Development of Novel Diagnostics for Profiling Drug Resistance in Clinically Important Bacterial Pathogens

Prof. Dr. Liang Wang
healthscience@foxmail.com
Research interests: Microbiology, Intelligent Medicine, Clinical Microbiology, Rapid Diagnostics, Raman Spectroscopy

Yongxuan Hong[a,#], Jie Chen[b,#], Xiuling Song[a], Liang Wang[a,b,c,d,e,f,*]

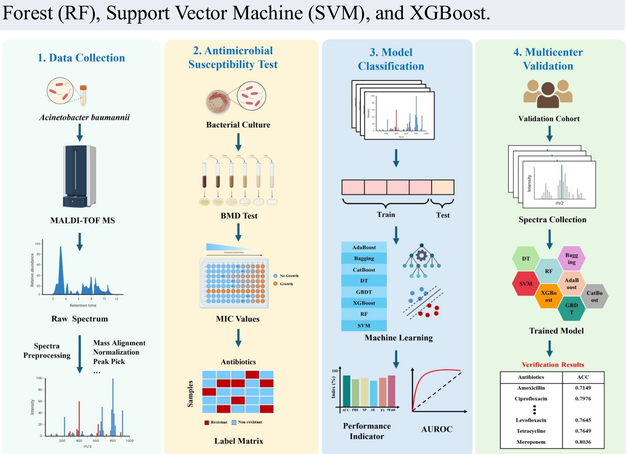
[a]Laboratory Medicine, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong Province, China
[b]School of Medicine, South China University of Technology, Guangzhou, Guangdong Province, China
[c]Division of Microbiology and Immunology, School of Biomedical Sciences, The University of Western Australia, Perth, Western Australia, Australia
[d]Center for Precision Health, School of Medical and Health Sciences, Edith Cowan University, Perth, Western Australia, Australia
[e]School of Agriculture and Food Sustainability, University of Queensland, Brisbane, Queensland, Australia
[f]Department of Applied Biology and Chemical Technology, Faculty of Science, The Hong Kong Polytechnic University (PolyU), Hong Kong SAR, China

Summary

Acinetobacter baumannii is a significant pathogen responsible for healthcare-acquired infections (HAIs), posing challenges due to its rising resistance to multiple antibiotics. Traditional diagnostic methods like bacterial cultures and antibiotic susceptibility testing (AST) are slow, necessitating the development of faster alternatives. This study aimed to create a predictive model for antibiotic resistance phenotypes of *A. baumannii* using clinical MALDI-TOF mass spectrometry (MS) spectra combined with machine learning techniques. A total of 3644 *A. baumannii* strains were analyzed for resistance patterns to nine antibiotic classes. Eight machine learning models were trained and evaluated, with the best model showing over 83% accuracy in predicting resistance to carbapenems, penicillin, and quinolones. The model performed especially well for imipenem and ceftazidime, with accuracies of 84.90% and 84.64%, respectively. Multi-center validation confirmed the model's robustness, achieving 81.68% and 80.72% accuracy for imipenem and ceftazidime. These findings demonstrate the potential of integrating machine learning with MALDI-TOF MS for rapid, accurate profiling of *A. baumannii's* antimicrobial resistance in clinical settings.

Methods

A total of 3644 *A. baumannii* strains were included in this study, which were isolated from blood, urine, respiratory tract, and wound specimens between 2019 and 2024 in the Department of Laboratory Medicine, Guangdong Provincial People's Hospital, and were subjected to a MALDI-TOF mass spectroscopy analysis. Mass spectra were acquired from all *A. baumannii* strains, and the spectra that were not identified as *A. baumannii* by MALDI-TOF MS were excluded. Additionally, to assess the robustness of the model, we collected MALDI-TOF mass spectral data as external validation dataset from three independent hospitals during the 2023-2024 period, which include Guangzhou Medical University First Affiliated Hospital (N=648); Sun Yat-sen University Affiliated First Hospital (N=450); and Sun Yat-sen University Affiliated Cancer Center (n=531). MICs of all *A. baumannii* strains were determined using the gold-standard broth microdilution method (BMD) according to the Clinical and Laboratory Standards Institute (CLSI) M7-A10 standard. A total of 21 antimicrobial agents were included in this study: amikacin, gentamicin, tobramycin, imipenem, meropenem, cefepime, cefoperazone/sulbactam, ceftazidime, ceftriaxone, tigecycline, ampicillin/sulbactam, piperacillin, piperacillin/tazobactam, ticarcillin, colistin, polymyxin, ciprofloxacin, levofloxacin, sulfamethoxazole/trimethoprim, doxycycline, and minocycline. All experiments were performed in triplicate to ensure reliability. The dataset was randomly split into a training set (80%) and a testing set (20%) to ensure unbiased model training and evaluation. To preprocess the sample labels, we used the *LabelEncoder* function and the *to_categorical* method from the Scikit-learn package (version 0.21.3) to convert the sample labels into a label-encoded format suitable for machine learning analysis. The following machine learning algorithms were employed for model development: Adaptive Boosting (AdaBoost), Bootstrap Aggregating (Bagging), Categorical Boosting (CatBoost), Decision Tree (DT), Gradient Boosting Decision Tree (GBDT), Random Forest (RF), Support Vector Machine (SVM), and XGBoost.



Results

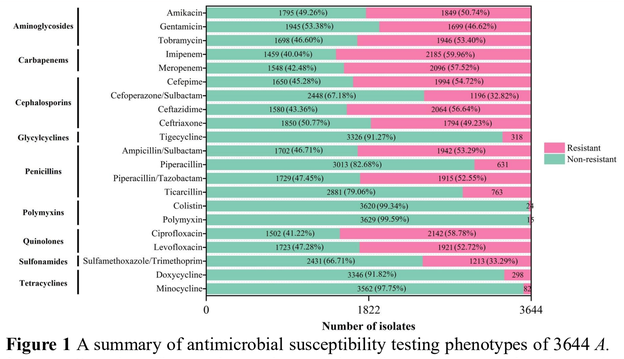


Figure 1 A summary of antimicrobial susceptibility testing phenotypes of 3644 *A. baumannii* strains. The number and proportion of non-resistant and resistant *A. baumannii* isolates for each antimicrobial phenotype are visualized.

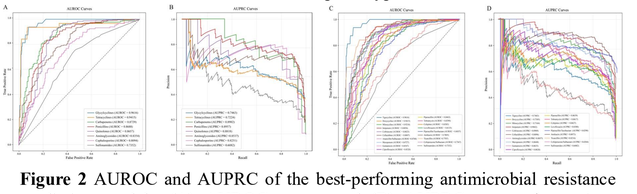


Figure 2 AUROC and AUPRC of the best-performing antimicrobial resistance prediction models. (A) AUROC curves and (B) AUPRC curves for various antibiotic types. (C) AUROC curves and (D) AUPRC curves for individual antibiotics.

Table 1 The Optimal predictive models for A. baumannii resistance to individual antibiotics.

Classification	Antibiotics	Best Model	Accuracy (95% CI)	Precision (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1-score (95% CI)	5-fold CV (95% CI)
Aminoglycosides	Amikacin	XGBoost	75.45% (72.29-78.46%)	68.57% (63.70-73.57%)	72.97% (67.69-78.06%)	77.14% (73.13-81.04%)	70.70% (66.44-74.63%)	75.45% (74.00-76.91%)
	Gentamicin	GBDT	78.88% (75.86-81.76%)	70.36% (65.18-75.50%)	77.42% (72.56-82.27%)	79.78% (76.03-83.44%)	73.72% (69.52-77.61%)	78.85% (77.58-80.13%)
	Tobramycin	XGBoost	80.52% (76.68-83.40%)	71.53% (66.01-76.95%)	76.43% (71.12-81.44%)	82.83% (79.25-86.31%)	73.90% (69.55-77.98%)	80.69% (79.37-82.02%)
Carbapenems	Imipenem	XGBoost	84.91% (82.30-87.38%)	78.65% (73.68-83.40%)	79.85% (75.00-84.49%)	87.77% (84.79-90.62%)	79.25% (75.34-82.89%)	84.68% (83.60-85.76%)
	Meropenem	XGBoost	82.44% (79.70-85.19%)	76.57% (71.65-81.27%)	80.28% (75.56-84.75%)	83.86% (80.36-87.26%)	78.38% (74.49-81.80%)	82.44% (81.38-83.50%)
	Cefepime	CatBoost	80.66% (77.78-83.54%)	76.57% (74.07-83.01%)	79.28% (74.79-83.63%)	81.82% (78.00-85.64%)	78.92% (75.39-82.28%)	80.26% (79.24-81.28%)
Cephalosporins	Cefoperazone/ Sulbactam	XGBoost	65.16% (61.59-68.59%)	43.62% (37.99-49.16%)	60.19% (53.77-66.67%)	67.25% (63.07-71.29%)	50.58% (45.27-55.58%)	63.53% (61.04-66.02%)
	Ceftazidime	CatBoost	84.64% (81.89-87.24%)	78.44% (73.18-83.33%)	79.62% (74.61-84.41%)	87.50% (84.28-90.52%)	79.03% (74.95-82.64%)	84.71% (83.62-85.79%)
	Ceftriaxone	CatBoost	79.15% (76.27-82.17%)	70.57% (65.46-75.69%)	76.73% (71.75-81.79%)	80.62% (76.97-84.26%)	73.52% (69.35-77.46%)	78.71% (77.40-80.02%)
Glycylcyclines	Tigecycline	CatBoost	95.34% (93.69-96.71%)	94.37% (88.46-98.68%)	69.07% (59.75-77.89%)	99.37% (98.72-99.84%)	79.76% (72.72-85.71%)	95.19% (93.82-96.55%)
Penicillins	Ampicillin/ Sulbactam	CatBoost	79.97% (77.09-82.72%)	74.58% (69.64-79.41%)	76.11% (71.08-80.95%)	82.57% (78.96-86.05%)	75.34% (71.38-78.98%)	79.47% (78.21-80.72%)
	Piperacillin	GBDT	82.44% (79.56-85.19%)	77.50% (72.70-81.79%)	81.58% (77.05-85.81%)	83.06% (79.43-86.44%)	79.49% (75.83-82.79%)	82.30% (81.33-83.28%)
	Piperacillin/Tazobactam	GBDT	78.33% (75.31-81.34%)	70.53% (65.23-75.71%)	75.53% (70.50-80.51%)	80.09% (76.17-83.72%)	72.95% (68.75-76.82%)	77.81% (76.47-79.16%)
	Ticarcillin	XGBoost	79.15% (76.13-82.17%)	69.86% (64.44-75.00%)	76.12% (71.00-81.19%)	80.91% (77.38-84.49%)	72.66% (68.58-76.80%)	79.39% (78.04-80.75%)
Quinolones	Ciprofloxacin	GBDT	83.13% (80.38-85.87%)	75.85% (70.80-80.77%)	77.31% (72.14-82.21%)	86.35% (83.19-89.32%)	76.57% (72.50-80.42%)	82.73% (81.50-83.96%)
	Levofloxacin	CatBoost	79.70% (76.68-82.58%)	70.89% (65.40-76.03%)	76.67% (71.43-81.58%)	81.48% (77.83-84.99%)	73.67% (69.33-77.65%)	79.84% (78.53-81.15%)
Sulfonamides	Sulfamethoxazole/ Trimethoprim	XGBoost	66.67% (63.24-70.10%)	44.17% (38.49-50.18%)	59.52% (52.88-65.88%)	69.56% (65.50-73.55%)	50.71% (45.27-55.91%)	65.67% (63.14-68.19%)
Tetracyclines	Doxycycline	GBDT	95.61% (94.10-96.98%)	92.16% (83.72-98.28%)	62.67% (51.90-73.42%)	99.39% (98.75-99.85%)	74.60% (65.57-82.50%)	95.36% (93.65-97.08%)
	Minocycline	GBDT	95.20% (93.55-96.71%)	92.73% (85.11-98.48%)	62.20% (51.46-72.94%)	99.38% (98.75-99.85%)	74.45% (65.55-82.19%)	94.31% (92.59-96.04%)

Conclusion

In this study, we systematically evaluated the resistance of multiple antibiotics in *A. baumannii* using a larger cohort and performed multicenter external validation, which enhanced the robustness and clinical relevance of our findings. While our findings highlight the advantages of MS-based machine learning models in predicting AMR, further improvements are necessary to maximize their clinical utility. In conclusion, ML-assisted analysis of MALDI-TOF mass spectra offers a promising, scalable solution for rapid resistance prediction in *A. baumannii* strains, addressing a critical need for timely and accurate diagnostics in the face of rising multidrug resistance. The integration of machine learning with MALDI-TOF MS enables the detection of subtle, resistance-associated proteomic patterns that are often imperceptible to conventional methods, making it particularly useful for the identification of resistant strains in real-time clinical settings. This approach has demonstrated strong performance in predicting resistance across multiple antibiotic classes, particularly β-lactams, aminoglycosides, and quinolones, offering immediate clinical utility for high-priority antibiotics. Future research efforts should focus on expanding the dataset to include a wider variety of resistance profiles, integrating genomic data for more precise resistance predictions, and developing advanced machine learning techniques to enhance model sensitivity and reduce overfitting. By addressing these challenges, this method can evolve into an applicable tool in the fight against antimicrobial resistance, ultimately improving patient outcomes and supporting the efforts of clinicians to optimize antibiotic stewardship.