models for predicting antibiotic resistance.

Dataset	Model	Method	AUROC			F1-score			Accuracy			Precision			Recall		
			CV	TS 1	TS 2	CV	TS 1	TS 2	CV	TS 1	TS 2	CV	TS 1	TS 2	CV	TS 1	TS 2
Gram+	LR	Raw	0.815±0.006	0.828	0.828	0.759±0.004	0.767	0.761	0.785±0.004	0.779	0.777	0.768±0.008	0.757	0.758	0.660±0.006	0.706	0.692
		SMOTE	0.815±0.005	0.827	0.827	0.728±0.006	0.707	0.714	0.715±0.006	0.697	0.703	0.688±0.005	0.706	0.707	0.726±0.005	0.737	0.739
	RF	Raw	0.854±0.004	0.850	0.848	0.783±0.008	0.771	0.765	0.805±0.004	0.794	0.791	0.803±0.009	0.816	0.818	0.688±0.006	0.697	0.687
		SMOTE	0.856±0.004	**0.858**	**0.857**	0.804±0.005	0.807	0.803	0.812±0.005	0.815	0.814	0.777±0.007	0.804	0.809	0.737±0.006	**0.754**	**0.742**
	XGBoost	Raw	**0.882±0.004**	0.849	0.844	0.827±0.005	0.803	0.801	**0.834±0.005**	0.814	0.813	**0.809±0.007**	0.807	0.809	0.764±0.006	0.746	0.738
		SMOTE	0.881±0.004	0.852	0.848	**0.828±0.005**	**0.807**	**0.804**	0.833±0.005	**0.816**	**0.815**	0.803±0.007	0.809	0.812	**0.770±0.006**	**0.752**	**0.742**
	TabNet	Raw	0.847±0.010	0.851	0.851	0.793±0.012	0.794	0.795	0.813±0.008	0.810	0.812	0.810±0.010	**0.824**	**0.831**	0.702±0.017	0.727	0.723
		SMOTE	0.831±0.122	0.817	0.814	0.787±0.013	0.740	0.739	0.791±0.015	0.735	0.734	0.745±0.022	0.703	0.698	0.725±0.012	0.718	0.714
Gram-	LR	Raw	0.779±0.005	0.787	0.774	0.760±0.004	0.767	0.755	0.785±0.003	0.778	0.769	0.737±0.006	0.735	0.724	0.646±0.005	0.692	0.677
		SMOTE	0.779±0.005	0.786	0.773	0.727±0.003	0.655	0.645	0.713±0.004	0.640	0.630	0.665±0.003	0.662	0.655	0.701±0.003	0.695	0.685
	RF	Raw	0.845±0.004	0.773	0.756	0.798±0.003	0.745	0.736	0.820±0.002	0.766	0.760	0.815±0.005	0.727	0.721	0.688±0.005	0.655	0.645
		SMOTE	0.826±0.003	0.797	**0.782**	0.799±0.004	0.766	**0.757**	0.811±0.003	0.776	0.769	0.769±0.005	0.731	0.723	0.706±0.005	0.692	**0.681**
	XGBoost	Raw	**0.878±0.003**	**0.874**	0.755	0.834±0.003	**0.796**	0.741	**0.841±0.003**	0.790	0.746	**0.808±0.005**	0.727	0.691	0.759±0.004	0.698	0.676
		SMOTE	**0.878±0.003**	0.869	0.752	**0.835±0.004**	0.751	0.737	0.839±0.004	**0.794**	0.739	0.799±0.005	**0.797**	0.683	**0.768±0.005**	**0.700**	0.677
	TabNet	Raw	0.813±0.006	0.787	0.777	0.791±0.005	0.761	0.751	0.808±0.004	0.782	**0.776**	0.773±0.011	0.756	**0.757**	0.688±0.008	0.672	0.658
		SMOTE	0.752±0.032	0.724	0.717	0.681±0.181	0.700	0.696	0.692±0.159	0.697	0.696	0.670±0.059	0.637	0.633	0.657±0.056	0.641	0.632
E. coli	LR	Raw	0.811±0.006	0.820	0.811	0.794±0.003	0.807	0.796	0.822±0.003	0.821	0.814	0.732±0.011	0.760	0.757	0.619±0.005	0.688	**0.675**
		SMOTE	0.811±0.006	0.817	0.810	0.746±0.004	0.671	0.673	0.721±0.004	0.644	0.647	0.657±0.004	0.655	0.655	0.731±0.006	0.718	0.714

Dataset	Model	Method	AUROC			F1-score			Accuracy			Precision			Recall		
			CV	TS1	TS2	CV	TS1	TS2	CV	TS1	TS2	CV	TS1	TS2	CV	TS1	TS2
	RF	Raw	0.883±0.004	0.840	0.791	0.838±0.004	0.786	0.774	0.861±0.002	0.835	0.817	**0.854±0.004**	0.873	0.865	0.676±0.007	0.643	0.622
		SMOTE	0.859±0.004	0.828	**0.811**	0.840±0.005	0.805	0.790	0.853±0.003	0.831	**0.820**	0.791±0.008	**0.818**	**0.817**	0.703±0.005	0.667	0.648
	XGBoost	Raw	**0.919±0.002**	0.844	0.786	**0.882±0.004**	0.843	0.789	**0.886±0.003**	**0.854**	0.815	0.838±0.007	0.781	0.779	0.786±0.007	**0.762**	0.654
		SMOTE	0.917±0.003	**0.854**	0.785	0.881±0.003	**0.865**	0.793	0.885±0.003	0.818	0.810	0.813±0.006	0.792	0.747	**0.789±0.005**	0.709	0.674
	TabNet	Raw	0.841±0.008	0.817	0.809	0.824±0.006	0.810	**0.800**	0.846±0.005	0.827	**0.820**	0.794±0.013	0.778	0.779	0.663±0.012	0.685	0.674
		SMOTE	0.771±0.031	0.768	0.755	0.709±0.135	0.758	0.746	0.698±0.142	0.759	0.749	0.651±0.041	0.660	0.650	0.670±0.029	0.660	0.644
E. faecalis	LR	Raw	0.879±0.014	0.810	**0.814**	0.840±0.013	0.813	0.832	0.856±0.011	0.838	**0.852**	0.847±0.023	0.793	**0.805**	0.723±0.019	0.651	0.673
		SMOTE	0.876±0.015	0.805	0.811	0.805±0.013	0.777	0.800	0.793±0.014	0.760	0.760	0.733±0.014	0.679	0.680	0.795±0.016	0.739	**0.752**
	RF	Raw	0.930±0.010	0.763	0.801	0.876±0.010	0.801	0.812	0.882±0.009	0.820	0.822	0.855±0.014	0.727	0.718	0.798±0.020	0.650	0.672
		SMOTE	0.926±0.010	0.781	0.800	0.876±0.010	0.802	0.810	0.879±0.008	0.816	0.816	0.843±0.008	0.716	0.707	0.807±0.021	0.657	0.682
	XGBoost	Raw	**0.938±0.009**	**0.889**	**0.814**	0.890±0.010	**0.836**	0.830	0.894±0.009	**0.866**	0.847	**0.873±0.017**	**0.870**	0.779	0.820±0.013	0.755	0.680
		SMOTE	0.938±0.011	0.879	0.813	**0.893±0.010**	0.829	**0.834**	**0.896±0.009**	0.851	0.848	0.870±0.017	0.867	0.776	**0.830±0.014**	**0.866**	0.693
	TabNet	Raw	0.841±0.017	0.682	0.678	0.824±0.017	0.773	0.779	0.842±0.012	0.819	0.822	0.821±0.027	0.778	0.760	0.702±0.029	0.585	0.583
		SMOTE	0.853±0.055	0.781	0.792	0.716±0.142	0.808	0.816	0.703±0.130	0.806	0.809	0.694±0.050	0.706	0.708	0.752±0.074	0.714	0.741
K. pneumoniae	LR	Raw	0.832±0.018	**0.820**	**0.797**	0.817±0.011	0.789	0.757	0.830±0.010	0.790	0.757	0.820±0.018	0.748	0.714	0.731±0.014	0.744	0.715
		SMOTE	0.832±0.019	0.811	0.794	0.792±0.017	0.593	0.578	0.788±0.018	0.584	0.572	0.738±0.019	0.656	0.651	0.757±0.018	0.672	0.660
	RF	Raw	0.882±0.009	0.720	0.667	0.849±0.010	0.707	0.667	0.856±0.009	0.726	0.684	0.842±0.012	0.661	0.608	0.782±0.015	0.621	0.587
		SMOTE	0.876±0.009	0.804	0.773	0.829±0.014	0.776	0.748	0.836±0.013	0.785	0.762	0.808±0.020	0.746	0.725	0.760±0.017	0.710	0.676
	XGBoost	Raw	**0.941±0.009**	0.810	0.765	**0.891±0.010**	**0.870**	0.771	**0.893±0.010**	**0.881**	0.777	**0.878±0.015**	**0.851**	0.738	0.847±0.013	**0.833**	0.716

Dataset	Model	Method	AUROC			F1-score			Accuracy			Precision			Recall		
			CV	TS 1	TS 2	CV	TS 1	TS 2	CV	TS 1	TS 2	CV	TS 1	TS 2	CV	TS 1	TS 2
		SMOTE	0.939±0.009	0.816	0.781	0.890±0.013	0.800	**0.784**	0.892±0.013	0.807	**0.796**	0.872±0.018	0.775	**0.775**	**0.852±0.016**	0.740	**0.717**
	TabNet	Raw	0.833±0.021	0.700	0.701	0.817±0.013	0.686	0.667	0.829±0.012	0.675	0.655	0.814±0.021	0.637	0.630	0.733±0.016	0.657	0.651
		SMOTE	0.833±0.018	0.782	0.772	0.820±0.014	0.572	0.560	0.823±0.015	0.563	0.556	0.784±0.023	0.635	0.637	0.761±0.016	0.648	0.643
P. aeruginosa	LR	Raw	0.840±0.011	0.708	0.692	0.794±0.010	0.518	0.528	0.798±0.009	0.586	0.586	0.792±0.011	0.717	0.710	0.770±0.010	0.594	0.605
		SMOTE	0.839±0.012	0.709	0.693	0.769±0.008	0.432	0.427	0.766±0.008	0.540	0.531	0.754±0.008	0.715	0.706	0.764±0.010	0.549	0.554
	RF	Raw	0.921±0.008	0.739	0.747	0.850±0.006	0.690	0.691	0.854±0.006	0.703	0.702	0.860±0.007	0.751	0.752	0.826±0.007	0.708	0.713
		SMOTE	0.917±0.007	0.739	0.745	0.845±0.007	0.697	0.695	0.849±0.007	0.709	0.705	0.850±0.008	0.753	0.752	0.823±0.007	0.713	0.716
	XGBoost	Raw	0.941±0.005	0.839	0.740	**0.876±0.008**	0.804	0.697	**0.877±0.008**	0.741	0.706	**0.873±0.009**	**0.849**	0.743	**0.862±0.009**	0.817	0.716
		SMOTE	**0.946±0.005**	**0.849**	**0.750**	0.870±0.007	0.802	0.697	0.871±0.007	**0.821**	0.706	0.868±0.008	**0.849**	0.748	0.855±0.008	0.816	0.717
	TabNet	Raw	0.871±0.013	0.811	0.675	0.824±0.011	**0.809**	**0.701**	0.831±0.009	0.781	**0.710**	0.844±0.010	0.793	0.749	0.794±0.013	**0.792**	**0.720**
		SMOTE	0.876±0.012	0.807	0.771	0.819±0.014	0.803	0.694	0.821±0.014	0.781	0.705	0.814±0.018	0.774	**0.751**	0.801±0.014	0.781	0.716
P. mirabilis	LR	Raw	0.825±0.017	0.781	0.791	0.750±0.017	0.694	0.715	0.753±0.017	0.702	0.719	0.740±0.019	0.688	0.695	0.727±0.018	0.670	0.682
		SMOTE	0.825±0.017	0.782	0.792	0.742±0.016	0.712	0.713	0.739±0.016	0.709	0.709	0.732±0.016	0.700	0.696	0.744±0.016	0.708	0.710
	RF	Raw	0.872±0.015	0.784	0.801	0.799±0.014	0.707	0.726	0.804±0.014	0.706	0.722	0.801±0.016	0.692	0.706	0.774±0.015	0.694	0.718
		SMOTE	0.873±.017	0.792	**0.808**	0.800±0.015	0.715	0.725	0.802±0.014	0.712	0.720	0.792±0.016	0.706	**0.715**	0.782±0.016	0.715	**0.733**
	XGBoost	Raw	**0.894±0.011**	**0.879**	0.804	**0.816±0.010**	**0.787**	0.707	0.817±0.010	0.783	0.701	**0.810±0.011**	**0.800**	0.701	0.798±0.011	0.757	0.717
		SMOTE	0.893±0.010	0.878	0.805	**0.816±0.008**	0.796	0.695	**0.818±0.008**	**0.793**	0.689	0.809±0.009	0.797	0.698	**0.800±0.009**	**0.766**	0.713
	TabNet	Raw	0.823±0.014	0.808	0.802	0.749±0.015	0.782	**0.732**	0.753±0.016	0.708	**0.734**	0.740±0.018	0.703	0.711	0.727±0.015	0.703	0.704
		SMOTE	0.826±0.015	0.792	0.778	0.748±0.015	0.720	0.701	0.746±0.015	0.700	0.696	0.735±0.013	0.698	0.683	0.743±0.014	0.704	0.697

Dataset	Model	Method	AUROC			F1-score			Accuracy			Precision			Recall		
			CV	TS 1	TS 2	CV	TS 1	TS 2	CV	TS 1	TS 2	CV	TS 1	TS 2	CV	TS 1	TS 2
S. aureus	LR	Raw	0.859±0.005	0.862	0.863	0.804±0.007	0.808	0.809	0.822±0.006	0.813	0.816	0.830±0.009	0.793	0.801	0.714±0.008	0.766	0.758
		SMOTE	0.859±0.005	0.861	0.862	0.766±0.004	0.737	0.742	0.756±0.004	0.728	0.733	0.722±0.004	0.726	0.726	0.762±0.004	0.758	**0.760**
	RF	Raw	0.874±0.004	0.867	0.861	0.826±0.006	**0.818**	**0.814**	0.837±0.005	**0.827**	**0.826**	0.829±0.007	0.828	0.836	0.751±0.007	0.763	0.751
		SMOTE	0.877±0.003	**0.873**	**0.871**	0.825±0.005	0.816	0.812	0.832±0.004	0.824	0.823	0.807±0.005	0.818	0.827	0.760±0.008	**0.764**	0.752
	XGBoost	Raw	0.890±0.004	0.849	0.844	0.839±0.004	0.801	0.798	**0.846±0.004**	0.805	0.804	0.826±0.006	0.784	0.781	0.778±0.006	0.757	0.750
		SMOTE	**0.891±0.003**	0.851	0.847	**0.840±0.005**	0.800	0.797	0.844±0.005	0.804	0.801	0.817±0.005	0.780	0.775	**0.785±0.006**	0.760	0.752
	TabNet	Raw	0.874±0.006	0.866	0.866	0.821±0.005	0.814	0.811	0.837±0.005	0.825	0.824	**0.855±0.014**	**0.832**	**0.837**	0.734±0.007	0.756	0.747
		SMOTE	0.861±0.017	0.868	0.869	0.806±0.019	0.813	0.811	0.813±0.024	0.820	0.820	0.786±0.040	0.809	0.812	0.739±0.016	0.765	0.756

Performance metrics of machine learning models to predict antibiotic susceptibility results (susceptible or non-susceptible) across various test sets. The prediction was performed according to the antibiograms available for each patient. XGBoost consistently achieved the highest discriminative metrics, outperforming all other classifiers. The highest score for each metric is highlighted in bold. Abbreviations: LR, Logistic Regression; RF, Random Forest; XGBoost, eXtreme Gradient Boosting; AUROC, Area Under the Receiver Operating Characteristic curve; CV: cross-validation; TS1: test set that includes patients appearing in the training set; TS2: test set where patients appearing in the training set were excluded.

Declaration of Interest Statement

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.