

RESEARCH

Open Access



# Combining mass spectrometry and machine learning models for predicting *Klebsiella pneumoniae* antimicrobial resistance: a multicenter experience from clinical isolates in Italy

Ettore Rocchi<sup>1,2\*†</sup>, Emanuele Nicitra<sup>3†</sup>, Maddalena Calvo<sup>4</sup>, Valeria Cento<sup>5,6</sup>, Laura Peiretti<sup>6</sup>, Zian Asif<sup>5</sup>, Giulia Menchinelli<sup>7</sup>, Brunella Posteraro<sup>8,9</sup>, Claudia Sala<sup>1,2</sup>, Claudia Colosimo<sup>1,10</sup>, Monica Cricca<sup>1,10</sup>, Vittorio Sambri<sup>1,10</sup>, Maurizio Sanguinetti<sup>7,8</sup>, Gastone Castellani<sup>1,2†</sup> and Stefania Stefani<sup>3,4†</sup>

## Abstract

**Background** Multidrug-resistant *Klebsiella pneumoniae* represents a significant challenge in healthcare settings, prompting numerous studies on the rapid detection of antimicrobial resistance. Mass spectrometry has been recently integrated into routine laboratory diagnostics, providing highly sensitive results for pathogen identification. Furthermore, previously published studies have demonstrated its potential application in predicting antimicrobial resistance.

**Materials and methods** The study collected 686 clinical isolates of *K. pneumoniae* from three Italian hospitals and used their MALDI-TOF mass spectra as input to machine learning models for predicting susceptibility profiles to amikacin and meropenem, which were selected as the most represented antibiotic molecules within the aminoglycoside and carbapenem classes, commonly used for the treatment of *K. pneumoniae* infections. After preprocessing, *K. pneumoniae* spectra were fed to machine learning classifiers within a nested cross-validation framework. Several performance metrics were computed to compare models and identify the most appropriate one for each antibiotic. Given the multicentric nature of the study, a batch-effect correction step was applied to reduce site-specific variability using the in-house developed Python package *combatlearn* (available on GitHub: <https://github.com/EttoreRocchi/combatlearn>).

<sup>†</sup>Ettore Rocchi and Emanuele Nicitra contributed equally to this work.

<sup>†</sup>Gastone Castellani and Stefania Stefani jointly supervised this work.

\*Correspondence:

Ettore Rocchi  
ettore.rocchi3@unibo.it

Full list of author information is available at the end of the article



**Results** The XGBoost model achieved the best performance for both antibiotics (AUROC =  $0.822 \pm 0.028$  for amikacin; AUROC =  $0.887 \pm 0.019$  for meropenem). A per-site performance analysis revealed that, while performances' variability was inherently linked to each center's sample size, *combatlearn*-based harmonization effectively aligned mean AUROC values across sites.

**Conclusions** Our study demonstrates the capability of MALDI-TOF mass spectra to predict amikacin and meropenem resistance in *K. pneumoniae* directly from clinical spectra, supporting its potential as a rapid and cost-effective approach for both antimicrobial resistance surveillance through machine learning models and clinical decision support in routine microbiology practice.

**Keywords** *Klebsiella pneumoniae*, Antimicrobial resistance, Amikacin, Meropenem, Mass spectrometry, MALDI-TOF, Machine learning, Batch-effect correction, Harmonization

## Background

Mass spectrometry techniques have recently been integrated into clinical diagnostic workflows, representing one of the most advanced frontiers in microbiological identification. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) provides broad-spectrum pathogen identification, allowing genus- or species-level recognition within a short time frame. This analytical technique enables microbial identification from bacterial colonies or biological samples such as blood cultures, which are directly inoculated onto a platinum target plate for spectrum acquisition [1].

Despite these advances, Gram-negative pathogens remain a major healthcare challenge due to high antimicrobial resistance (AMR) rates [2]. European epidemiological surveys have reported increasing trends in  $\beta$ -lactam resistance among Gram-negative bacteria, particularly in *Klebsiella pneumoniae* [3]. *K. pneumoniae* is a commensal and opportunistic pathogen of the human upper respiratory tract, skin, and gastrointestinal tract. Although usually part of the normal microbiota, it can cause severe healthcare-associated infections, especially in hospitalized or immuno-compromised patients. The most frequent clinical manifestations include respiratory tract infections and bacteremia, often associated with high morbidity and mortality.

Over the past decade, the emergence and dissemination of multidrug-resistant (MDR) *K. pneumoniae* isolates, particularly those resistant to last-line antibiotics such as carbapenems, has become a global public health concern. According to recent data from the ISS-ARISS surveillance network in Italy [4], after a moderate increase in 2019–2020 and a decline during 2021–2022, carbapenem resistance (e.g., to imipenem and meropenem) in *K. pneumoniae* showed a slight rise in 2023, with cumulative resistance rates decreasing from 33.2% in 2015 to 26.5% in 2023.

Similarly, third-generation cephalosporin resistance (cefotaxime, ceftazidime, ceftriaxone) showed a minor increase in 2023, though the overall trend from 2015 to 2023 remained stable. In contrast, aminoglycoside

resistance (gentamicin, amikacin) continued its gradual decline, from 42.4% in 2015 to 30.6% in 2023. However, following a reduction between 2020 and 2022, fluoroquinolone resistance (ciprofloxacin, levofloxacin) rose again in 2023 to 50.1%.

These epidemiological trends highlight the evolving dynamics of AMR in *K. pneumoniae* and emphasize the need for continuous surveillance, antimicrobial stewardship programs, and the development of new rapid methods for detecting resistance determinants, such as mass spectrometry combined with machine learning (ML).

Several studies have explored MALDI-TOF MS applications for AMR detection over the past decade [5]. Published data have demonstrated the potential of mass spectral peak analysis to identify  $\beta$ -lactamase-mediated hydrolysis of  $\beta$ -lactam antibiotics, where increasing peak intensities reflect hydrolysis products, and decreasing peaks indicate  $\beta$ -lactamase induction under antibiotic exposure [6–9]. Building on these principles, Oviano et al. proposed a similar approach for detecting carbapenem resistance in Gram-negative bacteria [10]. Other authors have applied a direct-on-target microdroplet growth assay to identify extended-spectrum  $\beta$ -lactamases (ESBLs), which involves incubating bacteria with and without antibiotics and  $\beta$ -lactamase inhibitors prior to MALDI-TOF analysis [11].

Moreover, Weis et al. [12] applied machine learning to MALDI-TOF spectra, demonstrating the potential of computational models for AMR classification. Together, these studies confirm the growing interest in detecting AMR through mass spectrometry-based approaches.

Based on this premise, the present study analyzes clinical isolates of *K. pneumoniae* via their MALDI-TOF spectra to predict susceptibility profiles to carbapenems and aminoglycosides using machine learning algorithms. This approach aims to advance mass spectrometry applications beyond species identification, extending its use to rapid, data-driven AMR prediction.

## Methods

### Collection of clinical isolates and corresponding spectra

The study involved three Italian hospital centers: the University Hospital “Policlinico G. Rodolico” (Catania), “Humanitas” Research Hospital (Rozzano), and the University Hospital “Fondazione Policlinico Universitario A. Gemelli” (Rome). Each center collected meropenem- and amikacin-susceptible and -resistant *K. pneumoniae* clinical isolates. These two antibiotics were selected due to their frequent inclusion in therapeutic regimens for both uncomplicated and severe infections.

Meropenem and amikacin resistance were defined as a minimum inhibitory concentration (MIC) greater than 8 mg/L, while susceptibility was defined as an MIC  $\leq$  2 mg/L for meropenem and  $<$  8 mg/L for amikacin, according to the latest EUCAST guidelines [13].

Isolate identification was performed using the MALDI Biotyper® Sirius System (Bruker, Billerica, MA, USA) with the MBT IVD Library (June 2021, Doc. No. 5023016). The methodology generated a mass spectrum, which was exported to the FlexAnalysis software platform (Bruker, Billerica, MA, USA). The software processed the spectral peaks, evaluating their consistency within the expected mass range.

Following FlexAnalysis processing, spectra were exported as individual.txt files for subsequent pre-processing and machine learning analyses.

### Spectra pre-processing

Mass spectra acquired via MALDI-TOF from clinical *K. pneumoniae* isolates were processed through a standardized pipeline, following the approach proposed by Weis et al. [14]. The objective of this pipeline was to extract a 6,000-dimensional feature vector from each spectrum, suitable for a machine-learning-based binary classification task.

The pre-processing phase began with a square-root transformation applied to each intensity value, a common technique to stabilize variance and reduce random noise. This step was followed by spectral smoothing using a Savitzky–Golay filter (polynomial order = 3; window size = 11) to preserve the true peak profile while attenuating high-frequency fluctuations. To remove background noise and baseline artifacts, the Statistics-sensitive Non-linear Iterative Peak-clipping (SNIP) algorithm [15] was applied.

After baseline correction, each spectrum was discretized through binning with a bin width of 3 Da – a favorable trade-off between preserving informative peaks and reducing the dimensionality of the feature set, as suggested by Weis et al. [14], considering only the m/z range from 2,000 to 20,000 Da, which is standard in clinical microbiology for bacterial species identification. The resulting feature vectors were normalized via Total

Ion Count (TIC) normalization, adjusting each intensity value ( $I_{ij}$ ) according to the formula:

$$I_{ij}' = I_{ij}/(TIC_j)$$

where.

$$TIC_j = \sum_{i=1}^n I_{ij}$$

Here,  $TIC_j$  represents the total ion count of the j-th spectrum,  $I_{ij}$  the i-th intensity (of the i-th bin) in the j-th spectrum, and n is the total number of bins. This normalization ensures comparability across spectra with varying total signal intensities.

All preprocessing steps were performed using the Mal-diAMRKit Python package (v0.3.0), available on GitHub at <https://github.com/EttoreRocchi/MaldiAMRKit>.

The resulting normalized vector representation of each MALDI-TOF spectrum was then used as input for supervised classification. When aggregating spectra from all centers, a harmonization step was introduced to mitigate inter-site variability and reduce batch effects that could bias results. This was achieved through a customized implementation of the ComBat algorithm, specifically adapted for integration into the machine learning pipeline. This tool ensures harmonization without data leakage and is publicly available at <https://github.com/EttoreRocchi/combatearn>. Specifically, harmonization is implemented as a transformation step within a scikit-learn *Pipeline*. Thus, ComBat parameters are fitted exclusively on the training portion of each outer cross-validation fold, and the harmonization model is subsequently applied to the corresponding validation split (within the inner cross-validation) or test split (within the outer cross-validation), so that data leakage is prevented both at the outer and inner cross-validation levels. Site-specific identifiers are the only covariates used by the ComBat model.

Finally, all feature vectors were standardized to zero mean and unit variance per feature within the training pipeline.

### Spectra analyses

MALDI-TOF spectra classification aimed at distinguishing between susceptible and resistant *K. pneumoniae* isolates was performed using multiple machine learning models, including logistic regression, random forest, XGBoost, and multilayer perceptron (MLP). Model training and evaluation were conducted within a nested cross-validation framework, consisting of a 10-fold outer loop for assessing generalization performance and a 5-fold

inner loop for hyperparameter tuning. Each antibiotic was analyzed independently.

We performed two complementary analyses to comprehensively evaluate the predictive capability of MALDI-TOF mass spectra in detecting antimicrobial resistance using machine learning models. These analyses were designed to capture distinct yet complementary aspects of the data and their clinical relevance.

The first analysis considered the entire dataset, aggregating spectra from all participating hospitals. This approach maximized sample size and diversity, allowing us to assess how well harmonized models generalize across epidemiologically distinct sites. Stratification during cross-validation was performed jointly by batch and binary outcome, the latter representing either resistance or susceptibility to the antibiotic under investigation. To mitigate class imbalance, all models were trained using class-balanced weighting to avoid favoring the majority class. Batch effect correction was implemented using a ComBat-based method [16] to mitigate site-specific biases, ensuring that performance estimates reflected biological rather than technical variation. Feature importance was assessed using SHapley Additive exPlanations (SHAP) values to interpret model behavior.

The second analysis focused on site-specific modeling strategies. For each clinical center, models were trained exclusively on local data using the same nested cross-validation scheme. In addition to within-site evaluation, each model was externally tested on datasets from the other two centers, enabling assessment of cross-site generalization.

This cross-validation strategy allowed us to evaluate both local discriminative power and external robustness, providing insight into the model's ability to generalize across different epidemiological and technical contexts. Such cross-site testing is particularly relevant for real-world deployment, where models may be applied outside the institution where they were developed.

Together, these two complementary strategies enabled the investigation of two key dimensions of model performance: (1) global robustness and scalability when using harmonized, multi-center data; and (2) local specificity and inter-site heterogeneity in resistance profiles and spectral characteristics. Combined, they offer a comprehensive understanding of the models' practical applicability in clinical microbiology.

## Results

Our study included a total of 686 *K. pneumoniae* isolates, which were collected from the participating hospital centers. Catania provided 86 isolates, Rome provided 471, and Milan provided 203 isolates. Figure 1 illustrates the number of isolates provided by each center, while Table

1 summarizes the susceptibility and resistance profiles of the collected *K. pneumoniae* isolates.

The performance evaluation of models trained on MALDI-TOF spectra, corrected for batch effects using the ComBat algorithm, revealed that XGBoost consistently outperformed other classifiers across both antibiotic targets – amikacin and meropenem. The reported values represent the mean performance obtained through 10-fold external cross-validation, with corresponding 95% confidence intervals. In the nested cross-validation procedure, both inner and outer folds were stratified by site of origin and resistance class (separately for amikacin and meropenem) to ensure that each fold accurately reflected the overall distribution with respect to site and prevalence.

### The amikacin target

For amikacin, the XGBoost model yielded a mean AUROC of  $0.822 \pm 0.028$  and a balanced accuracy of  $0.729 \pm 0.043$ . Figure 2 illustrates the per-site performance distribution across the 10 test iterations of the outer cross-validation for the harmonized datasets. The confusion matrix of the XGBoost classifier is presented in Figure 3, while the SHAP beeswarm plot summarizing feature importance within the framework is shown in Figure 4.

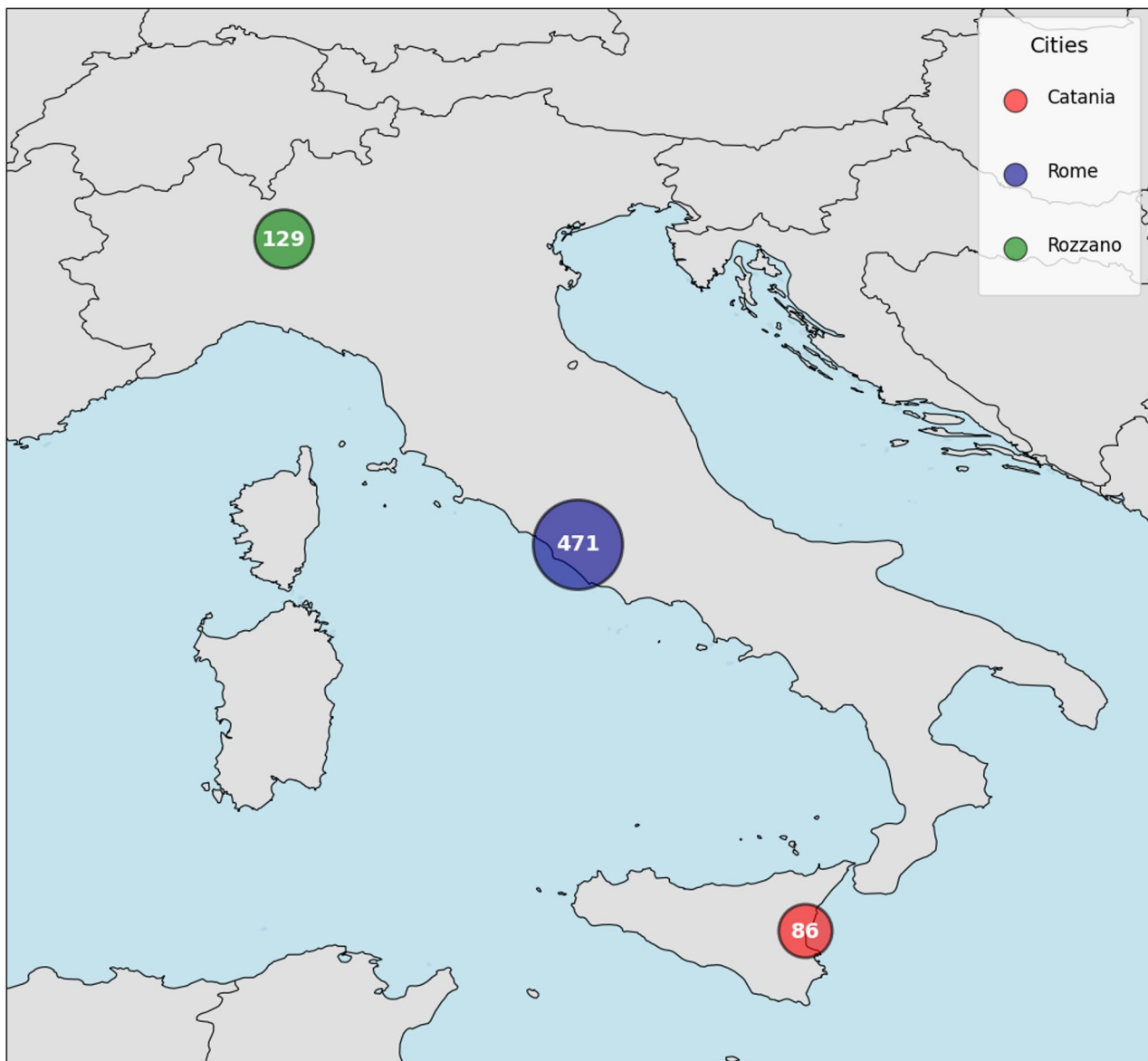
### The meropenem target

For meropenem, the XGBoost model achieved a mean AUROC of  $0.887 \pm 0.019$  and a balanced accuracy of  $0.768 \pm 0.015$ , indicating robust classification performance across the harmonized datasets. As for the amikacin analysis, per-site performances, the normalized confusion matrix, and the SHAP feature importance plots are presented in Figures 5, 6, and 7, respectively.

### Cross-site framework

For what concerns the cross-site evaluation, the heatmap illustrates the Area Under the Receiver Operating Characteristic (AUROC) scores obtained when training XGBoost models on data from individual clinical centers (Catania, Milan, and Rome) and evaluating them on both intra- and inter-center test sets. Each cell in the matrix represents the mean AUROC derived from a 10-fold cross-validation procedure, where models were trained exclusively on data from one center (rows) and tested on data from another – or the same – center (columns), while maintaining identical test folds across iterations to ensure a fair and consistent comparison.

Diagonal values represent within-center performance, highlighting how well each model captures local signal characteristics and resistance patterns. Notably, models trained and tested within Milan and Rome achieved strong discriminative performance (AUROC = 0.838



**Fig. 1** Number of *Klebsiella pneumoniae* isolates provided by the participating hospital centers

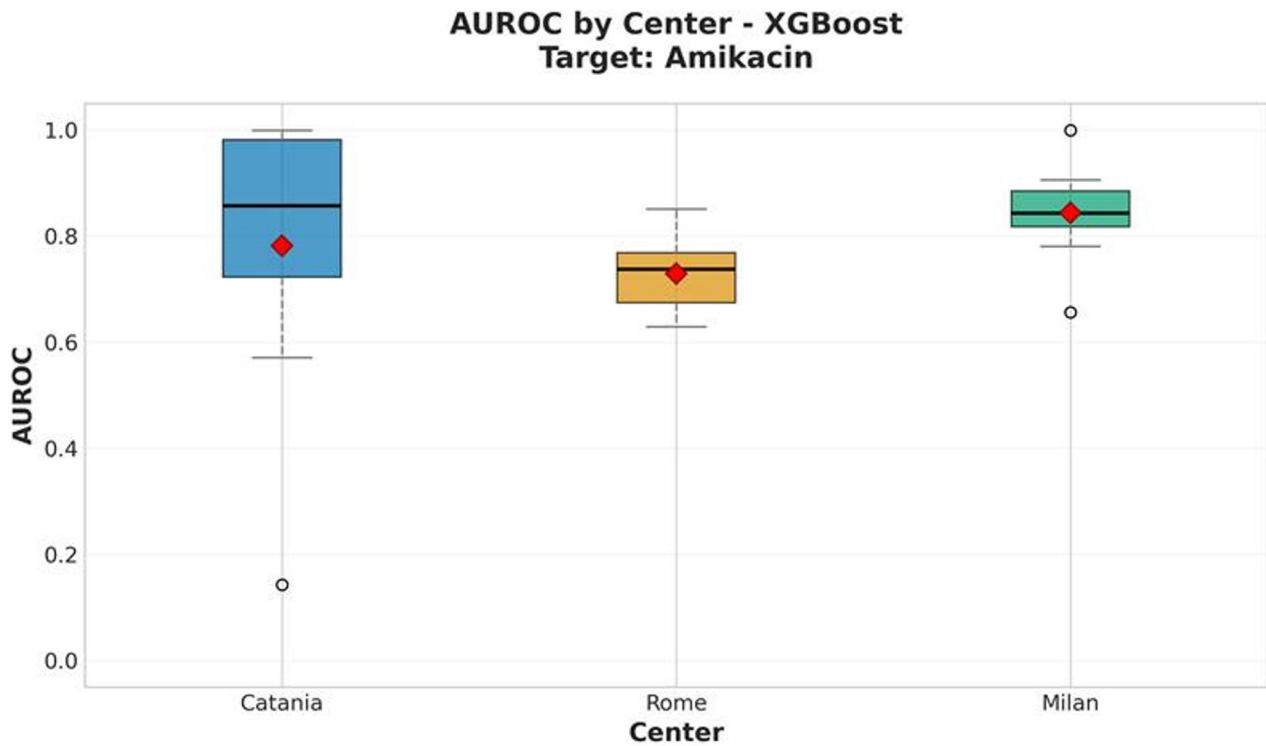
**Table 1** Summary of meropenem-susceptible, meropenem-resistant, amikacin-susceptible, and amikacin-resistant *Klebsiella pneumoniae* isolates provided by each participating hospital, along with their corresponding MIC ranges

Source	n° of MEM-susceptible strains	MEM-susceptibility MIC ranges (mg/L)	n° of MEM-resistant strains	MEM-resistance MIC ranges (mg/L)	n° of AK-susceptible strains	AK-susceptibility MIC ranges (mg/L)	n° of AK-resistant strains	AK-resistance MIC ranges (mg/L)
CATANIA	15	0.125–4.125	71	8–128	15	0.5–4.5	71	16–128
MILAN	99	0.25–4.25	30	16–64	104	4–8	25	16–32
ROME	199	0.125–128.125	272	16–128	338	2–8	133	16–128

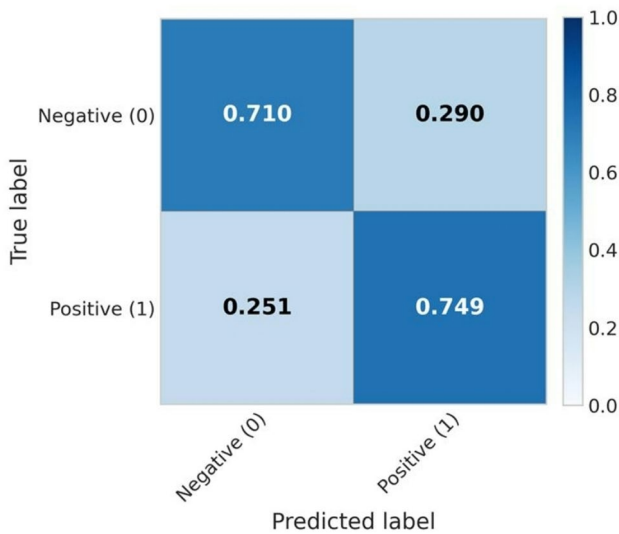
Abbreviations: MEM, meropenem; AK, amikacin

and 0.860, respectively), while Catania exhibited slightly lower within-site performance (AUROC = 0.792). Off-diagonal cells, on the other hand, reflect the model’s ability to generalize across sites. The generally lower

AUROC values in these positions suggest a reduction in cross-center generalizability, reflecting heterogeneity in local epidemiology or instrument calibration differences. Models trained on Rome data generalized better across



**Fig. 2** AUROC performance of the XGBoost model by center for amikacin resistance prediction. The red diamond indicates the mean value across all cross-validation iterations



**Fig. 3** Mean normalized confusion matrix across the 10-fold cross-validation for amikacin resistance prediction

other centers, likely due to the larger and more diverse dataset, which provided greater variability in the training phase and thus improved robustness.

It is worth noting that the comparability of AUROC values across cells was ensured by the design of the evaluation protocol. In each iteration of the cross-validation process, test folds were kept constant, allowing a direct

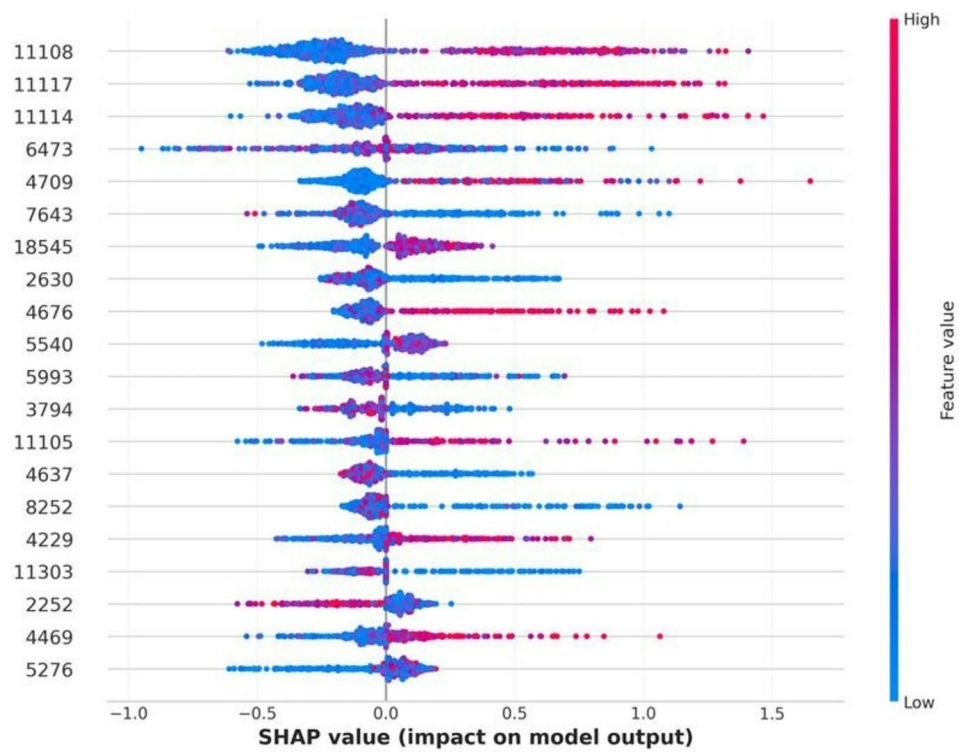
and fair comparison of results across centers and ensuring that observed differences reflected model generalization rather than test-set variance. A similar performance pattern was observed for amikacin, as detailed in the Supplementary Materials. These findings – summarized in Figure 8 – highlight both the opportunities and challenges of deploying predictive models across clinical centers, emphasizing the importance of harmonization or domain adaptation strategies to ensure robust performance in multi-site implementations.

**Comparative model performance summary**

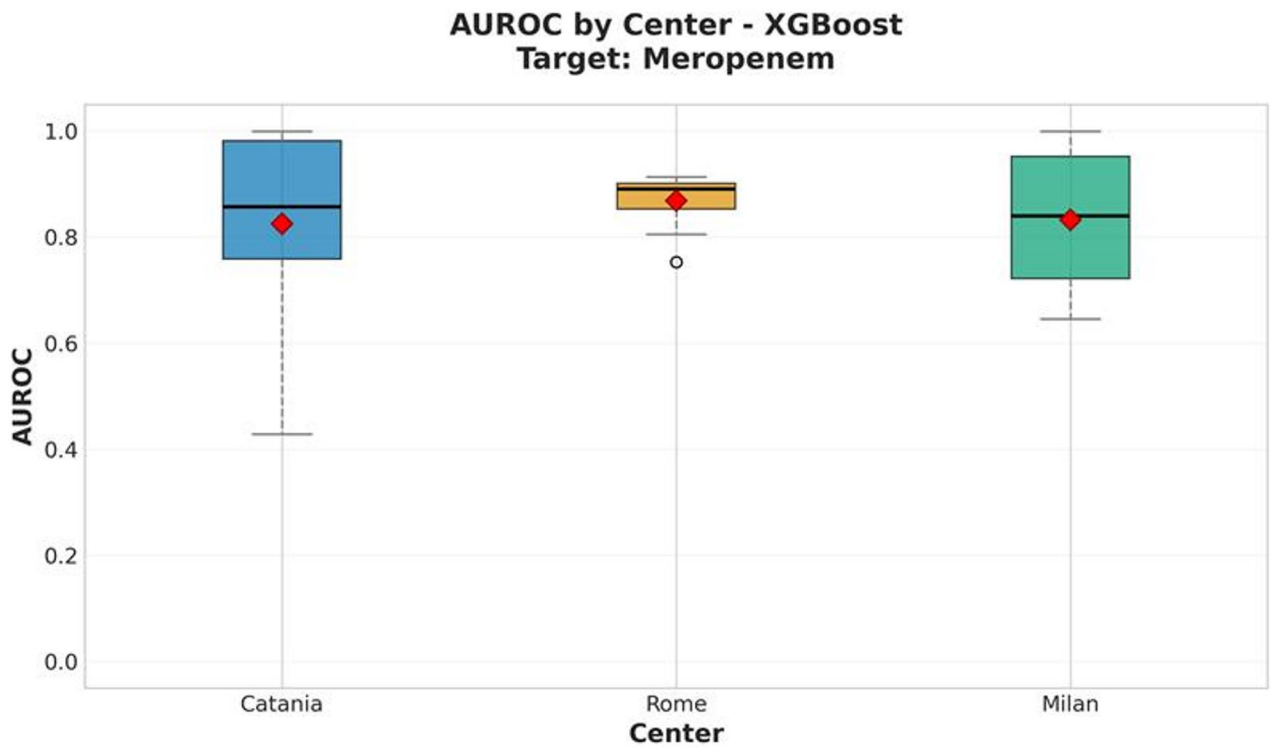
Regarding the comparative performance of the models, Table 2 and Table 3 summarize the balanced accuracy and AUROC values for each classifier trained on batch-corrected datasets. These values represent the mean performance across 10-fold external cross-validation, with 95% confidence intervals reported as error estimates. Detailed information on model selection, hyperparameter optimization, and additional performance metrics is provided in the Supplementary Materials.

**Discussion**

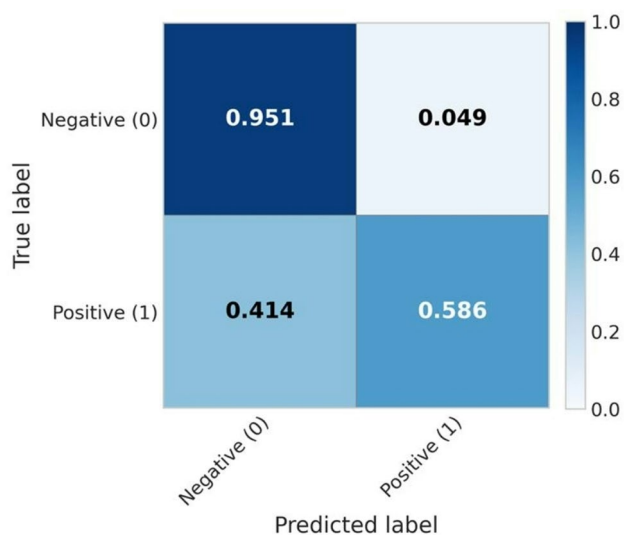
Multidrug-resistant (MDR) *K. pneumoniae* represents a major global threat, particularly in critical care settings and high-risk epidemiological areas. Various approaches have been explored to detect its resistance mechanisms,



**Fig. 4** SHAP beeswarm plot showing the top contributing features influencing the XGBoost model predictions for amikacin resistance



**Fig. 5** AUROC performance of the XGBoost model by center for meropenem resistance prediction. The red diamond indicates the mean value across all cross-validation iterations



**Fig. 6** Mean normalized confusion matrix across the 10-fold cross-validation for meropenem resistance prediction

including phenotypic, molecular, and rapid immunochromatographic assays [17]. Real-time polymerase chain reaction (PCR) technologies integrated the conventional resistance markers microbiological workflows, along with ultimate generation methods as whole genome sequencing (WGS) [18]. Moreover, microbiological laboratory routine often includes rapid phenotypic antimicrobial susceptibility testing (AST), which reveals minimum inhibitory concentration values directly from biological samples [19]. Certainly, rapid AST and molecular methods significantly reduced turn-around-time (TAT). However, they sometimes require expensive diagnostic tools and personnel expertise, limiting their potential extension to all the biological samples [17–19].

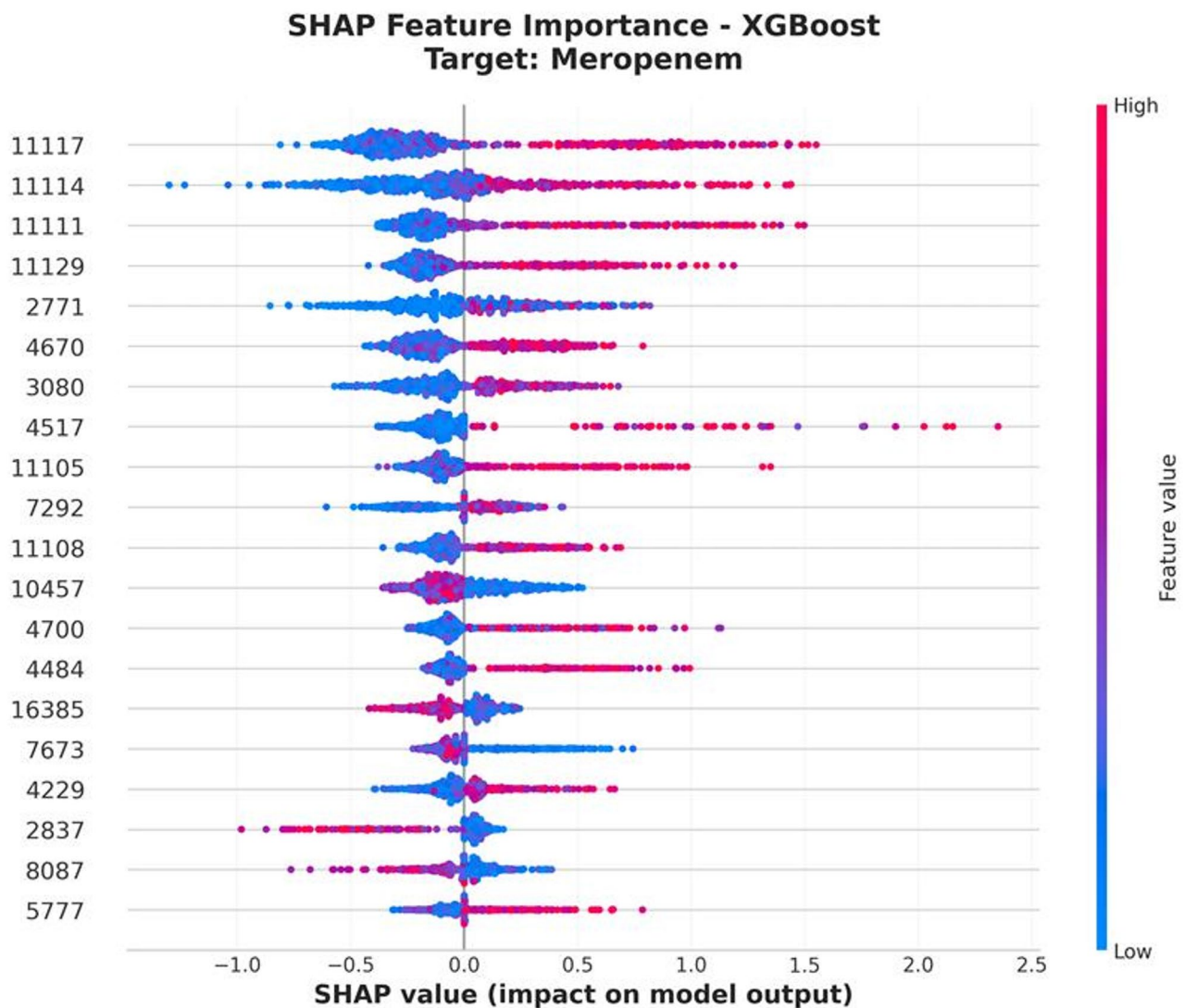
More recently, machine learning techniques have been applied to AMR prediction, showing considerable potential across both bacterial and yeast clinical isolates [14, 18, 19]. Mass spectrometry, already widely integrated into diagnostic workflows for microbial identification, further expands these possibilities. Its integration with ML could ultimately support automated, rapid AMR prediction among critical pathogens [20].

Given the widespread diffusion of MDR *K. pneumoniae* in Italy [3, 21], this multicenter study aimed to evaluate ML models for the rapid prediction of meropenem and amikacin resistance using MALDI-TOF MS spectra from clinical isolates. The three participating hospitals, located in Northern, Central, and Southern Italy, provided an ideal epidemiological representation for the study. MALDI-TOF MS spectra were analyzed using pre-processing and machine learning pipelines to identify spectral features associated with susceptibility or resistance patterns.

The superior performance of XGBoost across both antibiotics likely reflects its ability to capture complex, non-linear patterns and handle high-dimensional, col-linear spectral data. MALDI-TOF spectra are inherently variable and sparse, features that XGBoost can effectively manage through boosting. The SHAP-based interpretability analysis supported these findings, highlighting specific *m/z* peaks that consistently contributed to resistance prediction. Furthermore, harmonization with the ComBat algorithm (implemented via the *combat-learn* framework) successfully mitigated inter-site batch effects, as confirmed by consistent AUROC values across centers (Figs. 2 and 5). This adjustment preserved biological signals while minimizing technical bias, enhancing the reproducibility and generalizability of the models in multicenter settings.

At the site level, performance stability was highest in Rome, consistent with its larger dataset, while Catania exhibited slightly lower within-site AUROC scores due to its smaller sample size, which limits the reliability of site-specific conclusions. Nonetheless, it is worth pointing out that the harmonization-based framework enabled the predictive model to achieve promising performances even at this site, regardless of the limited training contribution it provided (as expected, these performances were associated with a higher variability, linked to the small sample size and to the inherent greater heterogeneity of the available isolates). More in depth, the performance degradation in the cross-site evaluation may be attributed to different insights. First, site-specific epidemiological differences must be acknowledged, including heterogeneity in the prevalence and distribution of resistance mechanisms which can alter spectral signatures and, ultimately, hinder transferability of models trained on a single site. Second, technical heterogeneities also arise, e.g. differences in mass spectrometers' calibration procedures. Third, the uneven sample sizes, particularly the predominance of isolates from Rome, may bias the model learning phase toward site-specific patterns, further reducing the generalization ability of models. Overall, cross-site testing confirmed that model generalization decreases when applied across centers, highlighting the importance of harmonization strategies to manage differences in local epidemiology and instrument calibration. Therefore, explicit modeling and correction of site-related batch effects within the learning pipeline enable the robustness and reliability of predictive performances and could represent a step forward clinical deployment of machine learning-based analyses of MALDI-TOF mass spectra for antimicrobial resistance prediction.

Our results align with those reported by López-Cortés et al. [22], who used CatBoost to predict ciprofloxacin-resistant *K. pneumoniae*, *Escherichia coli*, and *Staphylococcus aureus* based on MALDI-TOF spectra (AUROC =



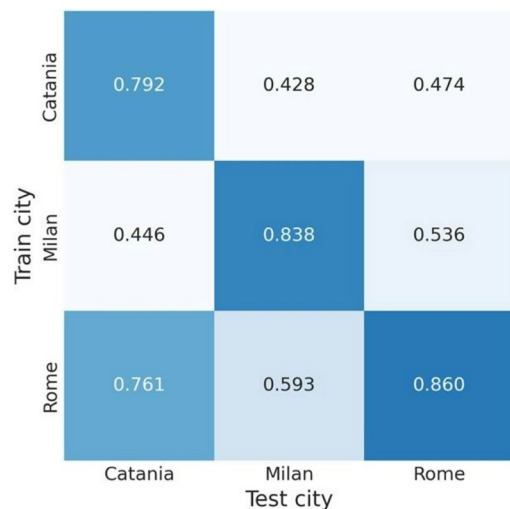
**Fig. 7** SHAP beeswarm plot showing the top contributing features influencing the XGBoost model predictions for meropenem resistance

0.73, F1 = 0.78) and similarly employed SHAP values to identify relevant m/z peaks [23]. Likewise, Nguyen et al. [24] applied AI models to *Pseudomonas aeruginosa*, using dynamic binning and transfer learning to improve cross-dataset generalizability, reporting strong performance for both  $\beta$ -lactam/ $\beta$ -lactamase inhibitors (AUROC = 0.869) and aminoglycosides such as amikacin (AUROC = 0.844) [23].

Compared with these studies, our work focuses specifically on meropenem and amikacin resistance in *K. pneumoniae*, achieving comparable or higher predictive performance. Moreover, our use of batch-effect correction through *combatlearn* demonstrates an additional methodological advancement, enabling reliable inter-site integration in ML workflows without data leakage. Regarding the achieved AUROC values (0.822 for amikacin and 0.887 for meropenem), interesting considerations

may be discussed. Unfortunately, there is not a current standardization about AUROC values and their applicability into clinical and therapeutical workflows. However, clinical diagnostic investigations previously recognized the AUROC interval 0.70–0.80 as acceptable, the 0.80–0.90 interval as excellent, and > 0.90 values as optimal [25, 26]. According to these assumptions, we could consider as promising the values obtained within our study, hypothesizing future implementations for both the tested molecules and their antimicrobial resistance prediction rates through machine learning models. On the other hand, further studies will be essential to clarify the obtained recall values, especially regarding the meropenem analysis.

According to our data, the 0.586–0.684 meropenem recall interval may be related to possible false susceptible



**Fig. 8** Cross-site AUROC performance of the XGBoost model for meropenem prediction

**Table 2** AUROC, balanced accuracy and resistance class's recall performances of the four machine learning models trained to classify amikacin resistance and susceptibility in *K. pneumoniae* isolates

Model	AUROC (Mean $\pm$ Std. Dev.)	Balanced Accuracy (Mean $\pm$ Std. Dev.)	Resistance class recall (Mean $\pm$ Std. Dev.)
XGBoost	<b>0.822 <math>\pm</math> 0.040</b>	<b>0.729 <math>\pm</math> 0.060</b>	0.749 $\pm$ 0.076
Random forest	0.802 $\pm$ 0.039	0.725 $\pm$ 0.036	<b>0.757 <math>\pm</math> 0.044</b>
MLP	0.739 $\pm$ 0.058	0.684 $\pm$ 0.061	0.692 $\pm$ 0.129
Logistic regression	0.741 $\pm$ 0.064	0.691 $\pm$ 0.067	0.642 $\pm$ 0.115

Highest results are highlighted in bold

**Table 3** AUROC, balanced accuracy and resistance class's recall performances of the four machine learning models trained to classify meropenem resistance and susceptibility in *K. pneumoniae* isolates

Model	AUROC (Mean $\pm$ Std. Dev.)	Balanced Accuracy (Mean $\pm$ Std. Dev.)	Resistance class recall (Mean $\pm$ Std. Dev.)
XGBoost	<b>0.887 <math>\pm</math> 0.027</b>	<b>0.768 <math>\pm</math> 0.022</b>	0.586 $\pm$ 0.044
Random forest	0.870 $\pm$ 0.036	0.673 $\pm$ 0.054	0.355 $\pm$ 0.107
MLP	0.768 $\pm$ 0.042	0.688 $\pm$ 0.050	0.465 $\pm$ 0.073
Logistic regression	0.829 $\pm$ 0.041	0.754 $\pm$ 0.044	<b>0.684 <math>\pm</math> 0.061</b>

Highest results are highlighted in bold

isolates, compromising a prompt clinical usage of the reported results.

Although the interesting data collected about meropenem and amikacin, most of the MDR *K. pneumoniae* isolates currently require the introduction of last resort antimicrobial choices. For instance, meropenem/vaborbactam demonstrated potent in vitro and in vivo activity

in the case of complicated *K. pneumoniae* infections [27]. Additionally, cefiderocol and aztreonam/avibactam became fundamental therapeutical choices against Gram-negatives severe infections, especially within high-risk epidemiological areas [28]. Finally, ultimate investigations involve cefepime/taniborbactam, which showed promising efficacy against *K. pneumoniae* clinical isolates [29, 30]. These newest antimicrobial resources may be furtherly analysed through machine learning models, particularly examining the models' evolution in presence of both the antibiotic and the  $\beta$ -lactamases inhibitor.

In conclusion, the integration of MALDI-TOF MS and ML represents a fast, cost-effective strategy for routine AMR surveillance in clinical microbiology. With further validation and expanded datasets, these models could be incorporated into existing laboratory workflows to provide real-time probabilistic outputs on susceptibility profiles, improving the speed and precision of antimicrobial therapy guidance.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12866-025-04657-2>.

Supplementary Material 1.

### Authors' contributions

E.R., E.N. and M.Ca. contributed to Conceptualization, Methodology, Investigation, Software, Visualization, Formal analysis, Validation and Writing – original draft. V.C., L.P., Z.A., G.M., B.P., C.S. and M.S. contributed to Data curation, Investigation, Resources, and Writing – Review & Editing. C.C., M.Cr., V.S. contributed to Writing – Review & Editing. G.C. and S.S. contributed to Supervision, Funding Acquisition, Project Administration, and Writing – Review & Editing.

### Funding

This research was supported by EU funding within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT).

### Data availability

The dataset analyzed in this study is publicly available in the MALDI-Kleb-AI Zenodo repository at: 10.5281/zenodo.17405072. The data processing and analysis pipeline used in the study is openly accessible on GitHub at: [<https://github.com/EttoreRocchi/MALDI-Kleb-AI>].

### Declarations

#### Ethics approval and consent to participate

This study involved bacterial isolates obtained from routine clinical diagnostic procedures and did not require any additional biological sampling. All isolates were anonymized, and no patient-related data were accessed. According to Italian regulations, studies based solely on anonymized bacterial isolates do not require ethical approval. In accordance with the Declaration of Helsinki, research using non-identifiable human material does not represent research involving human subjects; therefore, ethical review and informed consent were waived. The study was conducted in compliance with the General Data Protection Regulation (GDPR, EU 2016/679) and the Italian Legislative Decree 196/2003 (as amended by Legislative Decree 101/2018).

#### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Department of Medical and Surgical Sciences, Alma Mater Studiorum, University of Bologna, Bologna, Italy

<sup>2</sup>IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy

<sup>3</sup>Department of Biomedical and Biotechnological Sciences, University of Catania, Catania, Italy

<sup>4</sup>U.O.C. Laboratory Analysis, A.O.U. Policlinico "G. Rodolico-San Marco", Catania, Italy

<sup>5</sup>Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini 4, Pieve Emanuele, Milan 20072, Italy

<sup>6</sup>IRCCS Humanitas Research Hospital, Via Manzoni 56, Rozzano, Milan 20089, Italy

<sup>7</sup>Department of Laboratory and Hematology Sciences, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy

<sup>8</sup>Department of Basic Biotechnological, Intensive Care, and Perioperative Sciences, Università Cattolica del Sacro Cuore, Rome, Italy

<sup>9</sup>Precision Medicine in Clinical Microbiology Unit, Scientific Directorate, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy

<sup>10</sup>Unit of Microbiology, The Greater Romagna Area Hub Laboratory, Cesena 47522, Italy

Received: 30 October 2025 / Accepted: 18 December 2025

Published online: 19 January 2026

### References

- <https://asm.org/articles/2021/november/beyond-the-matrix-mass-spec-a-s-a-clinical-microbio>. (Accessed on 28th June 2025).
- Salam MA, Al-Amin MY, Salam MT, Pawar JS, Akhter N, Rabaan AA, et al. Antimicrobial resistance: a growing serious threat for global public health. *Healthcare*. 2023;11:1946. <https://doi.org/10.3390/healthcare111131946>.
- <https://www.ecdc.europa.eu/sites/default/files/documents/Antimicrobial%20Resistance%20surveillance%20in%20Europe%202023%20-%202021%20data.pdf>. (Accessed on 9th July 2025).
- [https://www.epicentro.iss.it/antibiotico-resistenza/ar-iss/RIS-5\\_2024.pdf](https://www.epicentro.iss.it/antibiotico-resistenza/ar-iss/RIS-5_2024.pdf). (Accessed on 10th July 2025).
- Palacios-Baena ZR, Gutiérrez-Gutiérrez B, De Cueto M, et al. Development and validation of the INCREMENT-ESBL predictive score for mortality in patients with bloodstream infections due to extended-spectrum- $\beta$ -lactamase-producing Enterobacteriaceae. *J Antimicrob Chemother*. 2017;72(3):906–13. <https://doi.org/10.1093/jac/dkw513>.
- Burckhardt I, Zimmermann S. Using matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry to detect carbapenem resistance within 1 to 2.5 hours. *J Clin Microbiol*. 2011. <https://doi.org/10.1128/jcm.00287-11>.
- Hrabák J, Walková R, Studentová V, Chudácková E, Bergerová T. Carbapenemase activity detection by matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry. *J Clin Microbiol*. 2011;49(9):3222–7. <https://doi.org/10.1128/JCM.00984-11>.
- Johansson A, Nagy E, Söki J, ESGAI (ESCMID Study Group on Anaerobic Infections). Detection of carbapenemase activities of *Bacteroides fragilis* strains with matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS). *Anaerobe*. 2014;26:49–52. <https://doi.org/10.1016/j.anaerobe.2014.01.006>.
- Jung JS, Eberl T, Sparbier K, et al. Rapid detection of antibiotic resistance based on mass spectrometry and stable isotopes. *Eur J Clin Microbiol Infect Dis*. 2014;33(6):949–55. <https://doi.org/10.1007/s10096-013-2031-5>.
- Oviaño M, Gato E, Bou G. Rapid detection of KPC-producing enterobacteriales susceptible to Imipenem/Relebactam by using the MALDI-TOF MS MBT STAR-Carba IVD assay. *Front Microbiol*. 2020;11:328. <https://doi.org/10.3389/fmicb.2020.00328>. (Published 2020 Feb 28).
- Correa-Martínez CL, Idelevich EA, Sparbier K, Kostrzewa M, Becker K. Rapid detection of Extended-Spectrum  $\beta$ -Lactamases (ESBL) and AmpC  $\beta$ -Lactamases in Enterobacteriales: development of a screening panel using the MALDI-TOF MS-based direct-on-target microdroplet growth assay. *Front Microbiol*. 2019;10:13. <https://doi.org/10.3389/fmicb.2019.00013>.
- Weis C, Cuénod A, Rieck B, et al. Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. *Nat Med*. 2022;28:164–74. <https://doi.org/10.1038/s41591-021-01619-9>.
- The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 15.0. 2025. <https://www.eucast.org>. Accessed on 10th February 2025 (Accessed on 3th July 2025).
- Weis C, Cuénod A, Rieck B, et al. Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. *Nat Med*. 2022;28(1):164–74. <https://doi.org/10.1038/s41591-021-01619-9>.
- Ryan CG, Clayton E, Griffin WL, Sie SH, Cousens DR. SNP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater Atoms*. 1988;34(3):396–402. [https://doi.org/10.1016/0168-583X\(88\)90063-8](https://doi.org/10.1016/0168-583X(88)90063-8).
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118–27. <https://doi.org/10.1093/biostatistics/kxj037>.
- Li J, Shi Y, Song X, Yin X, Liu H. Mechanisms of antimicrobial resistance in *Klebsiella*: advances in detection methods and clinical implications. *Infect Drug Resist*. 2025;18:1339–54. <https://doi.org/10.2147/IDR.S509016>.
- Kao K, Alocilja EC. A review of the diagnostic approaches for the detection of antimicrobial resistance, including the role of biosensors in detecting carbapenem resistance genes. *Genes*. 2025;16(7):794. <https://doi.org/10.3390/genes16070794>.
- Smith KP, Kirby JE. Rapid susceptibility testing methods. *Clin Lab Med*. 2019;39(3):333–44. <https://doi.org/10.1016/j.cll.2019.04.001>.
- Wang HY, Chung CR, Chen CJ, et al. Clinically applicable system for rapidly predicting *Enterococcus faecium* susceptibility to Vancomycin. *Microbiol Spectr*. 2021;9(3):e0091321. <https://doi.org/10.1128/Spectrum.00913-21>.
- Vella A, De Carolis E, Mello E, et al. Potential use of MALDI-ToF mass spectrometry for rapid detection of antifungal resistance in the human pathogen *Candida glabrata*. *Sci Rep*. 2017;7(1):9099. <https://doi.org/10.1038/s41598-017-09329-4>.
- López-Cortés XA, Manriquez-Troncoso JM, Sepúlveda AY, Soto PS. Integrating machine learning with MALDI-TOF mass spectrometry for rapid and accurate antimicrobial resistance detection in clinical pathogens. *Int J Mol Sci*. 2025;26(3):1140. <https://doi.org/10.3390/ijms26031140>.
- Yu J, Lin YT, Chen WC, et al. Direct prediction of carbapenem-resistant, carbapenemase-producing, and colistin-resistant *Klebsiella pneumoniae* isolates from routine MALDI-TOF mass spectra using machine learning and outcome evaluation. *Int J Antimicrob Agents*. 2023;61(6):106799. <https://doi.org/10.1016/j.ijantimicag.2023.106799>.
- Nguyen H-A, Peleg AY, Song J, et al. Predicting *Pseudomonas aeruginosa* drug resistance using artificial intelligence and clinical MALDI-TOF mass spectra. *mSystems*. 2024;9(9):e0078924. <https://doi.org/10.1128/mSystems.00789-24>.
- <https://qlik.qualitasiciliassr.it/anonimo/single/?appid=85ada16c-4b41-4bc6-9ca1-405b8243d0c2&sheet=6ad6f3ac-3369-41c5-bd72-792243f9091b&opt=c&txmenu,curssel> (Accessed on 7th July 2025).
- Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347–58. <https://doi.org/10.1056/NEJMra1814259>.
- Kim JI, Maguire F, Tsang KK, et al. Machine learning for antimicrobial resistance prediction: current practice, limitations, and clinical perspective. *Clin Microbiol Rev*. 2022;35(3):e0017921. <https://doi.org/10.1128/cmr.00179-21>.
- Sivori F, Francalancia M, Truglio M, et al. Meropenem/vaborbactam activity against carbapenem-resistant *Klebsiella pneumoniae* from catheter-related bloodstream infections. *Front Cell Infect Microbiol*. 2025;15:1616353. <https://doi.org/10.3389/fcimb.2025.1616353>. (Published 2025 Jul 31).
- Tozilyurt A. The effectiveness of aztreonam, aztreonam-avibactam, cefiderocol on *bla*<sub>NDM</sub>-producing *E. coli* isolates. *Mol Genet Microbiol Virol*. 2024;39:158–64. <https://doi.org/10.3103/S0891416824700198>.

30. Jacobs MR, Abdelhamed AM, Good CE, Mack AR, Bethel CR, Marshall S, et al. ARGONAUT-III and -V: susceptibility of carbapenem-resistant *Klebsiella pneumoniae* and multidrug-resistant *Pseudomonas aeruginosa* to the bicyclic boronate  $\beta$ -lactamase inhibitor taniborbactam combined with cefepime. *Antimicrob Agents Chemother.* 2024;68(9):e0075124. <https://doi.org/10.1128/aac.00751-24>.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.