From Days to Hours: Artificial Intelligence in Antimicrobial Resistance Diagnostics and Drug Discovery, and Why No Tool Has Yet Reached the Clinic

Nahid Amin[1*], Papiya Sultana Sharker Sharna[1]

Abstract

Background: Antimicrobial resistance (AMR) is projected to contribute to millions of deaths in the coming decades, and the conventional antibiotic-discovery pipeline has, by most accounts, not kept pace with it. Artificial intelligence (AI) is frequently proposed as a corrective, though whether that promise has translated into demonstrable clinical benefit is less often examined directly. **Methods**: We conducted a narrative-systematic review of peer-reviewed and preprint literature on AI applications in AMR diagnostics and antimicrobial discovery, searching PubMed/MEDLINE, Scopus, Web of Science, and the Cochrane Library through mid-2026, and organized findings across four domains: rapid phenotypic diagnostics, genomic and metagenomic resistome prediction, explainable AI, and de novo drug design. **Results**: AI-enabled diagnostics reduced susceptibility-testing turnaround from a conventional 36–72 hours to under 2–4 hours in several platforms; genomic language models such as DNABERT outperformed conventional classifiers by 12–18% in resistance-gene classification; explainable AI methods, SHAP in particular, linked model predictions to known resistance mechanisms; and generative frameworks yielded antimicrobial peptide candidates with confirmed in vitro and in vivo activity. Nearly all of this evidence, however, derives from retrospective, single-center validation, and no AI-based AMR tool has yet secured regulatory clearance anywhere. **Conclusion**: AI has moved convincingly beyond proof-of-concept in AMR diagnostics and discovery, but its path to the clinic now depends less on further algorithmic refinement than on prospective validation, equitable data representation, and interpretability standards that clinicians can reasonably trust.

Keywords

antimicrobial resistance; artificial intelligence; explainable AI; whole-genome sequencing; clinical translation

Significance | AI compresses AMR diagnosis and discovery timelines dramatically, yet no tool has reached regulatory clearance — validation and trust, not algorithms, remain the bottleneck.

**Correspondence.*

Nahid Amin, Department of Public Health, Anwer Khan Modern University, Dhaka, Bangladesh. E-mail: nahid2u@gmail.com.

Editor Armaghan Shafaei, Ph.D. And accepted by the Editorial Board Sep 07, 2026 (received for review Jun 22, 2026)

1. Introduction

Antimicrobial resistance has, by now, become one of those phrases that risks losing its force through sheer repetition — invoked so often in editorials and policy briefs that it can start to sound like background noise rather than an emergency. And yet the numbers behind it have not softened with familiarity. Bacterial AMR was associated with an estimated 4.95 million deaths worldwide in 2019,

Author Affiliation: 1Department of Public Health, Anwer Khan Modern University, Dhaka, Bangladesh.

Please cite this article as: Amin, N., Sharna, P. S. S. (2026). "From Days to Hours: Artificial Intelligence in Antimicrobial Resistance Diagnostics and Drug Discovery, and Why No Tool Has Yet Reached the Clinic", Integrative Biomedical Research, 10(1), 1-16, 10890

of which roughly 1.27 million were directly attributable to resistant infections (Murray et al., 2022), and later modelling has done little to reassure: the Global Research on Antimicrobial Resistance initiative projects a cumulative 39 million deaths from drug-resistant infections between 2025 and 2050 (Sardar et al., 2026). The economic picture tracks the human one — unmitigated resistance could add close to $1 trillion in healthcare expenditure by mid-century and subtract $1–3.4 trillion from global GDP by 2030 (Sardar et al., 2026; Scaglione et al., 2026; Bhattacharya et al., 2026). None of this is abstract to a clinician standing at a bedside; it shows up, ward by ward, as prolonged admissions, failed empirical regimens, and a steadily narrowing set of drugs that still work (Chines et al., 2026; Alatawi et al., 2025).

What makes the crisis so resistant to correction — no pun particularly intended, though it is hard to avoid — is that it has no single cause. Bacteria mutate and share resistance genes at a pace that outstrips most human institutions, and that biological agility is compounded by how antimicrobials are used, and overused, across medicine, animal husbandry, and agriculture alike (Branda & Scarpa, 2024). At the same time, the commercial pipeline for new antibiotics has more or less stalled (Nagre et al., 2025), a consequence, in part, of a discovery model built around slow, capital-intensive, largely trial-and-error screening that was never designed to keep up with horizontal gene transfer (Sanapalli et al., 2026). The result is a widening gap between how quickly resistance emerges and how quickly new options arrive — a gap that conventional methods, working alone, seem unlikely to close on their own terms (Salama et al., 2026; Mohammed et al., 2025).

It is against this backdrop that artificial intelligence has entered the conversation — not, we would argue, as a silver bullet, so much as a genuinely different way of approaching an old problem. Its arrival looks less like an incremental tool upgrade and more like a shift in what microbiology is for: moving from a largely descriptive discipline toward one that is predictive, and increasingly, data-driven (Salama et al., 2026; Chines et al., 2026). Drawing on high-throughput multi-omics data, clinical metadata, and computational architectures that would have seemed implausible only a decade ago, AI-driven workflows now offer early detection of resistance markers, support for prescribing decisions, and de novo design of candidate molecules (Branda & Scarpa, 2024; Scaglione et al., 2026). When genomic data, patient history, and epidemiological signal are woven together, AI-based diagnostics can, at least in principle, sidestep some of the oldest limitations of conventional testing — false negatives, unacceptable delay — even in settings where molecular infrastructure is thin on the ground (Ahmed & Alghamdi, 2026; Hassan et al., 2026).

Two threads of progress run through this review, and it is worth naming them early. The first concerns diagnostics. Conventional culture-based antimicrobial susceptibility testing still requires 36 to 72 hours in most settings — long enough to push clinicians toward empirical, broad-spectrum prescribing that, somewhat ironically, intensifies the very selective pressure driving resistance in the first place (Aggarwal et al., 2026; Scaglione et al., 2026). AI-enabled platforms are beginning to unsettle that status quo, offering rapid resistance prediction across both genomic and phenotypic modalities (Khangarot et al., 2026). On the genomic side, whole-genome and metagenomic sequencing, paired with machine learning, now bypass slow culture steps almost entirely, catching resistance determinants that classical alignment tools such as BLAST would likely miss (Sardar et al., 2026; Bhattacharya et al., 2026). On the phenotypic side, MALDI-TOF mass spectrometry and Raman spectroscopy — already embedded in many clinical laboratories — have been repurposed through machine learning to classify resistance phenotypes within minutes, and, when coupled with microfluidics, to deliver minimum inhibitory concentration estimates in under two to four hours (Aggarwal et al., 2026; Liao et al., 2025).

The second thread concerns discovery. Where traditional screening of physical compound libraries is slow and expensive almost by design, AI-driven virtual screening explores vast in silico chemical spaces instead, predicting binding affinity and resistance-development potential before a single molecule is synthesized (Sanapalli et al., 2026; Abo-Tbeak & AL-Hilali, 2026). The discovery of halicin — identified by a deep neural network trained on roughly 2,500 molecules and later shown to disrupt bacterial electrochemical gradients through a genuinely novel mechanism — remains perhaps the most cited proof of concept (Stokes et al., 2020). Abaucin, a narrow-spectrum antibiotic active against multidrug-resistant *Acinetobacter baumannii*, followed a broadly similar computational path (Awan et al., 2024; Sanapalli et al., 2026). Generative frameworks have since extended this logic to antimicrobial peptides, several of which show

confirmed efficacy in murine models (Sardar et al., 2026; Maldonado-Hernández et al., 2026).

And yet, for all this technical momentum, a considerable gap persists between what these models can do in silico and what they have been shown, so far, to do at the bedside. This review is organized around that tension. We synthesize the technical progress achieved in AI-driven AMR diagnostics and drug discovery, examine the methodological and regulatory obstacles that continue to constrain clinical translation — retrospective validation bias, data inequity, interpretability concerns, and the absence, to date, of regulatory clearance for any AI-based AMR tool — and propose a framework for what a credible path forward might look like (Sardar et al., 2026; Alatawi et al., 2025).

3. Methods

3.1 Study Design and Reporting Framework

We designed this work as a narrative-systematic review — a hybrid form that borrows the transparency of a systematic search process while retaining the interpretive latitude a purely narrative synthesis allows. That combination felt like the right fit here: the underlying literature is too methodologically heterogeneous (spanning diagnostic accuracy studies, computational drug-discovery pipelines, and health-services implementation reports) to support formal meta-analytic pooling, but a purely narrative account, unanchored to any reproducible search process, would have understated how much empirical work already exists. Reporting followed, as far as the review's hybrid scope reasonably allowed, the structure recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement (Page et al., 2021), so that a reader wishing to reproduce or extend this search could, in principle, do so from the description below. The protocol was not prospectively registered, given the review's hybrid narrative-analytical scope; however, the search strategy, eligibility criteria, and data-extraction fields were fixed in advance of full-text screening to limit post hoc selection bias.

3.2 Information Sources and Search Strategy

We searched four databases — PubMed/MEDLINE, Scopus, Web of Science, and the Cochrane Library — for records published within a multi-year window through mid-2026, supplemented by manual citation-chasing of key reviews and hand-searching of journals with a strong AMR/AI publication footprint (*Antibiotics*, *Briefings in Bioinformatics*, *Frontiers in Public Health*, *Frontiers in Microbiology*, and *Cell*). Search strings combined controlled vocabulary (MeSH terms, including "Drug Resistance, Microbial" and "Artificial Intelligence") with free-text keywords using Boolean operators. The core search string, adaptable across databases with minor syntax changes, was structured as:

("antimicrobial resistance" OR "antibiotic resistance" OR "drug resistance, bacterial") AND ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*") AND (diagnos* OR "drug discovery" OR "susceptibility testing" OR genomic* OR "clinical decision support")

Where a database's syntax did not support wildcard truncation, equivalent term variants were entered manually. English-language filters were applied where the database allowed it, and the reference lists of all eligible articles were hand-screened for additional relevant studies not captured by the primary electronic search — a step that, in practice, surfaced several foundational computational papers (e.g., DeepARG, HMD-ARG) that predate the review window but remain methodologically central to the field.

3.3 Eligibility Criteria

We considered a study eligible if it met all three of the following conditions: (a) it applied a machine learning or deep learning method to any stage of AMR diagnostics, resistome prediction, antimicrobial or antimicrobial-peptide discovery, or AI-enabled clinical decision support; (b) it reported at least one quantifiable performance metric — accuracy, AUROC, sensitivity, specificity, precision, or minimum inhibitory concentration accuracy, among others — or a clearly described discovery outcome; and (c) it was published as a peer-reviewed article, a systematic or narrative review, or a credible preprint with sufficient methodological detail to permit critical appraisal. We excluded studies that applied AI to questions unrelated to antimicrobial resistance, that reported no empirical or quantitative outcome at all, or that took the form of editorials, conference abstracts without accompanying full data, or non-English publications for which a reliable translation could not be obtained.

3.4 Study Selection and Data Extraction

Titles and abstracts were screened against the eligibility criteria above, with full-text review triggered whenever relevance could not be confidently determined from the abstract alone — a judgment call that, admittedly, leaves some room for reviewer discretion, though the pre-specified criteria were applied consistently throughout. For each included study, we extracted, wherever it was reported: pathogen species or genus studied; diagnostic or discovery modality (e.g., whole-genome sequencing, MALDI-TOF, Raman spectroscopy, virtual screening); the underlying AI/ML architecture; feature-representation strategy; dataset or sample size; validation scheme (internal cross-validation, external holdout, or prospective clinical validation); and the reported performance metric(s). These fields form the backbone of the synthesis tables presented in the Results (Tables 1–4) and allowed us to track, systematically, which studies relied on internal resampling alone versus those tested on genuinely external cohorts.

3.5 Data Synthesis Approach

Given the substantial heterogeneity in outcome metrics, pathogen types, and validation designs across the included literature, we judged a formal meta-analysis inappropriate and used a narrative, thematically organized synthesis instead. Findings were grouped into four analytical domains that emerged, reasonably naturally, from the literature itself rather than being imposed on it beforehand: rapid phenotypic diagnostics, genomic and metagenomic resistome prediction, explainable AI and model interpretability, and de novo antimicrobial discovery. Within each domain, we prioritized studies reporting external or prospective validation over those relying solely on internal cross-validation, and we noted explicitly, wherever apparent, indicators of retrospective bias, dataset imbalance, or limited generalizability — consistent with recommended practice for critically appraising AI-based diagnostic and prognostic research (Collins et al., 2015).

3.6 Quality and Risk-of-Bias Considerations

Because the included studies spanned diagnostic accuracy research, computational drug-discovery pipelines, and health-services implementation reports, no single standardized risk-of-bias instrument applied cleanly across all of them. We instead assessed each study qualitatively against domains adapted from the PROBAST (Prediction model Risk Of Bias ASsessment Tool) framework (Wolff et al., 2019) — participant/sample selection, predictor and outcome definition, and analysis methodology — flagging studies that relied exclusively on internal resampling, drew on small single-center cohorts, or did not report external validation as carrying a higher risk of overestimated performance. This appraisal directly informed the translational-gap analysis presented in the Discussion, and, we think, is what keeps this review from reading as simply a celebration of technical progress.

4. Results

4.0 Overview of the Empirical Landscape

Taken together, the convergence of high-throughput biotechnology, digital pathology, and machine learning has, over a relatively compressed span of years, established what looks like a genuinely robust empirical baseline for transforming AMR management (Sardar et al., 2026). Moving away from purely observational approaches, next-generation AI pipelines show reproducible progress across four fronts: compressing diagnostic timelines, predicting genomic resistance phenotypes, offering biological explainability, and executing de novo molecular discovery (Chines et al., 2026; Liao et al., 2025). Tables 1 through 4 summarize the underlying study-level evidence referenced throughout this section — model performance for resistance detection (Table 1), computational drug-discovery platforms (Table 2), explainable AI methods (Table 3), and multi-omics diagnostic paradigms (Table 4). What follows synthesizes the evidence documented across the reviewed literature, organized around these same four dimensions.

4.1 AI-Driven Rapid Phenotypic Diagnostics and Digital Microbiology

The clinical value of any diagnostic stewardship program is, in the end, bounded by how fast it delivers an answer — and in critical care, that bound carries real mortality weight: delayed appropriate antibiotic therapy against ESKAPE pathogens raises mortality risk by approximately 7.6% per hour (Sardar et al., 2026). AI chips away at that constraint by sidestepping the slow biology of overnight cell division altogether (Liao et al., 2025); Figure 1 illustrates just how far that compression now extends relative to conventional culture-based testing, and Table 4 situates these phenotypic gains within the broader multi-omics diagnostic landscape.

One clear line of empirical progress involves automating

Gram-stain interpretation and colony-morphology tracking — tasks that, done manually, are operator-dependent and prone to fatigue-related error (Hassan & Zughaier, 2026). Deep convolutional neural networks trained on roughly 5,000 images achieved bacterial classification accuracies up to 94.0% (Khangarot et al., 2026), and Vision Transformers applied to the same data pushed performance further still, to an AUC of 0.952 — a gain that seems attributable to their capacity for capturing global contextual patterns that local convolutional kernels tend to miss (Khangarot et al., 2026). Swin Transformers applied to more complex colony-morphology datasets (4,200 images across more than ten morphotypes) produced an 8.0–10.0% accuracy improvement over standard CNN baselines under strict external validation (Khangarot et al., 2026, Table 1).

Beyond image-based analysis, AI has repurposed routinely generated diagnostic data streams — MALDI-TOF mass spectrometry and single-cell Raman spectroscopy chief among them — into genuinely fast resistance-classification systems (Aggarwal et al., 2026). MALDI-TOF peak-intensity classifiers, paired with artificial neural networks or gradient-boosted trees, discriminate carbapenem-resistant from carbapenem-susceptible *Klebsiella pneumoniae* isolates with AUROC values exceeding 0.85–0.90 within minutes of colony isolation (Liao et al., 2025; Scaglione et al., 2026). Single-cell Raman spectroscopy captures dynamic vibrational shifts within one to two hours of antibiotic exposure, reflecting metabolic perturbation well before macroscopic growth becomes visible (Liao et al., 2025); deep CNNs processing these spectral signatures predict phenotypic susceptibility directly from pre-incubation or flagged blood cultures at accuracies exceeding 90.0%, compressing a once 48–72-hour susceptibility profile into a point-of-care result available in under two to four hours (Liao et al., 2025, Figure 1; Table 4).

4.2 Genomic and Metagenomic Prediction of the Resistome

Phenotypic assays remain, in most laboratories, the gold standard for susceptibility testing, but genotypic modelling through next-generation sequencing offers something phenotypic testing simply cannot: a comprehensive survey of the molecular resistome itself (Sardar et al., 2026). The shift from rule-based alignment (BLAST-style homology matching against CARD or ResFinder) toward deep sequence representation learning has expanded this predictive capacity considerably (Khangarot et al., 2026), as reflected across the genomic classifiers summarized in Table 1.

Classical supervised models — random forests and support vector machines trained on manually engineered k-mer frequency matrices — performed acceptably but proved fairly vulnerable to sequence-level noise, assembly error, and poor generalization across divergent strain lineages (Sardar et al., 2026). Species-aware nucleotide language models such as DNABERT have since addressed much of this fragility: through self-supervised pretraining on large unannotated genomic corpora, DNABERT captures long-range genomic dependencies and epistatic interactions that k-mer-based models simply cannot see (Khangarot et al., 2026). In head-to-head comparisons, DNABERT outperformed random-forest models by 15.0% in species-level differentiation and achieved a 12.0–18.0% absolute improvement in classification precision and recall over BiLSTM networks across 500 diverse bacterial genome datasets (Khangarot et al., 2026, Figure 2) — a gap wide enough to suggest a genuine architectural advantage rather than incidental variation, and broadly consistent with the pan-genome and k-mer-based classifiers reported in Table 1.

Metagenomic ARG detection presents its own bottleneck: identifying highly divergent resistance genes directly from fragmented clinical or environmental reads, a task where standard best-hit homology matching tends to produce a troubling rate of false negatives for novel variants (Sardar et al., 2026). Treating ARG detection as a multi-label classification problem, DeepARG, trained on over 50,000 reference sequences using deep convolutional architectures, achieved classification precisions exceeding 0.97 across 30 major antibiotic classes, though performance dropped noticeably for rare ARG categories with sparse database representation (Arango-Argoty et al., 2018; Khangarot et al., 2026). The HMD-ARG framework — a hierarchical convolutional network paired with attention mechanisms — was built specifically to address that class-imbalance weakness; on a dataset of 60,000 genomic and metagenomic contigs under rigorous external validation, it delivered a 10.0–15.0% precision-and-recall improvement over DeepARG, successfully separating closely overlapping subclass designations that DeepARG tended to conflate (Khangarot et al., 2026, Table 1).

Genotype alone, however, is rarely the whole story, since

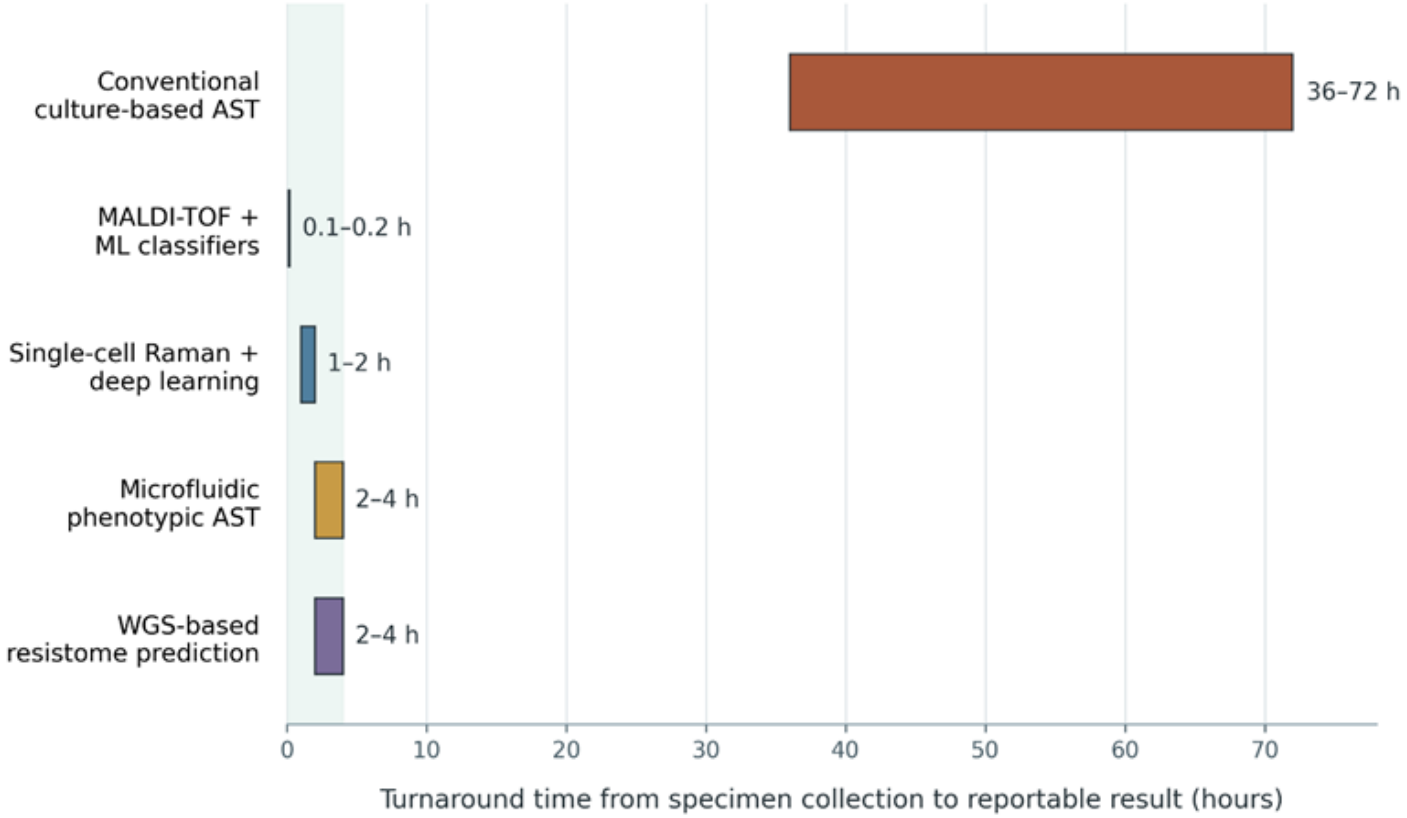

Figure 1. AI compresses antimicrobial susceptibility testing from days to hours. Comparison of turnaround time — from specimen collection to a reportable susceptibility result — across conventional culture-based antimicrobial susceptibility testing (AST) and four AI-enabled diagnostic modalities: MALDI-TOF mass spectrometry paired with machine-learning peak classifiers, single-cell Raman spectroscopy, microfluidic phenotypic AST, and whole-genome-sequencing-based resistome prediction. Conventional culture-based AST requires 36–72 hours because it depends on overnight bacterial growth; the AI-enabled modalities shown here return a result in under 2–4 hours by bypassing that growth step and instead classifying resistance from spectral, morphological, or sequence-level signal. Bars represent the typical reported turnaround-time range for each modality as documented across the studies synthesized in Table 4; this is an illustrative synthesis of reported ranges rather than a head-to-head prospective comparison, since no single study in this review tested all five modalities concurrently on the same specimen set.

gene presence does not guarantee functional expression (Rahman & Wakeman, 2026). Multimodal neural networks integrating genomic markers with transcriptomic RNA-seq data and clinical EHR metadata have begun to close that gap (Chines et al., 2026): modelling host–pathogen–drug dynamics at a systems level across 10,000 patients, multimodal transformers achieved AUROC values of 0.90–0.92 in predicting multidrug-resistant clinical outcomes — a meaningful improvement over genomics-only classifiers (Khangarot et al., 2026).

4.3 Explainable AI and Model Interpretability in Clinical Workflows

The adoption of high-capacity deep learning models in infectious disease medicine remains, in practice, constrained by their opacity (Sardar et al., 2026). Clinicians, who carry legal and ethical accountability for prescribing decisions, are understandably reluctant to act on recommendations derived from computational steps they cannot inspect (Sardar et al., 2026). Explainable AI (XAI), in this context, functions less as an optional add-on than as a strategic bridge between raw model accuracy and trustworthy, actionable stewardship (Chines et al., 2026); Table 3 summarizes the mathematical basis, clinical application, and inherent limitations of each method discussed below.

Feature-attribution methods remain the principal tool for taking these complex models apart in an interpretable way (Sardar et al., 2026). LIME, a post-hoc local explanation technique, approximates the decision boundary of a deep network locally using a simple, interpretable linear surrogate (Ribeiro et al., 2016); it is genuinely useful for explaining patient-specific recommendations but suffers from instability and offers limited global, dataset-wide insight (Sardar et al., 2026). SHAP, grounded in cooperative game theory, tends to fare better on both counts, providing more consistent, globally stable feature-importance attributions by calculating each genomic marker's, taxon's, or spectral peak's marginal contribution to the final prediction (Lundberg & Lee, 2017; Sardar et al., 2026).

Encouragingly, these methods have connected algorithmic output to established microbiological mechanisms rather than producing explanations that merely sound plausible. In genomic resistance prediction, SHAP-guided regularization frameworks reliably highlight biologically confirmed mutations — single-nucleotide polymorphisms within the quinolone-resistance-determining regions of *gyrA* or *parC* — as the dominant drivers of predicted fluoroquinolone resistance (Khangarot et al., 2026). In high-throughput single-cell spectral assays, attention maps visualize the exact mass-to-charge ratios or vibrational wave numbers associated with cell-wall peptidoglycan remodelling, letting clinicians confirm the model is responding to authentic drug-induced physiological change rather than background noise or batch artifact (Sardar et al., 2026, Table 3).

4.4 AI-Enabled De Novo Drug Discovery and Target Prioritization

The value of even the best diagnostic and genomic stewardship is ultimately capped by the therapeutic options available to act on it. To counter the stagnation of traditional small-molecule discovery, deep generative AI and structural biology models have become disruptive accelerators for both compound design and target identification (Abdulrazaq et al., 2025; Sanapalli et al., 2026); several landmark results are catalogued alongside their computational architectures in Table 2.

Halicin, discovered using deep neural network activity-prediction classifiers trained on a library of FDA-approved and natural compounds, exhibits a genuinely novel mechanism of action — disrupting transmembrane proton motive force — and has shown potent in vivo broad-spectrum efficacy in murine models against pan-resistant *A. baumannii* and *Clostridioides difficile* (Stokes et al., 2020; Sanapalli et al., 2026). Machine learning-guided virtual screening combined with molecular docking separately enabled the discovery of abaucin, a highly selective, narrow-spectrum therapeutic that targets *A. baumannii* while sparing host commensal microbiota (Awan et al., 2024).

In target discovery specifically, next-generation pipelines now draw on transformer-based structural prediction tools such as AlphaFold2 and RoseTTAFold to model high-resolution protein–ligand and host–pathogen interactions (Chines et al., 2026). Incorporating these high-accuracy 3D protein structures into molecular docking algorithms produced an 18.0% improvement in virtual hit-rates relative to traditional template-free approaches (Khangarot et al., 2026) — a gain substantial enough to meaningfully shift the economics of early-stage screening.

Deep generative architectures — variational autoencoders, generative adversarial networks, and

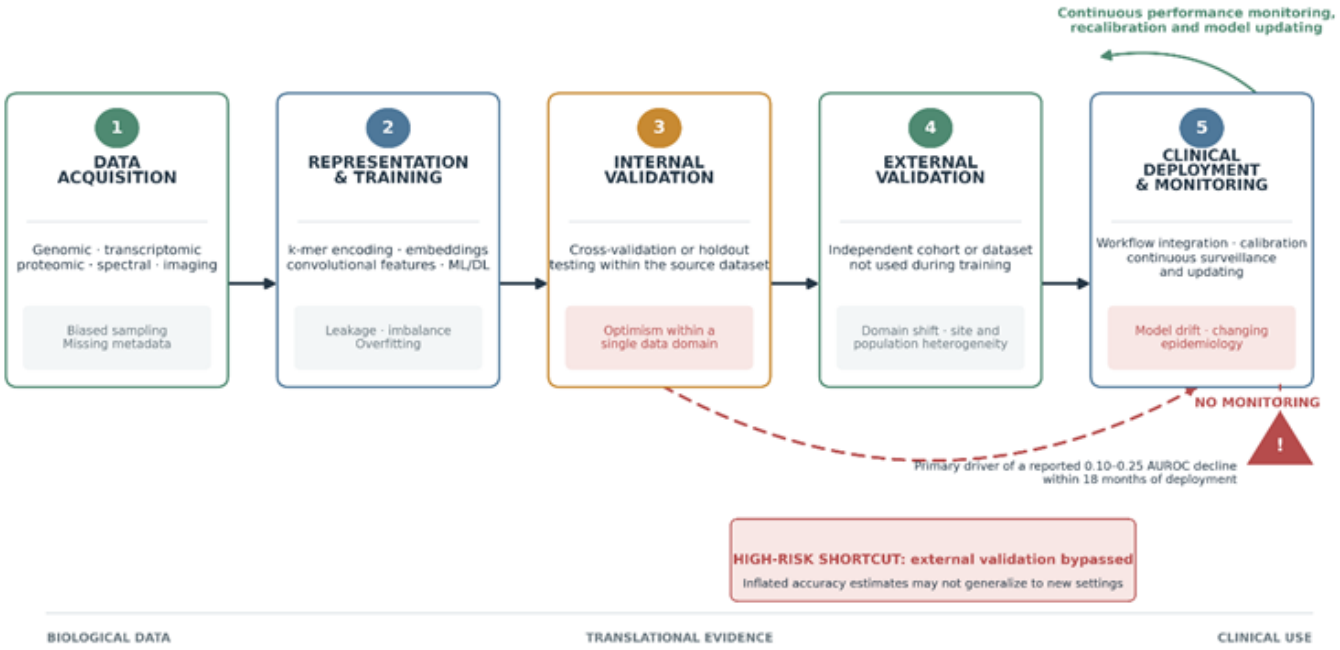

Figure 2. Conceptual pipeline for AI-driven AMR model development, with points of translational vulnerability flagged. A schematic of the standard workflow by which an AI/ML model moves from raw biological data to a clinically deployed diagnostic or discovery tool: (1) data acquisition (genomic, transcriptomic, proteomic, spectral, or imaging input); (2) feature representation and model training (e.g., k-mer encoding, sequence-language-model embedding, or convolutional feature extraction); (3) internal validation (cross-validation or holdout testing within the source dataset); (4) external validation (testing on an independent cohort or dataset not used in training); and (5) clinical deployment and post-deployment monitoring. Stages most vulnerable to overestimated performance or degraded real-world accuracy are marked: models that proceed directly from stage 3 to stage 5 without external validation (stage 4) are flagged as high risk of inflated accuracy estimates, and the absence of continuous post-deployment monitoring is flagged as the primary driver of the reported 0.10–0.25 AUROC decline observed within 18 months of deployment (Sardar et al., 2026). This figure is intended as an interpretive framework for the translational-gap analysis in Section 5, not as a depiction of any single published pipeline.

Table 1. Performance synthesis of AI/ML models for AMR detection and antimicrobial susceptibility prediction. Summarizes eight representative studies applying supervised machine learning or deep learning to resistance-phenotype or resistance-genotype prediction across clinically significant pathogens (Staphylococcus aureus, Pseudomonas aeruginosa, Escherichia coli, Klebsiella pneumoniae, Acinetobacter baumannii, Salmonella enterica, and multispecies PATRIC-database isolates). For each study, the table reports the input data type and model architecture, the validation scheme used (internal cross-validation, independent holdout, or prospective in vitro validation), and the headline performance metric(s) as reported by the original authors (accuracy, AUROC, Matthews correlation coefficient [MCC], sensitivity, specificity, or precision, depending on what each study reported). Studies are ordered by publication year rather than performance, since metrics are not directly comparable across differing pathogens, sample sizes, and validation designs. Validation scheme is highlighted deliberately: only two of the eight studies listed (Nguyen et al., 2019; Davis et al., 2016) report performance on a truly independent holdout set, while the remainder rely on internal cross-validation, a distinction discussed further in Section 5.1. WGS = whole-genome sequencing; RF = random forest; SVM = support vector machine; MCC = Matthews correlation coefficient; WT/NWT = wild-type/non-wild-type.

Study (APA)	Pathogen(s)	AI/ML Model	Validation Scheme	Reported Performance
Hyun et al. (2020)	S. aureus, P. aeruginosa, E. coli	SVM with Random Subspace Ensembles	Internal cross-validation (dissimilarity-based splits)	Accuracy: 79.3-99.5%; AUROC: 0.79-1.00; MCC: 0.39-0.95
Avershina et al. (2021)	K. pneumoniae, E. coli	Feed-Forward Artificial Neural Network	Internal cross-validation	Accuracy: 94.0-100% (WT/NWT); 91.0-99.0% (S/R status)
Jia et al. (2024)	A. baumannii	Deep Neural Network + Logistic Regression	Prospective in vitro validation (clinical MIC assays)	DNN accuracy: 87.6-98.6%; MIC accuracy: 86.2% (within +-1 dilution)
Kula et al. (2026)	P. aeruginosa	RF, LR, XGBoost, Naive Bayes (transcriptomic)	Internal cross-validation	Accuracy: 77.6-98.8%; Sensitivity: 68.4-100%; Specificity: 78.8-95.8%
Nguyen et al. (2019)	Salmonella enterica	Random Forest (WGS-derived gene matrix)	Independent holdout validation	Precision: 0.91-0.98
Moradigaravand et al. (2018)	Escherichia coli	Random Forest (pan-genome presence matrix)	Internal cross-validation	Accuracy >90.0% for beta-lactam susceptibility
Davis et al. (2016)	Multispecies (PATRIC database)	Set Cover Machine (reference-free k-mer counts)	Independent holdout test set	Accuracy >95.0% in designated drug-pathogen pairs

Table 2. AI-guided antibiotic and antimicrobial-peptide discovery: computational platforms and outcomes. Catalogues the principal machine-learning and deep-learning platforms used for de novo antimicrobial compound design, virtual screening, and antimicrobial-peptide (AMP) optimization discussed in Section 4.4, alongside the therapeutic target focus and headline outcome reported for each. Entries span three functional categories: general-purpose molecular-property prediction libraries (DeepChem, Chemprop), structure-based virtual screening tools (GNINA), and AMP-specific generative or discriminative architectures (AMPs-Net, AMPGen). Reported outcomes reflect in silico performance (e.g., bioactivity-prediction accuracy, pose-ranking improvement) unless otherwise indicated; where a platform's output has been validated experimentally (e.g., halicin, abaucin), that validation is described in the accompanying narrative text in Section 4.4 rather than restated here, to avoid duplicating in vitro/in vivo results that are not properties of the computational tool itself.

Tool / Platform	Architecture	Target Focus	Key Outcome	Reference
DeepChem	TensorFlow-based ML/DL library	Broad-spectrum pathogens	Standardized framework for QSAR modeling and toxicity prediction	Sanapalli et al. (2026)

Chemprop	Graph Neural Network (GNN)	Broad-spectrum pathogens	Outperformed traditional Random Forest in bioactivity prediction	Sanapalli et al. (2026)
GNINA	Convolutional Neural Network	Bacterial target proteins	Improved structural pose-ranking accuracy for ESKAPE targets	Sanapalli et al. (2026)
AMPs-Net	CNN + RNN hybrid	Antimicrobial peptides (AMPs)	High accuracy predicting cationic peptide motifs and activity	Sanapalli et al. (2026)
SyntheMol	Reinforcement learning + MCTS	Acinetobacter baumannii	Designed novel, easily synthesizable antibiotic leads	Sanapalli et al. (2026)
Halicin	Deep Neural Network (DNN)	A. baumannii, enterobacterales	Discovered structurally novel antibiotic with new mechanism of action	Stokes et al. (2020)
Abaucin	ML-guided screening + docking	Acinetobacter baumannii	Identified narrow-spectrum candidate highly potent against MDR strains	Awan et al. (2024)
AMP-Designer	LLM-based foundation model	Gram-negative & Gram-positive bacteria	Generated 18 candidates in 11 days; 94.4% active in vitro, 2 active in vivo	Sardar et al. (2026)
AMPGen	Evolutionary generator + XGBoost	ESKAPE pathogens & MDR fungi	Multi-level framework optimizing activity, charge, and host toxicity	Wu-Mo et al. (2026)

Table 3. Explainable AI (XAI) methods applied to AMR research: mechanism, application, and limitations. Compares eight interpretability techniques used to render AI-based AMR models auditable to clinicians and researchers, organized by explanation type — intrinsic (built into the model architecture, e.g., attention weights, rule-based classifiers) versus post-hoc (applied after training to an already-fit model, e.g., LIME, SHAP, integrated gradients). For each method, the table specifies the AMR-specific application in which it has been deployed (e.g., genotype-to-phenotype classification, WGS-derived biomarker discovery, chemical-scaffold rationalization), its principal technical limitation, and the source study. This table underpins the discussion of interpretability standardization in Section 5.3; note that no single method listed here has been formally adopted as a regulatory-submission requirement, which is itself part of the translational gap this review identifies. NCA = neighbouring component analysis; RFE = recursive feature elimination; LIME = local interpretable model-agnostic explanations; SHAP = Shapley additive explanations.

Method	**Type**	**AMR Application**	**Key Limitation**	**Reference**
Rule-Based (Kover)	Inherently interpretable	Genotype-to-phenotype classification	Low model expressiveness; misses complex epistatic pathways	Drouin et al. (2019)
LIME	Post-hoc local	Individual prescribing risk interpretation	Instability of local surrogates; lacks global interpretability	Ribeiro et al. (2016)
SHAP	Post-hoc global & local	WGS-derived ARG and biomarker discovery	High computational cost; assumes feature independence	Lundberg & Lee (2017)
Attention Weights	Intrinsic sequence-level	Sequence-level functional motif discovery	Weights do not mathematically equal explanation	Lee et al. (2023)
Graph Rationales	Post-hoc substructure	Active chemical scaffold discovery	High search-space constraints; limited to molecular graph inputs	Wong et al. (2024)
NCA (Neighbouring Component Analysis)	Intrinsic feature selection	Sequence genotype-to-phenotype mapping	Highly sensitive to noise; expensive for large genomes	Avershina et al. (2021)
Feature Weighting & RFE	Hybrid wrapper/intrinsic	Pan-genomic target-level gene discovery	Risk of population-structure (clonal) bias	Hyun et al. (2020)
Enhanced Integrated Gradients	Post-hoc global	Mutational analysis in M. tuberculosis	Requires differentiable architectures	Jha et al. (2020)

conditional GANs — are simultaneously being deployed for the rational design of antimicrobial peptides targeting ESKAPE pathogens and *Candida auris* (Sanapalli et al., 2026; Wu-Mo et al., 2026). Where natural-source AMP discovery is hindered by high hemolytic toxicity and poor proteolytic stability, generative models instead learn the latent grammar of peptide–membrane interactions directly, optimizing charge distribution, hydrophobicity, and amphipathicity in silico. The AMPGen platform illustrates this well: an evolutionary information-preserved generator pretrained on UniClust30, paired with an XGBoost discriminator and an LSTM scorer, optimizes activity against specific multidrug-resistant strains (Wu-Mo et al., 2026). Across these generative pipelines, resulting peptides have achieved minimal inhibitory concentrations as low as 2–8 μg/mL in vitro, reduced bacterial burdens by 2.3–3.0 log10, and measurably reduced murine mortality in experimental sepsis models (Chines et al., 2026) — findings that, taken together, suggest generative AMP design has moved well past a purely theoretical exercise (Table 2).

5. Discussion

5.0 Bridging the Translational Chasm

Reading back across the results assembled here, a somewhat paradoxical picture emerges. On one hand, the technical achievements are genuinely hard to dismiss: diagnostic timelines compressed from days to hours (Figure 1), resistome-prediction accuracies rivalling or exceeding classical alignment-based methods (Table 1), and generative pipelines that have produced antimicrobial peptides with confirmed in vivo activity (Table 2; Sanapalli et al., 2026; Chines et al., 2026). On the other hand — and this is the tension the field has not yet resolved — almost none of this evidence has been generated, or validated, under conditions that would satisfy a regulator, or, for that matter, a reasonably skeptical clinician standing at the bedside (Sardar et al., 2026). It seems worth sitting with that gap for a moment rather than rushing past it.

5.1 The Persistence of Retrospective, Single-Center Validation

Much of the literature synthesized here — and this is a limitation the field appears to acknowledge fairly openly — relies on internal cross-validation within homogeneous, single-center datasets (Aggarwal et al., 2026; Sardar et al., 2026). That is not a trivial methodological quibble; it inflates performance metrics in ways that dataset shift, temporal drift, and regional epidemiological variation can quietly unravel once a model is deployed elsewhere (Aggarwal et al., 2026), and the validation-scheme column in Table 1 makes plain just how few of the reviewed genomic classifiers were ever tested on truly external cohorts. The documented AUROC decline of 0.10–0.25 within eighteen months of deployment (Sardar et al., 2026) is, frankly, a sobering reminder that a model's reported accuracy at the moment of publication is closer to a snapshot than a guarantee. Continuous retraining pipelines and formal drift-detection protocols (Figure 2) look less like luxuries here than prerequisites for any AI-AMR tool intended for sustained clinical use.

5.2 Data Equity and the Geography of Resistance

A related, and arguably more troubling, concern is geographic. Global AMR databases skew heavily toward high-income, Western settings, leaving precisely the regions carrying the greatest resistance burden — sub-Saharan Africa and parts of Latin America among them — severely underrepresented in training data (Aggarwal et al., 2026; Sardar et al., 2026). This is not merely a fairness issue in the abstract; it carries direct clinical consequences, since a model trained predominantly on Western strain distributions may perform poorly, or worse, confidently and incorrectly, when applied to pathogen populations it has effectively never encountered (Sardar et al., 2026; Scaglione et al., 2026). Inconsistent breakpoint definitions across EUCAST and CLSI standards, together with a general lack of standardized wet-lab annotation, compound the problem further, introducing noise that even a sophisticated architecture cannot fully compensate for (Sardar et al., 2026; Branda & Scarpa, 2024; Alatawi et al., 2025).

5.3 The Interpretability Imperative

The black-box nature of deep learning architectures constitutes, in our reading of this literature, the single most consequential barrier to clinical adoption (Sardar et al., 2026; Branda & Scarpa, 2024). Given that delayed or incorrect antibiotic administration can raise septic shock mortality by approximately 7.6% per hour (Sardar et al., 2026; Alatawi et al., 2025), it is hard to fault clinicians for hesitating to trust recommendations they cannot interrogate. The XAI methods reviewed here — SHAP, attention maps, and feature-weighting approaches, summarized comparatively in Table 3 — offer a genuinely

Table 4. Multi-omics and multimodal diagnostic paradigms for pathogen profiling and outbreak surveillance. Outlines eight diagnostic and surveillance categories spanning the full multi-omics landscape — genomics, epigenomics, transcriptomics, proteomics, lipidomics, phenomics, integrated clinical-genomic multimodal systems, and One Health biosurveillance — together with the key enabling technology, primary clinical or public-health use case, and deployment-maturity stage for each. Deployment maturity is scored on a four-stage scale adapted from the translational framework used throughout this review: Stage 1 (retrospective proof-of-concept, tested only on archived data), Stage 2 (shadow mode, running alongside but not yet informing clinical decisions), Stage 3 (active influence, informing clinical or surveillance decisions under human oversight), and Stage 4 (regulated device, cleared for autonomous or semi-autonomous clinical use). As Section 5.4 discusses, only point-of-care diagnostic platforms and select phenomics tools have reached Stage 4 in the reviewed literature; the majority of multi-omics categories remain at Stage 1–3, underscoring that technical feasibility has, in most of these domains, outpaced regulatory clearance.

Omics / Diagnostic Category	Key Technology	Clinical Use Case	Deployment Maturity	Reference
Genomics	Whole-Genome Sequencing (WGS), SMRT	Rapid pathogen identification & resistome mapping	Stage 3: Active influence	Rahman & Wakeman (2026)
Epigenomics	Pacific Biosciences SMRT, Nanopore	Characterizing adaptive, non-sequence-based resistance	Stage 1: Retrospective PoC	Rahman & Wakeman (2026)
Transcriptomics	Shotgun RNA-Seq, RT-qPCR arrays	Monitoring immediate bacterial response to drug stress	Stage 2: Shadow mode	Rahman & Wakeman (2026)
Proteomics	LC-MS/MS, targeted proteomic profiling	Verifying functional translation of genomic targets	Stage 1: Retrospective PoC	Rahman & Wakeman (2026)
Lipidomics	LC-MS/MS lipid profiling, MALDIxin	Rapid detection of colistin/polymyxin resistance	Stage 2: Shadow mode	Rahman & Wakeman (2026)
Phenomics	Microfluidic AST, single-cell morphometrics	Ultra-rapid culture-free susceptibility testing	Stage 4: Regulated device	Rahman & Wakeman (2026)
Clinical-Genomic Multimodal	Integrated WGS + EHR spatial-temporal data	Real-time hospital-acquired outbreak surveillance	Stage 3: Active influence	Scaglione et al. (2026)
One Health Biosurveillance	Wastewater metagenomics + GIS mapping	Early-warning public health surveillance networks	Stage 3: Active influence	Scaglione et al. (2026)
Point-of-Care Diagnostics	Smartphone-connected PCR, paper microfluidics	Rapid resistance-gene detection in resource-limited settings	Stage 4: Regulated device	Hassan et al. (2026)

promising path forward, having already demonstrated the ability to map model predictions onto biologically confirmed mechanisms such as *gyrA* and *parC* mutations in fluoroquinolone resistance (Khangarot et al., 2026; Sardar et al., 2026). What remains missing, though, is standardization: no consensus yet exists on which XAI method, or combination of methods, should be required for regulatory submission, and this ambiguity likely contributes to the field's current stall at the pre-clearance stage (Sardar et al., 2026).

5.4 Regulatory and Infrastructural Bottlenecks

Perhaps the starkest fact to emerge from this synthesis is this: despite everything described above, no AI-based AMR tool has, to date, received regulatory clearance from a body such as the FDA (Sardar et al., 2026). Existing regulatory pathways were built for static software, and they are not well suited to evaluating continuously adaptive machine learning models — tools whose entire value proposition rests on their capacity to keep learning as resistance patterns shift (Sardar et al., 2026; Alatawi et al., 2025). Integrating these tools into existing hospital IT infrastructure and electronic health record systems introduces further friction: interoperability challenges, compatibility issues, and data-privacy obligations under frameworks such as GDPR and HIPAA that were, again, not designed with adaptive AI in mind (Alatawi et al., 2025).

5.5 Toward a Standardized Roadmap for Clinical Translation

None of this, it should be said, amounts to a reason for pessimism, so much as a reasonably clear list of what needs to happen next. Federated learning offers one plausible route around the data-scarcity and privacy problems simultaneously, allowing multi-institutional model training without centralizing sensitive patient data — an approach that could, in principle, address both the geographic representation gap and HIPAA/GDPR compliance concerns at once. One Health–linked surveillance networks, similarly, could extend the six-month early-warning capacity already demonstrated in wastewater resistome monitoring (Abdul Ghafur et al., 2026; Scaglione et al., 2026) into a genuinely global, continuously updated resistance-forecasting system (Table 4). And standardized XAI reporting requirements, akin to what CONSORT-AI and related frameworks have begun to establish for other clinical AI domains, would give regulators a consistent basis for evaluating whether a model's reasoning is auditable enough to trust. Taken together, these three pillars — federated data-sharing, One Health surveillance integration, and standardized interpretability reporting — sketch out what a realistic path from retrospective proof-of-concept to prospective clinical deployment might actually look like.

5.6 Limitations of This Review

This synthesis is, admittedly, not without its own limitations. As a narrative-systematic review rather than a formal meta-analysis, it cannot statistically pool effect sizes across the heterogeneous outcome metrics reported in the underlying literature, and its reliance on published, largely English-language sources may itself introduce a degree of publication and language bias — an irony not lost on us, given that data inequity is one of the central concerns raised throughout this discussion. Future prospective, multi-site validation studies, ideally registered and reported according to emerging AI-specific reporting standards, will be needed to confirm whether the performance gains documented here in fact translate to real-world, external clinical settings.

6. Conclusion

Taken together, the evidence assembled here suggests AI has already done what it was first asked to do in the AMR space: it has shown, repeatedly and across modalities, that resistance can be predicted, diagnosed, and in some cases outrun computationally. What it has not yet done, at least not convincingly, is prove itself at the bedside, across diverse populations, or under regulatory scrutiny. Closing that gap will likely depend less on ever more powerful architectures than on unglamorous, structural work — prospective multi-site trials, harmonized data standards, federated learning across institutions, and explainability frameworks that clinicians can actually act on rather than merely admire. None of this is a small undertaking, and none of it happens quickly. But if those pieces do come together, and there is at least reasonable cause for cautious optimism that they will, AI-AMR tools may finally cross the threshold from promising literature into the diagnostic workflows and hospital formularies where they are genuinely needed.

Acknowledgements

The authors N. Amin et al., thank the Department of Public Health, Anwer Khan Modern University, Dhaka,

Bangladesh, for institutional support during the preparation of this review. No dedicated funding was received for this work, and no professional medical writers were engaged.

Author Contributions

N. Amin: conceptualization, literature search, data extraction, writing – original draft, writing – review & editing, supervision. P.S.S. Sharna: literature search, data extraction, writing – original draft, writing – review & editing. Both authors read and approved the final manuscript.

Competing Financial Interests

The authors N. Amin et al., declare no competing financial interests or non-financial competing interests related to this work.

References

Abdul Ghafur, A., Nagao, M., Kanj, S. S., Abbara, S., Tattevin, P., Awad, M., Dagher, M., Ndegwa, L. K., Mundia, G., Schouten, J., Vu, H. T. L., Hara, G. L., Satyaseela, M. P., Aldeyab, M. A., Bhat, R. K., Bansal, N., & Wertheim, H. F. (2026). Artificial intelligence and antimicrobial stewardship: From clinical decision support to public health action. Journal of Global Antimicrobial Resistance, 28, 1–15. https://doi.org/10.1016/j.jgar.2026.08.003

Abdulrazaq, I., Elelu, S.-A., Ibrahim, G. O., Temitope, I. A., Avoswahi, A. H., Zakari, A. T., & Abdulsalam, M. (2025). Advancements in microbial drug discovery: Leveraging AI, CRISPR, and microbiome insights to overcome antimicrobial resistance. Biological Sciences, 1, 1–15. https://doi.org/10.55006/biolsciences.2025.5304

Abo-Tbeak, Z. J., & AL-Hilali, R. H. (2026). Artificial intelligence in microbiology: Applications in microbial data analysis and drug discovery. World of Medicine: Journal of Biomedical Sciences, 3(5), 45–52.

Aggarwal, R., Nagre, S., & Wardell, S. J. (2026). Artificial intelligence for antimicrobial resistance detection and prediction. Infection and Drug Resistance, 19, 614–240. https://doi.org/10.2147/IDR.S614240

Ahmed, N., & Alghamdi, M. (2026). Role of artificial intelligence in infectious diseases and antimicrobial resistance: A comprehensive review on diagnostic, treatment, and prevention aspects. Saudi Medical Journal, 47(2), 219–237. https://doi.org/10.15537/1658-3175.1003

Alatawi, A. D., Hetta, H. F., Ali, M. A. S., Ramadan, Y. N., Alaqyli, A. B., Alansari, W. K., Aldhaheri, N. H., Bin Selim, T. A., Merdad, S. A., Alharbi, M. O., Alatawi, W. A. H., & Algammal, A. M. (2025). Diagnostic innovations to combat antibiotic resistance in critical care: Tools for targeted therapy and stewardship. Diagnostics, 15, 2244. https://doi.org/10.3390/diagnostics15172244

Antonie, N. I., Gheorghe, G., Ionescu, V. A., Tiuca, L. C., & Diaconu, C. C. (2025). The role of ChatGPT and AI chatbots in optimizing antibiotic therapy: A comprehensive narrative review. Antibiotics, 14(1), 60. https://doi.org/10.3390/antibiotics14010060

Arango-Argoty, G., Garner, E., Pruden, A., Heath, L. S., Vikesland, P., & Zhang, L. (2018). DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome, 6(1), 23. https://doi.org/10.1186/s40168-018-0401-z

Avershina, E., Sharma, P., Taxt, A. M., Singh, H., Frye, S. A., Paul, K., Kapil, A., Naseer, U., Kaur, P., & Ahmad, R. (2021). AMR-Diag: Neural network based genotype-to-phenotype prediction of resistance towards -lactams in Escherichia coli and Klebsiella pneumoniae. Computational and Structural Biotechnology Journal, 19, 1896–1906. https://doi.org/10.1016/j.csbj.2021.03.027

Awan, R. E., Zainab, S., Yousuf, F. J., & Mughal, S. (2024). AI-driven drug discovery: Exploring Abaucin as a promising treatment against multidrug-resistant Acinetobacter baumannii. Health Science Reports, 7, e2150. https://doi.org/10.1002/hsr2.2150

Bhattacharjee, S., & Bhattacharya, S. (2025). Leveraging AI-driven nudge theory to enhance hand hygiene compliance: Paving the path for future infection control. Frontiers in Public Health, 12, 1522045. https://doi.org/10.3389/fpubh.2024.1522045

Bhattacharya, R., Bose, D., Rodriguez, R. V., Kaur, T., & Pillai, M. (2026). A One Health plan to combat antimicrobial resistance for improving global health through sustainable development. Discover Public Health, 23(10), 1–17. https://doi.org/10.1186/s12982-025-01302-1

Branda, F., & Scarpa, F. (2024). Implications of artificial intelligence in addressing antimicrobial resistance: Innovations, global challenges, and healthcare's future. Antibiotics, 13(6), 502. https://doi.org/10.3390/antibiotics13060502

Chines, E., Tempesta, A. A., Boscarelli, L., Parisi, M. F., Marcoccia, L., Capillo, A., Mezzatesta, M. L., Ledda, C., Chessari, M., & Cafiso, V. (2026). Next-generation target discovery in ESKAPE pathogens: An AI-driven framework from omics-based to systems-level modeling and clinical translation. Antibiotics, 15(5), 469. https://doi.org/10.3390/antibiotics15050469

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. BMJ, 350, g7594. https://doi.org/10.1136/bmj.g7594

Davis, J. J., Boisvert, S., Brettin, T., Kenyon, R. W., Mao, C., Olson, R., Overbeek, R., Santerre, J., Shukla, M., Wattam, A. R., Will, R., Xia, F., & Stevens, R. (2016). Antimicrobial resistance prediction in PATRIC and RAST. Scientific Reports, 6, 27930. https://doi.org/10.1038/srep27930

Drouin, A., Letarte, G., Raymond, F., Marchand, M., Corbeil, J., & Laviolette, F. (2019). Interpretable genotype-to-phenotype classifiers with performance guarantees. Scientific Reports, 9, 4071. https://doi.org/10.1038/s41598-019-40561-2

Hassan, M., Ali, R., & Zughaier, S. M. (2026). Role of artificial intelligence in bacterial diagnostics, surveillance, and clinical decision support. Digital Health, 12, 1–13.

Hudu, S. A., Alshrari, A. S., Abu-Shoura, E. J. I., Osman, A., & Jimoh, A. O. (2025). A critical review of the prospect of integrating artificial intelligence in infectious disease diagnosis and prognosis. Interdisciplinary Perspectives on Infectious Diseases, 2025, 6816002. https://doi.org/10.1155/2025/6816002

Hyun, J. C., Kavvas, E. S., Monk, J. M., & Palsson, B. O. (2020). Machine learning with random

subspace ensembles identifies antimicrobial resistance determinants from pan-genomes of three pathogens. PLoS Computational Biology, 16, e1007608. https://doi.org/10.1371/journal.pcbi.1007608

Jha, A., Aicher, J. K., Gazzara, M. R., Singh, D., & Barash, Y. (2020). Enhanced integrated gradients: Improving interpretability of deep learning models using splicing codes as a case study. Genome Biology, 21, 149. https://doi.org/10.1186/s13059-020-02055-7

Jia, H., Li, X., Zhuang, Y., Wu, Y., Shi, S., Sun, Q., He, F., Liang, S., Wang, J., & Draz, M. S. (2024). Neural network-based predictions of antimicrobial resistance phenotypes in multidrug-resistant Acinetobacter baumannii from whole genome sequencing and gene expression. Antimicrobial Agents and Chemotherapy, 68, e01446-24. https://doi.org/10.1128/aac.01446-24

Kasse, G. E., Cosh, S. M., Humphries, J., & Islam, M. S. (2025). Leveraging artificial intelligence for One Health: Opportunities and challenges in tackling antimicrobial resistance—A scoping review. One Health Outlook, 7(1), 51. https://doi.org/10.1186/s42522-025-00170-8

Khangarot, R., Kumari, V., Mishra, R., & Singh, A. (2026). Artificial intelligence in microbiology: Implications for metagenomics, diagnostics, and AMR surveillance. BioMed Engineering OnLine, 25, 1–18. https://doi.org/10.1186/s12938-026-01568-9

Kim, J. I., Manuele, A., Maguire, F., Zaheer, R., McAllister, T. A., & Beiko, R. G. (2024). Identification of key drivers of antimicrobial resistance in Enterococcus using machine learning. Canadian Journal of Microbiology, 70, 446–460. https://doi.org/10.1139/cjm-2024-0012

Kula, C., Erguven, I., Ozcelik, B., Gulfidan, G., & Arga, K. Y. (2026). Transcriptome-based machine learning models to predict antimicrobial resistance in Pseudomonas aeruginosa. OMICS: A Journal of Integrative Biology, 30, 94–104. https://doi.org/10.1089/omi.2026.0022

Lee, H., Lee, S., & Lee, I. (2023). AMP-BERT: Prediction of antimicrobial peptide function based on a BERT model. Protein Science, 32, e4529. https://doi.org/10.1002/pro.4529

Liao, H., Xie, L., Zhang, N., Lu, J., & Zhang, J. (2025). Advancements in AI-driven drug sensitivity testing research. Frontiers in Cellular and Infection Microbiology, 15, 1560569. https://doi.org/10.3389/fcimb.2025.1560569

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

Maldonado-Hernández, R., Ortiz-Gómez, V., Flores-González, A., Meléndez-Delgado, J., Meléndez-González, H., & Wu-Mo, C. (2026). Artificial intelligence-driven discovery and optimization of antimicrobial peptides targeting ESKAPE pathogens and multidrug-resistant fungi. Microorganisms, 14, 591. https://doi.org/10.3390/microorganisms14030591

Mohammed, A. M., Mohammed, M., Oleiwi, J. K., Osman, A. F., Adam, T., Betar, B. O., Gopinath, S. C. B., & Ihmedee, F. H. (2025). Enhancing antimicrobial resistance strategies leveraging artificial intelligence: Opportunities, challenges, and future prospects. South African Journal of Chemical Engineering, 51, 272–286. https://doi.org/10.1016/j.sajce.2024.12.005

Moradigaravand, D., Palm, M., Farewell, A., Mustonen, V., Warringer, J., & Parts, L. (2018). Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data. PLoS Computational Biology, 14, e1006258. https://doi.org/10.1371/journal.pcbi.1006258

Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., … Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. The Lancet, 399(10325), 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0

Nagre, S., Dole, V., Kharat, A., Kharat, R., Musadwale, S., Solanke, P., Sonone, R., & Nagrik, S. (2025). Artificial intelligence in infectious disease pharmacology: Transforming drug discovery, precision therapy, and global health. International Journal of Scientific Research and Technology, 2(9), 1–12.

Nguyen, M., Long, S. W., McDermott, P. F., Davis, J. J., Olsen, R. J., Sanderson, R., & Butler, J. E. (2019). Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella. Journal of Clinical Microbiology, 57, 1–15. https://doi.org/10.1128/JCM.01260-18

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Rahman, M. S., & Wakeman, C. A. (2026). Antibiotic resistance in Pseudomonas aeruginosa: Mechanisms, diagnostic challenges, and omics-based diagnostic solutions. Frontiers in Microbiology, 17, 1791577. https://doi.org/10.3389/fmicb.2026.1791577

Rahman, M. U., Shah, J. A., Khan, M. N., Bilal, H., Zhu, D., Du, Z., & Mu, D.-S. (2025). Innovative approaches to combat antimicrobial resistance: A review of emerging therapies and technologies. Probiotics and Antimicrobial Proteins. https://doi.org/10.1007/s12602-025-10676-2

Rahman, S. Z., Senthil, R., Ramalingam, V., & Gopal, R. (2023). Predicting infectious disease outbreaks with machine learning and epidemiological data. Journal of Advanced Zoology, 44, 110–121.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

Salama, R. A., Nagre, S., Dole, V., Kharat, A., Kharat, R., Musadwale, S., Solanke, P., Sonone, R., & Nagrik, S. (2026). Artificial intelligence in infectious disease pharmacology: Transforming drug discovery, precision therapy, and global health. Infection Prevention in Practice, 8, 100522. https://doi.org/10.1016/j.infpip.2026.100522

Sanapalli, B. K. R., Palit, S., Deshpande, A., Tokala, R., Sigalapalli, D. K., & Sanapalli, V. (2026). Artificial intelligence and the discovery of antibiotics: Reinventing with opportunities, challenges, and clinical translation. Antibiotics, 15, 233. https://doi.org/10.3390/antibiotics15020233

Scaglione, G., Mastroianni, N., Rizzo, A., Palomba, E., Carcione, D., Rimoldi, S. G., Brigante, G., Principe, L., Colaneri, M., Gori, A., & Borgonovo, F. (2026). Integrating artificial intelligence with genome sequencing against antimicrobial resistance: A narrative review. Frontiers in Public Health, 14, 1757161. https://doi.org/10.3389/fpubh.2026.1757161

Sardar, S., Dash, S., Roychowdhury, P., Solanki, J., Sikdar, P., Thangappan, J., & Dutta, S. (2026). Artificial intelligence for antimicrobial resistance: Advancing reproducibility, interpretability, and clinical deployment. Briefings in Bioinformatics, 27, bbag269. https://doi.org/10.1093/bib/bbag269

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., Tran, V. M., Chiappino-Pepe, A., Badran, A. H., Andrews, I. W., Chory, E. J., Church, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R., & Collins, J. J. (2020). A deep learning approach to antibiotic discovery. Cell, 180(4), 688–702.e13. https://doi.org/10.1016/j.cell.2020.01.021

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 170(1), 51–58. https://doi.org/10.7326/M18-1376

Wong, F., Zheng, E. J., Valeri, J. A., Donghia, N. M., Anahtar, M. N., Omori, S., Li, A., Cubillos-Ruiz, A., Krishnan, A., Jin, W., Manson, A. L., Friedrichs, J., Helbig, R., Hajian, B., Fiejtek, D. K., Wagner, F. F., Soutter, H. H., Earl, A. M., Stokes, J. M., … Collins, J. J. (2024). Discovery of a structural class of antibiotics with explainable deep learning. Nature, 626, 177–185. https://doi.org/10.1038/s41586-023-06887-8

Wu-Mo, C., Flores-González, A., Meléndez-Delgado, J., Ortiz-Gómez, V., Meléndez-González, H., & Maldonado-Hernández, R. (2026). Artificial intelligence-driven discovery and optimization of antimicrobial peptides targeting ESKAPE pathogens and multidrug-resistant fungi. Microorganisms, 14, 591. https://doi.org/10.3390/microorganisms14030591