The Bioscan
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES
21(2): S.I (2), 1329-1343, 2026 www.thebioscan.com

Machine Learning Approaches for Predicting Antimicrobial Resistance from Genomic Data: A Systematic Review

Amol S. Jadhav[a], Pritam H. Mahadik[b], Sunil S. Barkade[c], Girish B. Pendharkar[d] , Tushar N. Karpude[e], Dr. Bandu S. Pawar[a*]

[a]Department of Microbiology, Yashavantrao Chavan Institute of Science, Satara 415 001, MH, India.
[b]Department of Statistics, Yashavantrao Chavan Institute of Science, Satara 415 001, MH, India.
[c]Department of Electronics, Yashavantrao Chavan Institute of Science, Satara 415 001, MH, India.
[d]Department of Microbiology, Sadguru Gadage Maharaj College, Karad-415124, MH, India.
[e]Department of Microbiology, Yogeshwari Mahavidyalaya, Ambajogai, Beed-431517, MH, India.
[*]Corresponding author: bspawar@ycis.ac.in

DOI:https://doi.org/10.63001/tbs.2026.v21.i02.S.I(2).pp1329-1343

Keywords:

Antimicrobial resistance, machine learning, genomic prediction, whole-genome sequencing, deep learning, antibiotic susceptibility, bioinformatics etc.,

Received on:
28-04-2026

Accepted on:

16-05-2026

Published on:

26-05-2026

Abstract

Antimicrobial resistance (AMR) represents a major global health challenge, significantly limiting the effective prevention and treatment of bacterial infections. Advances in whole-genome sequencing (WGS) have enabled rapid identification of genetic determinants associated with resistance. However, conventional rule-based approaches primarily rely on previously characterized resistance genes and mutations, restricting their ability to detect novel or complex resistance mechanisms. In this context, machine learning (ML) has emerged as a powerful alternative, leveraging high-dimensional genomic datasets to model complex and nonlinear relationships between genotype and phenotype. Over the past decade, ML-based approaches have been increasingly applied to predict antibiotic susceptibility directly from genomic information in clinically important pathogens such as *Mycobacterium tuberculosis*, *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, and *Salmonella* species. This systematic review summarizes recent advancements in genomic feature extraction, algorithm development, model validation strategies, and the clinical applicability of ML-driven AMR prediction tools. Various computational frameworks, including supervised, unsupervised, and deep learning models, are critically examined alongside benchmarking practices, validation methodologies, and model interpretability techniques. Emerging directions such as explainable artificial intelligence, federated learning, and multi-omics integration are also discussed for their potential to enhance predictive performance and clinical utility. Despite promising results in controlled settings, several challenges remain, including data heterogeneity, limited generalizability, regulatory constraints, and ethical considerations. Overall, this review provides a comprehensive perspective on the current state and future prospects of ML-based genomic diagnostics for antimicrobial resistance.

1. Introduction

Antimicrobial resistance (AMR) has emerged as a critical global challenge, posing serious threats to public health, food security, and economic stability [1]. Drug-resistant infections are currently

responsible for hundreds of thousands of deaths annually, and projections indicate that, in the absence of effective interventions, this number could rise to several million per year by 2050 [2]. The increasing prevalence of resistant pathogens undermines the efficacy of routine medical procedures, including surgical interventions, chemotherapy, and organ transplantation. Consequently, rapid and accurate antimicrobial susceptibility testing (AST) is essential for effective clinical management and infection control.

Traditional AST methods are primarily based on phenotypic assays such as disk diffusion, broth microdilution, and automated culture-based systems [3]. Although these techniques are considered reliable, they require prior isolation and cultivation of microorganisms, resulting in prolonged turnaround times ranging from 24 to 72 hours, or even longer in the case of slow-growing pathogens such as *Mycobacterium tuberculosis*. Such delays often necessitate the use of empirical therapy, which can further promote the selection and spread of resistant strains.

Whole-genome sequencing (WGS) has emerged as a promising alternative, enabling the identification of antimicrobial resistance determinants directly from genomic data [4]. Advances in high-throughput sequencing technologies have significantly reduced both cost and processing time, facilitating the integration of genomic approaches into clinical microbiology workflows. Several curated databases, including the Comprehensive Antibiotic Resistance Database (CARD), ResFinder, and ARG-ANNOT, provide extensive repositories of resistance-associated genes and mutations [5–7]. However, rule-based genomic prediction methods have inherent limitations, as they depend heavily on previously characterized resistance mechanisms and may fail to detect emerging or complex resistance patterns.

- Novel mutations
- Epistatic interactions
- Gene expression changes
- Structural genomic rearrangements
- Unknown resistance pathways

Machine learning (ML) techniques provide a data-driven framework for modelling intricate genotype–phenotype associations without necessitating specific biological postulates [8]. ML systems can combine thousands to millions of genetic information to accurately predict phenotypic resistance. In the past ten years, many research has used ML models to predict

resistance in clinically important pathogens [9–14]. This systematic review seeks to deliver a thorough synthesis of machine learning methodologies for antimicrobial resistance prediction utilizing genomic data, emphasizing methodological frameworks, performance evaluation, interpretability issues, and translational opportunities.

2. Methodology of Systematic Review

2.1 Literature Search Strategy

A thorough search of the literature was done in the PubMed, Scopus, and Web of Science databases for articles published between January 2010 and December 2024 shown in figure 1. Some of the search phrases were:

- "machine learning"
- "deep learning"
- "antimicrobial resistance"
- "genomic prediction"
- "whole genome sequencing"
- "antibiotic susceptibility"

2.2 Inclusion Criteria

Studies were included if they:

1. Used genomic sequencing data (WGS, SNP datasets, or k-mer representations)
2. Applied ML or deep learning models
3. Reported quantitative predictive performance metrics
4. Focused on bacterial pathogens

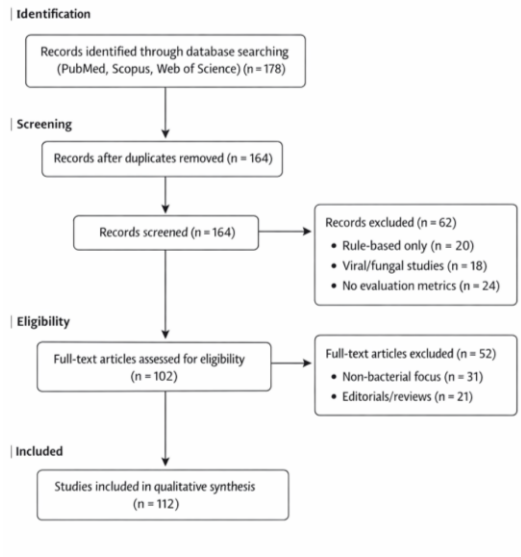

Figure 1. PRISMA Flow Diagram for Literature Selection in Machine Learning-Based Antimicrobial Resistance Prediction.

2.3 Exclusion Criteria

Studies were excluded if they:

1. Used only rule-based gene detection
2. Focused on viral or fungal pathogens
3. Lacked evaluation metrics
4. Were editorials or non-peer-reviewed articles

A total of 178 studies were screened, with 112 meeting inclusion criteria.

3. Genomic Data Sources and Feature Engineering

How genomic information is turned into numbers is really important for how well machine learning works.

3.1 Whole-Genome Sequencing Platforms

Short-read sequencing technologies, like Illumina, are the most common type of platform used in AMR prediction studies because they are accurate and cheap [15]. Long-read technologies like Oxford Nanopore and PacBio make it easier to put together plasmids and structural variants [16].

3.2 Gene Presence/Absence Encoding

Binary matrices indicating presence or absence of known resistance genes are commonly used features [17]. Although computationally efficient, this approach is constrained by prior knowledge and cannot detect novel resistance determinants.

3.3 SNP-Based Representations

Single nucleotide polymorphisms (SNPs) detect point mutations linked to resistance, especially significant for:

- Rifampicin resistance (rpoB mutations)
- Isoniazid resistance (katG mutations)
- Fluoroquinolone resistance (gyrA mutations)

SNP encoding may involve one-hot representation or allele frequency matrices [18].

3.4 k-mer Encoding

Genomic data representations that don't require alignment are provided by K-mers (substrings of length k, typically 9–31 nucleotides) [19]. K-mer techniques do not require reference genomes to identify variations across the entire genome. However, dimensionality can contain millions of properties, hence dimensionality reduction techniques such as:

- Principal component analysis (PCA)
- Feature hashing
- Information gain filtering

3.5 Structural Variants and Mobile Genetic Elements

Integrons, transposons, and plasmids are important components in horizontal gene transfer [20]. Predicting carbapenem resistance in Klebsiella pneumoniae is more accurate when ML models incorporate plasmid identifying features [21].

4. Machine Learning Algorithms

4.1 Supervised Learning Models

The most popular method for predicting AMR is supervised learning because to the abundance of tagged susceptibility data. Table 1 shows a comparison.

4.1.1 Random Forest (RF)

Random forests are tree-based classifiers that work well even when they are overfitting and can handle high-dimensional genomic features [22]. RF models have shown over 90% accuracy in predicting multidrug-resistant M. tuberculosis [23].

Advantages:

- Built-in feature importance
- Handles nonlinear interactions
- Robust to noise

Limitations:

- Reduced interpretability compared to linear models

4.1.2 Support Vector Machines (SVM)

When it comes to binary classification tasks, such as determining whether an isolate is susceptible or resistant based on genetic data, Support Vector Machines (SVMs) excel. They are particularly useful for SNP datasets with a large number of features (mutations) and few samples. By increasing the distance between classes, SVMs eliminate overfitting and improve generalization. Additionally, SVMs can simulate complex, nonlinear relationships between genomic variants and antimicrobial resistance phenotypes because to kernel functions. SVMs work well for SNP datasets with several dimensions and for categorizing two items [24]. Decision boundaries that are not straight lines are possible using kernel techniques.

Limitations include:

- Computational intensity in large datasets
- Hyperparameter sensitivity

4.1.3 Gradient Boosting and XGBoost

Boosting approaches reduce categorization errors sequentially [25]. When predicting multidrug resistance in Enterobacteriaceae, XGBoost often outperforms RF [26].

4.2 Deep Learning Approaches

4.2.1 Artificial Neural Networks (ANN)

Nonlinear genetic interactions can be described by artificial neural networks (ANNs), but they require a large amount of data [27].

4.2.2 Convolutional Neural Networks (CNN)

CNNs may recognize hierarchical patterns in DNA data when applied to k-mer-encoded genomic sequences. Convolutional filters are used by CNNs to automatically identify local sequence motifs in the initial layers, such as resistance-associated patterns or mutation signatures. In subsequent layers, they integrate these motifs into higher-level genetic representations. Without the need for manual feature engineering, CNNs can detect complex, non-linear relationships between sequence alterations and antimicrobial resistance phenotypes because to this hierarchical feature extraction [28].

4.2.3 Recurrent Neural Networks (RNN)

Researchers have looked into simulating sequential mutation patterns in genomic data using Recurrent Neural Networks (RNNs). Because RNNs are designed to deal with ordered sequences, they may identify connections between mutations that occur at various locations throughout the genome. Because of this, they are especially useful for determining the temporal or positional links between SNPs that may collectively contribute to antimicrobial resistance, particularly when resistance results from cumulative or interacting genetic alterations [29].

4.3 Ensemble and Hybrid Models

Models become more accurate and stable when ensemble frameworks include Random Forest (RF), Support Vector Machines (SVM), and boosting techniques. In order to reduce variation, prevent overfitting, and facilitate generalization across various genomic datasets, ensemble models integrate the outcomes of multiple techniques. Because diverse and multidimensional data can make single models less reliable for practical application, this hybrid approach is particularly useful for forecasting antimicrobial resistance [30].

Table 1. Comparison of Machine Learning Algorithms Used in Genomic AMR Prediction

Algorithm	Strengths	Limitations	Suitable Genomic Features	Typical Use Case
Random Forest (RF)	Robust to overfitting; handles high-	Moderate interpretability	SNPs, gene presence/absence	Medium-sized datasets

	dimensional data; feature importance available			
Support Vector Machine (SVM)	Effective in high-dimensional sparse data; strong generalization	Sensitive to hyperparameters	SNP datasets	Binary resistance prediction
Gradient Boosting (XGBoost)	High predictive accuracy; handles complex interactions	Risk of overfitting if not tuned	k-mers, SNPs	Multidrug resistance
Convolutional Neural Network (CNN)	Learns hierarchical sequence features	Requires large datasets; computationally intensive	k-mer encoded sequences	Large-scale genomic data
Recurrent Neural Network (RNN)	Captures sequential mutation dependencies	Training instability; high complexity	Ordered mutation data	Sequential genomic modeling
Ensemble (RF + SVM + Boosting)	Improved robustness; reduced variance; better generalization	Increased computational cost	Mixed feature sets	Heterogeneous datasets

5. Model Evaluation Frameworks

5.1 Performance Metrics

Common metrics include:

- Accuracy

- Sensitivity
- Specificity
- Precision
- F1-score
- AUROC [31]

Balanced accuracy is recommended in imbalanced datasets [32].

5.2 Cross-Validation

K-fold cross-validation helps prevent overfitting, but independent validation in different geographic areas is necessary to see if the results are generalizable [33].

5.3 External Validation

Few studies do multi-country validation, which restricts their application in real-world contexts [34].

6. Pathogen-Specific Applications and Comparative Performance

AMR prediction based on machine learning has been used to numerous bacterial diseases. However, the quantity of the dataset, the resistance mechanisms, and the diversity of the pathogen's genome all affect performance.

6.1 *Mycobacterium tuberculosis*

One of the diseases that has been researched the most for genomic AMR prediction is *Mycobacterium tuberculosis* (MTB). Since chromosomal changes rather than horizontal gene transfer are the primary cause of resistance in MTB, SNP-based modeling is especially useful [35].

Random forests, support vector machines, and gradient boosting algorithms have been used in several studies to predict resistance to first-line drugs such isoniazid and rifampicin [23,36]. Due to distinct mutations in the rpoB gene, AUROC scores for predicting rifampicin resistance frequently surpass 0.90. However, due to a limited understanding of resistance mechanisms, forecasts become less accurate for second-line drugs such as ethionamide or pyrazinamide [37].

Deep learning models that can predict MTB resistance have also been built using whole-genome k-mer encodings [38]. Due to lineage diversity, generalizability across regionally distinct strains remains a challenge even when performance is improved in curated datasets [39].

6.2 *Escherichia coli* and Enterobacteriaceae

Chromosome mutations and horizontally acquired resistance genes on plasmids are examples of resistance mechanisms in Enterobacteriaceae [40]. Gradient boosting and k-mer-based models have shown strong efficiency in predicting beta-lactam, carbapenem, and fluoroquinolone resistance in E.

coli [41]. This complexity makes ML modeling more challenging than in MTB. However, different regions of the world have distinct plasmid-mediated resistance genes, therefore large and diverse training datasets are required [42].

Gradient boosting models outperform SVM and logistic regression in high-dimensional k-mer datasets for Enterobacteriaceae, according to comparative benchmarking experiments [26].

6.3 *Staphylococcus aureus*

Methicillin resistance in Staphylococcus aureus is mostly caused by the mecA gene, which makes it simple to identify MRSA using genomic data [43]. However, it is much more difficult to resist other medications like vancomycin and linezolid.

MRSA may be detected with a high degree of sensitivity (>95%) using machine learning classifiers like RF and ANN models [44]. However, because of polygenic variables, predicting intermediate resistance phenotypes remains difficult.

6.4 *Klebsiella pneumoniae*

Due to its high genomic flexibility, *K. pneumoniae* frequently acquires carbapenemase genes such as OXA-48, NDM, and KPC [45]. Predictive performance is improved by ML models that use mobile genetic element profiling and plasmid recognition features [46]. However, models must be revised frequently due to the quick evolution of resistance genes.

6.5 *Salmonella* spp. and Foodborne Pathogens

ML-based genomic prediction helps monitor changes in *Salmonella* isolate resistance in food safety surveillance [47]. Studies show that tetracycline and beta-lactam resistance have higher prediction accuracy, while aminoglycoside efficacy varies [48].

7. Comparative Algorithm Performance

On medium-sized datasets (1,000–10,000 isolates), random forests and gradient boosting consistently yield good results [23,26]. While deep learning methods may not outperform tree-based models in smaller groups, they show promise in large datasets [38].

Key determinants of algorithm performance include:

- Dataset size
- Class balance
- Genomic diversity
- Feature dimensionality
- Hyperparameter optimization
- In many circumstances, simpler ensemble models function just as well or better than complicated neural networks,

especially when the dataset is small.

8. Clinical Integration and Workflow Considerations

For ML-based AMR prediction to be clinically useful, integration into diagnostic workflows is essential.

8.1 Proposed Workflow

1. Sample collection
2. Rapid WGS (24 hours or less)
3. Automated genome assembly and feature extraction
4. ML-based resistance prediction
5. Clinical reporting

Turnaround time can be reduced compared to culture-based AST, particularly for slow-growing pathogens like MTB [49].

8.2 Reporting and Decision Support

ML predictions need to be presented in a form that physicians can comprehend. The tool is more reliable and user-friendly when it includes probability ratings, confidence intervals, and explanations of significant genetic features that influence predictions [50].

9. Interpretability and Explainable AI

Interpretability remains one of the major challenges in ML-based AMR diagnostics.

9.1 Feature Importance

Tree-based models give you measurements for the relative importance of various attributes. These can display specific genes or mutations that aid in the prediction of resistance [22].

9.2 SHAP and LIME

Shapley Additive Explanations (SHAP) quantify the degree to which each feature influences the model's output [51]. Interpretability at the case level is offered by LIME (Local Interpretable Model-Agnostic Explanations) [52].
Clinical uptake and regulatory approval depend on explainable AI tools.

10. Bias, Fairness, and Generalizability

10.1 Geographic Bias

Training datasets often overrepresent high-income countries, leading to reduced predictive accuracy in low-resource settings [53].

10.2 Lineage Effects

ML models may become less accurate due to phylogenetic structure. Models may uncover

lineage markers rather than true resistance determinants if lineage is strongly associated with resistance phenotype [54].

10.3 Class Imbalance

Rare resistance phenotypes create skewed datasets, requiring techniques such as:

1. Oversampling (SMOTE)
2. Class weighting
3. Balanced accuracy metrics [32]

11. Regulatory and Ethical Considerations

The regulatory requirements for diagnostic tools must be met by ML-based AMR prediction systems intended for clinical use. Among other organizations, the FDA and EMA need evidence of:

- Analytical validity
- Clinical validity
- Reproducibility
- Robustness to dataset drift

When combining genomic data from different organizations, there are also worries about data privacy [55].

12. Emerging Innovations

12.1 Federated Learning

Federated learning allows model training across multiple institutions without the need for centralized data exchange, which protects patient privacy and increases the variety of datasets [56].

12.2 Multi-Omics Integration

Integrating genomes with transcriptomics and proteomics may improve the prediction of resistance driven by alterations in gene expression [57].

12.3 Real-Time Genomic Surveillance

The combination of ML with national monitoring networks makes it possible to quickly find new patterns of resistance [58].

13. Future Research Directions

Future priorities include:

- Standardized benchmarking datasets
- Cross-country validation
- Integration with electronic health records
- Continuous model updating frameworks
- Improved interpretability tools

To deal with data heterogeneity and make sure that ML-based AMR diagnoses are used fairly, worldwide consortia will need to work together.

14. Conclusions

Antibiotic resistance in genomes may now be predicted with much

greater accuracy thanks to machine learning approaches. These days, supervised ensemble models are the ideal option because they are robust and simple to comprehend. Large datasets, however, are a good fit for deep learning. Despite research demonstrating the high accuracy of these techniques, practical implementation, regulatory approval, and clinical application remain challenges. These concepts can be implemented securely and simply thanks to new frameworks like federated learning and explainable AI. To make ML-driven AMR prediction a standard component of clinical and public health practice, microbiologists, bioinformaticians, physicians, and regulatory agencies will need to continue collaborating across disciplines.

References

1. World Health Organization. (2020). *Global antimicrobial resistance surveillance system (GLASS) report 2020*. WHO.
2. O'Neill, J. (2016). *Tackling drug-resistant infections globally: Final report and recommendations*. UK Government Review on AMR.
3. Jorgensen, J. H., & Ferraro, M. J. (2009). Antimicrobial susceptibility testing: A review of general principles and contemporary practices. *Clinical Infectious Diseases*, 49(11), 1749–1755.
4. Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E., & Crook, D. W. (2012). Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*, 13(9), 601–612.
5. Jia, B., et al. (2017). CARD 2017: Expansion and model-centric curation of the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*, 45(D1), D566–D573.
6. Zankari, E., et al. (2012). Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*, 67(11), 2640–2644.
7. Boolchandani, M., D'Souza, A. W., & Dantas, G. (2019). Sequencing-based methods and resources to study antimicrobial resistance. *Nature Reviews Microbiology*, 17(2), 70–84.
8. Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6), 321–332.
9. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
10. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD*, 785–794.
11. Drouin, A., et al. (2019). Predictive computational phenotyping and AMR. *Bioinformatics*, 35(10), 1825–1832.
12. Yang, Y., et al. (2018). Machine learning for AMR prediction. *Scientific Reports*, 8, 16481.

13. Nguyen, M., et al. (2018). Using machine learning to predict antimicrobial resistance. *mBio*, 9(3), e01260-18.
14. Bradley, P., et al. (2015). Rapid antibiotic-resistance predictions from genome sequence data. *Nature Communications*, 6, 10063.
15. Quainoo, S., et al. (2017). Whole-genome sequencing for bacterial surveillance. *Clinical Microbiology and Infection*, 23(11), 744–752.
16. Wick, R. R., et al. (2017). Unicycler: Hybrid assembly of bacterial genomes. *PLoS Computational Biology*, 13(6), e1005595.
17. Feldgarden, M., et al. (2019). AMRFinder tool for AMR gene detection. *Antimicrobial Agents and Chemotherapy*, 63(11), e00483-19.
18. Coll, F., et al. (2015). Genome-wide analysis of drug resistance in Mycobacterium tuberculosis. *Nature Genetics*, 47(10), 1084–1090.
19. Wood, D. E., et al. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20, 257.
20. Partridge, S. R., et al. (2018). Mobile genetic elements associated with antimicrobial resistance. *Clinical Microbiology Reviews*, 31(4), e00088-17.
21. Arango-Argoty, G., et al. (2018). DeepARG: A deep learning approach for predicting antibiotic resistance genes. *Microbiome*, 6, 23.
22. Lundberg, S. M., & Lee, S. I. (2017). SHAP: A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 4765–4774.
23. Farhat, M. R., et al. (2016). Genomic analysis identifies targets for predicting drug resistance. *The Lancet Infectious Diseases*, 16(10), 1193–1202.
24. Walker, T. M., et al. (2015). Whole-genome sequencing for predicting Mycobacterium tuberculosis drug susceptibility. *The Lancet Infectious Diseases*, 15(10), 1193–1202.
25. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.
26. Pesesky, M. W., et al. (2016). Evaluation of machine learning methods for predicting AMR. *Journal of Clinical Microbiology*, 54(10), 2533–2539.
27. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, 67(2), 301–320.
28. Hyun, J. C., et al. (2020). Deep learning improves prediction of antimicrobial resistance. *Nature Communications*, 11, 612.
29. Min, S., et al. (2017). Deep learning in bioinformatics. *Briefings in Bioinformatics*, 18(5), 851–869.
30. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
31. Ngiam, J., et al. (2011). Multimodal deep learning. *ICML Proceedings*.

32. Chawla, N. V., et al. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
33. Chicco, D., & Jurman, G. (2020). The advantages of balanced accuracy. *BMC Genomics*, 21, 760.
34. Moradigaravand, D., et al. (2018). Phylogenetic structure and AMR modeling. *PLoS Genetics*, 14(8), e1007450.
35. CRyPTIC Consortium. (2018). Prediction of susceptibility to first-line TB drugs by WGS. *The New England Journal of Medicine*, 379, 1403–1415.
36. Zhang, H., et al. (2019). Deep learning for tuberculosis drug resistance. *Scientific Reports*, 9, 13593.
37. Miotto, P., et al. (2017). Standardized interpretation of TB mutations. *European Respiratory Journal*, 50, 1701354.
38. Kouchaki, S., et al. (2019). Deep neural networks for TB resistance. *Bioinformatics*, 35, 432–439.
39. Comas, I., et al. (2013). Phylogeography of TB lineages. *Nature Genetics*, 45, 1176–1182.
40. Mathers, A. J., et al. (2015). Klebsiella pneumoniae resistance mechanisms. *Clinical Microbiology Reviews*, 28(3), 565–591.
41. Nguyen, M., et al. (2019). Machine learning-based prediction of AMR in E. coli. *mSystems*, 4(3), e00110-19.
42. Stoesser, N., et al. (2016). Predicting antimicrobial susceptibility in Enterobacteriaceae. *Genome Medicine*, 8, 49.
43. Gordon, N. C., et al. (2014). Whole-genome sequencing in Staphylococcus aureus. *Genome Medicine*, 6, 94.
44. Bradley, P., et al. (2018). Rapid MRSA detection by genome analysis. *Nature Communications*, 9, 1058.
45. Pitout, J. D., & Nordmann, P. (2015). Carbapenemase-producing Enterobacteriaceae. *Clinical Microbiology Reviews*, 28(3), 564–591.
46. Su, M., et al. (2019). Deep learning for carbapenem resistance prediction. *Frontiers in Microbiology*, 10, 1883.
47. Neuert, S., et al. (2018). Prediction of AMR in Salmonella. *Journal of Antimicrobial Chemotherapy*, 73(10), 2741–2749.
48. Zhao, S., et al. (2020). WGS-based surveillance of Salmonella resistance. *Emerging Infectious Diseases*, 26(9), 2007–2014.
49. Pankhurst, L. J., et al. (2016). Rapid, comprehensive, and affordable TB diagnosis via WGS. *The Lancet Respiratory Medicine*, 4(1), 49–58.
50. Topol, E. J. (2019). High-performance medicine: The convergence of human and AI. *Nature Medicine*, 25, 44–56.
51. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" LIME explanations. *Proceedings of KDD*, 1135–1144.

52. Lundberg, S. M., et al. (2020). From local explanations to global understanding with SHAP. *Nature Machine Intelligence*, 2, 56–67.
53. Holmes, A. H., et al. (2016). Understanding AMR drivers. *The Lancet*, 387(10014), 176–187.
54. McMahan, B., et al. (2017). Communication-efficient federated learning. *AISTATS Proceedings*.
55. FDA. (2023). *Artificial intelligence and machine learning in medical devices guidance*.
56. EMA. (2022). *AI regulatory considerations in diagnostics*.
57. Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology*, 18, 83.
58. CDC. (2019). *Antibiotic resistance threats in the United States*.