scientific reports

Check for updates

OPEN

Repurposing MALDI-TOF MS for effective antibiotic resistance screening in *Staphylococcus epidermidis* using machine learning

Michael Ren[1]✉, Qiang Chen[2] & Jing Zhang[3]

The emergence of *Staphylococcus epidermidis* as a significant nosocomial pathogen necessitates advancements in more efficient antimicrobial resistance profiling. However, existing culture-based and PCR-based antimicrobial susceptibility testing methods are far too slow or costly. This study combines machine learning with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) to develop predictive models for various antibiotics using a comprehensive dataset containing thousands of *S. epidermidis* isolates. Optimized machine learning models utilized feature selection and achieved high AUROC scores ranging from 0.80 to 0.95 while maintaining AUPRC scores up to 0.97. Shapley Additive exPlanations were employed to analyze relevant features and assess the significance of corresponding protein biomarkers while also verifying that predictive power was derived from the detection of proteins rather than noise. Antimicrobial resistance models were validated externally to evaluate model performance outside the original data collection site. The approaches and findings in this study demonstrate a significant advancement in rapid, cost-effective antimicrobial resistance profiling, offering a promising solution for improving treatments for nosocomial infections and being potentially applicable to other microbial pathogens in the future.

Keywords MALDI-TOF MS, Machine learning, Antibiotic resistance, Resistance screening, *Staphylococcus epidermidis*, Diagnostics

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is an emerging technique that has solidified its usefulness in routine and rapid microbial species identification[1]. Despite its practicality, the implementation of MALDI-TOF MS in antimicrobial susceptibility testing (AST) has yet to be fully explored. Previous research attempted to predict resistance based on the existence of singular peaks, but such methods are inconsistent and do not consider significant patterns of multiple peaks that impart resistance in combination[2,3]. More recent works have demonstrated the potential of artificial intelligence in detecting such patterns and successfully predicting resistances for clinically relevant bacteria such as *Staphylococcus aureus*, *Escherichia coli*, *Klebsiella pneumoniae*, and *Enterococcus faecium*[2,4]. However, *Staphylococcus epidermidis*, a member of the coagulase-negative staphylococci bacterial family, has recently emerged as a formidable pathogen, gaining prominence as a leading agent behind nosocomial infections within the body[5]. To date, no studies have demonstrated a substantial MALDI-TOF MS approach for antimicrobial resistance (AMR) prediction in *S. epidermidis*.

Under normal circumstances, *S. epidermidis* is a standard component of human epithelial microflora and maintains a harmless relationship with humans. However, many *S. epidermidis* isolates can form biofilms—conglomerates of multiple cells that form on surfaces—and pose an issue due to the difficulty of eradicating such formations. Due to the ubiquity of *S. epidermidis* on various surfaces, *S. epidermidis* has become a common contaminant of medical devices despite the sanitation protocols typically implemented within hospital settings, leading to various implications[6]. For example, *S. epidermidis* stands out as the predominant cause of infections associated with indwelling medical devices. One study from the United States indicated that at least 22% of central intravenous catheter insertions were caused by *S. epidermidis* alone[7]. A different study found that *S. epidermidis* was behind approximately 13% of prosthetic valve endocarditis infections, which was followed by an intracardiac abscess rate and mortality rate of 38% and 24%, respectively[8]. Such infections are of serious concern

[1]Syosset High School, Syosset, NY 11791, USA. [2]Lieber Institute for Brain Development, Baltimore, MD 21205, USA. [3]Purdue University in Indianapolis, Indianapolis, IN 46202, USA. ✉email: michael.jiujiu.ren@gmail.com

because *S. epidermidis* is extremely difficult to treat once developed, which incurs annual expenditures of more than $2 billion in the United States healthcare system alone[9].

As such, it is imperative to develop more rapid and effective methods of diagnosis for *S. epidermidis* to treat infections before they can progress. By facilitating the early detection of microbes and gaining an in-depth understanding of their inherent resistance mechanisms, healthcare professionals can significantly improve the precision and effectiveness of their treatments[10]. This necessitates extensive AMR profiles to ensure optimal treatment plans are prescribed and tailored to the specific microbe at hand. These plans benefit not only the patient but also contribute to antibiotic stewardship efforts to combat the rise of multidrug-resistant bacteria[11].

However, existing methods for AMR profile creation come with many drawbacks that make diagnosis suboptimal, thereby increasing patient morbidity, mortality, and healthcare costs[9]. Resistance to antibiotics is commonly determined by culture-based methods, which are known to have significant drawbacks. For one, it can take up to 96 h for a complete AMR resistance profile to be reported, representing a significant time gap between sample collection and results[12]. In addition, these methods are expensive and require a high level of technical expertise to ensure that AMR assays are completed successfully[13]. Using culture-based methods can, therefore, delay the delivery of treatment decisions, which leads to the prescription of suboptimal antibiotics—either too narrow- or broad-spectrum—that can complicate treatment and put the patient's health at risk[14,15].

Another approach to determining antibiotic resistance involves amplifying genes and analyzing them via polymerase chain reaction (PCR). Although PCR can quickly detect resistance biomarkers directly from patient samples, it is generally limited to single, targeted genes. This narrow focus means PCR may not account for the full range of resistance mechanisms, such as non-genetically mediated resistance like efflux pump upregulation. Additionally, the need for multiple gene assays to cover various resistance mechanisms contributes to its higher cost compared to other diagnostic methods[4].

As such, looking towards other unexplored methods like MALDI-TOF MS to predict AMR is essential. MALDI-TOF-MS works by profiling bacterial extracts and constructing a mass spectral fingerprint of microorganisms, which can be subsequently matched with databases of known mass spectra patterns[16]. Recent studies have suggested that MALDI-TOF MS could also serve as a potential tool for AMR profiling[2,4]. Studies have found that the diagnostic turnaround time for MALDI-TOF MS is approximately 38 min and that MALDI-TOF MS takes a mere $0.50 per analysis[17]. As a result of these advantages, MALDI-TOF MS has become the most widespread method for species-level microbial identification in clinical laboratories, which means that efficient integration of the machine learning workflow for AMR prediction into existing sites would be feasible[8]. Furthermore, recent studies have already achieved groundbreaking results using machine learning to analyze MALDI-TOF MS data with microbes such as *Staphylococcus aureus*, reporting area under the receiver operating characteristic (AUROC) scores up to 0.89[2]. AUROC is a commonly used metric in medicine to summarize the effectiveness of a model in predicting a binary outcome, with scores closer to 1.0 indicating better predictive power.

Thus, to combat the emergence of *S. epidermidis* as a clinically relevant pathogen, a novel machine learning coupled with MALDI-TOF MS approach was taken to create prediction models for a variety of different antibiotics. Over 4,000 samples of *S. epidermidis* and their corresponding AMR profiles were taken from a mass spectra dataset compiled by Weis et al. from four different clinical sites between 2015 and 2018. In the study associated with the dataset, which contained data for multiple species, Weis et al. focused specifically on *Staphylococcus aureus*, *Escherichia coli*, and *Klebsiella pneumoniae* because it was found that species-specific predictions yielded high performance. However, they did not take *S. epidermidis* into account while developing and analyzing models for individual species. Furthermore, the priority pathogens evaluated in the study attained a maximum AUROC of 0.80, which does not indicate poor performance but suggests room for improvement through additional measures[4].

Given the limitations of previous work and the lack of MALDI-TOF MS studies on nosocomial infections like S. *epidermidis*, the present study proposes a multitude of robust models for predicting the susceptibility of *S. epidermidis* to several antibiotics from various antibiotic families. SHapley Additive Explanations (SHAP) values were analyzed to improve model interpretations. The most significant feature bins for several models were examined and aligned with proteins listed in UniProt, confirming that the model's selected features corresponded to combinations of relevant proteins associated with resistance. Results from the models exemplify not only a viable solution for improving existing methods for diagnosing and treating *S. epidermidis* infections but also a framework for developing MALDI-TOF MS diagnosis and prescription methodologies for other microbes in the future.

Methods

Data

The present study utilizes an extensive mass spectra data collection of 303,195 microbes known as the Database of Resistance against Antimicrobials with MALDI-TOF Mass Spectrometry (DRIAMS) that was created by Weis et al. between 2015 and 2018. The mass spectra were obtained using the Bruker Microflex MALDI-TOF MS system. Raw data was collected in .fid file format before preprocessing with MALDIquant v1.19 and binning[4]. For each microbe analyzed, a corresponding instance exists in the DRIAMS dataset that contains the results of the MALDI-TOF MS analysis on the microbe and a laboratory-confirmed AMR profile. The dataset is also broken down into four sub-collections labeled DRIAMS-A to -D, each corresponding to a different site from which the data was collected. Of these sub-collections, DRIAMS-A has been identified to be the largest and most comprehensive subcollection, containing over 145,341 mass spectra in total[18]. As such, the present study focuses on DRIAMS-A and uses the other data sites for external testing.

A processed version of the DRIAMS dataset was obtained from Kaggle with raw and preprocessed files that were unnecessary for model training removed[19]. For each microbe sample, the downloaded dataset contains

metadata, including the identified species, an AMR profile, and a sample ID (e.g., a0a359b5-c00a-4e6e-a195-728af31f7666) that corresponds to the file containing the mass spectra of the instance. For the present study, resistant and intermediate samples were labeled as positive classifications and susceptible samples as negative classifications.

Preprocessing specifics

The original dataset was preprocessed using mass spectra bins of 3 Da before machine learning analysis to balance peak separation and computational efficiency. Preprocessing involved intensity variance stabilization, smoothing, baseline removal, total ion current calibration, and mass spectra trimming to a range of 2,000 to 20,000 Da. Subsequently, each mass spectra was separated with a bin size of 3 Da and turned into a 6,000-dimensional vector for model training. Moreover, AMR resistance profiles for each antibiotic were initially reported by the assays to be either resistant, intermediate, or susceptible[4].

Project workflow

The proposed workflow for the present study involves various processing, training, and interpretation steps (Fig. 1). After obtaining the dataset, a random forest model was fitted to the data and combined with a meta-transformer to filter out irrelevant features. Six different classifiers were trained and tuned using filtered data, and then interpreted with biomarker analysis and external validation. The source code used for this study is available on GitHub.

Feature selection

Datasets with too many irrelevant features can lead to redundant and overcomplicated algorithms with low accuracy and unnecessarily long training times. With over 6,000 mass spectra bins that profile the composition of each *S. epidermidis*, not every mass spectrum value generated by MALDI-TOF MS necessarily corresponds to a protein involved in imparting AMR to the bacterium. As such, feature selection becomes necessary to decrease the time required to train models, reduce the likelihood of overfitting, and avoid the curse of dimensionality that affects high-dimensional datasets like that of DRIAMS-A[20]. In contrast to the original study conducted by Weis et al. that did not employ feature selection, the workflow of the current study considers the need to filter out irrelevant features to improve performance. A meta-transformer based on the Random Forest estimator was used to isolate relevant features and eliminate all other irrelevant ones[21]. The meta-transformer selects features by fitting a Random Forest model to the entire dataset, calculating the importance of each feature using the average reduction in impurity, and selecting relevant features based on a threshold with the 1.25 times the mean feature importance score as a baseline.

Initial modeling

Each feature-selected mass spectra file was grouped into a single collection containing all mass spectra files for each unique antibiotic. These collections were subsequently split into a 70/30 training and test dataset with a stratification based on the susceptibility to the attributed antibiotic. For each antibiotic, models were constructed with the training datasets, while the test dataset was used for external validation on each trained model. A total of six different algorithms were implemented for each antibiotic collection: Random Forest (RF), Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), LightGBM, and Multilayer Perceptron (MLP), as these models have performed well in previous machine learning studies on MALDI-TOF MS[4,21]. All

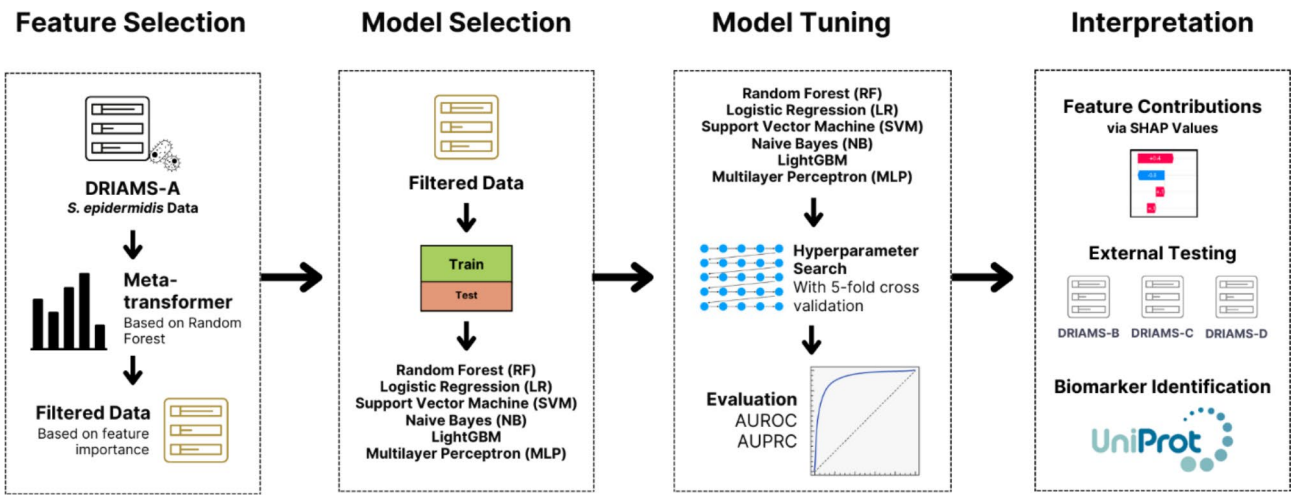

Fig. 1. Summary of modeling workflow. *S. epidermidis* data is filtered with a meta-transformer and quickly tested with general, unoptimized models. This was followed by an extensive tuning process that aimed to optimize AUROC. After tuning, models were evaluated with SHAP values, which were subsequently matched with possible or known biomarkers. Finally, external testing on sites outside of DRIAMS-A was conducted.

classifiers were created with Python 3.11.3 and LightGBM —installed separately—was the only library not based on Scikit-learn.

Model tuning

After initial model selection, the best-performing model for each antibiotic was selected and optimized based on AUROC. AUROC was the evaluation metric of choice because it is not influenced by class ratio and provides an interpretable, overarching summary of performance with positive and negative susceptibility predictions for various classification thresholds. In addition, while many of the modeled antibiotics had relatively balanced datasets four antibiotic datasets had a positive (intermediate/resistant instances) class prevalence (PCP) of < 0.20 or > 0.80. Since AUROC is relatively invariant to the ratio of positive or negative classes, it is an acceptable metric for comparing the performance of different antibiotic models with different class ratios[4]. Tuning was performed via hyperparameter optimization using Scikit-learn by exhaustively searching through a specified subset of the hyperparameter space—known as a grid—that is individually defined for each classifier. Validation of the model for each combination of hyperparameters in the grid was done using 5-fold cross-validation with a scoring method aiming to optimize AUROC. After the model tuning process, the area under the precision-recall curve (AUPRC) metric was also used to evaluate the tuned models. AUPRC is another holistic metric used to assess model performance based on precision and recall across various thresholds and is especially useful for models trained on imbalanced datasets.

Interpretation of SHAP values

SHapley Additive exPlanations (SHAP) values can provide a more interpretable framework by quantifying the contribution of each feature to the classification outcome using a game theory approach[22]. Unlike Random Forest feature importances, which are usually based on the decrease in Gini impurity across all decision trees for each feature, SHAP values calculate the marginal contribution of each feature for every individual prediction. This can provide a more holistic view of all feature importances and illustrate the change in the expected model prediction for each specific feature. As such, SHAP values ensure fair interpretations in complex models where interactions between features can be non-linear and interdependent[23]. SHAP values were calculated for selected models to analyze how AMR models arrived at their predictions.

Biomarker identification

Unlike decision trees, where the feature bins are ranked between 0 and 1, the range for possible SHAP values can vary widely depending on the model and data. However, higher absolute SHAP values always indicate a greater impact on model decisions. The most significant feature bins based on calculated SHAP values were investigated using UniProt and matched to the closest significant proteins around the range of Da values for each feature bin. This process helps identify novel resistance mechanisms and verify if the features identified by the model correspond to well-known and reputable mechanisms of antimicrobial resistance. UniProt analysis ensures that models are not simply capturing unclear patterns or noise within mass spectra profiles. In addition, it can also help identify connections with other related microbes like *S. aureus*.

External validation

Models for each antibiotic are primarily created using DRIAMS-A because it is the most extensive and viable dataset out of DRIAMS- A-D. Given the independent construction of each DRIAMS dataset at distinct sites, this presents an ideal opportunity to evaluate the transferability of models across different clinical centers. The models created based on antibiotics from DRIAMS-A were evaluated for their performance on resistance predictions using each of DRIAMS-B -D if data for the antibiotic was available at the external site. AUROC was used as the primary performance metric.

Results

Data processing summary

After all *S. epidermidis* instances were preprocessed, a summary of all antibiotics of interest was obtained (Table 1). For 19 of the 20 antibiotics, the number of instances available for training and testing exceeded 4,000 samples. In the case of teicoplanin, only 488 samples were found in DRIAMS-A. The positive class prevalence (PCP) was between 0.20 and 0.80 for 16 of the 20 antibiotics. However, the susceptibility status of most instances is heavily skewed towards one side for rifampicin and especially for ampicillin-amoxicillin, penicillin, and tigecycline. As such, it is necessary to avoid biased performance metrics such as accuracy and opt for metrics less sensitive to class distribution, such as AUROC and AUPRC. Another anomaly to consider with the summarized data is groups of antibiotics with an identical number of instances and PCP. In particular, one group of β-lactams that includes ceftriaxone all have a sample size of 4,261 and a PCP of 0.7055. In addition, both ampicillin-amoxicillin and penicillin have 4,371 instances and a PCP of 0.9938. Further analysis of DRIAMS-A indicates that for each instance that describes an antibiotic in either of the groups, the susceptibility of the instance to other antibiotics in the same group is identical. As such, only ceftriaxone and penicillin were considered to represent each of the respective groups because a model trained on any of the antibiotics within the same group would yield identical results.

After selecting one antibiotic from each identical instance group, a meta-transformer based on the Random Forest estimator was applied. The meta-transformer filtered out features with an information gain less than 1.25 times the mean. After feature selection, a summary of the antibiotic datasets (Table 2) indicates an average feature size of 1,158 across all antibiotic models after applying the meta-transformer.

Antibiotic class	Antibiotic	Number of instances	Positive class prevalence
Aminoglycosides	Gentamicin	4221	0.4414
Fusidanes	Fusidic acid	4589	0.5245
β-lactam (Penicillins)	Oxacillin	4664	0.7127
	Ampicillin-Amoxicillin*	4371	0.9938
	Penicillin*	4371	0.9938
	Amoxicillin-Clavulanic acid	4262	0.7053
	Piperacillin-Tazobactam**	4261	0.7055
β-lactam (Cephalosporins)	Ceftriaxone**	4261	0.7055
	Cefepime**	4261	0.7055
	Cefuroxime**	4261	0.7055
	Cefazolin**	4261	0.7055
β-lactam (Carbapenems)	Meropenem**	4261	0.7055
	Imipenem**	4261	0.7055
Lincosamides	Clindamycin	4192	0.5060
Tetracyclines	Tetracycline	4132	0.5620
	Tigecycline	4365	0.0032
Sulfonamides	Cotrimoxazole	4545	0.3400
Fluoroquinolones	Ciprofloxacin	4679	0.5664
Glycopeptides	Teicoplanin	488	0.2377
Rifamycins	Rifampicin	4758	0.0839

Table 1. Summary of data for a variety of antibiotics evaluated in DRIAMS-A with sufficient sample size for creating admissible models to predict resistance in *S. epidermidis*. One group of β-lactam antibiotics is highlighted with a single asterisk because these antibiotics have the same number of instances, and each instance has the same resistance status for each antibiotic in the group. Another group of β-lactam antibiotics is highlighted with two asterisks for the same reasons.

Antibiotic	Original feature size	Number of selected features
Gentamicin	6000	989
Fusidic acid	6000	1375
Oxacillin	6000	1182
Penicillin*	6000	1324
Amoxicillin-Clavulanic acid	6000	1206
Ceftriaxone**	6000	1195
Clindamycin	6000	1376
Tetracycline	6000	489
Tigecycline	6000	923
Cotrimoxazole	6000	956
Ciprofloxacin	6000	848
Teicoplanin	6000	1645
Rifampicin	6000	1545

Table 2. Summary of reduced model size after applying a Random Forest-based meta-transformer on the mass spectra. Originally a 6,000-dimensional vector, the number of features was reduced on average from 6,000 to 1,158. The rows highlighted with asterisks indicate antibiotics selected to represent their respective groups from Table 1.

Predictive modeling

Six different machine learning models were initially assessed for each antibiotic in this study, including Random Forest (RF), Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), LightGBM, and Multilayer Perceptron (MLP). Initially, the datasets for *S. epidermidis* were stratified by susceptibility and randomly partitioned into a 70/30 split for training and testing, respectively, to maintain the integrity of class proportions. Default hyperparameters were applied to train each model on training data, while testing data was used to evaluate the models' effectiveness on microbial instances the model had not encountered. For the most part, the best-performing models were RF, SVM, and LightGBM when considered across every collection of antibiotic data (Table 3). Interestingly, MLP was not able to keep up with these classifiers, despite being a neural

Classifier	Default hyperparameters	Average AUROC score
Random Forest (RF)	n_estimators = 100	0.81
	max_features="sqrt"	
Logistic Regression (LR)	C = 1.0	0.69
	penalty="L2"	
	solver="lbfgs"	
	max_iter = 100	
Support Vector Machine (SVM)	C = 1.0	0.84
	gamma="scale"	
	kernel="rbf"	
Naïve Bayes (NB)	–	0.71
LightGBM	boosting_type="gdbt"	0.82
	n_estimators = 100	
	learning_rate = 0.1	
Multilayer Perceptron (MLP)	hidden_layer_sizes=(100,)	0.73
	max_iter = 200	

Table 3. Summary of default hyperparameters that were tuned later in the study. Average AUROC scores associated with the default hyperparameters for each classifier across all evaluated antibiotic data collections are also reported.

network model. However, such a summary can only be used as a reference, as the best-performing model can vary and necessitates individual tuning for each antibiotic data collection.

After initial model evaluation, a multi-faceted model tuning process was carefully conducted, extensively searching a variety of hyperparameter combinations to improve model performance. A summary of tuned hyperparameters is available (Supplementary Table S1). All six models mentioned above were tuned for each antibiotic data collection using 5-fold cross-validation based on finding a more optimized AUROC score. Tuned hyperparameters for each classifier are listed in the previous table. Generally, LightGBM and SVM were the best-performing classifiers, but other models also performed relatively well (Fig. 2). All models performed exceptionally well, achieving mean AUROC scores ranging from 0.80 to 0.95 after ten separate, shuffled, and stratified 70/30 train-test splits were performed on each (Fig. 3). The best predictive performance was achieved by either LightGBM (7 models) or SVM (6 models) for all antibiotic data collections. Additionally, the standard deviations were generally low between each split except for penicillin and tigecycline.

Although ROC curves were similar for the best-performing classifiers, there were some noticeable differences in the PR curves for the classifiers that performed well with respect to AUROC (Fig. 4). For example, although the best-performing model for rifampicin achieved a high AUROC of 0.94, the same model attained an AUPRC of 0.80. Compared to gentamicin, whose model had a similar AUROC of 0.93, an AUPRC of 0.80 is noticeably lower than gentamicin's AUPRC of 0.91. However, because the PCP of rifampicin is low at 0.08, an AUPRC of 0.80 is commendable and indicates adequate model performance.

Among the selected antibiotics, ciprofloxacin, rifampicin, and gentamicin performed the best, with AUROC scores ranging from 0.93 to 0.96. Other antibiotic models including the models for tetracycline, amoxicillin-clavulanic acid, ceftriaxone, and oxacillin also performed well, maintaining AUROC scores of at least 0.90. On the other hand, ciprofloxacin, ceftriaxone, and oxacillin performed the best for AUPRC, with scores ranging from 0.96 to 0.97. The performance of such a multitude of antibiotics, when analyzed with machine learning, illustrates the potential for MALDI-TOF MS to emerge as a leading technique in diagnosing bacterial infections like *S. epidermidis*. A summary of all performance metrics is also available (Supplementary Table S2).

SHAP interpretation and biomarker identification

Evaluation of feature importance based on SHAP was conducted on the best-performing models of ciprofloxacin, gentamicin, and rifampicin. SHAP values help explain the relevance of a feature bin and provide a deeper understanding of the biological mechanisms behind antibiotic resistance in *S. epidermidis*. In the case of the ciprofloxacin, the feature bins of 1,916 and 1,604 appear to have contributed the most to the output of the model (Fig. 5). A pseudogel plot of a randomly selected susceptible ciprofloxacin instance and resistant ciprofloxacin instance illustrates the locations of these two feature bins, where there is a noticeable difference in the intensity of protein peaks at each feature bin can be observed (Fig. 6). As such, the approximate location of these feature bins likely contains a biomarker associated with the impartment of antibiotic resistance. Closely associated proteins to these feature bins have been identified around mass values of 7,753 Da and 6,812 Da. The same process was used to analyze the five most important feature bins for gentamicin and rifampicin, with associated protein names and UniProt IDs recorded (Table 4). Out of the fifteen feature bins, five of them did not contain a documented *S. epidermidis* protein within the range, so a closely associated protein in *S. aureus* was instead identified. In addition, six out of the fifteen feature bins corresponded to an uncharacterized protein in *S. epidermidis*, indicating the possibility for further research into these biomarkers.

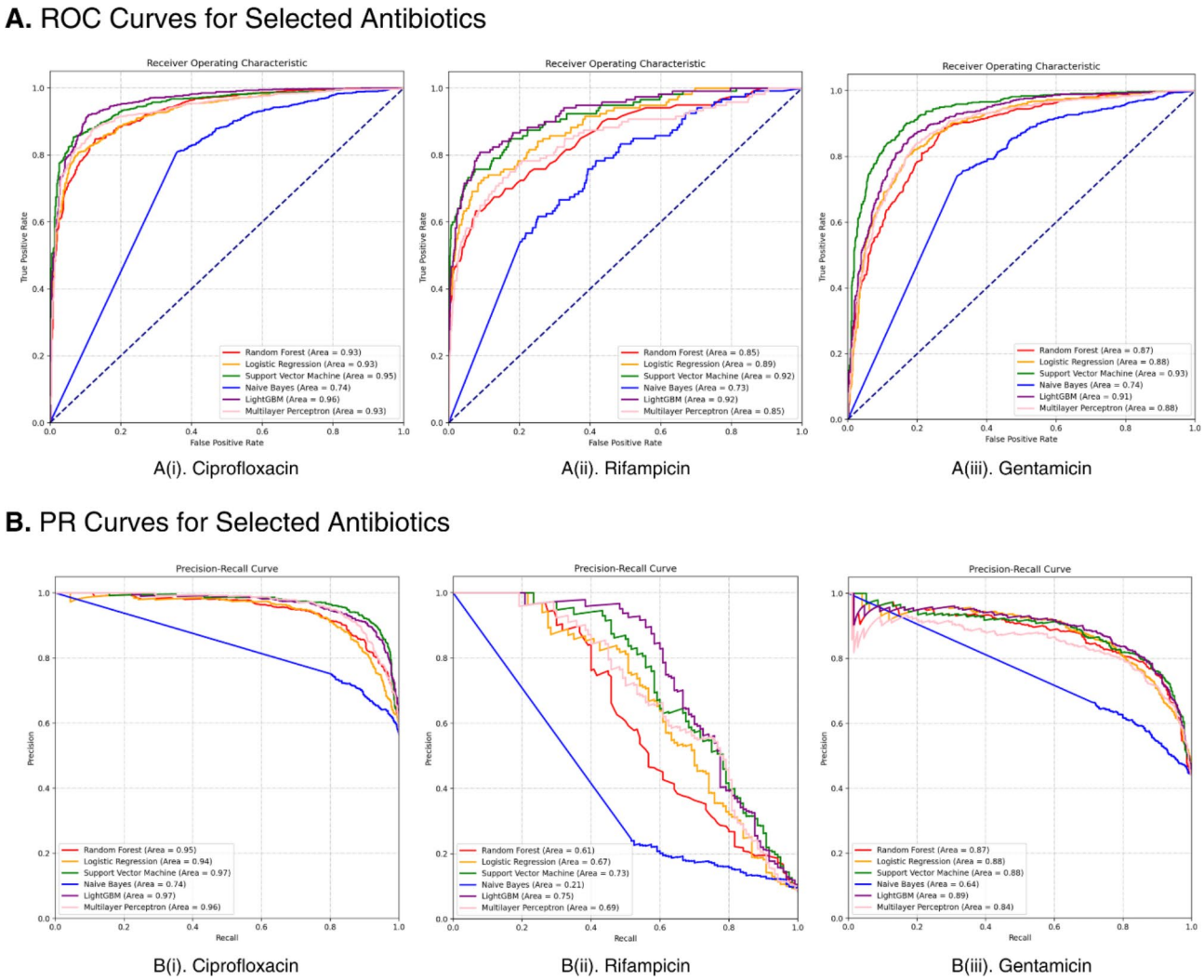

Fig. 2. Performance of the six classifiers was evaluated for three selected antibiotics—ciprofloxacin, rifampicin, and gentamicin—and yielded high performance for both AUROC and AUPRC metrics in a single 70/30 train-test split. The graphs represent the performance of tuned models in predicting S. epidermidis resistance to the drug on each respective antibiotic data collection. (**A**) Comparison of receiving operating characteristic curves for models under varying thresholds. For ciprofloxacin, LightGBM had the highest predictive performance (AUROC = 0.96). On the other hand, both LightGBM and SVM performed equally well for rifampicin (AUROC = 0.92), while SVM alone performed the best for gentamicin (AUROC = 0.93). (**B**) Comparison of precision-recall curves for models under varying thresholds. Unlike AUROC, which is evaluated using 0.5 as the baseline, AUPRC is evaluated using the positive class prevalence (PCP) as the baseline. LightGBM performed the best for all three selected antibiotics for AUPRC. Ciprofloxacin (PCP = 0.57) was able to achieve high predictive performance (AUPRC = 0.97). The overall performance for rifampicin (PCP = 0.08) was also extremely high (AUPRC = 0.75) despite being noticeably lower than ciprofloxacin because of rifampicin's low PCP baseline. Gentamicin (PCP = 0.44) also performed relatively well using AUPRC as a metric (AUPRC = 0.89).

External validation

After tuned models for each antibiotic were optimized, isolates from DRIAMS- B-D were tested on models trained on their entire corresponding DRIAMS-A dataset. Overall, performance varied greatly from site to site, with some models carrying over their performance between sites and others performing less optimally (Table 5). Other times, there was not sufficient or workable data for *S. epidermidis* in other sites to analyze the performance of different antibiotic models properly.

Discussion

Many studies have shown that MALDI-TOF MS is not only a faster approach than existing AST methods but also much more cost-effective[17]. In order to leverage the full capabilities of MALDI-TOF MS in AST, combining it with other analysis techniques, such as machine learning after dimensional reduction, is necessary to extract

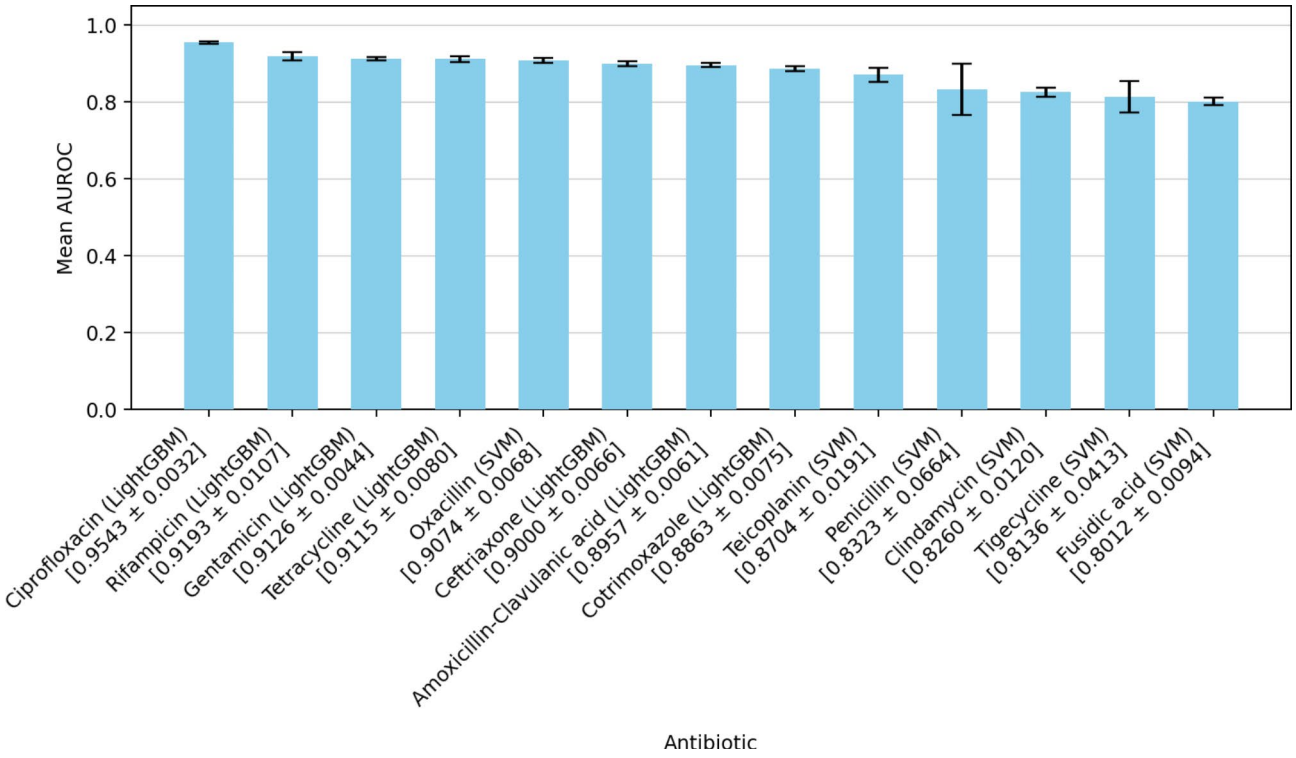

Fig. 3. Average performance based on AUROC for machine learning models across all antibiotic data collections. Bars report the mean ± standard deviation for ten separate, shuffled, and stratified 70/30 train-test splits.

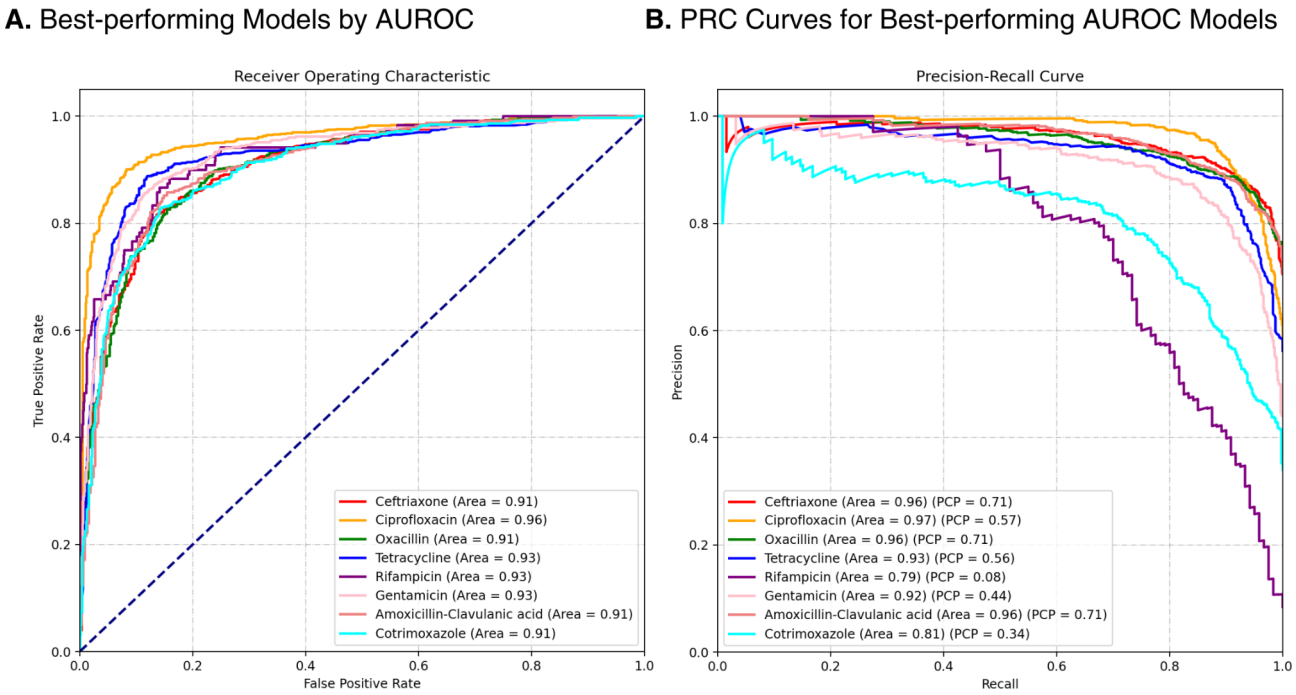

Fig. 4. Performance of eight different antibiotics plotted against one another under the ROC and PR curves. (**A**) ROC curves and corresponding areas for the best-performing model for each antibiotic data collection. The model was selected by finding the train-test split with the highest AUROC performance across selected antibiotics. (**B**) Corresponding AUPRC for the same models selected previously. Positive class prevalence (PCP) is also listed to serve as a baseline for comparing model performances based on AUPRC.

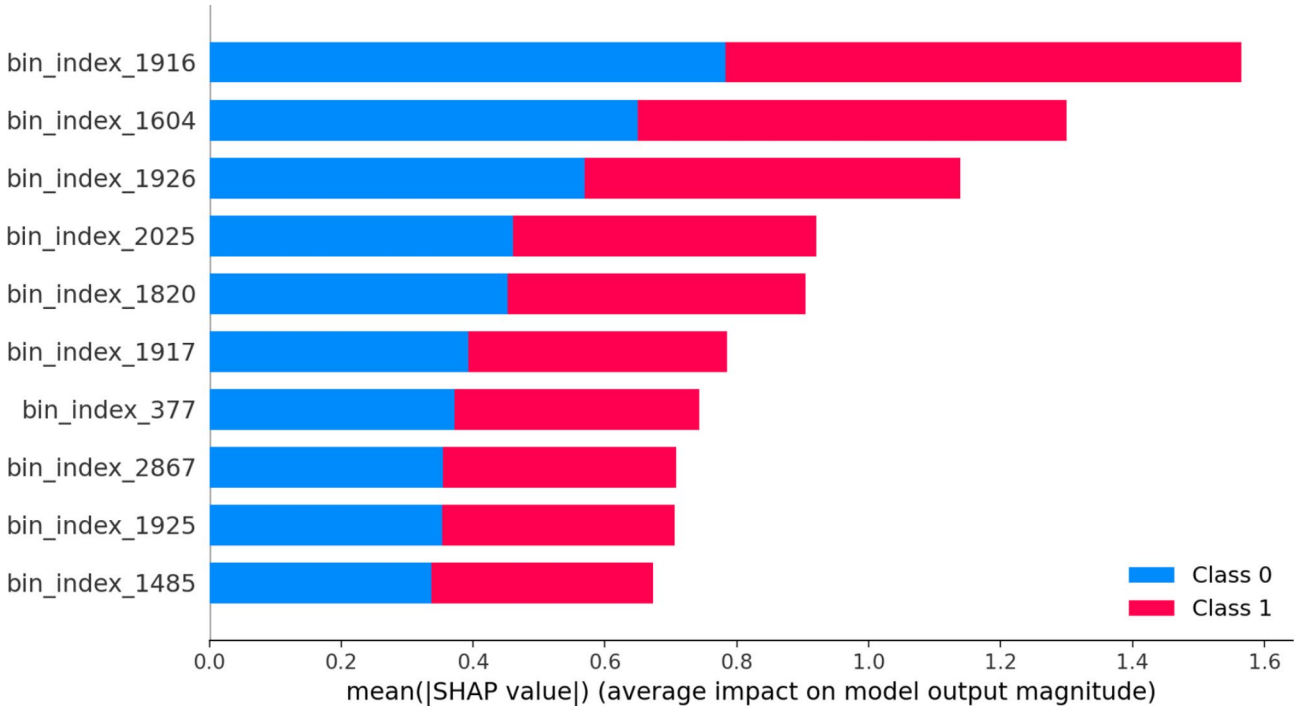

Fig. 5. List of most important features by contribution to ciprofloxacin model based on SHAP values. Longer bars generated by the program indicate that a feature has a more considerable influence on model output. Most features appear to be located around feature bin 2,000, which corresponds to a Da value of approximately 8,000.

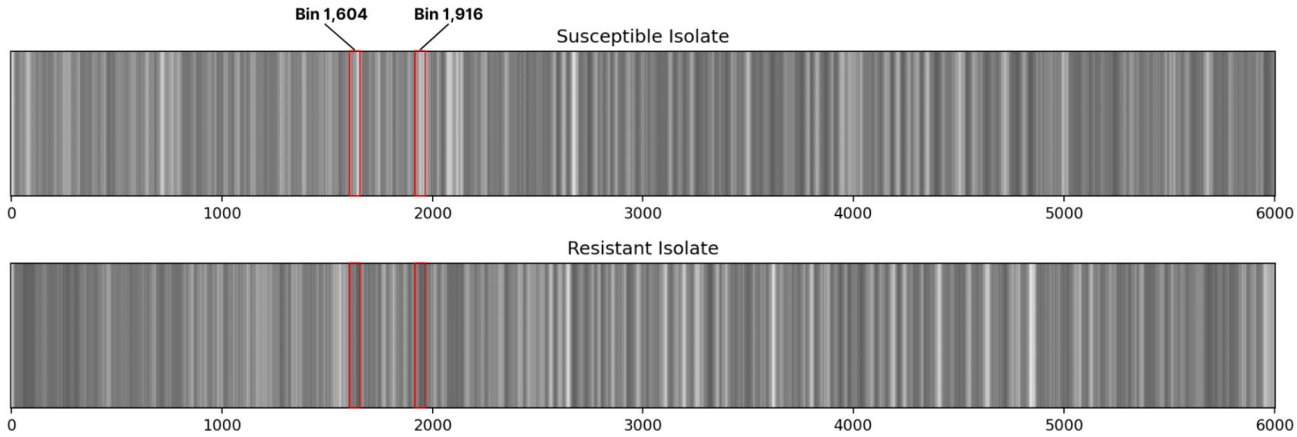

Fig. 6. Pseudogel plot of two randomly selected ciprofloxacin isolates of different classes from DRIAMS-A. The approximate locations for two bins are shown. For bins 1,916 and 1,604, there appears to be a peak (darker intensity) in the resistant isolate but not the susceptible, indicating the existence of a significant protein biomarker for antibiotic resistance.

meaningful information for resistance predictions. With high-performing models, such a workflow would be relevant and reasonably surpass traditional AST methods in practice. Although several other studies on machine learning applications in MALDI-TOF MS have already been conducted on various pathogens, there are no studies at the time of writing that specifically address nosocomial infections such as *S. epidermidis*. Furthermore, most publications focus only on one antimicrobial class at a time (e.g., glycopeptides like vancomycin) rather than a more diverse group of differing antimicrobials necessary for a holistic evaluation[21]. Frequent and indiscriminate use of strong antibiotics like vancomycin increases the prevalence of resistant bacteria, so using these drugs only when necessary helps slow the emergence of superbugs[24].

As such, this study serves as a consideration of the potential for mass spectrometry to be combined with machine learning to rapidly predict AMR for a variety of different antibiotics in nosocomial infections. The high AUROC scores up to 0.95 and AUPRC scores up to 0.97 achieved by this study can be attributed

Antibiotic	Rank	Feature bin	Mass (Da)	UniProt annotation	UniProt ID
Ciprofloxacin	1	1,916	7,748	Uncharacterized protein (M453_0201920)	A0A829M6R5
	2	1,604	6,812	UPF0337 protein SE_0604	Q8CTA8
	3	1,926	7,780	Uncharacterized protein (I3V53_12125)	A0A8I1BC98
	4	2,025	8,077	Pathogenicity island family protein	A0A0E8JV84*
	5	1,820	7,460	Uncharacterized protein	A0A125SJI8
Gentamicin	1	1,917	7,753	Transposase	A0A2G7I0T7
	2	2,411	9,233	RecG wedge domain-containing protein	A0A0E0VNH2*
	3	326	2,979	Delta-hemolysin	P0C1V1*
	4	2,857	10,572	LPXTG-motif cell wall anchor domain protein	A0A0E1VM26*
	5	625	3,877	Uncharacterized protein (SERP0759)	Q5HPZ8
Rifampicin	1	2,025	8,077	Pathogenicity island family protein	A0A0E8JV84*
	2	2,035	8,107	Pathogenicity island protein	G9D5B0
	3	1,242	5,726	Uncharacterized protein (I3V53_09740)	A0A6B9V0A1
	4	4,736	16,210	Uncharacterized protein (I3V53_10930)	A0A8I0WAC7
	5	1,922	7,766	DNA-directed RNA polymerase subunit omega	Q5HPY0

Table 4. Most important feature bins and closest significant protein biomarkers for the best-performing models. The closest significant protein was found by searching for the best-annotated *S. epidermidis* protein within the feature bin and identifying the best-annotated protein within the feature bin for *S. aureus*, a closely related bacteria if an entry could not be found for *S. epidermidis*. Listed masses and annotations were acquired directly from the UniProt entry for each protein. Any proteins identified from *S. aureus* are marked with an asterisk in the UniProt ID column.

Antibiotic	DRIAMS-B ($n=213$)	DRIAMS-C ($n=21$)	DRIAMS-D ($n=374$)
Gentamicin	0.452	0.702	0.704
Fusidic acid	0.557	0.250	-
Oxacillin	0.374	0.613	-
Penicillin	-	0.150	0.442
Amoxicillin-Clavulanic acid	0.353	0.963	-
Ceftriaxone	-	-	-
Clindamycin	0.328	1.000	0.605
Tetracycline	0.306	-	0.813
Tigecycline	-	-	0.946
Cotrimoxazole	-	0.683	0.748
Ciprofloxacin	0.598	1.000	0.879
Teicoplanin	-	-	-
Rifampicin	0.776	-	0.717

Table 5. Performance of models trained on DRIAMS-A on the external sites given by DRIAMS-B -D. The number of samples available for testing is given in the title of each column. When the prevalence of the positive class is either 0 or 1, or if there is no data available on the antibiotic in the dataset, nothing is recorded because calculating AUROC would not be meaningful.

either to the quality and scale of the *S. epidermidis* dataset or the usage of feature selection as a technique to eliminate irrelevant features and prevent the creation of an overcomplicated model. Indeed, the original study conducted by the authors of the DRIAMS-A dataset achieved maximal AUROC scores of 0.74, 0.74, and 0.80 for *Staphylococcus aureus*, *Escherichia coli*, and *Klebsiella pneumoniae*, respectively[4], which are all lower than or equal to the lowest average AUROC score of 0.80 for fusidic acid in the antibiotics modeled for *S. epidermidis*. In addition, their corresponding AUPRC scores were 0.49, 0.30, and 0.33, respectively, which are significantly overshadowed by the near-perfect AUPRC scores achieved by the best-performing *S. epidermidis* models. Although the results achieved by this study were significantly more performative than those achieved in the original study, it is essential to remember that the microbe of focus is different in this study, and other factors may have been involved in the performance of these models.

RF was specifically chosen for feature selection due to its advantages over other methods. For one, previous studies have shown that the performance of models trained with features filtered by RF has outperformed those filtered by other classifiers like K-Nearest Neighbors, Linear Discriminant Analysis, and SVM[20]. In addition, although SHAP was later used for model explanation, it was not used for feature selection because the time required for SHAP analysis increases significantly with each added column, making it unsuitable for extremely

high-dimensional datasets. RF provides sufficient performance for preliminary feature filtering and reduces the number of features to a size that can be more efficiently handled by SHAP analysis.

As shown in Fig. 3, despite high average AUROC scores, the classifiers for penicillin and tigecycline had relatively high standard deviations for AUROC scores between the ten separate train-test splits. Such performance is likely explained by the fact that these classifiers are subject to highly imbalanced class prevalences, with both approaching nearly no susceptible or resistant isolates, respectively. In many cases, imbalanced datasets can be fixed using oversampling techniques such as Synthetic Minority Oversampling TEchnique (SMOTE). However, due to the high dimensionality of each mass spectra instance, consequential data points are often sparsely distributed, making it difficult for SMOTE to generate meaningful synthetic samples. Past studies with other high dimensional datasets like genome data have demonstrated this, showing negligible improvement in model performance after applying SMOTE[25]. On the other hand, undersampling techniques can lead to significant information loss. Due to the overwhelming difference in numbers between the positive and negative classes within imbalanced datasets in DRIAMS-A, discarding majority class instances to match a small minority can result in a significant loss of diversity and meaningful patterns. As such, holistic metrics like AUROC and AUPRC are crucial for capturing and evaluating model performance based on both the majority and minority classes at the same time.

SHAP analysis provided insight into functions of proteins that have already been documented as well as evidence for uncharacterized biomarkers—if a specific mass value corresponds to a protein belonging to the same or related species that is scarcely annotated, then it could indicate a resistance protein for future investigation. For example, the feature bin of 1,916 in the ciprofloxacin model contributed significantly to model performance and was identified to be a known but minimally documented protein at 6,812 Da. Moreover, the second most relevant feature for predicting ciprofloxacin resistance was calculated to be the feature bin of 1,604, which corresponds to UPF0337 protein SE_0604 at 6,812 Da. A different MALDI-TOF MS study conducted on *S. aureus* found that another UPF0337 protein that underwent a mutation reduced antibiotic binding affinity through a molecular docking simulation, confirming that mutations may also play a major role in antibiotic resistance[26]. On the other hand, another possibility that explains the significance of particular feature bins is the horizontal transfer of resistance-encoding genes between Staphylococci or other bacterial species, such as the closely related *S. aureus*. For instance, another relevant feature for the ciprofloxacin model corresponding to the feature bin of 2,025 was not identified for *S. epidermidis* in UniProt but was identified at 8,077 Da as a pathogenicity island family protein in *S. aureus*.

In the case of gentamicin, the most relevant feature bin of 1,917 was identified as transposase at 7,753 Da. Transposase is responsible for the movement of transposons—which may encode antibiotic resistance—from one plasmid to another[27]. Another interesting feature identified by SHAP analysis was the feature bin of 326, for which, although no documented protein exists for *S. epidermidis* in UniProt, it corresponds to a delta-hemolysin biomarker at 2,979 Da in S. aureus. Delta-hemolysin is one of the primary virulence agents of *S. aureus*, is known to increase bacterial resistance to antibiotics indirectly, and has also been identified in some isolates of *S. epidermidis*[28,29]. This illustrates the significance of gene transfers between Staphylococci in AMR while also emphasizing the ability of MALDI-TOF MS machine learning models to capture relevant biomarkers rather than noise.

Finally, the same pathogenicity island family protein identified by SHAP analysis in ciprofloxacin was also the most relevant feature of the rifampicin model, demonstrating that AMR resistance is multifaceted and similar resistance mechanisms exist between different families of antibiotics. Additionally, the rifampicin model also detected the presence of DNA-directed RNA polymerase subunit omega—encoded by the *rpoZ* gene—at 7,766 Da for the feature bin of 1,922. Recently, the *rpoZ* gene and the subunit encoded by it were validated to be a significant factor in biofilm formation, motility, and antibiotic resistance, although in *E. coli* and not *S. epidermidis*[30].

However, interpretation of the roles of the features within the bacteria should be made cautiously due to limitations. For one, if a potential biomarker is identified, it may not necessarily be the protein conferring resistance, but rather a protein that is encoded on the same plasmid as the resistance protein[31]. Previous studies have demonstrated that the primary mutation responsible for ciprofloxacin resistance are mutations in the *gyrA* and *parC* genes, which have masses of 100,144 Da and 91,145 Da, respectively[32]. Such proteins are outside of the detection range and are therefore not considered when building the model, demonstrating that although model performance may not be optimized to detect the main resistance biomarker directly, highly accurate results can still be achieved by detecting other smaller but related proteins. Additionally, due to the restrictions of MALDI-TOF MS, it is more challenging to detect proteins with larger masses because the current dataset is limited to 20,000 Da.

There are several other limitations in the present study. First, dealing with imbalanced datasets is challenging due to the dimensionality of mass spectra. As such, constructing models to analyze MALDI-TOF MS may not be adequate when an antibiotic is nearly always resisted or effective against *S. epidermidis*. In addition, models trained on one site may not easily transfer to another—even though processes can be nearly identical between sites. There are often other factors, such as differing calibrations of the same instruments, that lead to compounding differences between the mass spectra generated at one site compared to the mass spectra generated at another. This can be seen in the ciprofloxacin resistance model, where the model trained on DRIAMS-A performed relatively well on DRIAMS-C and DRIAMS-D but was not as adequate in DRIAMS-B.

Conclusion

When combined with machine learning, MALDI-TOF MS has proven to be a powerful tool for rapid, economical, and efficient AMR prediction in *S. epidermidis* for a variety of different antibiotics and antimicrobial families. With the amount of predictive power that can be derived from a single MALDI-TOF MS diagnostic for so

many antibiotics, MALDI-TOF MS has the potential to serve as a holistic, go-to tool for species identification, AMR profile prediction, and diagnostic for identifying novel biomarkers for future investigation. As long as sufficient data is available with ample representation of both the susceptible and resistant isolates, MALDI-TOF MS can be combined with machine learning to benefit the patient-clinician workflow and contribute to antibiotic stewardship efforts. This study serves as a significant example for future research to follow, which should aim to investigate a broader range of species, increase the mass spectra range, integrate patient-specific data such as family history to create an all-encompassing model, and investigate potential biomarkers identified by the machine learning models.

Data availability

A processed version of DRIAMS with unnecessary files removed was analyzed for this study and is available on Kaggle (https://www.kaggle.com/datasets/drscarlat/driams). The original version of DRIAMS is available on Dryad (https://doi.org/10.5061/dryad.bzkh1899q). The code used in this study is available on GitHub (https://github.com/XFG16/StaphEpiMaldiML.git).

Received: 12 April 2024; Accepted: 1 October 2024
Published online: 15 October 2024

References

1. Candela, A. et al. Rapid and reproducible MALDI-TOF-based method for the detection of Vancomycin-resistant *Enterococcus faecium* using classifying algorithms. *Diagnostics* **12**(2), 328. https://doi.org/10.3390/diagnostics12020328 (2022).
2. Wang, H. Y. et al. Efficiently predicting Vancomycin resistance of *Enterococcus faecium* from MALDI-TOF MS spectra using a deep learning-based approach. *Front. Microbiol.* **13**, 821233. https://doi.org/10.3389/fmicb.2022.821233 (2022).
3. Wolters, M. et al. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant Staphylococcus aureus lineages. *Int. J. Med. Microbiol.* **301**(1), 64–68. https://doi.org/10.1016/j.ijmm.2010.06.002 (2011).
4. Weis, C. et al. Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. *Nat. Med.* **28**(1), 164–174. https://doi.org/10.1038/s41591-021-01619-9 (2022).
5. National Nosocomial Infections Surveillance System. National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004. *Am. J. Infect. Control* **32**(8), 470–485. https://doi.org/10.1016/S0196655304005425 (2004).
6. Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: A common cause of persistent infections. *Science* **284**(5418), 1318–1322. https://doi.org/10.1126/science.284.5418.1318 (1999).
7. O'Grady, N. P. et al. Guidelines for the prevention of intravascular catheter-related infections. Centers for Disease Control and Prevention. *MMWR Recomm. Rep.* **51**(RR-10), 1–29 (2002).
8. Chu, V. H. et al. Coagulase-negative staphylococcal prosthetic valve endocarditis—A contemporary update based on the International collaboration on endocarditis: Prospective cohort study. *Heart* **95**(7), 570–576. https://doi.org/10.1136/hrt.2008.152975 (2009).
9. Otto, M. Staphylococcus epidermidis—The 'accidental' pathogen. *Nat. Rev. Microbiol.* **7**(8), 555–567. https://doi.org/10.1038/nrmicro2182 (2009).
10. Huang, A. M. et al. Impact of rapid organism identification via matrix-assisted laser desorption/ionization time-of-flight combined with antimicrobial stewardship team intervention in adult patients with bacteremia and candidemia. *Clin. Infect. Dis.* **57**(9), 1237–1245. https://doi.org/10.1093/cid/cit498 (2013).
11. CDC. Core elements of antibiotic stewardship. *Centers for Disease Control and Prevention* (2019). http://www.cdc.gov/antibiotic-use/core-elements/index.html
12. Banerjee, R. & Humphries, R. Rapid antimicrobial susceptibility testing methods for blood cultures and their clinical impact. *Front. Med.* **8**, 635831. https://doi.org/10.3389/fmed.2021.635831 (2021).
13. McLain, J. E., Cytryn, E., Durso, L. M. & Young, S. Culture-based methods for detection of antibiotic resistance in agroecosystems: Advantages, challenges, and gaps in knowledge. *J. Environ. Qual.* **45**, 432–440. https://doi.org/10.2134/jeq2015.06.0317 (2016).
14. Banerjee, R. et al. Randomized trial of rapid multiplex polymerase chain reaction-based blood culture identification and susceptibility testing. *Clin. Infect. Dis.* **61**(7), 1071–1080. https://doi.org/10.1093/cid/civ447 (2015).
15. Kommedal, Ø., Aasen, J. L. & Lindemann, P. C. Genetic antimicrobial susceptibility testing in Gram-negative sepsis - impact on time to results in a routine laboratory. *APMIS* **124**(7), 603–610. https://doi.org/10.1111/apm.12549 (2016).
16. Carbonnelle, E. et al. MALDI-TOF mass spectrometry tools for bacterial identification in clinical microbiology laboratory. *Clin. Biochem.* **44**(1), 104–109. https://doi.org/10.1016/j.clinbiochem.2010.06.017 (2011).
17. Dhiman, N., Hall, L., Wohlfiel, S. L., Buckwalter, S. P. & Wengenack, N. L. Performance and cost analysis of matrix-assisted laser desorption ionization-time of flight mass spectrometry for routine identification of yeast. *J. Clin. Microbiol.* **49**(4), 1614–1616. https://doi.org/10.1128/JCM.02381-10 (2011).
18. Weis, C. et al. DRIAMS: Database of resistance information on antimicrobials and MALDI-TOF mass spectra [Dataset]. *Dryad*. https://doi.org/10.5061/dryad.bzkh1899q (2022).
19. Scarlat, A. DRIAMS - Resistance to antibiotics. *Kaggle* (2022). www.kaggle.com/datasets/drscarlat/driams
20. Chen, R. C., Dewi, C., Huang, S. W. & Caraka, R. E. Selecting critical features for data classification based on machine learning methods. *J. Big Data* **7**, 52. https://doi.org/10.1186/s40537-020-00327-4 (2020).
21. Feucherolles, M. et al. Combination of MALDI-TOF mass spectrometry and machine learning for rapid antimicrobial resistance screening: The case of Campylobacter spp. *Front. Microbiol.* **12**, 804484. https://doi.org/10.3389/fmicb.2021.804484 (2022).
22. Shapley, L. A value for n-person games. In *Contributions to the Theory of Games II* (eds Kuhn, H. & Tucker, A.), 307–317 (Princeton University Press, Princeton, 1953). https://doi.org/10.1515/9781400881970-018.
23. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* 30 (eds Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. & Garnett, R.) (Curran Associates, Inc., 2017). https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
24. Remschmidt, C. et al. The effect of antibiotic use on prevalence of nosocomial Vancomycin-resistant enterococci- an ecologic study. *Antimicrob. Resist. Infect. Control* **6**. https://doi.org/10.1186/s13756-017-0253-5 (2017).
25. Blagus, R. & Lusa, L. Evaluation of SMOTE for high-dimensional class-imbalanced microarray data. In *Proceedings – 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012*, Vol. 2. https://doi.org/10.1109/ICMLA.2012.183 (2012).
26. Yu, J. et al. Rapid identification of methicillin-resistant Staphylococcus aureus using MALDI-TOF MS and machine learning from over 20,000 clinical isolates. *Microbiol. Spectr.* **10**(2), e0048322 (2022). 10.1128/spectrum.00483 – 22.
27. Babakhani, S. & Oloomi, M. Transposons: The agents of antibiotic resistance in bacteria. *J. Basic. Microbiol.* **58**(11), 905–917. https://doi.org/10.1002/jobm.201800204 (2018).

28. Motamedi, H., Asghari, B., Tahmasebi, H. & Arabestani, M. R. Identification of hemolysine genes and their association with antimicrobial resistance pattern among clinical isolates of Staphylococcus aureus in west of Iran. *Adv. Biomed. Res.* **7**, 153. https://doi.org/10.4103/abr.abr_143_18 (2018).
29. Pinheiro, L. et al. M.de L. Staphylococcus epidermidis and *Staphylococcus haemolyticus*: Molecular detection of cytotoxin and enterotoxin genes. *Toxins* **7**(9), 3688–3699. https://doi.org/10.3390/toxins7093688 (2015).
30. Patel, U. R., Gautam, S. & Chatterji, D. Validation of omega subunit of RNA polymerase as a functional entity. *Biomolecules* **10**(11), 1588. https://doi.org/10.3390/biom10111588 (2020).
31. Lau, A. F. et al. A rapid matrix-assisted laser desorption ionization-time of flight mass spectrometry-based method for single-plasmid tracking in an outbreak of carbapenem-resistant Enterobacteriaceae. *J. Clin. Microbiol.* **52**(8), 2804–2812. https://doi.org/10.1128/JCM.00694-14 (2014).
32. Kang, J. Y. et al. Fluoroquinolone resistance of Staphylococcus epidermidis isolated from healthy conjunctiva and analysis of their mutations in quinolone-resistance determining region. *Antimicrob. Resist. Infect. Control* **9**(1), 177. https://doi.org/10.1186/s13756-020-00841-3 (2020).

Author contributions

M.R. conceived the project, processed the MALDI-TOF MS data, undertook machine learning modeling, completed experimental analysis, and wrote the manuscript. Q.C. and J.Z. supervised the project and edited the manuscript. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-75044-6.

Correspondence and requests for materials should be addressed to M.R.

Reprints and permissions information is available at www.nature.com/reprints.

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