

Evaluation of Ensemble Machine Learning Model Utilizing Biomarkers to Predict Multidrug-Resistant Enterobacteriaceae Urosepsis: A Narrative Review

Sivakumar VELLUSAMY^{1*}, Santhosh PERIYAKARUPPAN², and Ragesh SARAVANAN²

¹Professor and Head of Department, Center of Excellence in Computational Intelligence in Pharmacy, Department of Pharmacy Practice, PSG College of Pharmacy, Peelamedu, Coimbatore – 641004, India.

² Doctor of Pharmacy, Center of Excellence in Computational Intelligence in Pharmacy, Department of Pharmacy Practice, PSG College of Pharmacy, Peelamedu, Coimbatore – 641004, India.

E-mails: sivakumar7868@psgpharma.ac.in; psanthoshpharmd06@gmail.com; rageshchess@gmail.com

* Author to whom correspondence should be addressed; Tel.: 04224587320

Received: 3 March 2026 / Accepted: 24 September 2026 / Published online: 30 September 2026

Abstract

Urosepsis primarily caused by multidrug-resistant Enterobacteriaceae is a major public health problem due to high mortality rates and a large diagnostic delay using conventional culture methods. This delay often translates to inappropriate empirical antibiotic treatment, exacerbating the problem of antimicrobial resistance. This review determines if ensemble-based machine learning models, especially those that use inflammatory markers, can help bridge this diagnostic gap. Ensemble based tree models, such as Extreme Gradient Boosting, might outperform traditional models when used to analyze structured clinical data to predict risk more effectively. Using model interpretability, such as SHapley Additive exPlanations, is a key factor to determine that C-reactive protein, procalcitonin, and the ratio of procalcitonin to albumin are key predictors. Implementing these transparent, data-driven systems may facilitate implementation of precision antimicrobial stewardship. By closing this diagnostic window ultimately enables individualized therapeutic interventions, minimizing the inappropriate second-line antibiotic use and improving patient survival rates.

Keywords: Sepsis; Enterobacteriaceae; Machine Learning (ML); Drug resistance; Biomarkers.

Introduction

Urosepsis is a life-threatening systematic inflammatory response to infections arising from the urinary system or genitalia which often leads to organ dysfunction. It represents approximately 25% of all cases of sepsis and has a high mortality rate of 30-40% [1]. Which can escalate to 60% in specific, highly vulnerable populations [2]. The microbiological spectrum of urosepsis is overwhelmingly dominated by the Enterobacteriaceae family. These pathogens are the most significant cause of morbidity and mortality in patients with urosepsis and are responsible for the majority of systemic urologic infections [3]. Over time, the complex evolutionary adaptations of this bacterial family have driven the emergence of multidrug-resistant organisms (MDROs), which are considered to be bacteria that are resistant to three or more antimicrobial classes. The emerging resistance is a significant global public health threat with 700,000 deaths annually worldwide, with 200,000 of those deaths being newborns [4]. When looking at the particular group of septicemia cases, *Escherichia coli* (*E. coli*) is the predominant causative pathogen associated with urinary tract infections acquired both in the hospital and the community, accounting for 80% of all cases of uncomplicated urinary tract infections [5]. *E. coli* comprises 61% of the blood culture isolates of specific septicemia cases. Although *E. coli* appears to be the most common pathogen, especially among female patients, the antimicrobial resistance profile of *E. coli* appears to be worsening rapidly. In fact, 50% of the *E. coli* isolates display resistance to at least one antibiotic, while 13% to 25.4% display resistance to trimethoprim-

sulfamethoxazole. From a mechanistic perspective, the development of extended-spectrum beta-lactamase (ESBL) resistance appears to represent a significant change in the therapeutic paradigm for Enterobacteriaceae. This appears to develop primarily as a result of the acquisition of plasmids that produce beta-lactamase, which breakdown the amide bond of the beta-lactam ring [6]. Furthermore, these plasmids, especially those carrying the genes for cefotaximases, tend to be horizontally transmitted along with resistance genes for aminoglycosides and fluoroquinolones, creating a robust multidrug-resistant (MDR) profile [7]. Indeed, the rising trend of resistance is substantially triggered by the dissemination of the ST131 lineage and ESBL genes globally [8]. Thus, the levels of ESBL production among urinary *E. coli* infections have reached alarming rates, with resistance rates as high as 79% in Türkiye and 64% in Egypt [9]. *Klebsiella pneumoniae* is the second most frequently isolated uropathogen, with rates ranging from 11.1% to 15.7% in different clinical settings [10]. In intensive care unit (ICU) settings, *Klebsiella pneumoniae* has established itself as a major concern, responsible for 65.9% of bacteriuria that often progresses to systemic shock. The high virulence and resistance capacity of *Klebsiella pneumoniae* are attributed to specific enzymes such as *Klebsiella pneumoniae* carbapenemase (KPC) and New Delhi metallo-beta-lactamase (NDM-1), which have been directly linked to the failure of carbapenem-based regimens in treating life-threatening infections. The widespread effect of this pathogen on a global scale is reflected in the 790,000 deaths recorded in 2019, underscoring the recurring inability of conventional empiric regimens to treat this infection effectively [11]. *Proteus mirabilis* adds another layer of complexity to the problem of urosepsis, being one of the most prominent members of this group and the major contributor of catheter-associated urinary tract infections (CAUTIs). This bacterium also presents unique challenges for therapy, given its inherent resistance and capacity for the acquisition of additional resistance mechanisms [12]. In contrast, less virulent species, such as *Enterococci* and *Pseudomonas aeruginosa*, are more commonly seen in immunocompromised patients, while the remaining 16% of blood culture isolates consists of other minor Enterobacteriaceae species.

Infections from these MDR Enterobacteriaceae are very unpredictable, with mortality rates for patients with MDR urinary tract infections ranging from 40 to 50%. In cases involving the bloodstream, the mortality rate can increase to as high as 70% [1]. Since these organisms can tolerate conditions that are exposed to significant antimicrobial pressure, the effectiveness of traditional antibiotics can be diminished. This changing environment creates a critical diagnostic gap, as the identification of MDROs can be difficult. Traditional phenotypic and culture-based diagnostic methods, although considered the gold standard, work-intensive and require approximately more than 48 hours to yield actionable antimicrobial susceptibility results [13]. Consequently, clinicians are often required to use broad-spectrum empirical therapy on an appropriate basis during this waiting time period, may have a chance of negative patient outcomes [14]. Conversely, delays in intervention and treatment not only worsen clinical outcomes but also contribute to the escalating global antimicrobial resistance crisis, which highlighting the need for rapid, predictive, data-driven diagnostic tools [15]. Early recognition of systemic inflammation is therefore critical to guide timely targeted therapy, prevent disease transmission, and reduce the high mortality burden associated with these conditions [16]. Inflammatory laboratory biomarkers such as C-Reactive Protein (CRP) and Procalcitonin (PCT), are integral to this early clinical decision-making process. Notably, the PCT/Albumin ratio (PAR) has been shown to act as a reliable predictor with an 84.62% sensitivity and 96.0% specificity at identifying bacteriuria [17]. Nevertheless, conventional scoring systems such as the Sequential Organ Failure Assessment (SOFA) score have shown considerable limitations in the diagnosis of early sepsis due to their low sensitivity and high false-negative rate [18]. Although the clinical utility of individual biomarkers and scoring systems is increasing, the complexity in diagnosing sepsis using these biomarkers alone still cannot be resolved effectively, leading to considerable gaps in diagnosis that hinder the implementation of appropriate early interventions [19].

In order to overcome these limitations, advanced ensemble-based machine learning (ML) models have emerged as potential solutions. Unlike traditional linear scoring systems, ML models process high-dimensional electronic health records—including vital signs, historical antibiotic exposure, and dynamic biomarker fluctuations to identify complex, non-linear clinical patterns. By iteratively learning from vast datasets, these algorithms optimize clinical decision support by calculating specific outcome probabilities and may support earlier risk stratification while conventional culture-based results are pending [20]. Among these methods, Extreme Gradient Boosting (XGBoost) has shown promising predictive performance in urosepsis and related urinary tract infection outcomes. In a study of critically ill patients with urosepsis, XGBoost demonstrated the highest predictive performance

among the evaluated models, with an Area Under the Receiver Operating Characteristic (AUROC) of 0.904 in the training cohort and 0.869 in the internal validation cohort [2]. Its ability to model non-linear relationships and interactions among clinical variables, together with regularization strategies that help control overfitting, makes it a relevant candidate for clinical prediction. Its ability to model non-linear relationships and interactions among clinical variables, together with regularization strategies that help control overfitting, makes it a relevant candidate for clinical prediction. Thus, comparative evaluation of ensemble machine learning architectures is important for assessing their performance in predicting multidrug-resistant Enterobacteriaceae urosepsis using clinical, laboratory, and relevant biomarker data. The incorporation of data-driven predictive systems into the clinical workflow and electronic health records systems has the potential to enhance precision antibiotic stewardship through the early prediction and optimization of patient care. This review analyze the potential clinical usefulness of ensemble machine learning models for the early prediction of urosepsis due to MDR Enterobacteriaceae. Specifically, the present objectives are related to finding out how first obtaining dynamic inflammatory markers and employing clever methods of feature selection enhance risk assessment before traditional culture results are available and considers the potential role of explainable artificial intelligence systems will help in converting complex predictions into clear, practical clinical decision support systems. Ultimately, this review seeks to outline pathways for bridging the diagnostic delay and advancing precision antimicrobial stewardship.

Materials and Methods

Search Strategy

A comprehensive literature review was performed to explore the use of ensemble machine learning architectures for the prediction of multidrug-resistant Enterobacteriaceae urosepsis. Relevant studies were identified through a search of the major electronic data bases, including PubMed, Scopus, Web of science, ScienceDirect and Google Scholar. To provide a comprehensive understanding of the basic biomarker literature, as well as the prevailing trends in prediction algorithms, the investigation has been limited to eligible literature published between January 2010 and February 2026 was considered. Each research database has been interrogated with the utilization of relevant combinations of the standardized vocabulary (if existent), keywords used in the literature, and Boolean filtering combined with language filtering and filtering relating to the type of publications offered by the respective platform. A comprehensive analysis of the search formulations used in each database is presented in Table 1, MeSH terms were employed only in the PubMed search since MeSH is unique for PubMed, and so we utilized equivalent free-text terms in other databases using the same Boolean operators.

Selection and Description of Literature

To ensure a focused and reproducible review, we followed strict eligibility criteria for literature selection that are relevant to the review. we included English-language research, clinical validations, and reviews evaluating tree-based ensemble architectures for predicting infection, septic shock, or multidrug resistance. Eligible studies were required to integrate structured electronic health records or dynamic inflammatory biomarkers, such as C-Reactive protein, Procalcitonin, or the Procalcitonin/albumin ratio into their predictive models. Random Forest and XGBoost were selected specifically as search terms because they are well known types of bagging- and boosting-based ensemble methods which are frequently tested in predictive tasks involving structured clinical data and electronic health records. We excluded non-English texts, preprints, isolated case reports, and purely theoretical computer science papers without direct clinical applications to infectious disease management.

Table 1: Database-specific search strategies, Boolean combinations, MeSH terminology, and filters applied in the literature search

Database	Search String	MeSH Terms Applied	Filters Applied
PubMed	("Sepsis" OR "Urinary Tract Infections") AND "Enterobacteriaceae" AND "Drug Resistance, Multiple, Bacterial" AND ("Machine Learning" OR "Ensemble models" OR "XGBoost" OR "Random Forest")	Sepsis; Urinary Tract Infections; Enterobacteriaceae; Drug Resistance, Multiple, Bacterial; Machine Learning; Procalcitonin; C-Reactive Protein; Antimicrobial Stewardship	Language: English; Document type: Clinical Study, Review; Year range: 2010–2026
Scopus	TITLE-ABS-KEY(("sepsis" OR "urinary tract infection") AND "enterobacteriaceae" AND "multidrug resistant" AND ("machine learning" OR "ensemble model" OR "XGBoost" OR "random forest"))	Not applicable	Language: English; Document type: Article, Review; Year range: 2010–2026
Web of Science	TS=(("sepsis" OR "urinary tract infection") AND "enterobacteriaceae" AND "multidrug resistant" AND ("machine learning" OR "ensemble model" OR "XGBoost" OR "random forest"))	Not applicable	Language: English; Document type: Article, Review; Year range: 2010–2026
ScienceDirect	(sepsis OR "urinary tract infection") AND enterobacteriaceae AND "multidrug resistance" AND ("machine learning" OR "ensemble model" OR XGBoost OR "random forest")	Not applicable	Language: English; Document type: Research article, Review; Year range: 2010–2026
Google Scholar	"urosepsis" "multidrug resistant" "Enterobacteriaceae" "machine learning" ("XGBoost" OR "Random Forest")	Not applicable	Language: English; Year range: 2010–2026; patents and citations excluded

Applications of Ensemble Machine Learning Architectures

The application of ML models has become increasingly vital for addressing the critical more than 48 hours diagnostic gap in urosepsis by enabling early risk stratification and precision antibiotic stewardship through the integration of inflammatory biomarkers, such as CRP, PCT, and the PCT/Albumin ratio. These models evaluate dynamic clinical indicators including urea nitrogen, urine output, and alkaline phosphatase alongside prior antibiotic exposures to predict significant adverse prognostic events and resistance profiles. For the specific prediction of MDR Enterobacteriaceae urosepsis, ML classifications predominantly feature tree-based ensemble architectures, notably Random Forest (RF) and XGBoost, alongside baseline algorithms like logistic regression and support vector machines. From the body of the literature, a set of primary validation studies were identified as directly meeting the clinical criteria for predicting urosepsis, urinary tract infections, and associated multidrug-resistant Gram-negative bacterial infections using structured electronic health records and laboratory biomarkers. These primary studies are compiled, evaluated, and contrasted in Table 2.

Recent studies have demonstrated the application of machine learning to predict infection-related outcomes, antimicrobial resistance, and clinical risk in UTI and urosepsis. Among the evaluated methods, tree-based ensemble algorithms, particularly RF and XGBoost, have shown promising predictive performance across several clinical applications. Wei et al. developed and internally validated algorithms for predicting prognosis in critically ill patients with urosepsis. Random Forest, LASSO, Boruta, and XGBoost were used for variable identification and feature selection, followed by comparison of seven machine learning algorithms. XGBoost demonstrated comparatively better performing model, achieving an AUROC of 0.904 in the training cohort and 0.869 in the internal validation cohort, with SHAP analysis used to interpret the model predictions [2]. Tolan et al. evaluated Random Forest and CatBoost models using 691 clinical *E. coli* isolates to predict antibiotic resistance, with Random Forest showing comparatively higher predictive performance among the evaluated models [9]. Pan et al. developed and validated a model to predict MDR-K. pneumoniae-related septic shock in 1,385 patients. LASSO regression and Boruta

were used to identify seven predictors, after which seven machine learning methods were evaluated. Logistic regression was selected as the optimal model, achieving an AUROC of 0.906 and 0.865 in the training and validation cohorts, respectively [14]. Similarly, Almalki et al. evaluated XGBoost, LightGBM, Random Forest, and HistGradientBoostingClassifier for predicting antimicrobial resistance using (Electronic Health Record) EHR-derived variables, with XGBoost achieving an AUROC of 0.902 [20]. Rich et al. developed decision tree, boosted logistic regression, and Random Forest models using EHR data from 9,990 patients to predict resistance to sulfamethoxazole-trimethoprim (SXT), nitrofurantoin (NIT), ciprofloxacin (CIP), and multidrug resistance, with AUROCs ranging from 0.58 to 0.66 across the evaluated outcomes [21]. Yang et al. have developed interpretable machine learning models by means of logistic regression, XGBoost, TabNet, and TabNetself to predict non-susceptibility to antibiotics including NIT, SXT, CIP, and levofloxacin (LVX) in complicated UTIs with XGBoost showing the best overall predictive performance among the other models [22].

Ultimately, Choi et al. validated AI models for predicting UTI and urine culture results and the risk of bloodstream infection in the course of an urinary tract infection with XGBoost yielding AUROC values of 0.967 and 0.955, respectively [23]. Collectively, these studies illustrate the potential in the use of machine learning, particularly XGBoost for infection-related risk and antimicrobial resistance prediction. XGBoost appears promising across several studies, although its performance varies with the dataset and modelling strategy. Thus, no single algorithm can be considered universally superior. XGBoost may represent a strong candidate when it comes to biomarker-integrated prediction and early risk stratification, while its clinical utility in antibiotic selection and patient care requires prospective validation.

Table 2: Comparative Summary of Machine Learning Studies in Infection Management: Data Characteristics, Predictors, Model Performance, Validation, and Clinical Outcomes.

Study	Data Source	Sample Size	Feature Selection	Predictors	Model/ AUROC / SHAP	Validation	Outcome
Wei et al. (2025) [2]	MIMIC-IV v3.1 (Beth Israel BIDMC, 2008-2022)	n=1,389 urosepsis ICU patients (70% train [n=986], 30% test [n=403])	LASSO regression and Boruta algorithm for variable reduction	Age, vital signs, prior comorbidities, urine output, biochemical biomarkers	XGBoost AUROC: 0.904, F1: 0.791. SHAP global/local interpretability	Internal validation on 30% holdout test set; cross-validation in LASSO	Major adverse prognostic events & 28-day mortality in ICU
Tolan et al. (2025) [9]	Ümraniye General Surgery Clinic, Istanbul (10-yr EHR)	n=691 postoperative <i>E. coli</i> isolates	Clinical domain selection based on postoperative metadata	Demographics, patient age, surgery ward, previous antibiotic use	Random forest; CatBoost AUROC: 0.99 for some antibiotics SHAP global feature rankings	10-fold cross-validation on test partitions	Multidrug resistance (MDR) classification in postoperative <i>E. coli</i>
Lee et al. (2024) [10]	Multi-centre cohort, Hong Kong (teaching and community hospitals)	n=26,087 adult UTI patients (77.6% train [2015-19], 22.4% test [2020])	Multivariable regression selection + clinical expert consensus	105 variables: prior UTI history, previous drug susceptibility	RF AUROC: 0.941 (NIT), 0.939 (SXT), 0.937 (CIP). SHAP XAI	Prospective temporal validation on independent 2020 cohort	Pathogen susceptibility to NIT, SXT, CIP, LVX
Khorasani et al. (2024) [12]	Isfahan University clinical laboratory database	n=150 <i>P. mirabilis</i> CAUTI catheter strains	Wavelet-based feature extraction and radiomics selection	Radiomics features, microscopic morphological metrics	RF ACC: 0.95, Spec: 0.97, Sens: 0.88, AUC: 0.98	5-fold cross-validation	Biofilm-forming capacity & antibiotic resistance profiles

Study	Data Source	Sample Size	Feature Selection	Predictors	Model/ AUROC / SHAP	Validation	Outcome
Mengsir i et al. (2025) [13]	Thatphano-m Crown Prince Hospital, Thailand	n=624 MDR-GNB patients (80/20 train/test split)	RFECV, SFS, SelectKBest, and Variance Threshold	41 features: prior drugs, length of stay, devices, ward	Ensemble AUROC: 0.933, TabNet AUROC: 0.883. SHAP local XAI	5-fold cross-validation with GridSearch hyperparameter tuning	Carbapenem-resistant and MDR-GNB carriage prediction
Pan et al. (2025) [14]	Huai'an No. 1 People's Hospital, Nanjing Med Univ	n=1,385 MDR Klebsiella pneumoniae infected patients	LASSO regression and Boruta feature selection	99 variables: demographic, SO ₂ , pH, HCO ₃ ⁻ , CRP, PCT, BUN	XGBoost AUROC: 0.913, RF: 0.931. SHAP feature importances	Internal 10-fold cross-validation on test partition	Septic shock risk and hemodynamic deterioration
Almalk-i et al. (2025) [20]	Stanford Dryad ARMD EHR database (2025)	n=1,000 cohort (Merged dataset: 4,982,642 rows) (80/20 split)	Delta feature exclusion (>99% missing); clinical reduction	11 comorbidity domains, medication history, median vitals	XGBoost, LightGBM, and RF AUROC: 0.902. SHAP XAI	Independent 20% holdout test partition validation	Bacterial susceptibility to beta-lactams and fluoroquinolones
Rich et al. (2022) [21]	UF Health Integrated Data Repository, Florida	n=9,990 clinics cohort EHR records (2-decade database)	Univariable screening followed by backward stepwise elimination	Sex, UTI history, catheterization, renal disease, dementia, HTN	Moderate BLR AUROC: 0.58-0.66. No SHAP implemented	K-fold cross-validation with caret package	Predicting SXT, NIT, CIP, and multi-drug resistant UTIs
Yang et al. (2023) [22]	Mass General and Brigham & Women's Hospital, Boston	n= 101,096 samples complicated and uncomplicated UTI cohorts	Feature selection based on prior exposures and EHR covariates	Prior antibiotic resistance, prior exposure windows, comorbidities	XGBoost AUROC: 0.686–0.793 across NIT, SXT, CIP, and LVX; TabNet: 0.662	Temporal validation across clinical cohorts	NIT, SXT, CIP, and LVX non-susceptibility
Choi et al. (2024) [23]	Gangnam Severance Hospital, Yonsei, Seoul	Dev: n=259,187; Ext Val: n=62,255 adults	10-parameter reduced clinical model feature selection	Automated urinalysis, daily CBC (monocyte, WBC, CRP, vitals)	XGBoost UTI AUROC: 0.967, UT-BSI: 0.955. SHAP plots	External validation on independent secondary cohort	Urine culture positivity and secondary bloodstream infection

Methodological Considerations for Ensemble Models in Biomarker Integration

Evaluation of Ensemble Machine Learning Models

Extreme Gradient Boosting, an advanced version of Gradient Boosted Decision Trees (GBDT), has emerged in several reviewed studies as a strong-performing architecture for high-dimensional clinical integration in the prediction of antimicrobial resistance. According to systematic review results, the GBDT model has a mean AUROC of 0.80; however, the performance range is consistent at 0.77-0.90 for a real-world setting [24]. For example, a study was conducted on the development of a prognosis prediction model for urosepsis patients using the MIMIC-IV (Medical Information Mart for Intensive Care) database. In the study, the developed model showed the highest performance with an AUROC of 0.904; the model was able to capture complex interactions among variables such as urine output, age, and biochemical markers compared to other models such as RF and logistic regression [5]. This capability is further validated in comparative clinical trial studies in which XGBoost showed

better prediction capability in resistance to co-amoxiclav and piperacillin/tazobactam compared to conventional clinical evaluation and five-factor risk score approaches [25].

The Random Forest architecture, although robust in design, has a lower mean AUROC score of 0.75 with a broader range from 0.58 to 0.98 in different clinical studies [24]. Despite this variance, large-scale measure confirm that RF performs significantly better than logistic regression in approximately 69% of binary classification tasks [26]. In studies predicting *E. coli* resistance and urosepsis encounters, RF models achieved median accuracy and F1-scores of approximately 0.90, outperforming Naive Bayes and neural networks. While RF can reach a peak AUROC of 0.90 specific narrow phenotypes, such as ampicillin and ertapenem resistance, it remains more susceptible to performance variance when integrated with broad EHR data [9]. However, recent advancements in deep learning have introduced competitive alternatives, Graph Neural Networks (GNNs), which have outperformed individual RF models by incorporating topological features of patient-healthcare worker interactions, achieving an AUROC of 0.93 compared to the 0.86 achieved by RF in similar cohorts [27].

While logistic regression has the advantage of being more robust against overfitting, especially for smaller data sets such as the prediction of septic shock from MDR *Klebsiella pneumoniae* infections, larger data sets have consistently shown the superiority of the ensemble methods (RF and XGBoost) for the prediction of septic shock and other rare outcomes such as MDR infections [6].

Even though the ensemble machine learning methods have shown considerable advantages compared to other algorithms, XGBoost showed comparatively better-performing method for the management of MDR urosepsis prediction because of the use of L1 and L2 regularization for the prevention of overfitting, especially for high-dimensional datasets; however, its performance may vary across datasets and modelling strategies. Perhaps the most critical advantage of the XGBoost method, however, has been the demonstration of the highest recall rate of 88% for resistant cases, ensuring that such infections are addressed promptly and the catastrophic consequence of failure to diagnose resistance is avoided [12]. In a direct comparison using the Antibiotic Resistance Microbiology Dataset, XGBoost produced the better performance in terms of AUROC of 0.902, outperforming the Random Forest model, which produced a performance of 0.876 [5]. Furthermore, the GBDT model is technically more appropriate to capture the complex relationships between inflammatory markers like PCT and CRP, making the model a good engine for data-driven precision medicine in diagnosis and antibiotic stewardship.

Performance Metrics across Tabular Clinical Manifolds

Ensemble learning models outperform Deep neural networks (NNs) on the structured, tabular manifolds that are characteristic of clinical datasets. There exists evidence that, while deep learning techniques are known to perform exceptionally well on unstructured data, such as images that display smooth spatial patterns, NNs are generally found to be overparameterized, lacking the specific inductive biases that are needed to converge to the optimal solution on tabular datasets [22]. This problem is further compounded by the fact that NNs are “black-box” models, meaning that they are not transparent about how they operate, unlike the tree-based structures that are characteristic of ensemble learning techniques [28]. Moreover, the heterogeneous features and irregular target functions of tabular clinical data are difficult for neural networks to learn. Neural networks have rotational invariance and an inherent bias towards learning over smooth functions. This limits their performance on features with unique semantic meaning, like age and biomarker levels. In contrast, tree-based ensemble models have been successful in learning this task because they have independent feature learning, preserving the orientation of the input data to learn the piecewise constant decision boundaries characteristic of antimicrobial resistance [29]. This same independent feature learning also make these models are naturally robust to uninformative features and multicollinearity, allowing them to fit the “rough” patterns typical of healthcare datasets where NNs often fail to converge or generalize.

Several studies verify this claim of superiority through comparative analysis of the paradigms, and most of them report the better performance of GBDT and RF in comparison to logistic regression and neural networks on structured electronic health records data. While accuracy, sensitivity, and specificity are commonly reported, the AUROC is commonly cited as the most important evaluation metric for these machine learning-based models [30]. Therefore, XGBoost and RF may perform competitively in some medium-sized clinical datasets, it is important to assess their performance against other alternative methods in the specific population and prediction task.

Feature Selection: Boruta and LASSO in Biomarker Integration

The use of dimensionality reduction techniques, such as LASSO regression and the Boruta algorithm, plays a vital role in the identification of high-impact predictors for high-dimensional urosepsis data. LASSO regression helps in the identification of vital parameters by reducing the coefficients of less informative variables to zero, which helps avoid multicollinearity and overfitting. In a similar term, the use of Lasso shrinkage has been successful in identifying high-value predictors for electronic health records for the estimation of the risk of antibiotic resistance for hospitalized patients [31]. In comparative evaluations of variable selection methods for omics data, Boruta was identified as the most powerful and sensitive approach for detecting causal variables [32]. Consequently, by using SHAP to interpret the results, we can see that the model heavily relies on specific biomarker concentrations, with high starting PCT levels and the PAR frequently ranking as the top predictors. During the process validation of the XGBoost model, removal of prior antibiotic resistance data and biomarkers resulted in the reduction in predictive performance of AUPRC (Area Under the Precision-Recall Curve), indicating that historical resistance data constitutes the most influential predictor MDR Enterobacteriaceae [22]. Therefore, by integrating these selected features, ensemble models bridge the diagnostic gap and facilitate data-driven precision antibiotic stewardship.

Synergistic Biomarkers: The Raw Signal for ML Models

The recognition of systemic inflammation is fundamental for to urosepsis management, as the worsening of a simple urinary tract infection to life-threatening organ failure is primarily being facilitated by the host's pro-inflammatory cascade. While cultures remain the gold standard for definitive diagnosis, their inherent time delay necessitates the use of rapid laboratory biomarkers to provide a “raw signal” for both clinical decision-making and machine learning integration. Among these biomarkers, CRP and PCT have been the most widely used biomarkers for the diagnosis of bacteriuria and rule-out testing in ICU settings. PCT has shown an AUROC of 0.884 for the detection of bacteriuria, while CRP has shown significantly higher sensitivity (90.1%) and negative predictive value (93.0%) compared with PCT (83.1% and 90.2%, respectively). This indicates the potential of CRP as an effective biomarker for rule-out testing [33]. However, taking into consideration the equivalent diagnostic capabilities of the two markers within the general population, clinical consensus has been reached that the simultaneous detection of the two markers is more advantageous than the individual detection of the two markers. This enhances the diagnostic accuracy significantly and allows clinical interventions up to 48 hours before the culture results.

Apart from these standard markers, PAR has appeared as a unique and reliable independent biomarker for the distinction of urosepsis from acute pyelonephritis. The ideal clinical cutoff value is >0.44 , which shows a sensitivity of 84.6% and a specificity of 96.0%. The exceptionally high specificity is of considerable value for the proper use of antibiotics as it helps in the accurate stratification of high-risk patients and avoids the misuse of broad-spectrum antibiotics in low-risk patients [34]. Moreover, PCT is an essential factor in predicting the occurrence of bacteremia and the load of bacteria in pyelonephritis, and it is observed that a PCT of ≤ 0.25 $\mu\text{g/L}$ is adequate to exclude bacteremia [35]. Furthermore, studies carried out in tertiary healthcare centers observed that 97.1% of patients with urosepsis have PCT levels of more than 2.0 ng/mL, thus proving the reliability of PCT as an indicator of bacterial load [36].

In order to complement these biochemical markers, the SOFA score has commonly been utilized for assessing the extent of organ dysfunction in patients with sepsis, providing a quantitative measure for assessing the progression and resolution of disease. The incorporation of PCT levels with the SOFA score has the potential for improving the prediction accuracy for therapeutic interventions in uroseptic shock patients [37]. These biomarkers serve as a highly precise raw signal for advanced machine learning models, which include gradient boosting decision trees that have utilized Postoperative PAR as a major predictive feature for model development, with model AUROC values as high as 0.90. The SHAP analysis has identified these features as part of the top five predictors for urosepsis, providing a highly transparent and clinically relevant model for assessing disease progression [27]. This integration with ML frameworks allows for the precision medicine paradigm, which enables the move from the empirical standardized approach to a personalized risk stratification approach, thus reducing the diagnostic delay.

Clinical Utility and Decision Support Systems

Real-Time Risk Categorization

The clinical utility of machine learning improves the prediction and management of urosepsis by converting complex information from the electronic health record data into a risk stratification framework that may provide earlier risk stratification before conventional culture results become available and could potentially inform empiric antimicrobial decision-making. These frameworks categorize clinical isolates by the probability of non-susceptibility to a specific antimicrobial agent, based on predetermined probability thresholds. The inclusion of a low-risk category ensures a 20% false negative rate, which safeguards the initial empiric therapies. Patients with resistant isolates have a significantly higher probability of being classified as high risk, which provides a robust warning for clinicians. Such models may have potential implications for antimicrobial stewardship and resource utilization; however, their effects on treatment decisions, healthcare costs, and patient outcomes require prospective evaluation [34]. The predictive capability for outbreak detection provides a positive impact on the reduction of economic costs and patient survival. By using real-time data, it allows for accurate and specific diagnosis and treatment planning in both community and hospital setting.

Interpretability via SHAP

In order to encourage the use of clinical models and develop trust among healthcare practitioners, it is vital to address the problem of the “black box” that exists with complex algorithms, which is achieved with the help of advanced interpretability tools like SHAP. The values generated by SHAP help explain the clinical model by showing how each clinical feature, like the Procalcitonin/Albumin ratio, age, and urine output, contributes to a particular prediction made by the model. Global interpretability is achieved with the help of summary plots that show how each clinical feature contributes, on average, to the model, highlighting postoperative Procalcitonin/Albumin ratio and Interleukin-6 as key contributors for urosepsis risk [38]. This also suggests that clinical and non-clinical factors have a significant influence on the prediction made by the model for Enterobacteriaceae colonization. The use of SHAP allows healthcare practitioners to understand the rationale for high-risk scores by relating them to relevant biological variables and informing hemodynamic optimization strategies.

Quantifying Impact on Antimicrobial Stewardship

The current global pandemic of antimicrobial resistance is a direct result of diagnostic delays, which in turn lead to the overuse of broad-spectrum antimicrobials, thus highlighting the need to improve the current antimicrobial stewardship program. However, the addition of machine learning-based systems significantly strengthens this program by reducing the misuse of empirical antimicrobial agents and moving towards a more personalized rather than standardized antimicrobial prescribing program. Constrained machine learning allows the healthcare system to minimize the rate of mismatched treatment by over 30%, while maintaining the target distribution of drug usage [23]. Specific algorithmic applications, such as the XGBoost model, have further demonstrated the potential to decrease the over-prescribing of broad-spectrum agents by over 41% while maintaining rigorous safety levels for high-risk patients [24]. Apart from the elimination of over-prescribing, the predictive tools also act as decision aids that incorporate optimized second-line agents, such as fluoroquinolones, and have shown better therapeutic selection compared with conventional practices [22]. In addition, the predictive tools also reduce the overall unnecessary treatment with antibiotics by 18%, while the coverage rates are maintained at the same level. The benefits are also evident, especially for vulnerable groups, with the predictive tools allowing for the escalation of treatment for 20.83% of pediatric and complex urinary tract infections, ensuring that high-risk patients receive the appropriate treatment up to 48 hours earlier than conventional culture methods [39]. Furthermore, this is supported by the value of these interventions through the use of a decision curve analysis, which shows that machine learning-informed approaches provide a greater net benefit within a certain range of probabilities compared to universal prophylaxis [4]. Thus, these developments improve patient safety on the individual level and resource allocation on the institutional level through a precision-based framework for the management of infectious diseases.

Discussion

Methodological Limitations and Reporting Standards

The combination of ensemble machine learning strategies with inflammatory biomarkers marks a big improvement over the traditional scoring systems, but a thorough assessment of evidence shows that important methodological limitations still persist. The methods used in the studies included in Table 2, is mostly based on electronic health record data, which brings the risk of selection bias and incomplete data. In most cases, the methods used for handling missing biomarker variables are often inadequate and poorly elaborated, since any wrong imputation will lead to the detection of data leakage during the process of tuning. Moreover, several methods lack appropriate external validation, obtaining a maximum estimate of forecasting accuracy only through means of internal cross-validation. To convert in to practical clinical application of these predictive models in future research, standardized principles of reporting should be strictly adhered to TRIPOD-AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis–Artificial Intelligence) and PROBAST-AI (Prediction model Risk of Bias Assessment Tool for AI) with the focus on transparency of the calibration and the validity issues of the models being developed.

Biomarker Heterogeneity and the Cutoff challenges

Inflammatory biomarkers such as CRP, PCT, and the PCT/Albumin ratio provide significant raw signals, their clinical interpretation is very complex. It is hard to determine just one perfect cut-off point for these biomarkers because biological processes do not shift radically and instead are seen as a risk continuum absolutely. Therefore, several retrospective datasets will lead to totally different cut-offs depending on the type of cohort. Moreover, CRP and albumin levels are influenced by multiple clinical confounders, including hepatic function, nutritional status, fluid resuscitation, and non-infectious systemic inflammation. There are many variations in the hepatic function, nutritional status, fluid resuscitation volume, and non-infectious systemic inflammation. It is important to take all the confounding variables into consideration.

Variation in ML Architectures and Performance Metrics

The assertion that tree ensemble models always outperform neural networks when it comes to tabular clinical data is not straightforward. Although XGBoost and Random Forest are indeed very useful tools, recent developments in deep learning models for tabular data like TabNet and TabPFN have proven to have great performance that incorporates decision tree-like elements into neural networks. Additionally, most of the evaluation of those models focuses too much on AUC and accuracy. In MDR Enterobacteriaceae urosepsis case, where the significant class imbalance affects evaluation (multidrug resistant cases represent a smaller class), those metrics may not be useful. So rigorous evaluation must prioritize calibration curves, Positive and Negative Predictive Values (PPV/NPV), and threshold-dependent decision metrics that account for prevalence dependence to ensure clinical safety.

Rethinking Explainability in AI

The application of explainable AI (XAI), in particularly SHAP, is frequently oversimplified in clinical literature. Recent appraisals argue that XAI research is often built on ill-defined objectives, warning that popular explanation methods are frequently incapable of answering causal clinical questions. In high-dimensional data, SHAP value may attribute high importance to variables that are not direct causal relationship with urosepsis or resistance, but confounded with them via indirect statistical associations, collider effects, or suppressor variables. while SHAP provides transparency into the mathematical functioning of the model, the identified features may be scientifically irrelevant, necessitating cautious clinical interpretation.

Clinical Deployment and Regulatory Governance

The ultimate utility of these predictive algorithms hinges on overcoming significant deployment hurdles. Retrospective high performance must be substantiated through prospective clinical trials and decision curve analysis to quantify the true net clinical benefit compared to standard universal prophylaxis. Furthermore,

integration into live clinical workflows requires robust AI governance to foster clinician trust and mitigate alert fatigue. From a regulatory perspective, navigating the deployment of AI as a Software as a Medical Device (SaMD) requires adherence to emerging legal frameworks, such as the EU AI Act (European Union Artificial Intelligence Act), which classifies medical AI as high-risk. Similarly, rigorous clinical evaluation may require alignment with strict investigational frameworks, such as navigating FDA regulations under 21 CFR 312.40 for investigational new drugs (IND) if the AI directly guides the administration of unapproved therapeutic regimens or novel investigational dosing strategies. Thus, to overcome existing translational gap requires a multidisciplinary approach that integrate data science, clinical pharmacology, and regulatory compliance.

Conclusion

Current evidence indicates that ensemble machine learning methods, including XGBoost in several reviewed studies, show promise for the early prediction and risk stratification of multidrug-resistant Enterobacteriaceae urosepsis through integration of clinical, laboratory, and relevant biomarker data. However, their performance varies across datasets and modelling techniques, and the predominantly retrospective evidence is insufficient to establish clinical utility or improved patient outcomes. Prospective multi-center validation and clinical trials are therefore necessary to determine generalizability and make these models applicable in practice for routine antimicrobial decision-making process and potential optimization of individualized patient care.

List of Abbreviations: MDROs - Multidrug-resistant organisms; *E. coli* - *Escherichia coli*; UTIs - Urinary tract infections; ESBL - Extended-spectrum beta-lactamase; MDR - Multidrug-resistant ; ICU - Intensive Care Unit; KPC - *Klebsiella pneumoniae* carbapenemase; NDM - New Delhi metallo-beta-lactamase; CAUTIs - Catheter-associated urinary tract infections; CRP - C-Reactive Protein; PCT - Procalcitonin; PAR - Procalcitonin/Albumin ratio; SOFA - Sequential Organ Failure Assessment; ML - Machine learning; XGBoost - Extreme Gradient Boosting; RF - Random Forest; NIT- Nitrofurantoin; SXT - sulfamethoxazole-trimethoprim; CIP - Ciprofloxacin; LVX - levofloxacin; LASSO - Least Absolute Shrinkage and Selection Operator; SHAP - SHapley Additive exPlanations; CatBoost - Categorical Boosting; LightGBM - Light Gradient Boosting Machine; EHRs - Electronic health records; GBDT - Gradient Boosted Decision Trees; AUROC - Area Under the Receiver Operating Characteristic; MIMIC - Medical Information Mart for Intensive Care; GNNs - Graph Neural Networks; ROC-Receiver Operating Characteristic; NNs - Neural networks; AUPRC - Area Under the Precision-Recall Curve; TRIPOD-AI - Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis–Artificial Intelligence; PROBAST-AI - Prediction model Risk of Bias Assessment Tool for Artificial Intelligence; PPV/NPV - Positive and Negative Predictive Values; XAI - Explainable Artificial Intelligence; EU AI Act - European Union Artificial Intelligence Act

Author Contributions: All the authors were involved in the design, searching strategy, literature review, manuscript drafting, manuscript preparation and revision.

Funding: This research received no funding.

Ethics Statement: The authors are accountable for all aspects of the work and ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Data Availability Statement: Not applicable.

Acknowledgments: None

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

References

1. Sarwal V, Rakocz N, Dominique G, Chiang JN, Ackerman AL. Machine learning for the prediction of urosepsis using electronic health record data. *PLOS Digit Health*. 2025;4(7):e0000896. <https://doi.org/10.1371/journal.pdig.0000896>.
2. Wei Y, Xu W, Yang S, Zhang C, Wang J, Wan X. Significant adverse prognostic events in patients with urosepsis: a machine learning based model development and validation study. *Front Cell Infect Microbiol*. 2025;15:1623109. <https://doi.org/10.3389/fcimb.2025.1623109>.

3. Wagenlehner FM, Lichtenstern C, Rolfes C, Mayer K, Uhle F, Weidner W, Weigand MA. Diagnosis and management for urosepsis. *Int J Urol*. 2013;20(10):963-70. <https://doi.org/10.1111/iju.12200>.
4. Zhao W, Sun P, Li W, Shang L. Machine Learning-Based Prediction Model for Multidrug-Resistant Organisms Infections: Performance Evaluation and Interpretability Analysis. *Infect Drug Resist*. 2025;18:2255-2269. <https://doi.org/10.2147/IDR.S459830>.
5. Ku JH, Bruxvoort KJ, Salas SB, Varley CD, Casey JA, Raphael E, et al. Multidrug Resistance of *Escherichia coli* From Outpatient Uncomplicated Urinary Tract Infections in a Large United States Integrated Healthcare Organization. *Open Forum Infect Dis*. 2023;10(7):ofad287. <https://doi.org/10.1093/ofid/ofad287>.
6. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat Rev Microbiol*. 2015;13(5):269-84. <https://doi.org/10.1038/nrmicro3432>.
7. Al-Khlifeh EM, Alkhazi IS, Alrowaily MA, Alghamdi M, Alrashidi M, Tarawneh AS, et al. Extended Spectrum beta-Lactamase Bacteria and Multidrug Resistance in Jordan are Predicted Using a New Machine-Learning system. *Infect Drug Resist*. 2024;17:3225-3240. <https://doi.org/10.2147/IDR.S469877>.
8. Goodman KE, Lessler J, Cosgrove SE, Harris AD, Lautenbach E, Han JH, et al. A Clinical Decision Tree to Predict Whether a Bacteremic Patient Is Infected With an Extended-Spectrum β -Lactamase-Producing Organism. *Clinical Infectious Diseases*. 2016;63(7):896–903. <https://doi.org/10.1093/cid/ciw425>.
9. Tolan HK, Aydın İ, Tanyildizi-Kokkulunk H, Karakuş M, Akkaya Y, Kaya O, et al. Machine Learning Model for Predicting Multidrug Resistance in Clinical *Escherichia coli* Isolates: A Retrospective General Surgery Study. *Antibiotics*. 2025;14(10):969. <https://doi.org/10.3390/antibiotics14100969>.
10. Lee ALH, To CCK, Chan RCK, Wong JSH, Lui GCY, Cheung IYY, et al. Predicting antibiotic susceptibility in urinary tract infection with artificial intelligence-model performance in a multi-centre cohort. *JAC Antimicrob Resist*. 2024;6(4):dlac121. <https://doi.org/10.1093/jacamr/dlae121>.
11. Zhou X, Yang M, Chen F, Wang L, Han P, Jiang Z, et al. Prediction of antimicrobial resistance in *Klebsiella pneumoniae* using genomic and metagenomic next-generation sequencing data. *J Antimicrob Chemother*. 2024;79(10):2509-17. <https://doi.org/10.1093/jac/dkac248>.
12. Khorasani A, Moghim S, Wagemans J, Lavigne R, Mirzaei A. Antibiotic profile classification of *Proteus mirabilis* using machine learning: An investigation into multidimensional radiomics features. *Comput Biol Med*. 2024;182:109131. <https://doi.org/10.1016/j.combiomed.2024.109131>.
13. Mengsiri P, Ungcharoen R, Gertphol S. A machine learning and neural network approach for classifying multidrug-resistant bacterial infections. *Healthcare Analytics*. 2025;7:100388. <https://doi.org/10.1016/j.health.2025.100388>.
14. Pan S, Shi T, Ji J, Wang K, Jiang K, Yu Y, et al. Developing and validating a machine learning model to predict multidrug-resistant *Klebsiella pneumoniae*-related septic shock. *Frontiers in Immunology*. 2025;15: 1539465. <https://doi.org/10.3389/fimmu.2024.1539465>.
15. Pennisi F, Pinto A, Ricciardi GE, Signorelli C, Gianfredi V. The Role of Artificial Intelligence and Machine Learning Models in Antimicrobial Stewardship in Public Health: A Narrative Review. *Antibiotics*. 2025;14(2):134. <https://doi.org/10.3390/antibiotics14020134>.
16. Moffa L, Tana C, Ucciferri C, Moffa S, Jacopo Vecchiet, Falasca K. URO-RESIST: A Real-World Retrospective Study on Multidrug-Resistant Sepsis with Urinary Tract Infection, Clinical Predictors, Inflammatory Biomarkers, and Patient Outcomes. *Antibiotics*. 2025;14(10):1036–6. <https://doi.org/10.3390/antibiotics14101036>.
17. Luo X, Yang X, Li J, Zou G, Lin Y, Qing G, et al. The procalcitonin/albumin ratio as an early diagnostic predictor in discriminating urosepsis from patients with febrile urinary tract infection. *Medicine*. 2018;97(28):e11078. <https://doi.org/10.1097/MD.00000000000011078>.
18. Bassetti M, Kanj SS, Kiratisin P, Rodrigues C, Van Duin D, Villegas MV, et al. Early appropriate diagnostics and treatment of MDR Gram-negative infections. *JAC Antimicrob Resist*. 2022 Sep 13;4(5):dlac089. <https://doi.org/10.1097/10.1093/jacamr/dlac089>.
19. Guliciuc M, Porav-Hodade D, Chibelea BC, Voidazan ST, Ghirca VM, Maier AC, et al. The Role of Biomarkers and Scores in Describing Urosepsis. *Medicina (Kaunas)*. 2023;59(3):597. <https://doi.org/10.3390/medicina59030597>.

20. Almalki Z, Althagafi A, Al-Shareef S. Predicting Antimicrobial Resistance Using Machine Learning on Electronic Health Records: A Comparative Study of Ensemble Models. *Research Square* 2025;preprint. <https://doi.org/10.21203/rs.3.rs-7529240/v1>.
21. Rich SN, Jun I, Bian J, Boucher C, Cherabuddi K, Morris JG Jr, Prosperi M. Development of a Prediction Model for Antibiotic-Resistant Urinary Tract Infections Using Integrated Electronic Health Records from Multiple Clinics in North-Central Florida. *Infect Dis Ther.* 2022 Oct;11(5):1869-1882. <https://doi.org/10.1007/s40121-022-00677-x>.
22. Yang J, Eyre DW, Lu L, Clifton DA. Interpretable machine learning-based decision support for prediction of antibiotic resistance for complicated urinary tract infections. *NPJ Antimicrob Resist.* 2023;1(1):14. <https://doi.org/10.1038/s44259-023-00015-2>.
23. Choi MH, Kim D, Park Y, Jeong SH. Development and validation of artificial intelligence models to predict urinary tract infections and secondary bloodstream infections in adult patients. *J Infect Public Health.* 2024;17(1):10-17. <https://doi.org/10.1016/j.jiph.2023.10.021>.
24. 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. <https://doi.org/10.1371/journal.pone.0319460>.
25. Moran E, Robinson E, Green C, Keeling M, Collyer B. Towards personalized guidelines: using machine-learning algorithms to guide antimicrobial selection. *J Antimicrob Chemother.* 2020;75(9):2677-2680. <https://doi.org/10.1093/jac/dkaa222>.
26. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics.* 2018;19(1):270. <https://doi.org/10.1186/s12859-018-2264-5>. & Gouareb R, Bornet A, Proios D, Pereira SG, Teodoro D. Erratum to "Detection of Patients at Risk of Multidrug-Resistant Enterobacteriaceae Infection Using Graph Neural Networks: A Retrospective Study". *Health Data Sci.* 2023;4:0216. <https://doi.org/10.34133/hds.0216>. Erratum for: *Health Data Sci.* 2023;3:0099. <https://doi.org/10.34133/hds.0099>.
27. Gouareb R, Bornet A, Proios D, Pereira SG, Teodoro D. Detection of Patients at Risk of Multidrug-Resistant Enterobacteriaceae Infection Using Graph Neural Networks: A Retrospective Study. *Health Data Sci.* 2023;3:0099. <https://doi.org/10.34133/hds.0099>. Erratum in: *Health Data Sci.* 2023;4:0216. <https://doi.org/10.34133/hds.0216>.
28. Molnar C. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. Leanpub; [online] 2019 [accessed on 20 January 2026]. Available from: URL: <http://leanpub.com/interpretable-machine-learning>
29. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? *arXiv.org.* 2022. <https://doi.org/10.48550/arXiv.2207.08815>
30. Sakagianni A, Koufopoulou C, Feretzakis G, Kalles D, Verykios VS, Myrianthefs P, et al. Using Machine Learning to Predict Antimicrobial Resistance—A Literature Review. *Antibiotics (Basel).* 2023;12(3):452. doi: 10.3390/antibiotics12030452.
31. Kim C, Park RW, Rhie SJ. Development of a Machine Learning Prediction Model to Select Empirical Antibiotics in Patients with Clinically Suspected Urinary Tract Infection using Urine Culture Data. *Open Forum Infect Dis.* 2021;8(1):S196. <https://doi.org/10.1093/ofid/ofab466.376>.
32. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics.* 2019;20(2):492-503. <https://doi.org/10.1093/bib/bbx124>.
33. Zhang GM, Guo XX. Combining PCT with CRP is better than separate testing for patients with bacteriuria in the intensive care unit: a retrospective study. *Eur J Med Res.* 2024;29(1):441. <https://doi.org/10.1186/s40001-024-02036-7>.
34. Shields RK, Cheng WY, Kalé Kponee-Shovein, Indacochea D, Gao C, Kuwer F, et al. Development of Predictive Models to Inform a Novel Risk Categorization Framework for Antibiotic Resistance in *Escherichia coli*—Caused Uncomplicated Urinary Tract Infection. *Clin Infect Dis.* 2024;79(2):295-304. <https://doi.org/10.1093/cid/ciae171>.

35. Van Nieuwkoop C, Bonten TN, van't Wout JW, Kuijper EJ, Groeneveld GH, Becker MJ, et al. Procalcitonin reflects bacteremia and bacterial load in urosepsis syndrome: a prospective observational study. *Critical Care*. 2010;14(6):R206. <https://doi.org/10.1186/cc9328>.
36. Reddy VGK, Uma MA, Jagadish KL, Reddy VK, Kumar KA. Utility of Serum Procalcitonin and Serum Albumin Ratio as Predictors in Discriminating Urosepsis from Patients with Febrile Urinary Tract Infection. *International Journal of Pharmacy Research & Technology*. 2025;15(1):531-541.
37. Hang Z. Predictive value of procalcitonin for the therapeutic response of patients with uroseptic shock: a retrospective case-control study. *Am J Transl Res*. 2025;17(2):992-1004. <https://doi.org/10.62347/VFFF7133>.
38. Wang Z, Li X, Li Q, Zhou R, Pan D, Su Z, et al. Leveraging laboratory biomarkers to predict urosepsis after upper urinary tract stone surgery: an explainable machine learning approach. *BMC Med Inform Decis Mak*. 2025;26(1):27. <https://doi.org/10.1186/s12911-025-03314-y>
39. Pan F, Han P, Wu Q, Hou W, Rao G, Ma Z, et al. Rapid prediction of antibiotic resistance in *Enterobacter cloacae* complex using whole-genome and metagenomic sequencing. *mSystems* 10:e00584-25. <https://doi.org/10.1128/msystems.00584-25>.