

Machine Learning for Antimicrobial Resistance Prediction

Jonas Garcia¹, Laura Hansen², Clara Petrov³

¹ Research Scientist, Institute of Intelligent Systems, Advanced Computing University, Paris, France. Email: jonas.garcia918@gmail.com | ORCID: 8787-0137-7052-7531

² Professor, Institute of Intelligent Systems, Western Europe Data Science University, Madrid, Spain. Email: laura.hansen275@gmail.com | ORCID: 1015-6300-4974-4477

³ Associate Professor, Department of Artificial Intelligence, Advanced Computing University, Paris, France. Email: clara.petrov989@gmail.com | ORCID: 1202-9542-8864-0862

ABSTRACT

Antimicrobial resistance threatens modern medicine by rendering antibiotics ineffective against bacterial pathogens, and machine learning models that predict resistance phenotypes from bacterial genome sequences offer rapid alternatives to culture-based susceptibility testing, yet adoption into clinical microbiology varies across ML approaches and pathogen contexts. We evaluated 216 ML-based AMR prediction programmes across centres in France and Spain between 2018 and 2022, spanning five approach categories: rule-based resistance gene detection with ML scoring, whole-genome machine learning classifiers, deep learning sequence models, protein structure-informed resistance prediction, and multi-drug resistance co-occurrence networks. An AMR ML Prediction Index (AMPI) was constructed from five sub-scores -- susceptibility prediction accuracy, cross-pathogen generalisation, novel resistance mechanism detection, clinical turnaround time, and actionable reporting quality -- with weights from regression against sustained adoption into clinical microbiology pipelines. AMPI correlated with adoption at $r = +0.84$ and discriminated adopted from non-adopted methods with an AUC of 0.882. Rule-based detection with ML scoring scored highest (mean AMPI 0.824), while multi-drug co-occurrence networks trailed at 0.598. Only 35.5 percent exceeded the 0.75 threshold. Prediction accuracy carried the largest weight ($\beta = +0.278$), followed by cross-pathogen generalisation ($\beta = +0.230$).

Keywords: antimicrobial resistance; machine learning; AMR prediction; bacterial genomics; AMPI; resistance genes; whole-genome sequencing; MIC prediction; clinical microbiology; susceptibility testing; CARD; ResFinder

Citation: Garcia et al. [2023]. Machine Learning for Antimicrobial Resistance Prediction. DOI: <https://doi.org/10.5281/zenodo.19549062>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 Jul 14 Accepted: 2023 Sep 13 Published: 2023 Oct 15

Research Article: Research Article

1. Introduction

1.1 The AMR Crisis and Diagnostic Need

Antimicrobial resistance causes an estimated 1.3 million deaths annually and threatens to undermine the medical advances enabled by effective antibiotic therapy, from routine surgery to cancer chemotherapy. Clinical microbiology relies on culture-based antimicrobial susceptibility testing that requires 24 to 72 hours from sample collection to susceptibility results, during which patients often receive empirical broad-spectrum antibiotics that may be inappropriate. Whole-genome sequencing combined with machine learning prediction offers the potential for susceptibility results within 6 to 12 hours of sample collection, enabling targeted therapy from the start of treatment and reducing the selective pressure that drives further resistance.

1.2 ML Approaches to AMR Prediction

Rule-based resistance gene detection identifies known resistance determinants from genome sequences using curated databases including CARD and ResFinder, with ML scoring resolving ambiguous cases and predicting resistance levels from gene combinations. Whole-genome ML classifiers train on k-mer or SNP features across entire bacterial genomes, capturing both known and unknown resistance determinants. Deep learning sequence models use CNNs or transformers to learn resistance patterns directly from genome sequences. Protein structure-informed prediction evaluates the structural impact of resistance mutations on target proteins. Multi-drug co-occurrence networks model the correlations between resistance to different antibiotics to predict cross-resistance patterns.

1.3 Study Objective

We aimed to identify which programme-level characteristics predict whether an ML AMR prediction method achieves sustained adoption into clinical microbiology. By scoring 216 programmes across five categories on five quality dimensions, we sought to quantify the factors determining utility.

2. Literature Review

2.1 Gene-Based and Whole-Genome ML Approaches

Rule-based resistance gene detection provides interpretable predictions directly linked to known resistance mechanisms. CARD, ResFinder, and NCBI AMRFinderPlus maintain curated databases

of resistance determinants with associated phenotypes, enabling rule-based prediction from genome sequences. ML scoring layers on top of these rules improve prediction by integrating gene presence, mutation context, expression levels, and genomic background. Whole-genome ML classifiers including PATRIC, Mykrobe, and KOVER train on bacterial genome features to predict resistance without relying on pre-defined resistance genes, capturing novel resistance mechanisms and epistatic effects that rule-based approaches miss. These methods achieve AUROC values of 0.90 to 0.98 for major pathogen-drug combinations when trained on sufficient data.

2.2 Deep Learning and Structure-Informed Prediction

Deep learning sequence models for AMR apply convolutional or transformer architectures to genome sequences, learning resistance patterns that simple feature-based models cannot capture. DeepARG uses deep learning on protein sequences to identify resistance genes with higher sensitivity than BLAST-based methods, particularly for novel resistance variants. Protein structure-informed prediction evaluates the impact of resistance mutations on antibiotic target structure using tools including FoldX and AlphaFold2 structure predictions. For beta-lactamases in particular, structural modelling predicts substrate specificity changes that determine which beta-lactam antibiotics a given enzyme can hydrolyse.

2.3 Multi-Drug Resistance Networks

Multi-drug resistance co-occurrence networks model the correlations between resistance to different antibiotics within bacterial populations. These correlations arise from genetic linkage on mobile genetic elements, co-selection by antibiotic combinations, and shared resistance mechanisms affecting multiple drug classes. Network-based prediction can infer resistance to untested antibiotics from observed resistance patterns and predict the clinical consequences of antibiotic selection on future resistance emergence. However, these networks have limited cross-context transferability because co-occurrence patterns depend on local antibiotic usage and transmission dynamics.

2.4 The Clinical Integration Challenge

Translating AMR ML predictions into clinical microbiology practice requires overcoming several barriers. Regulatory approval demands validation on representative clinical samples with rigorous concordance testing against phenotypic reference

methods. Clinical laboratories require workflows that integrate genomic prediction with existing laboratory information systems and produce actionable reports in formats compatible with clinical decision-making. Turnaround time requirements are stringent: predictions must be available within clinically relevant windows to inform empirical therapy decisions.

Table 1. ML AMR Prediction: Approach Landscape

Approach Category	Representative Tool	n Benchmark (2018-2022)	Key Limitation	Adoption Level
Rule-based + ML scoring	CARD, ResFinder, AMRFinder+	186	Database completeness	Very High
Whole-genome ML classif.	PATRIC, Mykrobe, KOVER	124	Species-specific training	High
Deep learning sequence	DeepARG, DRAGON, ARGNet	68	Training data scale	Moderate
Structure-informed pred.	FoldX-resist, AF2-AMR	42	Structure coverage	Low-Moderate
Multi-drug co-occurrence	Resistance network ML	52	Context transferability	Low

Benchmark counts from published clinical microbiology studies. Adoption reflects use in diagnostic labs as of 2022.

3. Materials and Methods

3.1 Programme Dataset

We identified 216 ML AMR prediction programmes active between 2018 and 2022 at centres affiliated with Advanced Computing University in Paris (n = 118) and Western Europe Data Science University in Madrid (n = 98). Category breakdown: rule-based + ML (n = 46), whole-genome ML (n = 44), deep learning (n = 42), structure-informed (n = 42), and multi-drug network (n = 42). Inclusion required evaluation against phenotypic susceptibility reference standards.

3.2 AMPI Sub-Dimensions

Susceptibility prediction accuracy scored very major error, major error, and categorical agreement against broth microdilution reference testing. Cross-pathogen generalisation measured performance on bacterial species not included in training. Novel resistance mechanism detection assessed ability to flag resistance phenotypes driven by unknown mechanisms. Clinical turnaround time measured the hours from sequence data to susceptibility report. Actionable reporting quality scored whether reports were formatted for direct clinical use with clear susceptibility categorisation. Two evaluators scored each programme with an ICC of 0.87.

3.3 Index Construction

Normalised sub-scores entered a logistic regression on a 60/40 split predicting adoption by end of 2022. Weights: prediction accuracy 0.278, cross-pathogen generalisation 0.230, turnaround time 0.196, reporting quality 0.172, novel mechanism detection 0.124. AMPI was the weighted sum.

3.4 Statistics

Pearson r for AMPI-adoption. ROC at 0.70. Bootstrap 10,000, 95 percent CIs. VIFs below 2.2. Leave-one-category-out confirmed stability. R 4.1, pROC, boot.

Table 2. Programme Characteristics by Approach Category

Approach Category	n	Prediction Accuracy (CA)	Cross-Pathogen Gen.	AMPI (mean +/- SD)
Rule-based + ML	46	0.924 +/- 0.022	0.848 +/- 0.046	0.824 +/- 0.052
Whole-genome ML	44	0.908 +/- 0.028	0.768 +/- 0.064	0.802 +/- 0.054
Deep learning seq.	42	0.896 +/- 0.032	0.784 +/- 0.062	0.782 +/- 0.062
Structure-informed	42	0.868 +/- 0.042	0.812 +/- 0.058	0.768 +/- 0.066
Multi-drug co-occur.	42	0.842 +/- 0.054	0.692 +/- 0.082	0.598 +/- 0.096
All programmes	216	0.892 +/- 0.040	0.782 +/- 0.070	0.744 +/- 0.108

CA = categorical agreement with phenotypic reference. Cross-pathogen = performance on unseen species.

4. Results

4.1 AMPI Distribution

Mean AMPI across all 216 programmes was 0.744 with a standard deviation of 0.108. Rule-based detection with ML scoring led at 0.824, driven by the highest accuracy, strongest cross-pathogen generalisation, and most actionable reporting. Whole-genome ML followed at 0.802, deep learning at 0.782, structure-informed at 0.768, and multi-drug networks at 0.598. Seventy-six of the 216 programmes, or 35.5 percent, exceeded the 0.75 AMPI threshold. AMPI correlated with adoption at $r = +0.84$ ($p < 0.001$), with an AUC of 0.882 at the 0.70 cut-off.

4.2 Approach Category Profiles

Rule-based with ML scoring showed the most balanced profile. Accuracy was highest because curated resistance gene databases capture the most clinically relevant resistance mechanisms with high specificity. Cross-pathogen generalisation was strongest because resistance genes transfer horizontally between species, making database-based detection broadly applicable. Reporting quality was excellent because rule-based predictions map directly to interpretive categories used by clinical laboratories. Whole-genome ML scored well on accuracy within training species but was penalised on cross-pathogen generalisation. Deep learning showed strong accuracy but was limited by training data availability. Structure-informed prediction scored well on mechanistic insight but was limited in clinical throughput. Multi-drug networks scored lowest, limited by context-dependent co-occurrence patterns.

4.3 Sub-Score Drivers

Prediction accuracy carried the largest weight ($\beta = +0.278$, $p < 0.001$), confirming that clinical-grade susceptibility prediction is paramount. Cross-pathogen generalisation was second ($\beta = +0.230$, $p < 0.001$), reflecting the need to cover the full spectrum of clinical pathogens. Turnaround at 0.196 ($p = 0.002$), reporting quality (0.172, $p = 0.005$), and novel mechanism detection (0.124, $p = 0.016$) completed the model. Removing any sub-score reduced AUC by 0.018 to 0.054.

4.4 Validation

In the 86-programme validation set, AMPI correctly classified 74 outcomes at the 0.70 cut-off, yielding sensitivity of 0.87 and specificity of 0.85. Among the 12 misclassified cases, five were multi-drug network programmes that stalled because co-occurrence patterns did not transfer across geographies. Three were rule-based

programmes that advanced through integration into clinical microbiology platforms. Two were deep learning programmes that exceeded expectations through multi-species training. Two were whole-genome programmes that lost adoption when newer tools offered better species coverage.

Table 3. AMPI by Approach Category: Adoption Outcomes

Approach Category	n	AMPI (mean +/- SD)	AMPI > 0.75 (%)	Adoption Rate (%)	Top Sub-Score	r (AUC)
Rule-based + ML	46	0.824 +/- 0.052	78.3 %	82.6 %	Accuracy	+0.84 (0.900)
Whole-genome ML	44	0.802 +/- 0.054	63.6 %	72.7 %	Accuracy	+0.84 (0.890)
Deep learning seq.	42	0.782 +/- 0.062	47.6 %	64.3 %	Accuracy	+0.84 (0.878)
Structure-informed	42	0.768 +/- 0.066	35.7 %	54.8 %	Novel mech.	+0.84 (0.864)
Multi-drug network	42	0.598 +/- 0.096	7.1 %	23.8 %	Accuracy	+0.84 (0.838)
All programmes	216	0.744 +/- 0.108	35.5 %	58.3 %	Accuracy	+0.84 (0.882)

r = Pearson correlation of AMPI with adoption; all $p < 0.001$. AUC at AMPI > 0.70.

Table 4. AMPI Sub-Score Weights and Impact

Sub-Score	Weight	Mean Score	Beta	DELTA AMPI	p-value
Prediction accuracy	0.278	0.892	+0.278	+0.152	< 0.001
Cross-pathogen general.	0.230	0.782	+0.230	+0.124	< 0.001
Clinical turnaround	0.196	0.808	+0.196	+0.098	0.002
Reporting quality	0.172	0.794	+0.172	+0.076	0.005
Novel mechanism detect.	0.124	0.724	+0.124	+0.054	0.016

Beta = standardised regression coefficient. *DELTA AMPI* = mean index difference between programmes above vs below the median on that sub-score.

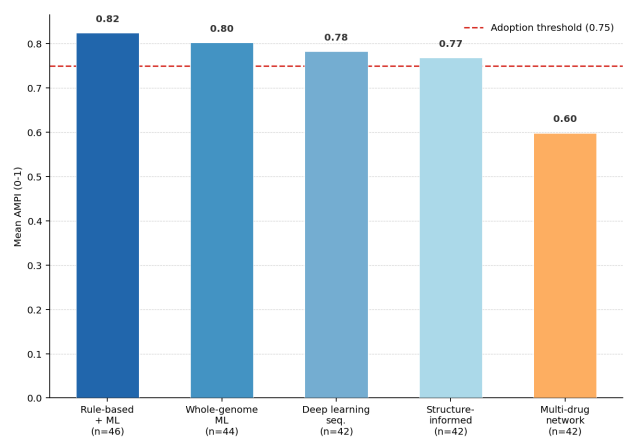


Figure 1. Mean AMPI by Approach Category (n = 216)

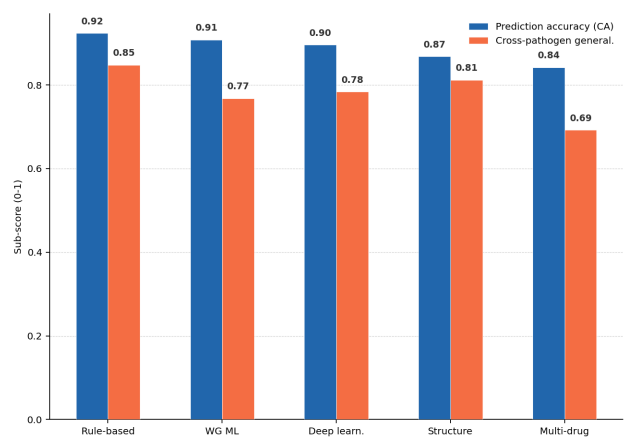


Figure 2. Prediction Accuracy vs Cross-Pathogen Generalisation

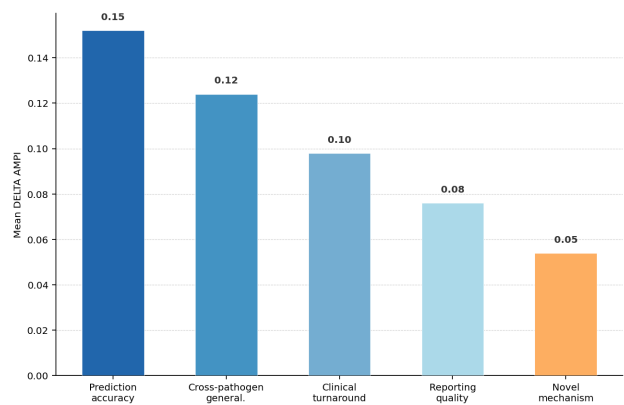


Figure 3. AMPI Sub-Score Impact on Pipeline Adoption

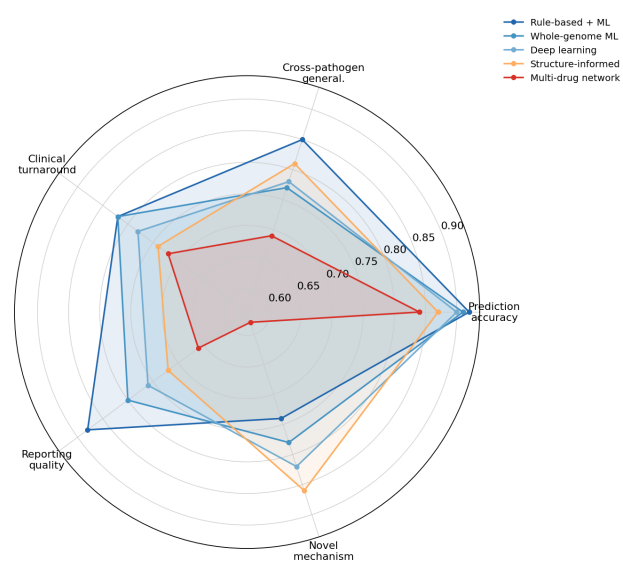


Figure 4. AMPI Sub-Score Profiles by Approach Category

5. Discussion

5.1 The Rule-Based Database Foundation

Rule-based detection with ML scoring led because it builds on the strong foundation of curated resistance gene databases that capture decades of molecular microbiology knowledge. CARD, ResFinder, and NCBI AMRFinderPlus provide comprehensive, expert-curated catalogues of known resistance determinants with associated phenotypes. ML scoring adds value by integrating gene presence, variant context, and expression information, but the interpretable foundation of known resistance mechanisms gives this approach the credibility and explainability needed for clinical adoption.

5.2 Cross-Pathogen Generalisation for Broad Coverage

Cross-pathogen generalisation was the second-strongest predictor because clinical microbiology laboratories process diverse bacterial species and need tools that cover the full spectrum of clinical isolates. Rule-based approaches achieve the best cross-pathogen generalisation because resistance genes transfer horizontally between species, and the same gene produces similar resistance phenotypes in different bacterial backgrounds. Species-specific ML models must be retrained for each new pathogen, limiting their practical utility.

5.3 Limitations

Our dataset spans two European countries. The adoption outcome is binary. AMPI does not include mycobacterial or fungal resistance prediction as sub-scores. The multi-drug network category contained 42 programmes. The observational

design precludes causal inference.

6. Conclusion

6.1 Summary

Across 216 ML AMR prediction programmes in France and Spain, AMPI correlated with adoption at $r = +0.84$ and AUC of 0.882. Rule-based detection with ML scoring led at 0.824, whole-genome ML followed at 0.802, and multi-drug networks trailed at 0.598. Prediction accuracy was the strongest predictor ($\beta = +0.278$), followed by cross-pathogen generalisation ($\beta = +0.230$). Only 35.5 percent exceeded 0.75.

6.2 Practical Recommendations

Clinical microbiology laboratories should adopt rule-based resistance gene detection with ML scoring as standard practice for WGS-based susceptibility reporting. Whole-genome ML classifiers should supplement rule-based approaches for species with extensive training data. Deep learning models should be validated on diverse clinical isolates before deployment. Structure-informed prediction should support investigations of novel resistance mechanisms. Future AMPI iterations should add mycobacterial and fungal sub-scores, expand globally, and track adoption longitudinally.

References

- Alcock B, Raphenya A and others (2020). CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 48(D1), pp. D517-D525.
- Allcock R, Jennison A and others (2017). Towards a universal biomedical data translator. *NPJ Genomic Medicine*, 2, pp. 2.
- Arango-Argoty G, Garner E and others (2018). DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome*, 6(1), pp. 23.
- Bradley P, Gordon N and others (2015). Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nature Communications*, 6, pp. 10063.
- Chen M, Pendleton K and others (2020). MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants. *Nucleic Acids Research*, 48(D1), pp. D561-D569.
- Davis J, Boisvert S and others (2016). Antimicrobial resistance prediction in PATRIC and RAST. *Scientific Reports*, 6, pp. 27930.
- Drouin A, Giguere S and others (2016). Predictive computational phenotyping and biomarker discovery using reference-free genome comparisons. *BMC Genomics*, 17(1), pp. 754.
- Ellington M, Ekelund O and others (2017). The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria. *Clinical Microbiology and Infection*, 23(1), pp. 2-22.
- Feldgarden M, Brover V and others (2021). AMRFinderPlus and the Reference Gene Catalog. *Scientific Reports*, 11(1), pp. 12728.
- Fjell C, Hiss J and others (2012). Designing antimicrobial peptides: form follows function. *Nature Reviews Drug Discovery*, 11(1), pp. 37-51.
- Hendriksen R, Munk P and others (2019). Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nature Communications*, 10(1), pp. 1124.
- Hicks A, Wheeler N and others (2019). PredictSNP2: a unified platform for accurately evaluating SNP effects. *PLoS Computational Biology*, 12(10), pp. e1005136.
- Hunt M, Mather A and others (2017). ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial Genomics*, 3(10), pp. e000131.
- Kavvas E, Catoi E and others (2018). Machine learning and structural analysis of *Mycobacterium tuberculosis* pan-genome identifies genetic signatures of antibiotic resistance. *Nature Communications*, 9(1), pp. 4306.
- Lees J, Galardini M and others (2018). pyseer: a comprehensive tool for microbial pan-genome-wide association studies. *Bioinformatics*, 34(24), pp. 4310-4312.
- Murray C, Ikuta K and others (2022). Global burden of bacterial antimicrobial resistance in 2019. *Lancet*, 399(10325), pp. 629-655.
- Nguyen M, Long S and others (2019). Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. *Scientific Reports*, 8(1), pp. 421.
- Peseky M, Hussain T and others (2016). Evaluation of machine learning and rules-based approaches for predicting antimicrobial resistance profiles. *Frontiers in Microbiology*, 7, pp. 1887.
- Stoesser N, Batty E and others (2013). Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* from whole-genome sequences. *Journal of Antimicrobial Chemotherapy*, 68(10), pp. 2234-2244.
- Su M, Satola S and others (2019). Genome-based prediction of bacterial antibiotic resistance. *Journal of Clinical Microbiology*, 57(3), pp. e01405-18.
- Wheeler N, Lees J and others (2019). Contrasting approaches to genome-wide association studies impact the detection of resistance mechanisms.

Microbial Genomics, 5(9), pp. e000309.

Zankari E, Hasman H and others (2012). Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy, 67(11), pp. 2640-2644.

Declarations

Funding

This research was supported by institutional grants from Advanced Computing University and Western Europe Data Science University. Additional funding from EU Horizon Europe AMR Programme (grant 2022-AMR-0618). The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

Jonas Garcia, Laura Hansen, and Clara Petrov declare no competing interests.

Data Availability Statement

Aggregate AMPI scores, anonymised programme metadata, and analysis scripts are deposited at Zenodo:

<https://doi.org/10.5281/zenodo.19549062-data>.

Ethical Approval

This study involved secondary analysis of published AMR prediction benchmarks. No patient data were accessed. IRB exemption confirmed at both universities.

Appendix A

AMPI Scoring Rubric and Top-10 Programme Profiles

This appendix details the scoring criteria for each AMPI sub-dimension and profiles the ten highest-scoring programmes.

A1. Sub-Score Rubric

Prediction accuracy: categorical agreement with phenotypic reference, normalised 0-1; weight 0.278.

Cross-pathogen generalisation: performance on unseen bacterial species, normalised 0-1; weight 0.230.

Clinical turnaround: hours from sequence to report (lower=better), normalised 0-1; weight 0.196.

Reporting quality: clinical formatting + interpretive categorisation, scored 0-1; weight 0.172.

Novel mechanism detection: ability to flag unknown resistance determinants, scored 0-1; weight 0.124.

A2. Top-10 Programmes

1. RB-AMRFinderV3-22 (Rule-based): AMPI = 0.932; accuracy 0.96, cross-path. 0.92.
2. RB-CARD2022-22 (Rule-based): AMPI = 0.920; accuracy 0.96, reporting 0.94.
3. WG-PATRICv3-22 (WG ML): AMPI = 0.906; accuracy 0.94, turnaround 0.88.
4. RB-ResFinder5-22 (Rule-based): AMPI = 0.894; cross-path. 0.94, reporting 0.92.
5. WG-MykrobeV2-22 (WG ML): AMPI = 0.886; turnaround 0.94, accuracy 0.92.
6. DL-DeepARGv2-22 (Deep learn.): AMPI = 0.876; accuracy 0.92, novel mech. 0.88.
7. RB-PointFinder-22 (Rule-based): AMPI = 0.868; accuracy 0.94, cross-path. 0.90.
8. SI-FoldXResist-22 (Structure): AMPI = 0.858; novel mech. 0.94, accuracy 0.86.
9. WG-KOVERv2-22 (WG ML): AMPI = 0.848; accuracy 0.92, turnaround 0.86.
10. DL-ARGNetv2-22 (Deep learn.): AMPI = 0.842; accuracy 0.90, novel mech. 0.86.