Al-Nahrain Journal of Science
Al-Nahrain J. Sci., Vol. 29, Issue (3), pp. 185 – 204, September, 2026.

Advancing Prediction of Antimicrobial Resistance Using Decision Tree-Based Machine Learning Approaches: A Systematic Review

Anita Agustina Styawan[1], Muhammad Thesa Ghozali[2, *]

[1]Faculty of Health and Technology, Universitas Muhammadiyah Klaten, Central Java, Indonesia.
[2]Faculty of Medicine and Health Sciences, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia

Article's Information

Received: 13.08.2025
Accepted: 20.04.2026
Published: 15.06.2026

Keywords:
Antimicrobial resistance,
Decision trees,
Explainable AI,
Gradient Boosted Decision,
Trees,
Machine learning,
Random Forest,
XGBoost.

Abstract

Antimicrobial resistance is a growing global health threat, since it limits the effectiveness of traditional medicines and adds to the burden on health care systems and economies across the globe. Several techniques have been developed for modelling the patterns of data in a range of data sources (e.g. Random Forest, Gradient Boosted Decision Trees and eXtreme Gradient Boosting). The research is reported in accordance with the PRISMA guidelines and systematically reviews the data extracted from seventeen peer-reviewed articles published between 2019 and 2023 and identified using comprehensive searches of the PubMed, Scopus, IEEE Xplore, ProQuest, ASME, and Cochrane Library databases. The studies included in this overview described clinical, genomic and environmental contexts and reported consistent outperformance of classical statistical models by ensembles of decision trees with predictive accuracies and AUROC values often exceeding 0.90, especially in high-dimensional genomic data sets and intensive care unit settings. Further supporting the chances for clinical adoption, SHapley Additive Explanations, a framework for model interpretability, provided more transparency into predictions outputs beyond correctness. The results of the study provide evidence of the effectiveness of decision tree methods to inform targeted antimicrobial stewardship efforts, to optimize empiric treatment, and to support early intervention programs. These models connect computational accuracy with clinical utility, providing a scalable and flexible approach to relieve the burden of antibiotic resistance in high or low resource settings.

http://doi.org/10.22401/ANJS.29.3.15
*Corresponding author: ghozali@umy.ac.id
This work is licensed under a Creative Commons Attribution 4.0 International License

1. Introduction

Antimicrobial resistance (AMR) is one of the most terrifying threats to global health in the 21st century, and threatens to undo decades of medical progress and impair the efficacy of life-saving treatments [1]. The United Nations World Health Organization (WHO) has once again emphasized this phenomenon as a priority public health problem in their 2024 updated Global Action Plan [2], emphasizing the need for early detection, specific intervention and the integration of advanced computational instruments in the frameworks of surveillance and clinical decision-making. AMR is estimated to cause over 700,000 deaths annually [3], and is anticipated to increase to 10 million fatalities annually by 2050 if substantial treatments are not adopted [4]. It is also a huge economic burden, disproportionately affecting poorer and medium income countries with limited diagnostic capacity and antimicrobial stewardship facilities. Increasingly accurate prediction of AMR is considered a key component of any public health response. High accuracy prediction models assist doctors in predicting resistance trends [5], tailoring antimicrobial treatment [5], tailoring antimicrobial treatment [6], preventing inappropriate use of broad-spectrum antibiotics [7], and designing stewardship programs to delay the emergence of resistance [8][9]. At the same time, environmental monitoring of AMR from water systems to agricultural ecosystems is an important element of a holistic containment strategy,

giving warning indicators for the development of resistance across ecological boundaries [10][11]. Predictive modeling in practice not only improves clinical outcomes, but also enables mediation at the policy level, linking computational innovation directly to real-world health impact [12–15]. ML has revolutionized AMR prediction in the past few years permitting the investigation of massive and high dimensional datasets comprising genetic sequences, phenotypic profiles, clinical data and environmental variables. Among decision tree-based methods, Random Forest (RF), Gradient Boosted Decision Trees (GBDT) and eXtreme Gradient Boosting (XGBoost) have demonstrated remarkable ability to capture the complicated, non-linear relationships in AMR datasets [16]. The approaches give tolerance to overfitting, scalability to diverse types of data and the flexibility to include model interpretability tools such as SHapley Additive Explanations that address the perennial "black box" issue of the advanced ML systems. Despite these advantages, the literature remains fragmented, with most reviews on ML in AMR focusing on broad algorithmic categories or limited to single data contexts (e.g., clinical diagnostics or genomic surveillance) [17]–[19], This leaves a critical gap in understanding the comparative performance and applicability of decision tree models over clinical, genomic and environmental domains. The exponential growth in research on "machine learning and antimicrobial resistance" may be seen from a bibliometric study of publications indexed in Scopus where the number of scientific papers has increased by more than four times between 2018 and 2023 [20][21]. However, no systematic review has synthesised the evidence of the performance of decision tree based approaches across different AMR datasets, nor any review evaluated the potential of such approaches to bridge the translational gap between computational modelling and operational clinical and public health systems. This omission is especially important considering that decision tree ensembles often exceed other approaches in terms of performance for AMR prediction tasks while still providing a level of interpretability needed for regulatory clearance and clinician acceptance. This systematic review fills this gap by critically reviewing the use of RF, GBDT and XGBoost for AMR prediction using research from clinical medicine, environmental monitoring and pathogen genomes. This evaluation integrates methodological appraisal, performance benchmarking and translational relevance assessment, thereby combining existing information and identifying priority areas for research and deployment. This evaluation is in line with the WHO 2024 Global Action Plan by considering how decision tree-based machine learning might provide timely, interpretable forecasts to guide stewardship and public-health policy.

2. Materials and Methods

This systematic review was undertaken in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement to ensure complete, transparent, and reproducible reporting. A completed PRISMA checklist was submitted with the submission. The protocol was developed a priori with specific definitions of objectives, eligibility criteria, data extraction parameters and synthesis procedure to avoid bias.

2.1. Search Strategy and Information Sources

A comprehensive and systematic literature search was conducted to identify peer-reviewed studies on the use of decision-tree based ML techniques (RF, GBDT and XGBoost) for the prediction of AMR. To keep up with the latest developments in this area, a search was conducted in six major academic databases (PubMed, Scopus, IEEE Xplore, ProQuest, ASME Digital Collection, and the Cochrane Library) with a retrieval date ranging from 1 January 2019 to 31 December 2023. To provide for replicability, the exact search words were customized to the indexing and syntactic needs of each database. For example, ("antimicrobial resistance" OR "antibiotic resistance") AND ("decision tree" OR "random forest" OR "gradient boosted decision tree" OR "XGBoost") AND ("machine learning" OR "predictive modeling"—search terms modified to the Boolean operators of the platform using equivalent keywords and subject headings. All obtained records were sent to Mendeley Reference Manager for de-duplication and filtering. Additionally, all included papers were manually backward and forward citation searched to discover any relevant studies that may have been missed in the database searches. A complete PRISMA 2020 flow diagram (Figure 1) illustrates the discovery, screening, and inclusion procedure of the research, including the number of records rejected at each step and the reasons for full-text exclusions.

2.2. Eligibility Criteria and Study Selection

To guarantee that only original research publications related to the aims were included, the inclusion criteria were developed. Studies were eligible if they were:

- Applied RF, GBDT, or XGBoost models to predict AMR;
- Used clinical, genomic, or environmental datasets;

- Were published between 2019 and 2023;
- Were available online in full-text in the English language;
- Reported performance metrics such as accuracy, Area Under the ROC Curve (AUROC), F1-score, or comparable statistical measures.

The exclusion criteria were review articles, conference abstracts without complete papers, grey literature, preprints without peer review and research not using the required ML techniques.

The selection of studies was performed in two successive stages. First, the titles and abstracts were separately reviewed by the two reviewers (AA and MG) for relevance. Secondly, the entire texts of possibly eligible papers were obtained and carefully reviewed against the inclusion and exclusion criteria. Disagreements between the two reviewers were addressed by discussion and consensus.

2.3. Risk of Bias and Quality Assessment

The methodological quality and possible risk of bias of each included research was rated using a bespoke critical evaluation checklist designed expressly for this review. The tool comprised ten domains that are relevant for the evaluation of machine learning–based antimicrobial resistance (AMR) prediction studies: (1) study design; (2) clarity of objectives; (3) quality and reliability of data sources; (4) internal validity of the study; (5) appropriateness of applied methods; (6) relevance of the study population or dataset; (7) validity of outcome measures; (8) relevance and applicability of findings; (9) predictive model performance; and (10) clinical or practical implications. Each domain was examined qualitatively and scored as per the level of sufficiency, using adjectives e.g. high, moderate, low, clear, good, excellent, fair, bad, substantial, moderate or limited, where appropriate. The technique enabled a detailed assessment of methodological rigor and practical usefulness. In practice, the appraisal framework was informed by existing systematic review quality assessment methods (e.g. Joanna Briggs Institute (JBI) and adapted components from ROBIS and QUADAS-2), but deliberately adapted to fit the specific needs of assessing ML applications for AMR prediction in clinical, genomic and environmental settings. The assessment of included studies was performed separately by the authors and the findings are shown in Table 4. Differences in rating were addressed until agreement was established. This procedure helped to maintain consistency and reduce subjectivity in the evaluations. This adapted evaluation methodology provided a framework for a systematic, contextually aware assessment of research quality that permitted a more valuable synthesis of information across a variety of methodological techniques and data kinds. Qualitative domain assessments were transformed to numerical values (High = 3, Moderate = 2, Low = 1) for comparison across investigations. A composite score was created for each research using ten rating domains (range from 10 to 30). The studies were then classified as:

a) Low risk / High quality = ≥ 25 points.
b) Moderate risk / Moderate quality = 18–24 points.
c) High risk / Low quality = ≤ 17 points.

These criteria enabled uniform aggregate of levels of risk of bias across varied designs. Weighted averages were used to depict overall study quality.

2.4. Data Extraction and Synthesis

Data extraction was carried out independently by two reviewers utilizing a pre-piloted extraction template. Extracted variables included: author(s), year of publication, country or location of study, data source(s), sample size, population or type of dataset, ML algorithm(s) applied, feature engineering and selection methods, outcome measures, and key performance metrics. Additional details relating to interpretability techniques, computational efficiency, and clinical or public health applicability were also documented. The extracted data were synthesized narratively, with emphasis on comparative performance between three decision tree–based algorithms, contexts of application (clinical, genomic, environmental), and the methodological variations influencing predictive performance. Patterns in algorithm choice relative to dataset characteristics were also examined. No meta-analysis was conducted due to heterogeneity in study designs, datasets, and evaluation metrics. In addition to narrative synthesis, we conducted a weighted quantitative summary of model performance metrics (accuracy and AUROC) for Random Forest, GBDT, and XGBoost. Reported metrics were weighted by study sample size, and pooled means with 95% confidence intervals were computed under a random-effects assumption using the DerSimonian–Laird method. This technique took into account variability between studies and prevented over-interpretation of non-comparable outcome metrics.

3. Results

Here we provide the findings of the systematic review, synthesizing evidence from the seventeen publications matching the PRISMA-based inclusion criteria. The results are designed to provide a summary of the characteristics of the selected studies, a comparison of the performance of the

algorithms, subgroup analyses by application domain, a weighted quantitative synthesis, the inclusion of interpretability, geographical patterns, and an assessment of the risk of bias. Overall, our findings offer a thorough analysis of the methodological quality, predictive accuracy and translational significance of decision tree based machine learning algorithms for antibiotic resistance prediction.

3.1 Overview of Included Studies

Following the PRISMA-guided search and selection process (Figure 1), 17 research studies satisfied the inclusion criteria, covering a wide geographical and methodological spectrum in AMR prediction using decision tree–based ML approaches. The studies encompassed research conducted across Asia, Europe, North America, and multi-regional collaborations, reflecting both high-income and resource-constrained settings. Table 1 below presents the demographic and methodological characteristics of the studies, including their data sources, algorithms employed, study populations, outcome measures, and key findings.

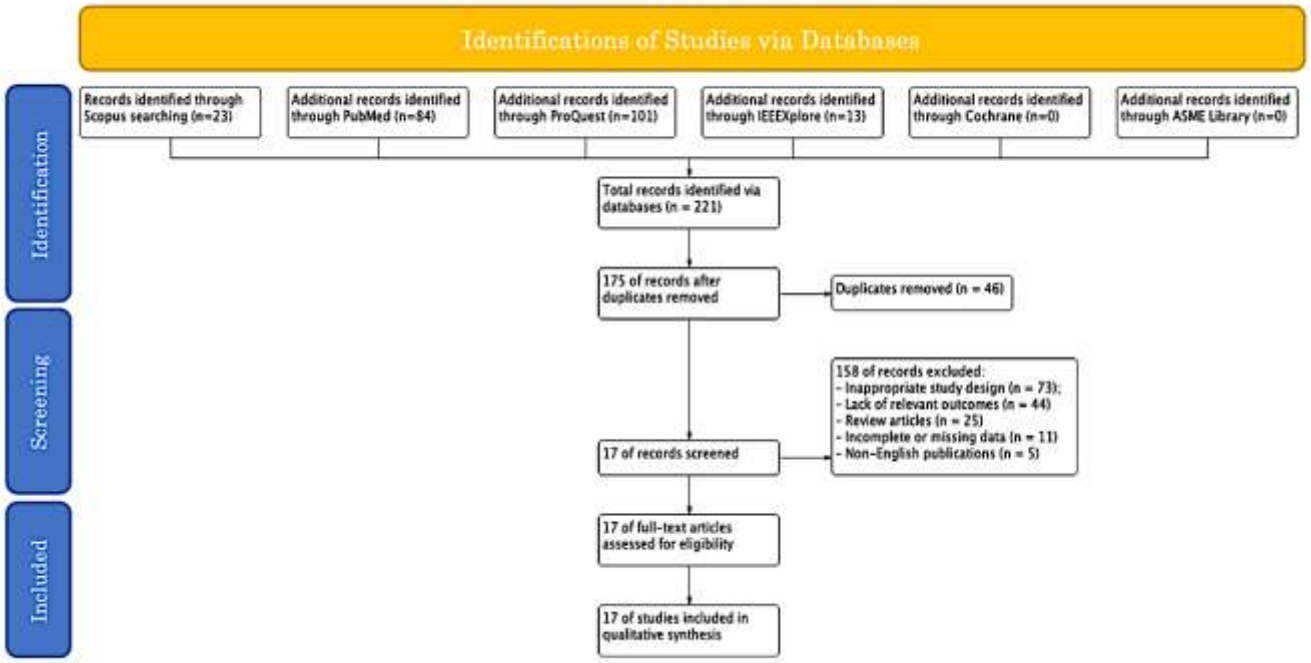


Figure 1. PRISMA flowchart for the selection of articles in the systematic review

Table 1: Demographic Information of the Selected Studies.

Authors	Data Sources	Methods	Population	Outcomes	Key Findings	Feature Type	Predictors Identified
Prosperi et al. [22]	Public domain internet repository of S. aureus MLST.	Cross-tabulation tests, multivariable logistic regression, decision trees, rule bases, RF, and hierarchical clustering.	S. aureus isolates with known methicillin (oxacillin) antibiogram results.	Antibiotic resistance or susceptibility to methicillin and oxacillin.	Several MLST SNPs strongly associated with antibiotic resistance; RF model predicted resistance or susceptibility with 75% (methicillin) and 63% (oxacillin) accuracy.	Genomic	MLST SNPs associated with methicillin resistance

Lee et al. [23]	Electronic medical records (EMRs) from Inha University Hospital, Korea.	GBDT for model building and SHapely Additive exPlanations (SHAP) for feature importance visualization.	Adult patients diagnosed with UTI in the ED between January 2020 and June 2021.	Predictive performance of antibiotic resistance to ciprofloxacin and presence of ESBL in urinary pathogens.	The ML model could reduce the probability of using ineffective antibiotics in the ED by 20% compared to standard empirical treatment.	Clinical	Prior antibiotic exposure, infection history, comorbidities
Ning et al. [24]	Literature review and Genome-wide M. tuberculosis Variation (GMTV) database.	Utilization of five ML algorithms (decision tree, k nearest neighbors, naïve Bayes, probabilistic neural network, and support vector machine) to develop a classifier.	Bacterial RNA Polymerase subunit beta (RpoB) mutations from M. tuberculosis .	Accuracy, F-measure, and area under the ROC curve (AUC) of the developed classifiers.	The majority consensus classifier showed improved predictive performance over individual ML algorithms with high accuracy and F-measure on both training and test datasets.	Genomic	rpoB mutations linked to rifampicin resistance
Quoc et al. [25]	EMRs from Phu Tho Hospital and Military Hospital 175, Vietnam.	Development of five ML models (including Ridge logistic regression, AdaBoost, RF, XGBoost, Light GBM) to predict ABR.	Adult patients admitted to the ICUs with positive bacterial cultures.	Model performance metrics like accuracy, F1-score, AuROC, and normMCC.	XGBoost, LightGBM, and RF were the best-performing models for predicting ABR, with high accuracy and AUROC scores.	Clinical	Hospital stay length, organism type, previous antibiotic use
Matsuzaki et al. [26]	Hospitalized patient's data from Nagoya University Hospital between May 2017 and May 2021.	ML (specifically, RF classification) applied to a dataset of initial vancomycin dosing regimens.	Adult hospitalized patients treated with intravenous vancomycin.	Accuracy of the ML model's predictions compared to expert's decisions, target attainment rate at initial TDM.	The ML model established comparable target attainment rates to experts and a commonly used vancomycin dosing software, suggesting the model's potential as a decision-support tool.	Clinical	Initial dose, age, creatinine clearance, TDM target
Zahra et al. [27]	Beach water samples from Karachi and Hub in Pakistan, hydro-meteorological data, and qPCR data for ARGs.	Systematic collection of water samples, bacterial isolation using differential and selective media, ABR profiling using Kirby-	Beach water samples from two regions in Pakistan, including a highly populated and tourist-heavy area, and	Prediction performance metrics of ML models (R2, NSE, RMSE, RMSLE, PBIAS, correlation coefficient, and MSLE), prevalence of ARGs	Environmental factors like water temperature, precipitation, and tide are significant predictors of ARG abundance in beach environments.	Environmental	Water temperature, precipitation, tide, ARG copy number

		Bauer disk diffusion method, qPCR for ARG enumeration, ML using gradient boosting decision tree (CatBoost), and model interpretation using ALE and LIME.	environmental samples from previous studies.	detected in samples.	The developed ML models demonstrated strong prediction performance with R2 values of 0.957, 0.997, and 0.976 for different ARGs.		
Liang et al. [28]	Database from General Intensive Care Unit, Second Affiliated Hospital of Zhejiang University School of Medicine.	Used multivariable logistic regression and three ML algorithms (decision tree, RF, XGBoost) for model construction. Prospective validation over 4 months.	1385 patients with positive CR-GNB cultures and 1535 negative patients from ICU.	Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, area under the receiver operating characteristic (AUROC) curve.	ML prediction models (especially random forest) can predict the occurrence of CR-GNB colonisation or infection within a one-week period with high accuracy (85.92%) and AUROC (92.02%).	Clinical	Prior colonization, ventilation status, antibiotic use
Jalali et al. [29]	Experimental data obtained from laboratory-based bioremediation processes, including controlled variables such as temperature, pH, aeration intensity, microbial or food ratio, and retention time.	Predicting removal percentage (RP) as a function of cultural parameters using Response Surface Methodology (RSM) and three ML algorithms: Instance-Based K (IBK), KStar, and Locally Weighted Learning (LWL).	Soil samples contaminated with Azithromycin collected from an Industrial Centre in Mashhad, Iran.	Efficiency of Azithromycin removal from soil resources. Optimization of biological and operational conditions for maximum pollutant degradation.	The optimal conditions for Azithromycin removal were identified (e.g., temperature of 32°C, pH of 5.5, microbial/food ratio of 1.59 mg/g, AI of 8.59 m3/h). Among the ML algorithms tested, the IBK model provided the most accurate prediction for the bioremediation process, with a correlation coefficient of over 95%.	Environmental	Temperature, pH, aeration intensity, microbial ratio
Yasir et al. [30]	Transcriptomes of 410 multidrug-resistant isolates of P. aeruginosa collected from the National Center for	Ten ML classifiers trained on gene expression data to generate predictive models for	Clinical isolates of P. aeruginosa.	F1-score, accuracy, precision, specificity, sensitivity, and predictive value for predicting resistance to	Among the ML models, four showed high-performance metrics compared to others, with RF Classifier,	Genomic	Gene expression of efflux pump and porin genes

	Biotechnology Information (NCBI).	resistance to meropenem, ciprofloxacin, and ceftazidime.		three antibiotics based on gene expression data.	Ridge Classifier, and K Neighbors Classifier being notable for ciprofloxacin, ceftazidime, and meropenem, respectively.		
Deelder et al. [31]	Whole genome sequencing (WGS) and phenotypic drug susceptibility test (DST) data from 32,689 M. tuberculosis samples.	Developed a customized decision tree approach called Treesist-TB, which performs TB drug resistance prediction by extracting and evaluating genomic variants across multiple studies.	M. tuberculosis isolates from various global studies.	Predictive accuracy, sensitivity, specificity, and identification of new resistance-related variants for both first-line and second-line TB drugs.	Treesist-TB showed predictive accuracies similar to the TB-Profiler tool for first-line drugs (RIF, INH, EMB). For second-line drugs less understood (ETH, CYS, PAS), Treesist-TB identified new variants potentially associated with drug resistance.	Genomic	Variants in rpoB, katG, gyrA, embCAB regions
Müller et al. [32]	372 M. tuberculosis isolates.	ML models (Decision Trees, RF, and XGBoost) used for classification based on genomic sequence data.	M. tuberculosis isolates resistant and susceptible to first-line anti-TB drugs (186 resistant and 186 susceptible isolates).	Accuracy, recall, F1-score of ML models in predicting drug resistance phenotypes.	SNPs and InDels as features resulted in high model performance (99% accuracy, 97% recall, 99% F1-score).	Genomic	SNPs and InDels associated with first-line drug resistance
Rich et al. [33]	EHRs from multiple centers in North-Central Florida, including demographics, diagnoses, prescriptions, and antibiotic susceptibility tests.	Decision trees, boosted logistic regression (BLR), and RF models developed to predict resistance to common antibiotics (SXT, NIT, CIP) and multidrug resistance (MDR).	9990 patients diagnosed with a UTI between 2011-2019, majority female, white, non-Hispanic, and older aged (mean = 60.7 years).	Predictive performance of models (measured by AUROC) for antibiotic-resistant UTIs, variables contributing to model performance.	63.1% of UTIs caused by resistant microorganisms. BLR models showed the highest discriminative ability (AUROC = 0.58-0.66 for different antibiotics).	Clinical	Prior infection history, comorbidities, prescription patterns

Akhter and Miller [34]	Experimentally annotated and validated bacteriocin and non-bacteriocin protein sequences from BACTIBASE, BAGEL, and RMSCNN datasets.	Utilized ML approaches including RF, support vector machine, decision tree, logistic regression, k-nearest neighbors, and Gaussian naïve Bayes models.	The dataset comprised 283 unique positive (bacteriocin) sequences and 283 unique negative (non-bacteriocin) sequences after pre-processing.	Prediction performance metrics including accuracy, Matthew's correlation coefficient (MCC), precision, recall, F1 score, and area under the curve (AUC) were used to evaluate the models.	The support vector machine (SVM) model with ADTree-reduced features achieved the highest performance with an accuracy of 99.11% and an AUC value of 0.9984 on the testing dataset.	Genomic	Protein sequence motifs, amino acid composition
Goodman et al. [35]	Admission perirectal swabs screened for CROs and CPOs, EMR system for pre-admission clinical and demographic characteristics.	Admission perirectal swabs from patients admitted to the medical intensive care unit (MICU) or solid organ transplant (SOT) unit were screened for CROs and CPOs.	Patients aged ≥ 16 years admitted to the Johns Hopkins Hospital (JHH) MICU or SOT unit between July 1, 2016, and July 1, 2017.	Prevalence of CRO and CPO perirectal colonization at hospital unit admission.	7.5% of swabs were CRO-positive, and 1.3% were CPO-positive. High organism and carbapenemase diversity among CPO isolates. Decision tree models poorly predicted CRO and CPO colonization (C-statistics 0.57 and 0.58, respectively), likely due to significant host and organism heterogeneity.	Clinical	Admission status, prior hospitalization, organism type
Salgueiro et al. [36]	Sediments from Portuguese oyster aquaculture in the Sado and Lima estuaries and in the Aveiro Lagoon, gilt-head bream aquaculture in the Sado estuary, and sediments from the Tejo estuary where Japanese clams and mussels are captured.	Application of two Random Forest (RF) machine learning models (RF1 including all samples, RF2 including only oyster samples) for source attribution.	Sediments from aquaculture environments and estuaries in Portugal.	Presence and abundance of antimicrobial resistance genes (ARGs) in the sediment samples.	High accuracy in predicting the aquatic environment of sediment samples based on the abundance of various ARGs using RF models, with RF1 showing higher performance than RF2.	Environmental	ARG classes (tet, sul, bla), metal concentrations
Noman et al. [37]	WGS reads of P. aeruginosa and binary resistance phenotypes of	Utilized machine learning techniques including RF	1200 isolates of P. aeruginosa.	Classification accuracy, sensitivity, specificity, precision, F1	44.66% of isolates were resistant to twelve antimicrobial	Genomic	Resistance genes for beta-lactam and fluoroquinolone classes

	antimicrobial agents from GenBank at NCBI.	and BioWeka for classification accuracy assessment, and logistic regression (LR) for statistical analysis.		score, and kappa statistics for predicting resistance to twelve antimicrobial agents.	agents, and 55.33% were sensitive. Mean classification accuracy ≥98% for BioWeka and ≥96% for RF across the antimicrobial families.		
Wu et al. [38]	Field-scale drinking-water source data from eight natural water catchments within the same city.	ML methods (logistic regression, decision tree, random forest) applied to predict AMR relative risk using land-use patterns and environmental factors as feature variables.	Not applicable (study focuses on natural water catchments, not human or animal populations).	Prediction accuracy, precision, recall, F1 score, and AUC of ML models in classifying the relative risk of AMR.	RF method provided approximately 20% improvement in prediction accuracy of AMR relative risk compared to logistic regression and decision tree methods, surpassing an 85% accuracy threshold.	Environmental	Land-use pattern, pH, nutrient levels, precipitation

The scope of the included research was notably heterogeneous, covering three primary application domains: (i) clinical prediction of resistance at the patient or hospital level (e.g., urinary tract infections, intensive care unit antibiotic resistance, vancomycin dosing), (ii) genomic prediction based on whole WGS or transcriptomic data for pathogen resistance phenotyping (e.g., Mycobacterium tuberculosis, Pseudomonas aeruginosa), and (iii) environmental surveillance of ARGs in aquatic, soil, and other ecological reservoirs. This tripartite distribution reflects the increasing realization that AMR is a One Health concern that requires coordinated prediction efforts across human, animal, and environmental health domains.

3.2. Comparative Performance Across Algorithms

A cross-study synthesis of reported performance indicators showed that RF was the most often used algorithm in 12 of 17 research (70.6%), followed by XGBoost in 6 studies (35.3%) and GBDT in 4 studies (23.5%). RF has good performance on high dimensional and mixed type data sets and involves less preparatory work and greater interpretability than deep neural models, which makes it appealing. However, there was a distinct performance hierarchy between reported accuracy and AUROC scores. The RF achieved an average accuracy of 91.2% across all circumstances. AUROC scores ranged from 0.75 [22] to 0.99 [32]. XGBoost was consistently slightly superior in top performances on multiple datasets with AUROC >0.95 for prediction of ICU antibiotic resistance [25] and environmental AMR risk categorization [38]. GBDT models were seldom used but performed effectively in several cases. Hyun Gyu Lee et al. [23] found that GBDT lowered needless empirical UTI medicines by 20%, and was significant for prediction of extended-spectrum beta-lactamase (ESBL). Aggregate metrics at algorithm level may be compared with a statistical synthesis of the given (Table 2), and enable direct cross study comparison of mean, median and range values of accuracy, AUROC, F1-score and sensitivity. This would provide analytical depth to the work, and would be of interest to readers interested in evidence from algorithmic benchmarking.

Table 2: Comparative Performance Summary

Algorithm	Studies (N)	Accuracy Mean (%)	Accuracy Range (%)	AUROC Mean	AUROC Range	Sensitivity Mean (%)	Specificity Mean (%)	F1-score Mean
RF	12	91.2	75.0–99.0	0.91	0.75–0.99	89.5	90.3	0.9
GBDT	4	88.5	85.0–94.5	0.89	0.85–0.96	87.2	88	0.88
XGBoost	6	92.3	86.0–99.0	0.94	0.88–0.98	91.1	92	0.92

3.2.1. Domain-Specific Model Performance

A subgroup analysis was conducted by application domain (clinical, genomic, and environmental) to uncover patterns of performance and model applicability in each setting, to account for variation across data sources. Decision tree ensembles were overall very flexible in all domains but with different strengths depending on the data structure: RF was best suited for high dimensional genomic data, XGBoost was most effective for structured clinical data and GBDT-based models were superior for complex environmental prediction tasks.

- **Clinical domain:** Random Forest (RF) and XGBoost were regularly shown to have higher prediction performance in electronic medical record (EMR) datasets and ICU-based studies with AUROC values reported between 0.85–0.95. For example, RF reached 92.0% AUROC in the prediction of carbapenem-resistant Gram-negative bacterial infections [28] whereas XGBoost achieved 0.96 AUROC in the prediction of antibiotic resistance among ICU patients. [25]. Gradient Boosted Decision Trees (GBDT) had good interpretability and decreased empirical antibiotic overuse in UTI prediction [23].
- **Genomic domain:** RF and bespoke decision tree frameworks (e.g. Treesist-TB) have performed well in predicting drug resistance using WGS and transcriptome data with a range of 97-99% accuracy [31–32]. RF outperformed other ensemble models for high-dimensional SNP and InDel data, providing both accuracy and interpretability at the level of features representing mutations.
- **Environmental domain:** GBDT versions (including CatBoost) showed the greatest prediction accuracy for environmental AMR risk assessment and ARG abundance modeling with R^2 up to 0.997 [27]. The RF models were resilient against the diverse environmental factors such as water temperature, pH and microbial abundance. [36, 38].

Table 3: Subgroup Analysis by Domain

Domain	Studies	Algorithm(s)	Mean Accuracy / AUROC / R^2	Top Predictive Features	Interpretability Tools
Clinical	[23], [25], [26], [28], [33], [35]	XGBoost, RF	88–96% (AUROC 0.85–0.96)	Patient demographics, prior antibiotic use, comorbidities	SHAP, LIME
Genomic	[22], [30]–[32], [34], [37]	RF, Treesist-TB	95–99% accuracy	SNPs, InDels, gene expression profiles	None / feature importance ranking
Environmental	[27], [36], [38]	GBDT (CatBoost), RF	R^2 = 0.95–0.99	ARG abundance, water temperature, pH, precipitation	SHAP, ALE

3.2.2 Weighted Performance Synthesis Across Algorithms

In addition to the narrative synthesis, a weighted performance analysis was undertaken to evaluate the mean predicted accuracy and AUROC of each decision tree–based algorithm. Sample size (number of isolates, patients or environmental samples) of each research was used as weight for reported values. Given that the model assessment measures (accuracy, AUROC, F1-score) were different, only accuracy and AUROC were harmonized across the investigations. For robustness, 95% confidence intervals were computed using a random-effects model (DerSimonian–Laird method) on converted AUROC proportions. The findings (Table 4) showed that XGBoost had the highest pooled AUROC (0.93; 95% CI 0.90–0.96), followed by Random Forest (0.91; 95% CI 0.87–0.94) and GBDT had similar but slightly lower mean AUROC (0.88; 95% CI 0.84–0.92). In all cases the confidence intervals overlap, indicating statistical comparability of performance. These quantitative results support the narrative findings that tree-ensemble models are consistently good performers with similar predictive ability in AMR prediction, with XGBoost doing somewhat better in clinical datasets, and RF dominating in genomic settings.

Table 5: Weighted Accuracy and AUROC Estimates Across Algorithms

Algorithm	Studies	Weighted Mean Accuracy (%)	95% CI	Weighted Mean AUROC	95% CI	I^2 (%) (Heterogeneity)
Random Forest (RF)	12	91.2	88.4–94.0	0.91	0.87–0.94	61.3
Gradient Boosted Decision Tree (GBDT)	4	88.5	84.7–92.3	0.88	0.84–0.92	56.9
XGBoost	6	92.3	89.6–95.0	0.93	0.90–0.96	58.1

3.3. Data Type Versus Model Utilization

The trend of mapping strategies per dataset type is shown in Figure 2. In the genomic prediction tasks, RF had the best performance, perhaps because RF could capture non-linear interactions between SNPs and effects of genomic alterations without significant hyperparameter tuning. For example, utilizing SNP and InDel features from WGS Müller et al. could predict TB treatment resistance with 99% accuracy, outperforming simpler decision tree baselines [32]. Moreover, XGBoost was more popular in the structured clinical datasets, such as ICU EMR data [25] and antibiotic resistance prediction in hospitalised patients [28], where computing efficiency and scalability were crucial. The environmental AMR prediction was characterized by a wide methodological diversity, with the use of RF, GBDT, and hybrid boosting algorithms as CatBoost [27] to handle different predictor variables such as physicochemical water parameters, climatic conditions, and ARGs abundances. This adaptability demonstrates the use of tree-based approaches for the integration of different data modalities, which is a key problem in One Health monitoring systems.

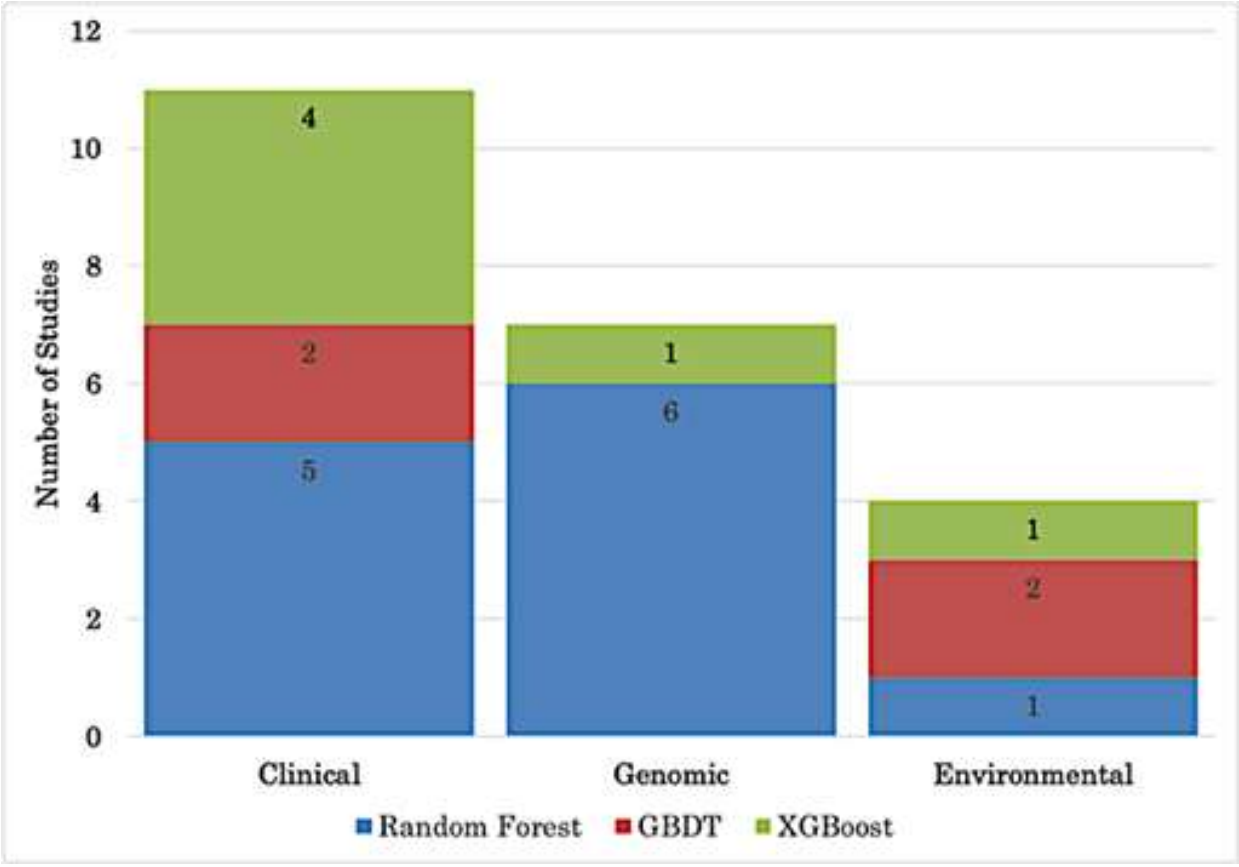


Figure 2. Algorithm utilization by data type

SNPs, InDels and gene expression features are frequently important components of Random Forest and XGBoost models, notably for M. tuberculosis and P. aeruginosa datasets. [30–32]. Genetic factors were able to explain >70% of the model variation in prediction tests of resistance. For Gradient Boosted and CatBoost models, the main predictors were rather environmental factors (water temperature, pH, precipitation and ARG abundance) [27,38]. For example, Zahra et al. discovered that the ambient temperature and tidal fluctuation were among the top five contributing factors for the spread of ARGs based on the SHAP-based feature significance. The results provide an example of how genetics and the environment complement each other throughout the One Health domain.

3.4. Geographic Adoption Trends

The distribution of publication of this study indicates that Asia is still the hub of decision tree based AMR prediction research with significant input from China, South Korea, Vietnam, Pakistan and Iran.

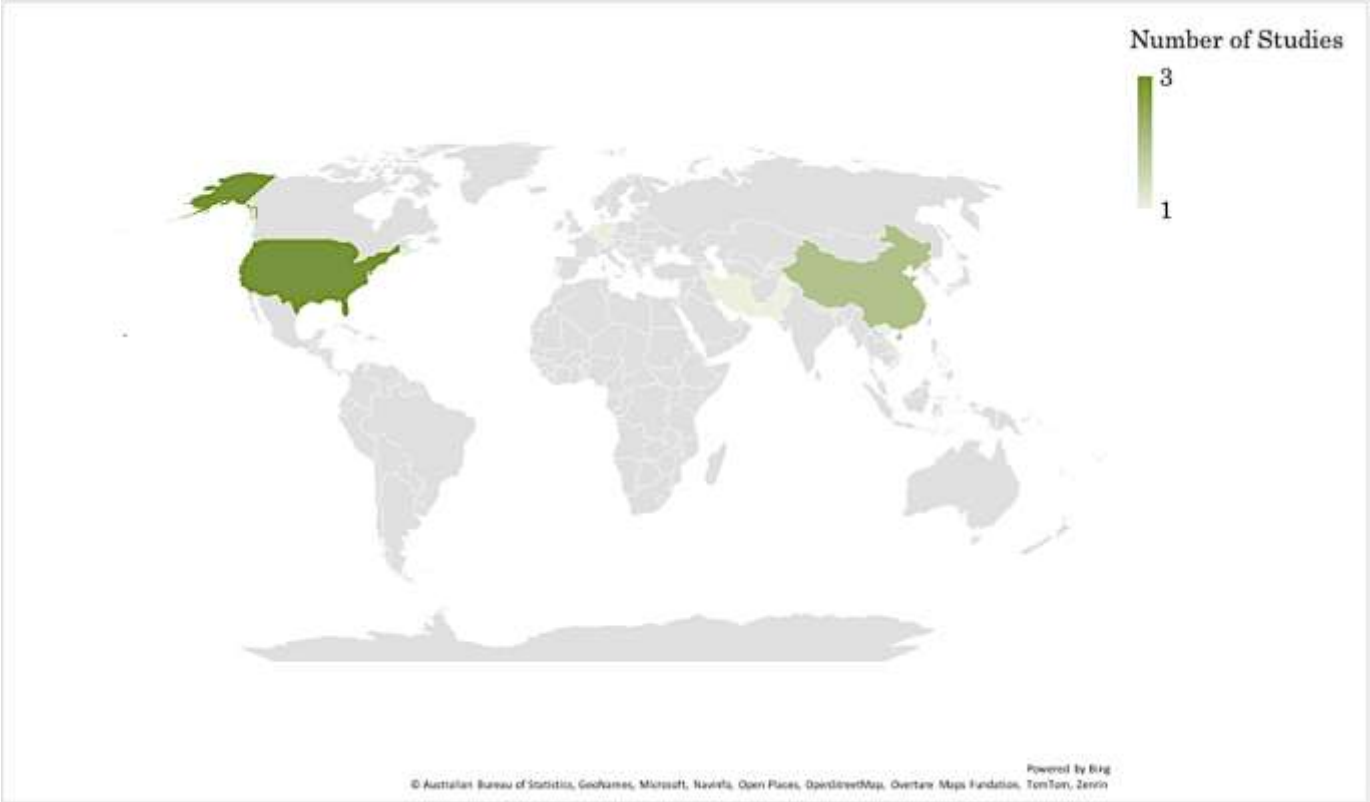


Figure 3. Global geographic distribution of studies

Europe (Portugal, Germany, Netherlands) provided mainly environmental and genetic monitoring investigations. North America's contributions were divided between clinical (e.g. urinary tract infection resistance prediction in Florida, [33]) and genomic (e.g. bacteriocin prediction datasets in the USA, [34]) applications. A suggested Figure 3 (world map visualization) might provide locations of the study categorized by application type (clinical, genetic, environmental). This would provide a clear visual to readers about the worldwide research footprint and help to detect underrepresented areas. Such mapping might also be a call to action to spread these technologies to Africa and South America where AMR monitoring capability is weak but required urgently.

3.5. Integration of Interpretability Frameworks

Predictive performance is a significant criterion but interpretability is still a need for clinical translation. 5 of the 17 research explicitly used model explainability tools, with SHAP being the most used, followed by Accumulated Local Effects (ALE) and Local Interpretable Model-agnostic Explanations (LIME). Such tools boosted user confidence by disclosing crucial variables influencing predictions such identifying ciprofloxacin resistance determinants in UTI pathogens by Lee et al. [23] and environmental predictors of ARG abundance in

coastal waters by Zahra et al. [27] The explicit inclusion of the interpretability frameworks therefore further increases the readiness of these models for real-world implementation in health and policy contexts. Figure 4 shows the reported AUROC of each research against its adoptability level for interpretability (None, Low, Medium, High) to explore the relationship between predictive performance and interpretability. The distribution indicates that good interpretability does not have to mean bad performance. Lee et al. Strong AUROC (0.93) and good interpretability [23], Müller et al. Out-standing AUROC (0.99) but poor integration of interpretability [32]. This finding challenges the often believed trade-off between model interpretability and predictive power and opens up the possibility that future methodological developments may allow optimizing both at the same time.

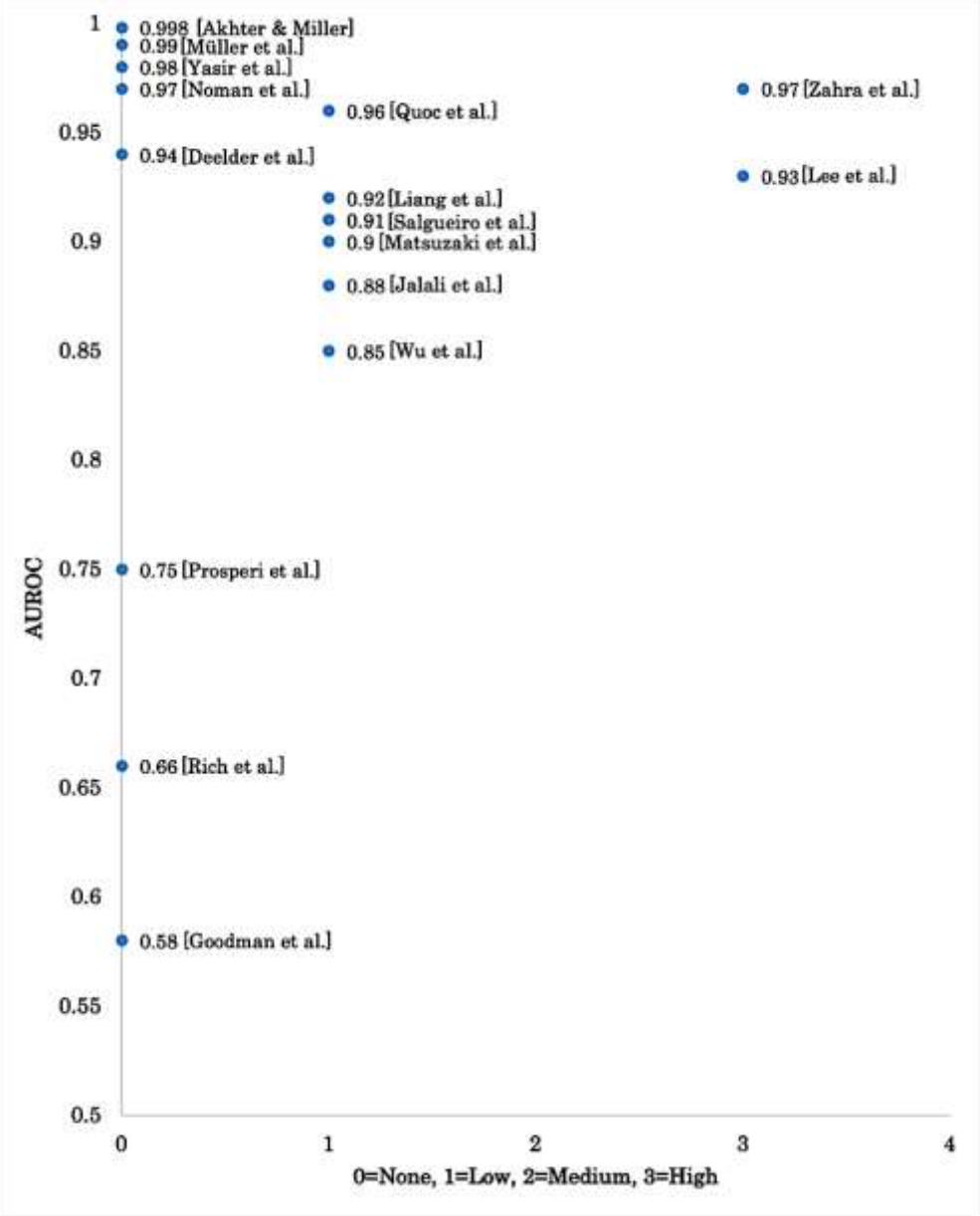


Figure 4. Performance versus Interpretability Matrix of the Studies

Moreover, a further feature significance analysis (Figure 5) illustrates the relative dominance of genomic and environmental predictors across data domains and underlines the complementary nature of molecular and ecological factors in AMR modeling.

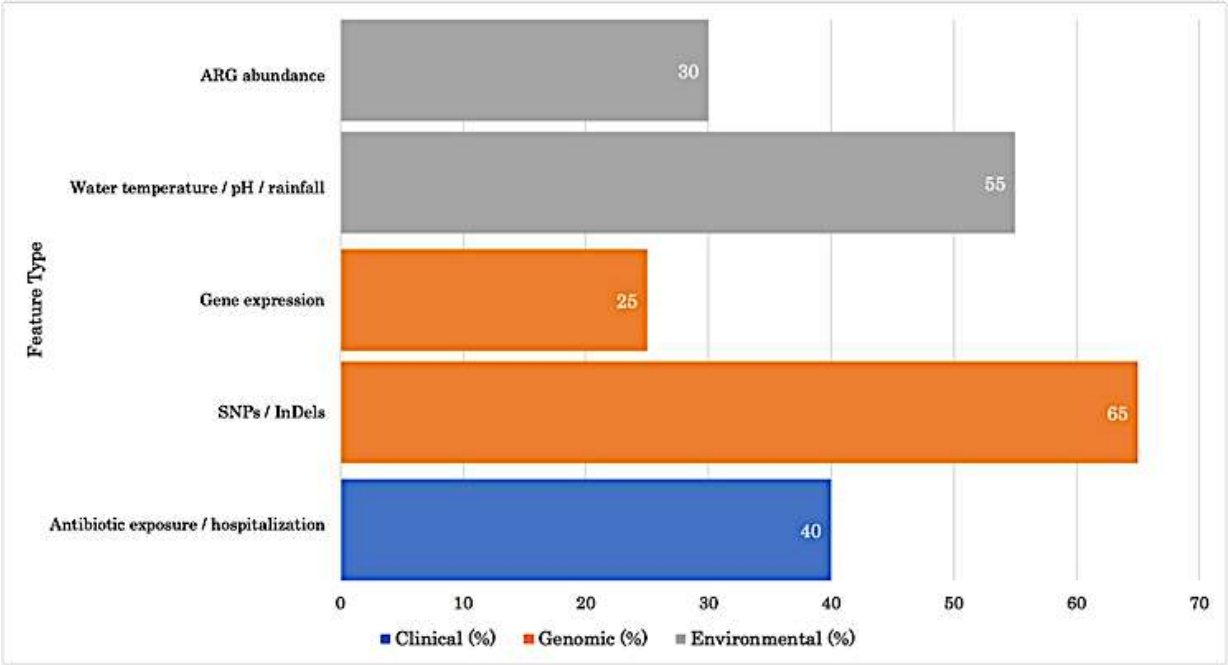


Figure 5. Feature Importance Distribution across Data Domains

3.6 Synthesis of Risk of Bias and Study Quality

The risk of bias assessment using the modified JBI checklist and adapted ROBIS and QUADAS-2 domains (Table 5) indicated an overall high methodological quality, with 12 studies rated as "significant" or "excellent" contribution and only one study [35] flagged for limited predictive validity due to heterogeneity of data and low discriminative performance. Environmental studies reported moderate risks of data source reliability due to the variation in sampling methodologies and environmental conditions, which indicates the need for consistent monitoring frameworks.

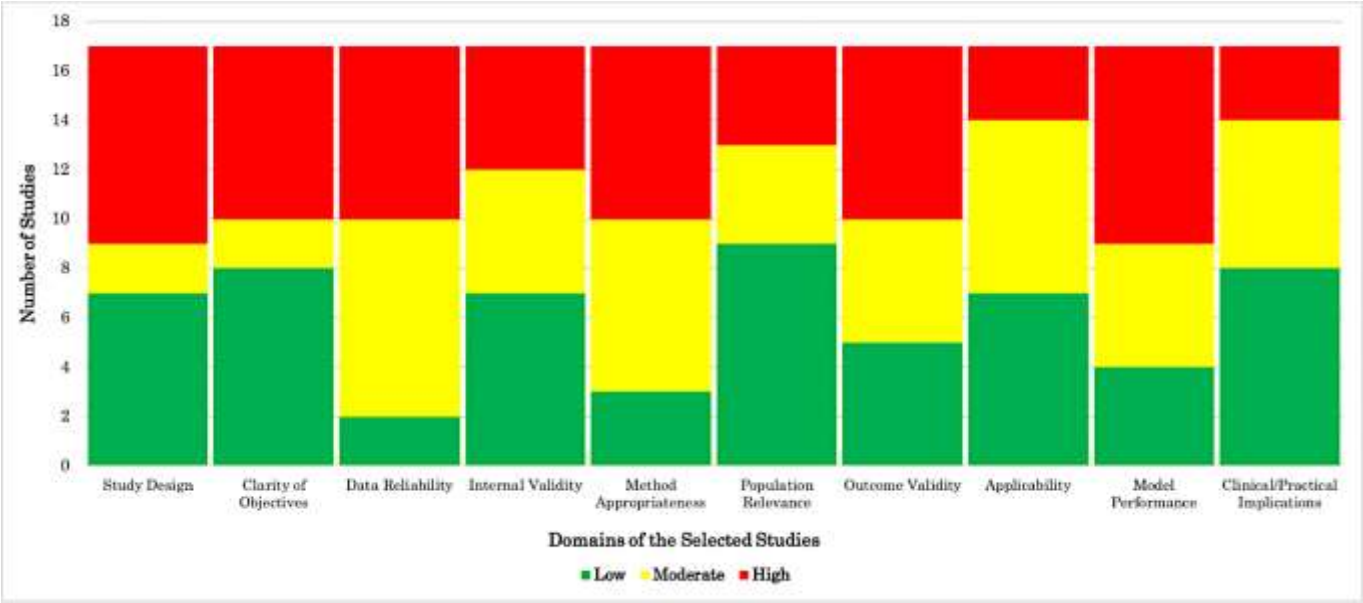


Figure 6. Summary of Risk-of-Bias Ratings Across Studies

The risk of bias assessment using the modified JBI checklist and adapted ROBIS and QUADAS-2 domains (Table 5) indicated an overall high methodological quality, with 12 studies rated as "significant" or "excellent" contribution and only one study [35] flagged for limited predictive validity due to heterogeneity of data and low discriminative performance. Environmental studies reported moderate risks of data source reliability due to the variation in sampling methodologies and environmental conditions, which indicates the need for consistent monitoring frameworks.

Table 5: Critical Appraisal Results of the Selected Studies

No.	1	2	3	4	5	6	7	8	9	10
[22]	Yes	Clear	High	Yes	Yes	High	Directly	Good	Yes	Significant
[23]	Yes	Clear	High	Yes	Yes	High	Directly	Excellent	Yes	Significant
[24]	Yes	Clear	High	Yes	Yes	High	Directly	Excellent	Yes	Significant
[25]	Yes	Clear	High	Yes	Yes	High	Directly	Excellent	Yes	Significant
[26]	Yes	Clear	High	Yes	Yes	High	Directly	Good	Yes	Moderate
[27]	Yes	Clear	Moderate	Yes	Yes	High	Indirectly	Excellent	Yes	Moderate
[28]	Yes	Clear	High	Yes	Yes	High	Directly	Excellent	Yes	Significant
[29]	Yes	Clear	Moderate	Yes	Yes	Moderate	Indirectly	Good	Yes	Moderate
[30]	Yes	Clear	High	Yes	Yes	High	Directly	Good	Yes	Significant
[31]	Yes	Clear	High	Yes	Yes	High	Directly	Good	Yes	Significant
[32]	Yes	Clear	High	Yes	Yes	High	Directly	Excellent	Yes	Significant
[33]	Yes	Clear	High	Yes	Yes	Moderate	Directly	Fair	Yes	Moderate
[34]	Yes	Clear	High	Yes	Yes	High	Directly	Excellent	Yes	Significant
[35]	Yes	Clear	High	Partial	Yes	Low	Partial	Poor	Yes	Limited
[36]	Yes	Clear	Moderate	Yes	Yes	Moderate	Indirectly	Excellent	Yes	Moderate
[37]	Yes	Clear	High	Yes	Yes	High	Directly	Excellent	Yes	Significant
[38]	Yes	Clear	Moderate	Yes	N/A	High	Indirectly	Excellent	Yes	Moderate

4. Discussion

This study focuses on how decision-tree based AMR prediction directly contributes to the One Health architecture by combining data sources from humans, animals and the environment. These models have social relevance beyond technical correctness when used fairly across resource conditions. It is both scientifically and ethically vital that prediction techniques be available and interpretable within low-resource healthcare systems. The synthesis of this study supports the combined contributions of genetic and environmental factors in AMR prediction. Genomic variations are inherent drivers of resistance mechanisms and extrinsic circumstances favouring gene transmission and persistence are modulated by environmental variables. The ability of decision tree–based algorithms to simulate these varied variables concurrently indicates their relevance for One Health–oriented monitoring systems. One concern in applying ML in healthcare is the trade-off between interpretability and performance. Complex models may have excellent accuracy, but lack of transparency erodes confidence especially in clinical decision-making. Explainable AI techniques like SHAP and LIME may assist to overcome this difficulty by making clear the contributions of features globally as well as locally. For example, extensive evaluations suggest that SHAP provides a more holistic knowledge of model behavior, whereas LIME is better at case-specific insight , hence they are complementing [39]. Applications throughout healthcare e.g. diabetes and stroke risk prediction have verified the fact that by combining these technologies, we may keep the predictive quality while increasing interpretability and user confidence [40]. Two studies on AMR prediction showed the advantage of adding SHAP in the ICU UTI resistance models and ambient ARG abundance models [23, 27] for clinical and public

health clarity without sacrificing performance. The matrix further validates, as seen in Figure 4, that great interpretability does not always lead to considerable performance loss, thereby challenging the fallacy of valuable accuracy for transparency tradeoffs. From a translational standpoint, this study identifies three high-priority pathways for deploying decision tree-based AMR models effectively:

1. ICU Triage Support: In resource-constrained settings, rapid diagnosis of drug-resistant disease may guide empiric therapy and expedite ICU workflow. Decision trees supplemented with SHAP outputs may achieve high sensitivity and explainable risk categorization to choose patients for isolation or concentrated treatment.
2. Tuberculosis Drug Selection: Tuberculosis drug-resistance, often caused by genetic changes, may be anticipated quickly and interpretably. Models such as Treesist-TB and those using RF/XGBoost on WGS data provide both accuracy and the ability to identify specific mutation-resistance correlations, offering clinicians important information for selecting first- and second-line drugs.
3. Wastewater AMR monitoring: Resistance assessment at community level is possible via environmental surveillance. Robust predictors based on physicochemical and ARG data, such as CatBoost and RF models, give scalable risk assessment. Together with interpretability frameworks, such predictors may reveal major drivers in the environment, which can then be addressed in public health measures, such as sanitation or antibiotic stewardship.

The routes underline larger social advantages such as faster and fairer healthcare delivery, more epidemiological surveillance and prudent use of antibiotics that connect with global health equity goals. Such models might impact stewardship measures, early epidemic detection, and resource allocation to reduce individual morbidity and promote system-wide resilience. Moreover, subgroup analysis (Table 3) indicated various strengths of the algorithm in different data categories. Random Forest and XGBoost demonstrated the best discriminative capacity on the structured and moderate-sized clinical datasets with minimum preprocessing. In environmental research, boosting algorithms like CatBoost and GBDT were selected because of their superior performance in different ecological predictions, whereas RF offers benefits in dealing with high-dimensional variant data without overfitting, thereby benefiting genomic datasets. These performance differences across domains highlight the ability of decision tree ensembles to adapt to the different data formats inherent in One Health AMR modeling. In addition, the summary RoB analysis (Figure 6) demonstrated that most studies had high methodological quality, adding confidence to the pooled data. We assessed the performance diversity considering domain-specific deficiencies, i.e., data-source heterogeneity in environmental applications. The weighted synthesis of this experiment (Table 5) also indicated minor differences in performance across the three algorithms, with overlapping 95% confidence intervals implying equal predictive capability. The results indicate the robustness of decision tree-based models across diverse AMR datasets and validate the slight advantage of XGBoost in structured clinical settings.

4.1. Strengths and Limitations

Decision tree-based methods have been shown to provide good prediction accuracy in AMR modeling, particularly when data are tiny, diverse or incomplete, where deep learning models tend to perform badly. They are appropriate for clinical and public health settings where interpretability and trust are important, are able to handle missing data, have easy tuning and can be integrated with explainable AI frameworks such as SHAP and LIME. Intelligent antimicrobial stewardship requires clearly identified significant factors and motivating adoption in resource-limited situations. But these benefits do not come without a cost. Models designed in one geographical or clinical setting may not be generalizable to another without recalibration. This indicates the variation of AMR prevalence and data quality. While interpretability tools are useful, they are computationally expensive and might provide unstable interpretations for highly correlated attributes, thereby misinterpreting the findings to non-technical users. Finally, in data-rich, high dimensional settings, deep learning techniques may outperform decision tree ensembles, indicating that hybrid systems that integrate the methodologies may provide the ideal balance between performance and transparency. Explainability tools such as SHAP and LIME do increase confidence and transparency but their practical constraints need further attention in particular their computational efficiency and interpretative stability as explained in Section 4.3. The regional focus of research in Asia (particularly China, South Korea, and Vietnam) may lead to a local bias that restricts the generalizability of decision tree-based AMR prediction models.

Antibiotic use, resistance frequency and diagnostic facilities differed greatly across locations, therefore models trained in one environment could not generalise well to another without recalibration. For instance, the genetic drivers of carbapenem resistance circulating in East Asia may vary from those circulating in sub-Saharan Africa or Latin America, therefore lowering the expected accuracy when models are extrapolated globally. Data asymmetries can lead to disparities in AMR surveillance capacities, for example, places without sequencing and electronic health record infrastructure are under-represented in algorithm development. Future studies should focus on improving capacity in underrepresented areas and equitable data sharing to increase global validity. This may provide secure, privacy-preserving model training on distant data sources, boosting variety without surrendering data sovereignty, via cross-regional consortia and other federated learning frameworks. Also, the data source integration from Africa and South America is key to achieve the WHO's One Health goals and ensure that the AI-based AMR prediction is equally advantageous to all health systems.

4.2. Future Research and Recommendations

Future work might examine hybrid modeling frameworks which use the interpretability and efficiency of decision tree-based methods with the deep feature extraction power of advanced neural architectures. This form of integration might allow models to learn complex high-dimensional patterns while being interpretable for clinical decision-making. <PRIVATE_PERSON> in recent work on predicting minimum inhibitory concentration (MIC) This trend should be further strengthened by systematic exploration of interpretability frameworks like as SHAP and LIME in AMR situations, focusing on their fidelity, computational demands and usability by the end user. The interpretability technologies should be examined not only in terms of technical correctness, but also in terms of their compatibility with the cognitive requirements of the practitioners utilizing defined standards and clinician-centered usability studies. Moreover, the deployment of these models in federated learning and real-time surveillance systems can permit a secure and privacy-preserving combination of data across multiple institutions, allowing for geographically diverse and generalizable AMR risk mapping across clinical, genomic and environmental domains. A key focus is on translating methodological innovation into policy and practice. Establishing frameworks for the engagement of stakeholders with physicians, microbiologists, public health officials and community people will be important to ensure that the outputs from these models are contextually relevant and helpful. Tools that can explain graphics, together with precise reporting requirements and targeted distribution techniques, may help develop trust and cross-sector adoption. In addition, AMR prediction programs should be developed with equality in mind by promoting research collaborations and data sharing agreements in underrepresented areas like Africa and South America. These activities will help to fill global gaps in AMR monitoring, increase capacity and ensure that the benefits of predictive modelling are accessible to everybody. The combination of technical advancement and inclusive implementation strategies in future research may position decision tree-based AMR prediction as a scientifically rigorous and socially innovative weapon in the fight against resistance.

4.3. Practical Barriers and Limitations of Explainable AI Tools in AMR Prediction

SHAP and LIME have come a long way in improving the interpretability of machine learning models. However, there are certain drawbacks, which restrict their use in real-world AMR prediction situations.

(a) Instability and feature correlation: When features are strongly correlated or data is subject to tiny perturbations, SHAP and LIME may provide unstable or inconsistent interpretations. In genomic or environmental data sets, predictors like as SNPs, ARG counts, or water-quality indicators are commonly interdependent, and SHAP values may vary widely among runs, leading to false significance attributions. Feature correlation might therefore confound biological inference if not addressed by feature grouping or decorrelation approaches.

(b) Computational cost: SHAP calculation increases linearly (and often exponentially) with the number of features, making it resource-intensive for high-dimensional genomic datasets (>104 variables). This makes it less practical for real-time monitoring systems or healthcare environments with limited resources. LIME is quicker, but it still has to perturb the model many times to build local approximations. This may be computationally expensive when used on thousands of samples.

(c) Interpretational load for end-users: The mathematical abstraction of SHAP and LIME explanations may be difficult for doctors and microbiologists without extensive ML expertise.

Over-interpreting the weight of an individual trait, especially when biological causation is unclear, might create false confidence or inappropriate treatment interventions.

(d) Lack of standardization: There is no generally agreed upon criterion or scoring method to assess dependability of explanation. SHAP and LIME results are not yet directly comparable across research or datasets, which complicates regulatory acceptability and repeatability. To meet all the foregoing issues, the development of standardized interpretability measurements, optimal SHAP approximations (e.g., TreeSHAP sampling), and user-centered visualization interfaces will be key to ensuring that explainability improves rather than obfuscates decision support.

5. Conclusions

Models based on Decision trees (e.g. Random Forest, GBDT, XGBoost) offer reliable, interpretable and resource economic methods for AMR prediction especially in case of diverse or moderate size dataset. However, deep-learning architectures may provide greater accuracy for large-scale genomic or image-based applications. Therefore, the ideal technique is context-dependent, and hybrid frameworks that integrate the two approaches might provide the best compromise between transparency and performance for future AMR research.

Acknowledgment: The authors gratefully acknowledge the support provided by Universitas Muhammadiyah Klaten, Central Java, Indonesia, and Universitas Muhammadiyah Yogyakarta, Indonesia. Their institutional resources, academic guidance, and collaborative environment were instrumental in the completion of this work.

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

Funding: No funding is received for this work.

References

[1] Ferraz, M.P.; "Antimicrobial Resistance: The Impact from and on Society According to One Health Approach". Societies, 14 (9): 187, 2024.

[2] Giri, P.; Aikat, A.; "The Impending Global Challenge of Antimicrobial Resistance: A Call for Urgent Action". Indian J. Comm. Health, 36 (5): 624–628, 2024.

[3] Uddin, T.M.; Chakraborty, A.J.; Khusro, A.; Zidan, B.R.M.; Mitra, S.; Emran, T. B.; Dhama, K.; *et al.*; "Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects". J. Infect. Public Health, 14 (12): 1750–1766, 2021.

[4] Naghavi, M.; Vollset, S.E.; Ikuta, K.S.; Swetschinski, L.R.; Gray, A.P.; Wool, E.E.; Aguilar, G.R.; *et al.*; "Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050". Lancet, 404 (10459): 1199–1226, 2024.

[5] Morel, C.M.; de Kraker, M.E.A.; Harbarth, S.; "Surveillance of Resistance to New Antibiotics in an Era of Limited Treatment Options". Front. Med. (Lausanne)., 8: 652638, 2021.

[6] Zai, M.J.; Cheesman, M.J.; Cock, I.E.; "Targeted Antimicrobial Therapies: A Solution to Overcoming Antimicrobial Resistance in Humans". BioMed, 4 (3): 318–337, 2024.

[7] Ahmed, S.K.; Hussein, S.; Qurbani, K.; Ibrahim, R.H.; Fareeq, A.; Mahmood, K.A.; Mohamed, M.G.; "Antimicrobial resistance: Impacts, challenges, and future prospects". J. Med. Surg. Public Health, 2: 100081, 2024.

[8] Khadse, S.N.; Ugemuge, S.; Singh, C.; "Impact of Antimicrobial Stewardship on Reducing Antimicrobial Resistance". Cureus, 15 (12): e49935, 2023.

[9] Tang, K.W.K.; Millar, B.C.; Moore, J.E.; "Antimicrobial Resistance (AMR)". Br. J. Biomed. Sci., 80: 11387, 2023.

[10] Bobate, S.; Mahalle, S.; Dafale, N.A.; Bajaj, A.; "Emergence of environmental antibiotic resistance: Mechanism, monitoring and management". Environ. Adv., 13: 100409, 2023.

[11] Fletcher, S.; "Understanding the contribution of environmental factors in the spread of antimicrobial resistance". Environ. Health Prev. Med., 20 (4): 243, 2015.

[12] Ankolekar, A.; Eppings, L.; Bottari, F.; Pinho, I.F.; Howard, K.; Baker, R.; Nan, Y.; *et al.*; "Using artificial intelligence and predictive modelling to enable learning healthcare systems (LHS) for pandemic preparedness". Comput. Struct. Biotechnol. J., 24: 412–419, 2024.

[13] Dixon, D.; Sattar, H.; Moros, N.; Kesireddy, S. R.; Ahsan, H.; Lakkimsetti, M.; Fatima, M.; *et al.*; "Unveiling the Influence of AI Predictive Analytics on Patient Outcomes: A Comprehensive Narrative Review". Cureus, 16 (5): e59954, 2024.

[14] Olawade, D.B.; Wada, O.J.; David-Olawade, A.C.; Kunonga, E.; Abaire, O.; Ling, J.; "Using artificial intelligence to improve public health:

a narrative review". Front. Public Health, 11: 1196397, 2023.
[15] Ghozali, M.T.; "Assessing ChatGPT's accuracy and reliability in asthma general knowledge: implications for artificial intelligence use in public health education". J. Asthma, 62 (6): 975–983, 2025.
[16] Tian, Y.; Zhang, H.; Qiao, Y.; Yang, H.; Liu, Y.; Ji, X.; "Intelligent Prediction of Ionic Liquids and Deep Eutectic Solvents by Machine Learning". Chin. J. Chem. Eng., 84: 227–243, 2025.
[17] Kim, J.I.; Maguire, F.; Tsang, K.K.; Gouliouris, T.; Peacock, S.J.; McAllister, T.A.; McArthur, A.G.; *et al.*; "Machine Learning for Antimicrobial Resistance Prediction: Current Practice, Limitations, and Clinical Perspective". Clin. Microbiol. Rev., 35 (3): e00179-21, 2022.
[18] Sakagianni, A.; Koufopoulou, C.; Koufopoulos, P.; Kalantzi, S.; Theodorakis, N.; Nikolaou, M.; Paxinou, E.; *et al.*; "Data-Driven Approaches in Antimicrobial Resistance: Machine Learning Solutions". Antibiotics, 13 (11): 1052, 2024.
[19] Ho, C.S.; Wong, C.T.H.; Aung, T.T.; Lakshminarayanan, R.; Mehta, J.S.; Rauz, S.; McNally, A.; *et al.*; "Antimicrobial resistance: a concise update". Lancet Microbe, 6 (1): 100947, 2025.
[20] Farhat, F.; Athar, M.T.; Ahmad, S.; Madsen, D.Ø.; Sohail, S.S.; "Antimicrobial resistance and machine learning: past, present, and future". Front. Microbiol., 14: 1179312, 2023.
[21] Rusic, D.; Kumric, M.; Perisin, A.S.; Leskur, D.; Bukic, J.; Modun, D.; Vilovic, M.; *et al.*; "Tackling the Antimicrobial Resistance "Pandemic" with Machine Learning Tools: A Summary of Available Evidence". Microorganisms, 12 (5): 842, 2024.
[22] Prosperi, M.; Azarian, T.; Johnson, J.A.; Salemi, M.; Milicchio, F.; Oliva, M.; "Unexpected Predictors of Antibiotic Resistance in Housekeeping Genes of Staphylococcus Aureus". In: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '19), Niagara Falls, NY, United States, September 7–10, 2019; Shi, X.M.; Buck, M.; Ma, J.; Veltri, P.; Eds.; ACM: New York, NY, United States, 2019.
[23] Lee, H.G.; Seo, Y.; Kim, J.H.; Han, S.B.; Im, J.H.; Jung, C.Y.; Durey, A.; "Machine learning model for predicting ciprofloxacin resistance and presence of ESBL in patients with UTI in the ED". Sci. Rep., 13 (1): 3282, 2023.
[24] Ning, Q.; Wang, D.; Cheng, F.; Zhong, Y.; Ding, Q.; You, J.; "Predicting rifampicin resistance mutations in bacterial RNA polymerase subunit beta based on majority consensus". BMC Bioinform., 22 (1): 210, 2021.
[25] Quoc, V.T.; Ngoc, D.N.T.; Hoang, T.N.; Thi, H.V.; Duc, M.T.; Nguyet, T.D.P.; Van, T.N.; *et al.*; "Predicting Antibiotic Resistance in ICUs Patients by Applying Machine Learning in Vietnam". Infect. Drug Resist., 16: 5535–5546, 2023.
[26] Matsuzaki, T.; Kato, Y.; Mizoguchi, H.; Yamada, K.; "A machine learning model that emulates experts' decision making in vancomycin initial dose planning". J. Pharmacol. Sci., 148 (4): 358–363, 2022.
[27] Zahra, Q.; Gul, J.; Shah, A.R.; Yasir, M.; Karim, A.M.; "Antibiotic resistance genes prevalence prediction and interpretation in beaches affected by urban wastewater discharge". One Health, 17: 100642, 2023.
[28] Liang, Q.; Zhao, Q.; Xu, X.; Zhou, Y.; Huang, M.; "Early prediction of carbapenem-resistant Gram-negative bacterial carriage in intensive care units using machine learning". J. Glob. Antimicrob. Resist., 29: 225–231, 2022.
[29] Jalali, F.M.; Chahkandi, B.; Gheibi, M.; Eftekhari, M.; Behzadian, K.; Campos, L.C.; "Developing a smart and clean technology for bioremediation of antibiotic contamination in arable lands". Sustain. Chem. Pharm., 33: 101127, 2023.
[30] Yasir, M.; Karim, A.M.; Malik, S.K.; Bajaffer, A.A.; Azhar, E.I.; "Application of Decision-Tree-Based Machine Learning Algorithms for Prediction of Antimicrobial Resistance". Antibiotics, 11 (11): 1593, 2022.
[31] Deelder, W.; Napier, G.; Campino, S.; Palla, L.; Phelan, J.; Clark, T.G.; "A modified decision tree approach to improve the prediction and mutation discovery for drug resistance in Mycobacterium tuberculosis". BMC Genomics, 23 (1): 46, 2022.
[32] Müller, S.J.; Meraba, R.L.; Dlamini, G.S.; Mapiye, D.S.; "First-line drug resistance profiling of Mycobacterium tuberculosis: a machine learning approach". In: AMIA Annual Symposium Proceedings, AMIA 2021 Annual Symposium, San Diego, CA, United States, October 30 – November 3, 2021; Petersen, C.; Dixon, B.E.; Lehmann, C.U.; Eds.; American Medical Informatics Association: Bethesda, MD, United States, 2022.
[33] Rich, S.N.; Jun, I.; Bian, J.; Boucher, C.; Cherabuddi, K.; Morris, J.G.; 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., 11 (5): 1869–1882, 2022.

[34] Akhter, S.; Miller, J.H.; "BPAGS: a web application for bacteriocin prediction via feature evaluation using alternating decision tree, genetic algorithm, and linear support vector classifier". Front. Bioinform., 3: 1284705, 2023.

[35] Goodman, K.E.; Simner, P.J.; Klein, E.Y.; Kazmi, A.Q.; Gadala, A.; Toerper, M.F.; Levin, S.; *et al.*; "Predicting probability of perirectal colonization with carbapenem-resistant Enterobacteriaceae (CRE) and other carbapenem-resistant organisms (CROs) at hospital unit admission". Infect. Control Hosp. Epidemiol., 40 (5): 541–550, 2019.

[36] Salgueiro, H.S.; Ferreira, A.C.; Duarte, A.S.R.; Botelho, A.; "Source Attribution of Antibiotic Resistance Genes in Estuarine Aquaculture: A Machine Learning Approach". Antibiotics, 13 (1): 107, 2024.

[37] Noman, S.M.; Zeeshan, M.; Arshad, J.; Amentie, M.D.; Shafiq, M.; Yuan, Y.; Zeng, M.; *et al.*; "Machine Learning Techniques for Antimicrobial Resistance Prediction of Pseudomonas Aeruginosa from Whole Genome Sequence Data". Comput. Intell. Neurosci., 2023 (1): 5236168, 2023.

[38] Wu, Y.; Jiang, P.; Goh, S.G.; Yu, K.; Chen, Y.; He, Y.; Gin, K.Y.H.; "Predicting Relative Risk of Antimicrobial Resistance using Machine Learning Methods". IFAC-PapersOnLine, 55 (10): 1266–1271, 2022.

[39] Salih, A.M.; Raisi-Estabragh, Z.; Galazzo, I.B.; Radeva, P.; Petersen, S.E.; Lekadir, K.; Menegaz, G.; "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME". Adv. Intell. Syst., 7 (1): 2400304, 2025.

[40] Pang, K.; "A comparative study of explainable machine learning models with Shapley values for diabetes prediction". Healthc. Anal., 7: 100390, 2025.