Machine Learning and Medical Authority Engagement for Antimicrobial Resistance Management: A Review of Surveillance, Prediction, and Stewardship

Joel Gijo
Department of Computer Science
Amal Jyothi College of Engineering (Autonomous)
Kanjirappally, India
joelgijo03@gmail.com

Bibin Kunnathettu Biju
Department of Computer Science
Amal Jyothi College of Engineering (Autonomous)
Kanjirappally, India
bibinbiju2005@gmail.com

K Ryan George
Department of Computer Science
Amal Jyothi College of Engineering (Autonomous)
Kanjirappally, India
ryangeorge1205@gmail.com

Bipin Dev B
Department of Computer Science
Amal Jyothi College of Engineering (Autonomous)
Kanjirappally, India
bipindevb@gmail.com

Dr. Anju J Prakash
Department of Computer Science
Amal Jyothi College of Engineering (Autonomous)
Kanjirappally, India
anjujprakash@amaljyothi.ac.in

***Abstract*—Antimicrobial resistance (AMR) is a critical global health challenge, with particularly severe consequences in low- and middle-income countries (LMICs) where surveillance infrastructure, diagnostic capacity, and stewardship resources remain constrained. This review synthesises recent advances in AMR surveillance, data integration, community-level antibiotic use, and the growing role of machine learning (ML) in resistance prediction and clinical decision support. We examine integrated digital platforms such as India's i-DIA and international initiatives like the Fleming Fund that are bridging data fragmentation across healthcare systems. We survey ML approaches from supervised classifiers to ensemble methods, with particular attention to resource-appropriate frameworks operating under minimal data requirements. As a contextual case study, we describe AMR-X—a calibrated probabilistic decision-support system designed to address the structural gap in empiric prescribing support for resource-limited environments. We further discuss socioeconomic barriers to antimicrobial access, community stewardship challenges, and evidence-based policy recommendations, aiming to assist researchers, clinicians, and policymakers in building effective, data-driven AMR control strategies.**

***Index Terms*—antimicrobial resistance, machine learning, surveillance, stewardship, low-resource settings, data integration, decision support, medical authority engagement**

I. Introduction

Antimicrobial resistance (AMR) has emerged as one of the defining public health crises of the twenty-first century. The progressive erosion of antibiotic efficacy threatens to reverse decades of medical progress, rendering once-treatable infections difficult or impossible to manage. AMR contributes to an estimated 33,000 deaths annually in Europe alone [4], and the burden in low- and middle-income countries is substantially higher yet systematically undercounted due to fragmented surveillance infrastructure [12].

The challenge is particularly acute in LMICs, where healthcare facilities operating at capacity, limited diagnostic access, weak regulatory enforcement over antibiotic sales, and inadequate stewardship programmes collectively create environments permissive to resistance emergence and spread. The WHO identifies AMR as a top-tier global health priority, calling for coordinated action spanning surveillance, stewardship, research, and equitable access to effective antimicrobials [2].

A key structural gap compounds the problem: clinicians and medical authorities in resource-limited settings routinely make prescribing decisions without access to local resistance data, relying on empiric therapy frequently misaligned with

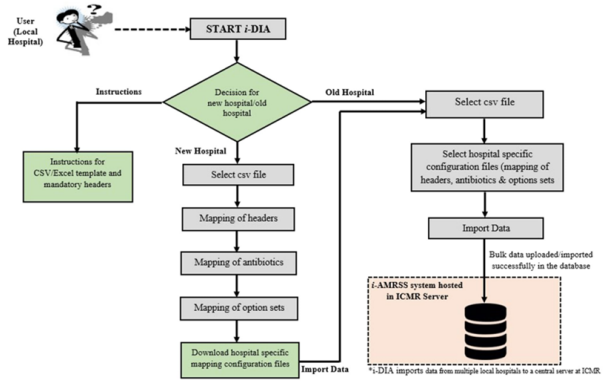

Fig. 1. Schematic workflow of the i-DIA platform for standardised AMR data import into national surveillance systems [16].

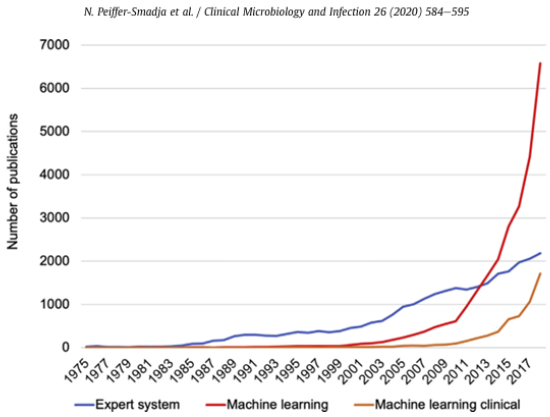

Fig. 2. Trends in ML-related publications in infectious disease research indexed in PubMed [4].

prevailing resistance patterns. Machine learning has emerged as a promising approach to closing this gap, capable of deriving actionable resistance estimates from routinely available clinical data at a scale and speed conventional methods cannot match [19].

This review covers four interconnected themes: AMR surveillance and data integration; ML applications in resistance prediction; community antimicrobial access and use; and stewardship strategies and policy. Throughout, we contextualise these themes through AMR-X—a practical ML-driven decision-support tool designed for resource-limited clinical environments.

II. AMR Surveillance and Data Integration

A. Integrated Surveillance Platforms

Effective AMR control requires reliable, timely knowledge of resistance patterns at local, national, and global levels. In high-income countries, well-established laboratory networks, integrated EHR systems, and standardised reporting frameworks permit meaningful cross-institutional comparisons [13]. In LMICs, fragmented healthcare infrastructure, poor interoperability between laboratory information systems (LIS) and hospital software, and inconsistent antimicrobial susceptibility testing (AST) methodologies produce data silos that significantly limit surveillance value [12].

India's integrated Data Import Application (i-DIA), developed by the Indian Council of Medical Research (ICMR), addresses this by automating bulk import of standardised AST data from diverse hospital systems into the national i-AMRSS surveillance system. By reducing manual data entry and enabling flexible taxonomic mapping, i-DIA substantially strengthens surveillance capabilities across heterogeneous institutions [16]. Fig. 1 illustrates its workflow. The Fleming Fund, operating across 14 countries in Africa and Asia, similarly builds comprehensive surveillance networks connecting laboratories, clinical departments, and national health agencies under the One Health approach [11].

Web-based platforms such as AMRmap illustrate the value of centralised surveillance, enabling retrospective resistance analysis with geographic visualisation [15]. However, such platforms remain fundamentally retrospective and do not generate real-time predictions for individual prescribing decisions. Modern platforms address data security through role-based access controls, cryptographic encryption, and API-driven interoperability [12]. Sustained expansion requires ongoing capacity building alongside mobile health tools supporting point-of-care data entry and rapid outbreak response in remote settings [21].

B. Persistent Challenges

Despite technological progress, critical gaps remain. Incomplete digitisation leaves significant holes in laboratory data, antimicrobial prescriptions, and patient outcomes. Variability in AST methodologies and interpretative criteria produces reporting inconsistencies difficult to resolve in aggregated analyses [3]. Data from clinical, microbiological, pharmacological, and demographic sources typically exist in disconnected silos, hindering comprehensive understanding of AMR drivers. Overcoming these barriers requires harmonised vocabularies, standardised taxonomies, and sustained investment in interoperable digital infrastructure [20].

III. Machine Learning for AMR Prediction

A. Approaches and Applications

Machine learning has rapidly established itself as a powerful analytical tool in AMR, providing the ability to derive predictive insights from complex, large-scale datasets beyond the capacity of conventional methods [19]. The volume of ML publications in clinical microbiology has grown substantially over the past decade, as shown in Fig. 2, reflecting growing recognition of ML's potential in this domain.

Supervised learning is the most widely applied category for AMR prediction. XGBoost is particularly favoured for its robustness, ability to handle diverse data types, and strong performance on structured tabular clinical data [1]. It employs gradient boosting to build ensembles of decision trees iteratively, each correcting the residual errors of its predecessor.

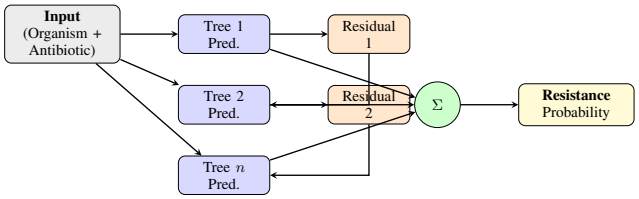

Fig. 3. Conceptual XGBoost gradient boosting mechanism. Each successive tree corrects the residual error of its predecessor; the final prediction is a weighted ensemble sum.

Fig. 3 illustrates this mechanism conceptually. Genomic ML approaches achieve high accuracy by identifying resistance genes and mutations [5] but require specialised infrastructure and significant per-sample costs, limiting applicability to reference laboratories. EHR-based approaches model institutional resistance patterns with high fidelity [13] but depend on comprehensive data infrastructure unavailable in much of the developing world.

B. Hospital Decision Support and LMIC Adaptation

In tertiary hospitals, ML models have demonstrated significant potential combining EHR, microbiology lab, and medication data to predict antibiotic susceptibility and patient outcomes. Peiffer-Smadja et al. reviewed 60 ML-based clinical decision support systems in infectious diseases, finding that algorithms like XGBoost achieve strong discriminative performance using routinely collected information [4]. However, most training data originate from well-resourced settings, raising concerns about generalisability to LMICs.

Adapting ML for resource-constrained environments requires utilising surveillance datasets from national networks such as India's ICMR-AMRSN [20] and regional programmes supported by the Fleming Fund, integrating demographic, clinical, and antimicrobial consumption data [3]. ML models for tuberculosis incorporating demographic risk factors and clinical imaging features demonstrate that gradient boosting methods can provide meaningful guidance where molecular diagnostics are delayed or unavailable [19].

Table I contrasts AMR-X against representative ML-based AMR prediction approaches across dimensions relevant to clinical deployment and LMIC applicability. AMR-X is distinguished by its minimal input requirement and deliberate design for resource-constrained settings, at the cost of the higher accuracy achievable by EHR-integrated or genomic approaches where richer infrastructure exists.

TABLE I
COMPARISON OF ML-BASED AMR PREDICTION APPROACHES

Approach	Input Requirements	AUC	LMIC Fit
AMR-X (this work)	Organism + antibiotic only	0.81	High
EHR-based [13]	Full EHR, lab, medication records	>0.85	Low
Genomic ML [5]	DNA sequencing data	High*	Very Low
CDSS review [4]	Various (EHR + lab)	0.73–0.97	Low–Med

*High accuracy; requires reference lab infrastructure.

IV. AMR-X: RESOURCE-APPROPRIATE ML DECISION SUPPORT

A. Motivation and Architecture

AMR-X was developed to address a structural stewardship gap: medical authorities in outpatient and rural settings lack a reliable mechanism to flag potentially ineffective antibiotic choices before dispensing. Static antibiogram-based lookup tables are geographically aggregated and infrequently updated, making them unreliable for facilities whose local resistance patterns diverge from regional trends [7].

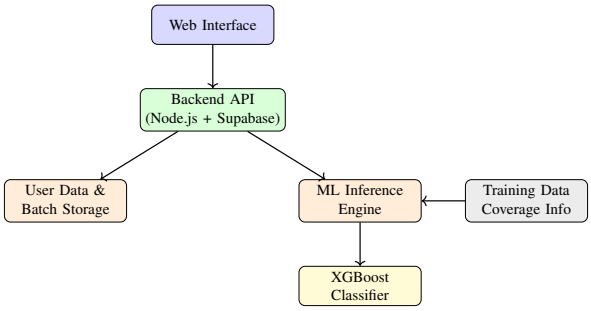

Fig. 4. AMR-X three-tier system architecture. The web interface collects input, the backend manages authentication and routing, and the ML inference engine returns calibrated resistance probabilities with evidence-strength indicators.

AMR-X is built on a modular three-tier architecture separating the web interface, application logic, and ML inference layer (Fig. 4), enabling independent component updates and compatibility with low-resource devices. The frontend allows authenticated users—medical authorities and clinicians—to submit organism-antibiotic pairs individually or via batch CSV/Excel uploads. The PostgreSQL backend stores structured susceptibility records enabling district-level resistance aggregation and retrospective surveillance queries while maintaining data provenance.

B. ML Methodology and Data Pipeline

The AMR-X classifier is trained on over one million susceptibility test records from the publicly available Antibiotic Resistance Microbiology Dataset (ARMD) [17], spanning 87 bacterial species and 45 antibiotics from North American and European clinical settings (2015–2023). The prediction task is formulated as binary classification: given only organism and antibiotic identifiers—the minimal data available from any standard susceptibility report—the model estimates resistance probability. This deliberate minimal feature design enables deployment without genomic data, EHR access, or patient demographics.

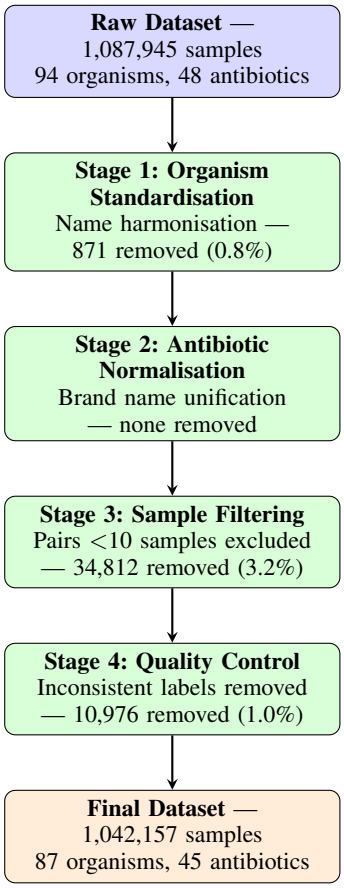

Fig. 5. AMR-X preprocessing pipeline showing four quality control stages and cumulative sample reduction.

Raw data underwent a four-stage preprocessing pipeline (Fig. 5): organism name harmonisation, antibiotic brand name unification, exclusion of organism-antibiotic pairs with fewer than 10 samples, and removal of records with inconsistent susceptibility labels—reducing 1,087,945 raw records to a final dataset of 1,042,157 samples. Class imbalance (32% resistant, 68% susceptible) was addressed via XGBoost's built-in `scale_pos_weight` parameter to preserve calibration integrity [10].

C. Uncertainty Communication and RWUI Extension

A key distinguishing feature of AMR-X is evidence-aware inference. Each prediction is accompanied by an evidence-strength indicator derived from the historical sample count for the corresponding organism-antibiotic pair: high ($>$1000 samples), medium (100–1000), or low ($<$100). Predictions are stratified into four clinical risk levels (low: $p<0.30$; moderate-low: $0.30\leq p<0.50$; moderate-high: $0.50\leq p<0.70$; high: $p\geq 0.70$). Probability calibration via isotonic regression ensures outputs correspond to observed resistance frequencies in practice [10], which distinguishes AMR-X from systems that output arbitrary confidence scores.

The system architecture supports a proposed Resistance-Weighted Usage Index (RWUI)—a proxy for antibiotic selective pressure derived from medical authority-reported prescription volumes at the healthcare unit level. Higher usage intensity correlates with elevated local resistance probability, a relationship well-established in epidemiological literature [11], [18]. Medical authorities contributing prescription data to AMR-X implicitly generate the volume data from which a district-level RWUI can be computed, enabling local adaptation of resistance estimates without requiring centralised laboratory coordination or EHR integration.

D. Performance and Deployment Scenarios

Model selection was conducted using 5-fold stratified cross-validation on the training partition, with a held-out test set of 104,215 organism-antibiotic pairs (10% of the final dataset) reserved exclusively for final evaluation. Calibration quality was verified via reliability diagrams confirming close alignment between predicted resistance probabilities and empirically observed resistance frequencies across probability bins, validating the practical utility of isotonic regression post-processing.

AMR-X achieves an ROC-AUC of 0.81 on the held-out test set, substantially outperforming logistic regression (0.73) and prevalence-based lookup (0.50). High sensitivity (0.83) and negative predictive value (0.95) support reliable detection of resistant cases and confident ruling out of susceptibility. Lower precision (0.33) reflects a deliberate conservative bias: in empiric prescribing, failing to detect resistance carries greater clinical risk than overestimating it.

Practical deployment scenarios include emergency department empiric prescribing when culture results are delayed 24–48 hours; outpatient settings without routine laboratory access; pharmacy prescription review for stewardship flagging; and regional resistance surveillance through aggregated authority-contributed uploads. AMR-X provides probabilities, not treatment recommendations; final prescribing decisions remain the responsibility of qualified healthcare professionals.

E. Geographic Generalisability and LMIC Adaptation

The ARMD training corpus originates predominantly from North American and European clinical settings, introducing a geographic distribution shift that warrants explicit consideration for LMIC deployment. Resistance epidemiology in South Asian and sub-Saharan African contexts differs substantially due to divergent antibiotic consumption patterns, prescribing norms, and locally circulating pathogen clones [3], [19]. AMR-X addresses this limitation through two design mechanisms. First, evidence-strength indicators explicitly flag organism-antibiotic pairs with sparse training support, directing clinicians to treat low-evidence predictions with appropriate caution—particularly relevant for locally prevalent pathogens underrepresented in North American or European surveillance data. Second, the RWUI mechanism enables district-level prescription volume data contributed by local medical authorities to progressively weight resistance estimates toward local epidemiology, without requiring centralised

laboratory infrastructure or EHR integration. Prospective validation on LMIC clinical datasets—particularly from India's ICMR-AMRSN network [20]—and incremental model fine-tuning using locally collected susceptibility records represent important future directions for strengthening geographic transferability.

V. Community Antimicrobial Access and Use

In many LMIC regions, significant barriers to formal healthcare—geographic remoteness, financial constraints, and shortages of trained professionals—cause informal providers such as pharmacies and drug shops to become the primary source of antibiotics. Weak enforcement of prescription-only rules leads to widespread self-medication, documented across India, Bangladesh, Vietnam, and sub-Saharan Africa in multi-country WHO collaborative reviews [6]. Qualitative studies consistently show widespread misconceptions about antibiotics: many individuals misunderstand indications, confuse viral and bacterial infections, and underestimate the risks of incomplete courses—driving premature treatment stoppage, dosage errors, and use of leftover antibiotics that accelerate resistance.

Medical authorities—particularly those operating at community dispensing points—occupy a structurally critical position as both stewardship actors and potential surveillance contributors. Research confirms that structured prescribing frameworks for medical authorities can reduce inappropriate antibiotic dispensing when supported by training, accessible clinical guidance tools, and regulatory backing [9]. ML approaches applied to community settings—such as predicting over-the-counter antibiotic use in rural India—demonstrate that data-driven methods can characterise not just resistance patterns but the upstream behavioural drivers that shape them [8]. Digital tools that create practical value at the point of dispensing while collecting resistance-relevant data offer a stewardship model that does not require top-down institutional mandates to sustain.

VI. Stewardship Strategies and Policy

A. Hospital and Community Stewardship

Effective antimicrobial stewardship programmes (ASPs) promote responsible antimicrobial use while improving patient outcomes, requiring the integration of surveillance data, laboratory diagnostics, and clinical decision support [2]. In hospital settings, ML algorithms are increasingly core ASP components, enabling automatic detection of inappropriate prescribing and real-time tailored treatment recommendations [14]. Fig. 6 illustrates the spectrum of clinician-ML interaction across automation levels—from full clinician oversight to near-autonomous prescribing adjustments—showing how ASPs can incorporate AI-driven insights at varying degrees of human involvement.

Stewardship must extend beyond hospitals into communities where most antibiotic use occurs [9]. Integrating medical authority engagement with platforms such as AMR-X creates a stewardship pathway independent of hospital infrastructure:

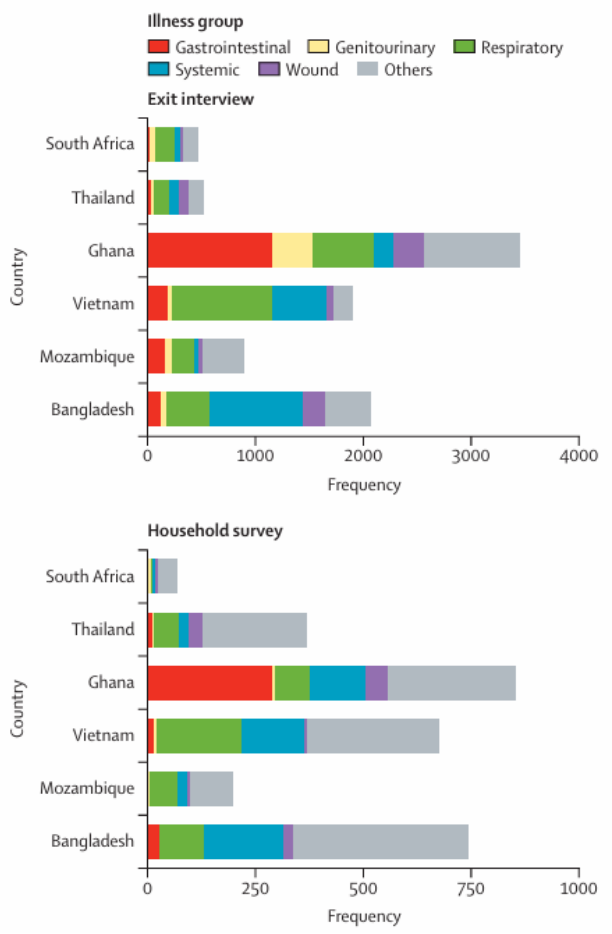

Fig. 6. Spectrum of clinician interaction with ML-powered systems across automation levels [6].

medical authorities performing prescription review simultaneously contribute dispensing data that powers district-level resistance monitoring, creating a bidirectional value exchange between clinical utility and public health surveillance.

B. Policy Recommendations

Table II summarises key recommended interventions across relevant policy and practice domains, synthesised from global and regional stewardship evidence [2], [11], [12].

VII. Conclusion

This review has synthesised evidence across AMR surveillance and data integration, machine learning for resistance prediction, community antimicrobial use patterns, and stewardship policy frameworks. Three interconnected themes emerge clearly. First, surveillance infrastructure investment is foundational: without reliable, integrated data, neither ML models nor stewardship programmes can function effectively. Second, ML-driven decision support can be designed for the minimal data conditions that characterise resource-limited settings—as demonstrated by AMR-X, which achieves clinically useful resistance risk estimates using only organism and antibiotic

TABLE II
KEY ANTIMICROBIAL STEWARDSHIP INTERVENTIONS AND POLICY RECOMMENDATIONS

Domain	Recommended Actions
Surveillance	Develop interoperable AMR/AMC platforms aligned with GLASS, ICMR-AMRSN, and Fleming Fund frameworks
Clinical	Integrate decision support tools and prescribing feedback into hospital and primary care workflows
Workforce	Train clinicians and medical authorities in stewardship principles, resistance mechanisms, and rational prescribing
Community	Engage medical authorities and drug vendors in education; regulate non-prescription antibiotic sales
Policy	Mandate prescription-only sales; enforce quality assurance; ensure access to diagnostics
Research	Support local AMR driver research and development of digital stewardship tools in LMICs
International	Foster public-private partnerships and knowledge exchange through global alliances

identifiers, without genomic or patient-level data. Third, medical authorities at community level represent an underexploited structural asset: positioned at the intersection of antibiotic dispensing and community health, they can simultaneously perform stewardship review and contribute to resistance surveillance when supported by accessible digital tools.

Geographic generalisability remains an open challenge; evidence-strength indicators and the RWUI local adaptation mechanism provide partial mitigation, while prospective clinical validation in diverse LMIC settings is an important future direction.

Collaborative global action, continuous education, and meaningful community engagement are essential to contain AMR spread and promote rational antibiotic use. The integration of AI-driven analytics must be accompanied by transparent uncertainty communication, ethical governance, and equitable deployment. By aligning scientific innovation with policy and practical implementation—and placing medical authorities at the centre of both stewardship and surveillance—these efforts will be essential to preserving antimicrobial effectiveness for future generations.

REFERENCES

[1] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[2] World Health Organization, "Global antimicrobial resistance and use surveillance system (GLASS) report," Geneva, Switzerland, 2023.

[3] V. Modgil *et al.*, "Mapping the AMR infection landscape in Bihar: Implications for strengthening policy and clinical practice," Preprint, Jun. 2025, doi:10.1101/2025.06.07.25329177.

[4] N. Peiffer-Smadja *et al.*, "Machine learning for clinical decision support in infectious diseases: a narrative review of current applications," *Clin. Microbiol. Infect.*, vol. 26, no. 5, pp. 584–595, 2020, doi:10.1016/j.cmi.2019.09.009.

[5] Y. Yang, K. Niehaus, T. M. Walker, and D. Wilson, "Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data," *Bioinformatics*, vol. 34, no. 10, pp. 1666–1671, 2018.

[6] N. T. T. Do *et al.*, "Community-based antibiotic access and use in six low-income and middle-income countries: a mixed-method approach," *Lancet Glob. Health*, vol. 9, no. 5, pp. e610–e619, 2021, doi:10.1016/S2214-109X(21)00024-3.

[7] T. Rawson *et al.*, "A systematic review of clinical decision support systems for antimicrobial stewardship," *Clin. Microbiol. Infect.*, vol. 23, no. 7, pp. 464–474, 2017.

[8] P. A. Sawant *et al.*, "Predicting over-the-counter antibiotic use in rural Pune, India, using machine learning methods," *Epidemiol. Health*, vol. 46, art. no. e2024044, 2024, doi:10.4178/epih.e2024044.

[9] J. H.-C. Wu *et al.*, "Community pharmacist prescribing of antimicrobials: A systematic review from an antimicrobial stewardship perspective," *Can. Pharm. J. (Ott.)*, vol. 154, no. 3, pp. 179–192, 2021, doi:10.1177/1715163521999417.

[10] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," *Proc. 22nd Int. Conf. Machine Learning*, pp. 625–632, 2005.

[11] S. Ajulo and B. Awosile, "Global Antimicrobial Resistance and Use Surveillance System (GLASS 2022): Investigating the relationship between antimicrobial resistance and antimicrobial consumption data," *PLoS One*, vol. 19, no. 2, art. no. e0297921, 2024, doi:10.1371/journal.pone.0297921.

[12] I. Frost *et al.*, "Status, challenges and gaps in antimicrobial resistance surveillance around the world," *J. Glob. Antimicrob. Resist.*, vol. 26, pp. 168–173, 2021, doi:10.1016/j.jgar.2021.02.015.

[13] K.-D. Vihta *et al.*, "Predicting future hospital antimicrobial resistance prevalence using machine learning," *Commun. Med.*, vol. 4, no. 1, art. no. 197, 2024, doi:10.1038/s43856-024-00502-8.

[14] ReAct Asia Pacific, "AI-Powered antibiotic stewardship: ReAct Asia Pacific's consultation," Mar. 2025. [Online]. Available: https://www.reactgroup.org/news-and-views/news-and-opinions/year-2025/ai-powered-antibiotic-stewardship-react-asia-pacifics-consultation

[15] A. Y. Kuzmenkov *et al.*, "AMRmap: An Interactive Web Platform for Analysis of Antimicrobial Resistance Surveillance Data in Russia," *Front. Microbiol.*, vol. 12, art. no. 620002, 2021, doi:10.3389/fmicb.2021.620002.

[16] J. Kaur *et al.*, "Design development of customizable web API for interoperability of antimicrobial resistance data," *Sci. Rep.*, vol. 11, art. no. 11226, 2021, doi:10.1038/s41598-021-90601-z.

[17] F. Nateghi Haredasht *et al.*, "Antibiotic Resistance Microbiology Dataset (ARMD): A resource for antimicrobial resistance from EHRs," *Scientific Data*, Dryad, 2025. [Online]. Available: https://doi.org/10.5061/dryad.jq2bvq8kp

[18] N. Taneja and M. Sharma, "Antimicrobial resistance in the environment: The Indian scenario," *Indian J. Med. Res.*, vol. 148, pp. 327–336, 2018, doi:10.4103/ijmr.IJMR-331-18.

[19] J. I. Kim *et al.*, "Machine Learning for Antimicrobial Resistance Prediction: Current Practice, Limitations, and Clinical Perspective," *Clin. Microbiol. Rev.*, vol. 35, no. 3, art. no. e00179-21, 2022, doi:10.1128/cmr.00179-21.

[20] K. Walia *et al.*, "Establishing Antimicrobial Resistance Surveillance Research Network in India: Journey so far," *Indian J. Med. Res.*, vol. 148, pp. 303–317, 2018, doi:10.4103/ijmr.IJMR22618.

[21] J. Mayito *et al.*, "Combatting Antimicrobial Resistance Through a Data-Driven Approach to Optimize Antibiotic Use and Improve Patient Outcomes: Protocol for a Mixed Methods Study," *JMIR Res. Protoc.*, vol. 13, art. no. e58116, 2024, doi:10.2196/58116.

AI Usage Disclosure: AI-assisted tools were used exclusively for language editing and grammar checking. All conceptual design, methodology, and interpretations represent original work by the authors.