MACHINE LEARNING–BASED PREDICTION OF CIPROFLOXACIN, CEFOTAXIME, CEFTA. ZIDIME, AND GENTAMICIN RESISTANCE USING BACTERIAL GENOMIC FEATURES

Dr. Meenu Mangal

Professor, Poddar International College, Email ID : drmeenumangal@gmail.com

Abstract

The rise in antimicrobial resistance has made the need for quick and accurate methods that can predict bacterial resistance to antibiotics from their genome data a pressing necessity. In this study, machine-learning models for the prediction of resistance to ciprofloxacin, cefotaxime, ceftazidime and gentamicin were developed from bacterial genomic characteristics. The phenotypic resistance data was matched to the corresponding matrix of genomic features, and individual binary classification models were created for each antibiotic. Stratified training and testing subsets were used to evaluate logistic regression, random forest, support vector machine and gradient boosting. The accuracy, balanced accuracy, precision, sensitivity, specificity, F1-score, ROC-AUC, and confusion matrices were used to measure the model's performance. Of the 668 matched isolates, prevalence of resistance was highest for ciprofloxacin (44.2%), followed by cefotaxime (40.7%), ceftazidime (32.8%) and gentamicin (20.8%). Over 50% of the isolates were resistant to at least one antibiotic and 12.7% were resistant to all four antibiotics. Random forest and support vector machine (SVM) had the highest ROC-AUC (0.948 and 90.3% respectively) for prediction of Ciprofloxacin resistance. The results indicated that the predictive performance of ceftazidime and cefotaxime was moderate while the resistance to gentamicin was harder to classify due to lower prevalence and class imbalance. The results show that the machine learning, based on genomic features, is effective in supporting antibiotic resistance prediction, but the predictive quality differs depending on the antibiotics and should be tested in external data before be used in clinical practice.

Keywords: *antimicrobial resistance; bacterial genomics; machine learning; whole-genome sequencing; antibiotic susceptibility*

1. Introduction

The spread of antimicrobial resistance (AMR) is a significant problem for treating infectious diseases as resistant bacterial pathogens can reduce the effectiveness of treatment, extend hospital stays and cause treatment failures. Early detection of resistance is thus vital to ensure the choice of suitable antibiotics and to prevent unnecessary empirical treatment. Recently, AI has shown potential to speed up AMR identification, facilitate antimicrobial drug discovery, and enhance clinical antimicrobial decision-making (Lau et al., 2021). The traditional method of antimicrobial susceptibility testing relies on bacterial culture and laboratory measurement of growth inhibition. Phenotypic testing, although still clinically relevant, could take a considerable amount of time before definitive susceptibility profiles are available. Another source of information is provided by whole genome sequencing, which can be used to identify resistance genes, point mutations, mobile genetic elements, etc. that are linked to resistant phenotypes (Waddington et al., 2022). When analysed using the appropriate computational approach, genomic data can therefore help to predict antimicrobial susceptibility earlier in diagnosis. Developed bioinformatics tools have helped to convert bacterial sequence data to resistance-associated variables.

The Comprehensive Antibiotic Resistance Database brings together genes, mutations, mechanisms, drug classes and computational detection models, appropriate for resistome analysis and machine learning (Alcock et al., 2023). ResFinder 4.0 is also able to predict antimicrobial phenotypes from acquired resistance genes and chromosomal mutations and has been shown to have a high genotype–phenotype concordance in several bacterial species (Bortolaia et al., 2020). However, there may be discrepancies between laboratories with respect to phenotypic definitions, analytical pipelines, reference databases and sequencing quality (Doyle et al., 2020). Machine learning is useful especially when the resistance is not determined by any one known factor. An algorithm can scan thousands of positions at once and find sets of variants that are linked to susceptibility or resistance. Modeling complex and nonlinear genomic relationships (Kim et al., 2022) can be achieved by random forest, support vector machine, gradient boosting, neural networks, and graph-based models, which are not easily interpretable, while logistic regression offers a baseline model.

Whole-genome sequencing of bacteria has been used to successfully predict resistance directly, in the case of Gram-negative bacteria, using machine-learning models (Van Camp et al., 2020). Similar strategies have been tested for Actinobacillus pleuropneumoniae and have been found to be effective in a non-human pathogen (Liu et al. 2020). Predictions of genomic resistance may vary in their reliability depending on the antibiotic and resistance mechanism. Whole genome analysis using rules has been useful to make predictions for resistance in Enterococcus faecium clinical isolates (Anahtar et al., 2022). Online genomic analysis has also been helpful in predicting resistance in enterococci, with the accuracy of predictions varying according to the completeness of known resistance determinants (Penven et al., 2023).

Cephalosporins have the potential to be complicated by various β-lactamases, regulatory changes, permeability changes, and genotype–phenotype relationships (Heydecke et al., 2023). Machine-learning models have, however, been shown to be able to predict cefepime susceptibility in Escherichia coli (Humphries et al., 2023). As a result, the combination of whole genome sequencing and machine learning has become a significant tool for AMR surveillance and susceptibility prediction to speed up the process (Ardila et al., 2024). For multidrug-resistant Acinetobacter baumannii, deep-learning models could also be used to further enhance the recognition of complex genomic signatures (Jia et al., 2024). But high-dimensional genomic data, class imbalance, small sample sizes, overfitting, population structure, and lack of external validation persist as clinical implementation barriers (Sakagianni et al., 2023).

This study aimed to create and evaluate machine-learning models for predicting resistance to ciprofloxacin, cefotaxime, ceftazidime and gentamicin based on genomic features of the bacteria. The scope included genomic-feature preprocessing, phenotype matching, feature selection, stratified model development and performance evaluation. The goals were to measure the prevalence of resistance, to describe the pattern of multidrug resistance, to compare logistic regression, random forest, support vector machine and gradient boosting models, and to see which of these models was best for each antibiotic.

2. Methodology

2.1 Research Design

In this study, the design was the quantitative machine-learning method that predicted the bacterial resistance to ciprofloxacin, cefotaxime, ceftazidime and gentamicin based on the genomic features. A series of binary classification models were built for each antibiotic resistance outcome.

2.2 Dataset Description

809 bacterial isolates were identified in the data set using the variable prename (Salman Ibne Eunus, 2025). There were four binary outcome variables that represented antimicrobial susceptibility: ciprofloxacin (CIP), cefotaxime (CTX), ceftazidime (CTZ), and gentamicin (GEN) (Kulkarni, 2025). Resistance was scored as 1 and susceptibility, 0. The phenotype records were then aligned with the genomic-feature matrix, based on the common isolate identifier.

2.3 Data Preprocessing

Data was checked for any missing data, duplicate isolates, invalid labels, and inconsistent identifiers. All resistance variables were coded as binary numbers. Genomic variables were filtered out where they had a variance of zero or almost zero as they gave little predictive information. Highly incomplete and non-informative features were also eliminated.

2.4 Feature Selection

As there were far more genomic predictors than isolates, feature selection was conducted prior to model training. The top most informative genomic variables selected for each antibiotic were kept using variance-based filtering and univariate ranking of the features. This made the calculation simpler and helped to prevent overfitting.

2.5 Data Partitioning

Data for each antibiotic were split into training dataset (80%) and testing dataset (20%). Stratified sampling was used to ensure that the proportions of resistant and susceptible isolates remained the same in each subset. The training data was used exclusively in all preprocessing and feature-selection process to avoid information leakage.

2.6 Machine-Learning Models

The classifiers tested were logistic regression, random forest, support vector machine and gradient-boosting classifiers. Separate models were developed to predict the resistance to ciprofloxacin, cefotaxime, ceftazidime and gentamicin. Where necessary, class weights were applied, as there was a more severe class imbalance for the gentamicin outcome.

2.7 Model Validation and Performance Evaluation

Five-Fold Stratified Cross Validation was used in the training data to select the model and tune the parameters. The independent test set was used to evaluate the final model with accuracy, balanced accuracy, precision, sensitivity, specificity, F1 score, and area under the receiver operating characteristic curve. For each antibiotic, a confusion matrix as well as a ROC curve were also created.

2.8 Ethical Considerations

The authors did not interact with patients and did not collect personal information in the study, which consisted of a de-identified, secondary dataset. Only data obtained for research purposes were analyzed and results were interpreted at the bacterial-isolate level.

3. Results

3.1 Distribution of Antibiotic Resistance

In total, 668 complete bacterial isolates (both phenotypic and genotypic) were taken into account. Overall, 295 isolates were resistant to ciprofloxacin (44.2%) and 272 isolates were resistant to cefotaxime (40.7%). Twenty-one nine (32.8%) isolates were found to be resistant to ceftazidime, while 139 (20.8%) were found to be resistant to gentamicin. So, ciprofloxacin had the greatest resistance prevalence while gentamicin had the least. The resistance prevalence to Ciprofloxacin, Cefotaxime, Ceftazidime and Gentamicin are shown in figure 1 where the lowest is for Gentamicin and the highest is for Ciprofloxacin. The prevalence of resistant isolates for each antibiotic is shown in Table 1, ranging from 20.8% for gentamicin to 44.2% for ciprofloxacin.

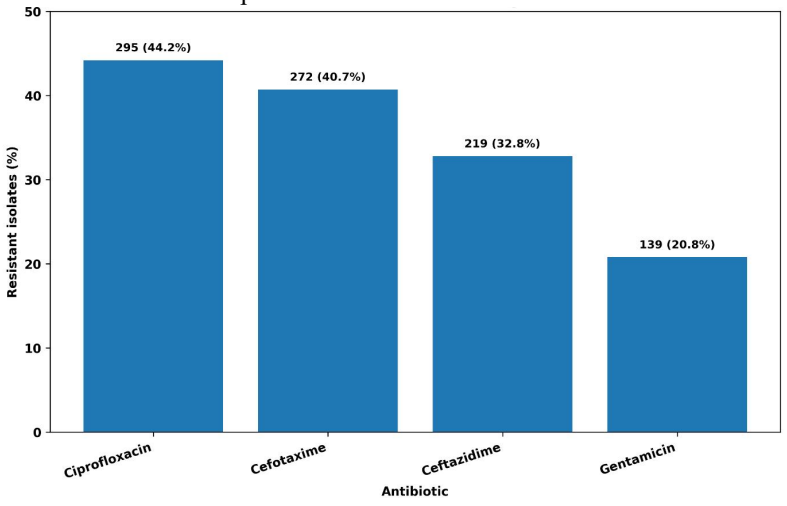

Figure 1. Prevalence of antibiotic resistance among bacterial isolates

Table 1. Distribution of antibiotic resistance among the bacterial isolate

Antibiotic	Susceptible, n	Susceptible (%)	Resistant, n	Resistant (%)
Ciprofloxacin	373	55.8	295	44.2
Cefotaxime	396	59.3	272	40.7
Ceftazidime	449	67.2	219	32.8
Gentamicin	529	79.2	139	20.8
Total isolates	668	—	668	—

3.2 Multidrug and Co-resistance Patterns

The total of 668 isolates, 297 (44.5%) of which were susceptible to all four antibiotics and 371 (55.5%) resistant to at least one antibiotic. 85 (12.7%) of the isolates were resistant to all 4 antibiotics. Furthermore, there was a large proportion (17.1%) of isolates resistant to three antibiotics, suggesting a significant level of multidrug resistance. Cefotaxime–ceftazidime resistance was the most common paired resistance pattern (219 isolates, 32.8%). This was

followed by ciprofloxacin–cefotaxime co-resistance in 202 isolates (30.2%). Co-resistance with gentamicin was not as common. The multidrug-resistance pattern distribution of all isolates is shown in Figure 2, which reveals that 55.5% of isolates were resistant to one or more antibiotics, and 12.7% were resistant to all four. The multidrug and paired co-resistance patterns are presented in Table 2 and cefotaxime–ceftazidime is the most common co-resistance combination.3.2 Multidrug and Co-resistance Patterns

The total of 668 isolates, 297 (44.5%) of which were susceptible to all four antibiotics and 371 (55.5%) resistant to at least one antibiotic. 85 (12.7%) of the isolates were resistant to all 4 antibiotics. Furthermore, there was a large proportion (17.1%) of isolates resistant to three antibiotics, suggesting a significant level of multidrug resistance. Cefotaxime–ceftazidime resistance was the most common paired resistance pattern (219 isolates, 32.8%). This was followed by ciprofloxacin–cefotaxime co-resistance in 202 isolates (30.2%). Co-resistance with gentamicin was not as common. The multidrug-resistance pattern distribution of all isolates is shown in Figure 2, which reveals that 55.5% of isolates were resistant to one or more antibiotics, and 12.7% were resistant to all four. The multidrug and paired co-resistance patterns are presented in Table 2 and cefotaxime–ceftazidime is the most common co-resistance combination.

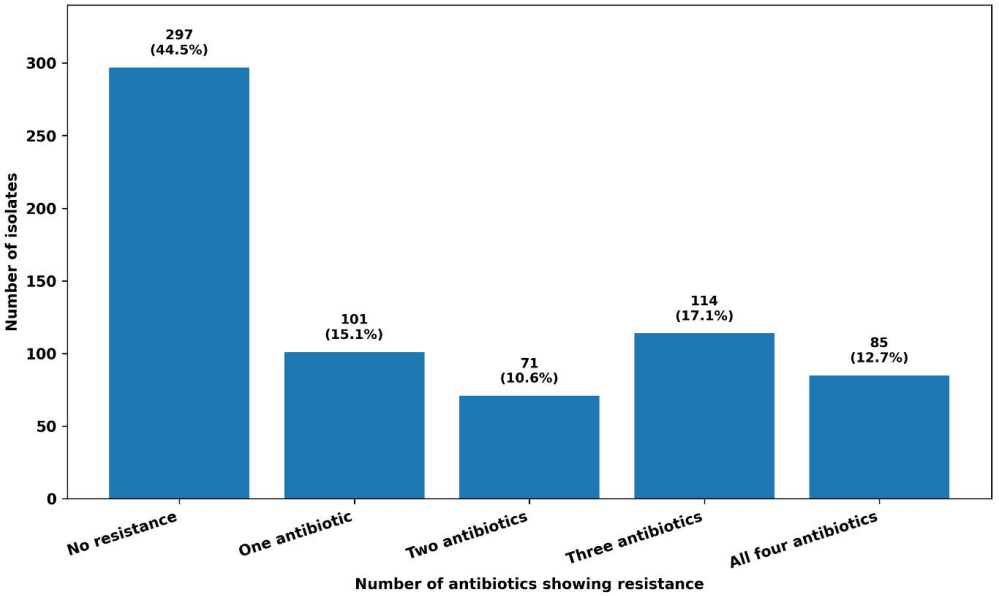

Figure 2. Distribution of multidrug-resistance patterns

Table 2. Multidrug and paired co-resistance patterns

Resistance pattern	Number of isolates	Percentage (%)
Susceptible to all four antibiotics	297	44.5
Resistant to one antibiotic	101	15.1
Resistant to two antibiotics	71	10.6
Resistant to three antibiotics	114	17.1
Resistant to all four antibiotics	85	12.7
Ciprofloxacin + Cefotaxime	202	30.2
Ciprofloxacin + Ceftazidime	172	25.7
Ciprofloxacin + Gentamicin	116	17.4
Cefotaxime + Ceftazidime	219	32.8
Cefotaxime + Gentamicin	113	16.9
Ceftazidime + Gentamicin	101	15.1

The most informative genomic features were identified after variance filtering for each antibiotic individually (100 features per antibiotic). The 80:20 split of train-test was performed on a stratified data set and the logistic regression, random forest, support vector machine and gradient boosting models were evaluated. All models performed well in predicting ciprofloxacin resistance. Random forest had the highest ROC-AUC of 0.948 and the highest accuracy of 90.3% was obtained by the support vector machine. Random forest had the highest ROC-AUC of 0.802 for resistance to cefotaxime, and the highest accuracy was 75.4% for gradient boosting (GB). Gradient boosting had the highest accuracy (79.1%) and ROC-AUC (0.781) for ceftazidime resistance. The gentamicin resistance was more challenging. Random forest performed best with the highest value of ROC-AUC of 0.681, while gradient boosting gave high values of overall accuracy but very low value of sensitivity, which means that it was not sensitive in detecting resistant isolates. The comparative performance of the four antibiotic-resistance outcomes on logistic regression, random forest, support vector machine and gradient boosting are shown in Table 3. The ROC curves for the machine learning models selected are presented in Figure 3, with the highest discriminatory model being ciprofloxacin random forest and the lowest being gentamicin.

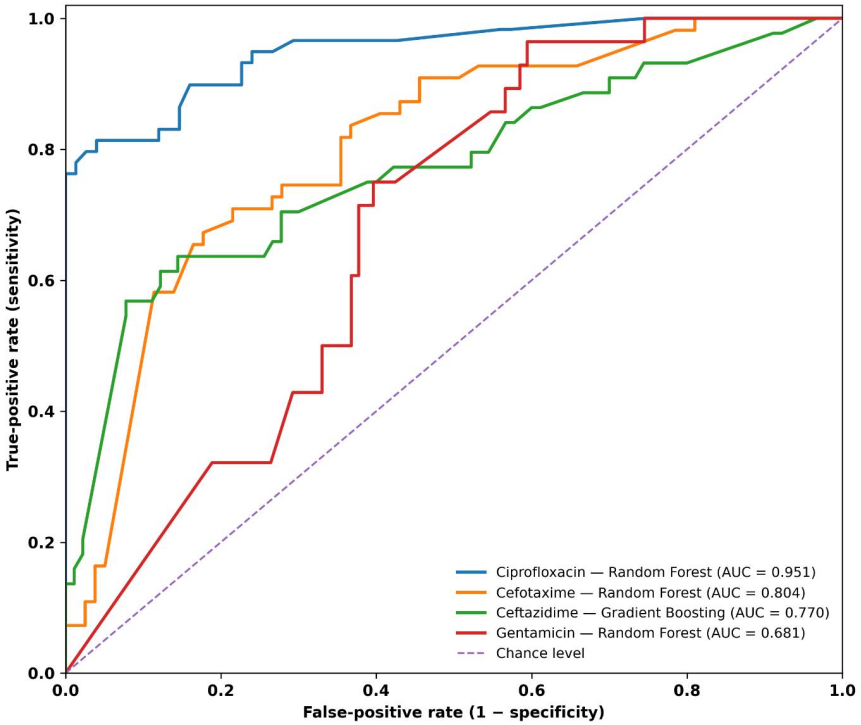

Figure 3. ROC curves of the selected antibiotic-resistance models

Table 3. Comparative performance of machine-learning models

Antibiotic	Model	Accuracy (%)	Balanced Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	ROC-AUC
Ciprofloxacin	Logistic Regression	88.1	87.3	90.6	81.4	93.3	85.7	0.933
Ciprofloxacin	Random Forest	89.6	88.3	97.9	78.0	98.7	86.8	0.948
Ciprofloxacin	Support Vector Machine	90.3	89.5	94.2	83.1	96.0	88.3	0.910
Ciprofloxacin	Gradient Boosting	88.1	86.8	95.7	76.3	97.3	84.9	0.940
Cefotaxime	Logistic Regression	73.1	73.1	65.6	72.7	73.4	69.0	0.762
Cefotaxime	Random Forest	73.1	72.8	66.1	70.9	74.7	68.4	0.802
Cefotaxime	Support Vector Machine	73.1	72.8	66.1	70.9	74.7	68.4	0.746
Cefotaxime	Gradient Boosting	75.4	74.4	70.4	69.1	79.7	69.7	0.777
Ceftazidime	Logistic Regression	69.4	67.9	52.8	63.6	72.2	57.7	0.735
Ceftazidime	Random Forest	74.6	71.2	61.4	61.4	81.1	61.4	0.763
Ceftazidime	Support Vector Machine	71.6	68.4	56.5	59.1	77.8	57.8	0.734
Ceftazidime	Gradient Boosting	79.1	74.0	72.2	59.1	88.9	65.0	0.781
Gentamicin	Logistic Regression	59.7	60.1	28.3	60.7	59.4	38.6	0.640
Gentamicin	Random	61.9	60.2	29.1	57.1	63.2	38.6	0.68

	Forest							1
Gentamicin	Support Vector Machine	58.2	61.8	28.8	67.9	55.7	40.4	0.612
Gentamicin	Gradient Boosting	78.4	50.8	33.3	3.6	98.1	6.5	0.653

3.4 Performance of the Best Predictive Models

Based on ROC-AUC, random forest model matched the best model between the three models (random forest, neural network and linear regression) for ciprofloxacin, cefotaxime and gentamicin. For the ceftazidime resistance, gradient boosting was chosen. The random forest model developed for ciprofloxacin correctly classified 120 of the 134 test isolates with just one false positive prediction. The model was correct with 39 resistant and 59 susceptible isolates. The ceftazidime gradient-boosting model could accurately classify 106 isolates, including 26 resistant ones. Gentamicin prediction was relatively poor, with 39 isolates classified as susceptible when they were actually resistant. The confusion matrices of the selected models are presented in figure 4 indicating a good classification for ciprofloxacin and more prediction errors for gentamicin resistance. The values of the confusion matrix and the predictive performance for ciprofloxacin, cefotaxime, gentamicin and ceftazidime of the selected models are given in Table 4 and it is seen that random forest is the best model for predicting ciprofloxacin, cefotaxime and gentamicin while gradient boosting is the best model for predicting ceftazidime.

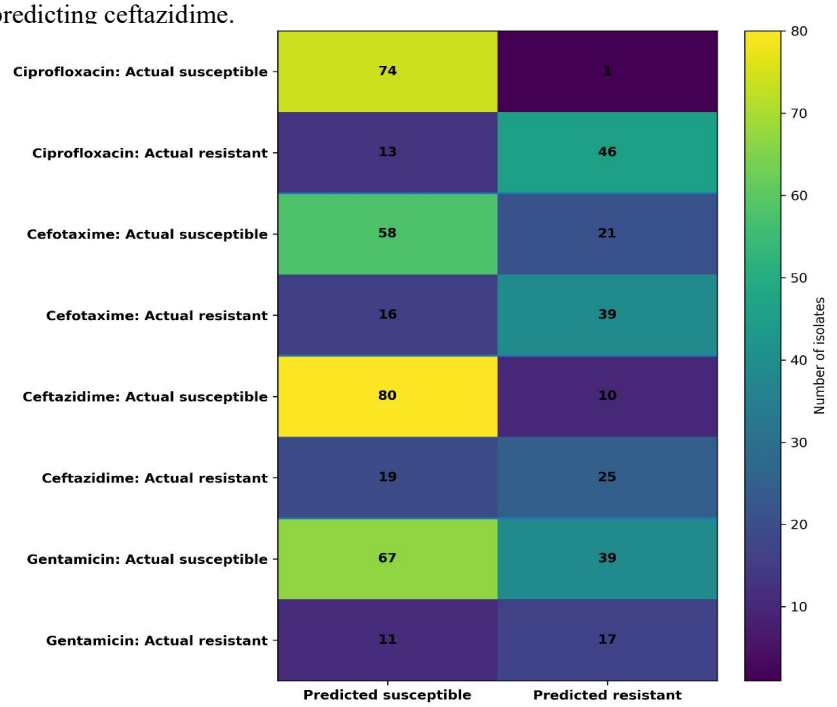

Figure 4. Confusion matrices of the selected resistance-prediction models

Table 4. Test-set confusion matrices and performance of the selected models

Antibiotic	Selected model	True negative	False positive	False negative	True positive	Accuracy (%)	ROC-AUC
Ciprofloxacin	Random Forest	74	1	13	46	89.6	0.948
Cefotaxime	Random Forest	59	20	16	39	73.1	0.802
Ceftazidime	Gradient Boosting	80	10	18	26	79.1	0.781
Gentamicin	Random Forest	67	39	12	16	61.9	0.681

4. Discussion

The present study has shown the possibility of predicting resistance to ciprofloxacin, cefotaxime, ceftazidime and gentamicin at antibiotic-specific levels of performance by analyzing the genome of the bacteria. Ciprofloxacin resistance had the highest degree of predictability with the random forest yielding a ROC-AUC of 0.948 and the support vector machine (SVM) achieving an accuracy of 90.3%. Our results corroborate the general trend that machine-learning systems identify resistance-associated genomic signatures which can be spread throughout mutations, genes and accessory genome components (Ren et al., 2022). 44.2% of isolates were ciprofloxacin resistant and this resistance was predicted with high sensitivity and specificity. It is often linked to easily identifiable mutations in quinolone resistance determinants and to other efflux or plasmid mediated determinants. This relatively well organized genomic signal could account for the good results seen for ciprofloxacin. Additionally, sequence-level patterns were found to be effective resistance classification methods in E. coli through deep-learning analysis of discriminative genomic positions (Jin et al.,

2024). Random forest was especially good for ciprofloxacin with a specificity of 98.7% and a single false-positive prediction (Nguyen et al., 2018).

However, ensemble tree methods may be able to detect nonlinear interactions without requiring the simple assumption of a relationship between individual genomic variables and resistance. This benefit applies to bacterial datasets where there are many features present in the genome, but not in every individual genome in the dataset. Similarly, machine-learning methods have been employed to find genomic markers and resistance-associated traits in E. coli populations (Moat et al., 2025). The prediction of resistance to cefotaxime and ceftazidime was medium, not great. Random forest gave ROC-AUC of 0.802 in case of cefotaxime, and gradient boosting in case of ceftazidime gave ROC-AUC of 0.781. The decrease in performance could be due to the heterogeneous mechanisms responsible for third generation cephalosporin resistance (ESBLs, AmpC enzymes, decreased expression, porin loss and multiple mechanism combinations). However, whole genome studies have emphasized that decreased cephalosporin susceptibility does not necessarily follow from the presence of known resistance genes (Heydecke et al., 2023). Gradient boosted the highest accuracy and specificity with the ceftazidime, but it only identified 59.1% of the resistant isolates (Vanalli et al., 2020). It reflects the fact that, in the case of fewer resistant cases, overall accuracy can overestimate the usefulness of a model, as it ignores an imbalance between specificity and sensitivity (Mohammad-Taheri et al., 2024).

Balanced accuracy, sensitivity, F1-score and ROC-AUC are therefore important measures to assess genomic AMR classifiers (Kim et al., 2022). The results also help to demonstrate the importance of analysing a confusion matrix instead of just choosing a model based on its accuracy. Prediction of gentamicin resistance proved to be the most challenging. The resistance rate of isolates was 20.8%, with the model (random forest) used for the selected random samples having a ROC-AUC of 0.681. Class imbalance and the limited amount of resistant examples to learn from will contribute to the decreased performance. Other factors, such as the variety of AMEs, target modification, target permeability changes, and gene-expression influences, may also complicate genomic prediction. Genomic surveys have revealed that bacteria isolated in a single location carry multiple AMR genes and integron-associated determinants, which can have differential phenotypic expression (Joddha et al., 2023). A high level of MDDR burden has been observed, with resistance to at least one antibiotic found in 55.5% of isolates and MDR (resistance to all four antibiotics) in 12.7% of isolates. The most significant paired pattern was between cefotaxime and ceftazidime, a pattern that seemed to be biologically plausible as both were third generation cephalosporins and might be susceptible to common mechanisms of β-lactam resistance. Genomic methods can be useful for the study of these overlapping resistance profiles as they allow both known and hitherto unknown marker combinations to be captured.

These present results should be taken with a grain of salt, however. With the aim of reducing overfitting, feature selection and internal stratified validation were performed, but the models were not tested with an external bacterial cohort. Generalizability may be influenced by population structure, species composition, sequencing platform, geographic origin, and/or laboratory breakpoints. Differences in bioinformatic pipelines have been demonstrated to produce varying bioinformatic predictions of AMR using the same sequence data between different laboratories (Doyle et al., 2020).

Furthermore, the quality of the prediction relies on the accuracy of the phenotype labelling and representation of the training isolates. Overall, the findings suggest that the use of machine learning in combination with the genomic features is most promising for ciprofloxacin resistance and moderately effective for cefotaxime and ceftazidime resistance. Gentamicin prediction needs more balanced and larger datasets, better feature engineering, and external validation. Before genomic models can be used routinely for antimicrobial susceptibility, future studies should include resistance-gene annotations, point mutations, population structure, explainable feature importance, and independent clinical isolates.

5. Conclusion

The machine-learning models developed in this study were shown to be useful for the prediction of resistance to the drugs: ciprofloxacin, cefotaxime, ceftazidime, gentamicin based on the bacterial genomic features. The four antibiotics, the resistance prediction was best performed by ciprofloxacin, and random forest and support vector machine exhibited good classification performance. The resistance against cefotaxime and ceftazidime was moderately well predicted, suggesting that genomic features reflected important, but more complex, resistance mechanisms. Gentamicin resistance was the most challenging resistance to be classified due to low resistance prevalence, high class imbalance and potentially different genetic determinants. The results also showed a significant level of multidrug resistance, with over half of the isolates resistant to at least one antibiotic and a substantial percentage multi-drug resistant to all four antibiotics. Co-resistance (cefotaxime and ceftazidime) was particularly common, suggesting that the resistance mechanisms to these two drugs may be shared. In general, the findings show that machine learning based on genomic features can help with the rapid prediction of antimicrobial resistance especially if the resistance is linked to stable, recognisable, and known genetic signatures. But the performance of the models was dependent on the type of antibiotic, and the accuracy of the models should not be used to judge model performance. Consider all of the following: sensitivity, specificity, balanced accuracy, F1 score, ROC-AUC, and confusion matrices. The models should be validated in the future with more species-specific and geographically expanded data sets, and with species collected from the outside. The integration of these resistance-gene annotations, point mutations, population structure, explainable feature importance and enhanced class imbalance treatment strategies could further enhance predictive reliability and clinical relevance.

Reference

1. Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., ... & McArthur, A. G. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the

Comprehensive Antibiotic Resistance Database. Nucleic acids research, 51(D1), D690-D699.
2. Anahtar, M. N., Bramante, J. T., Xu, J., Desrosiers, L. A., Paer, J. M., Rosenberg, E. S., ... & Kwon, D. S. (2022). Prediction of antimicrobial resistance in clinical Enterococcus faecium isolates using a rules-based analysis of whole-genome sequences. Antimicrobial Agents and Chemotherapy, 66(1), e01196-21.
3. Ardila, C. M., Yadalam, P. K., & González-Arroyave, D. (2024). Integrating whole genome sequencing and machine learning for predicting antimicrobial resistance in critical pathogens: a systematic review of antimicrobial susceptibility tests. PeerJ, 12, e18213.
4. Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., ... & Aarestrup, F. M. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. Journal of Antimicrobial Chemotherapy, 75(12), 3491-3500.
5. Doyle, R. M., O'sullivan, D. M., Aller, S. D., Bruchmann, S., Clark, T., Coello Pelegrin, A., ... & Harris, K. A. (2020). Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. Microbial genomics, 6(2), e000335.
6. Heydecke, A., Yin, H., Tano, E., & Sütterlin, S. (2023). Limitations in predicting reduced susceptibility to third generation cephalosporins in Escherichia coli based on whole genome sequence data. Plos one, 18(11), e0295233.
7. Humphries, R. M., Bragin, E., Parkhill, J., Morales, G., Schmitz, J. E., & Rhodes, P. A. (2023). Machine-learning model for prediction of cefepime susceptibility in Escherichia coli from whole-genome sequencing data. Journal of Clinical Microbiology, 61(3), e01431-22.
8. Jia, H., Li, X., Zhuang, Y., Wu, Y., Shi, S., Sun, Q., ... & Ruan, Z. (2024). Neural network-based predictions of antimicrobial resistance phenotypes in multidrug-resistant Acinetobacter baumannii from whole genome sequencing and gene expression. Antimicrobial Agents and Chemotherapy, 68(12), e01446-24.
9. Jin, C., Jia, C., Hu, W., Xu, H., Shen, Y., & Yue, M. (2024). Predicting antimicrobial resistance in E. coli with discriminative position fused deep learning classifier. Computational and Structural Biotechnology Journal, 23, 559-565.
10. Joddha, H. B., Mathakiya, R. A., Joshi, K. V., Khant, R. B., Golaviya, A. V., Hinsu, A. T., ... & Koringa, P. G. (2023). Profiling of antimicrobial resistance genes and integron from Escherichia coli isolates using whole genome sequencing. Genes, 14(6), 1212.
11. Kim, J. I., Maguire, F., Tsang, K. K., Gouliouris, T., Peacock, S. J., McAllister, T. A., ... & Beiko, R. G. (2022). Machine learning for antimicrobial resistance prediction: current practice, limitations, and clinical perspective. Clinical microbiology reviews, 35(3), e00179-21.
12. Kulkarni, V. (2025). Antimicrobial Resistance Dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/vihaankulkarni/antimicrobial-resistance-dataset
13. Lau, H. J., Lim, C. H., Foo, S. C., & Tan, H. S. (2021). The role of artificial intelligence in the battle against antimicrobial-resistant bacteria. Current genetics, 67(3), 421-429.
14. Liu, Z., Deng, D., Lu, H., Sun, J., Lv, L., Li, S., ... & Wang, G. (2020). Evaluation of machine learning models for predicting antimicrobial resistance of Actinobacillus pleuropneumoniae from whole genome sequences. Frontiers in microbiology, 11, 48.
15. Maguire, F., Rehman, M. A., Carrillo, C., Diarra, M. S., & Beiko, R. G. (2019). Identification of primary antimicrobial resistance drivers in agricultural nontyphoidal Salmonella enterica serovars by using machine learning. Msystems, 4(4), 10-1128.
16. Moat, J., Zovoilis, A., Steinkey, R., Zaheer, R., McAllister, T., & Laing, C. (2025). Machine learning methods to identify markers and predict antimicrobial resistance in Escherichia coli. Canadian Journal of Microbiology, 71, 1-15.
17. Mohammad-Taheri, S., Navada, P. P., Hoyt, C. T., Zucker, J., Sachs, K., Gyori, B. M., & Vitek, O. (2024). Eliater: a Python package for estimating outcomes of perturbations in biomolecular networks. Bioinformatics, 40(9), btae527.
18. Nguyen, M., Long, S. W., McDermott, P. F., Olsen, R. J., Olson, R., Stevens, R. L., ... & Davis, J. J. (2018). Using machine learning to predict antimicrobial minimum inhibitory concentrations and associated genomic features for nontyphoidal Salmonella. BioRxiv, 380782.
19. Penven, M., Zouari, A., Nogues, S., Collet, A., Lecourt, M., Birer, A., ... & Cattoir, V. (2023). Web-based prediction of antimicrobial resistance in enterococcal clinical isolates by whole-genome sequencing. European Journal of Clinical Microbiology & Infectious Diseases, 42(1), 67-76.
20. Ren, Y., Chakraborty, T., Doijad, S., Falgenhauer, L., Falgenhauer, J., Goesmann, A., ... & Heider, D. (2022). Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning. Bioinformatics, 38(2), 325-334.
21. Sakagianni, A., Koufopoulou, C., Feretzakis, G., Kalles, D., Verykios, V. S., & Myrianthefs, P. (2023). Using machine learning to predict antimicrobial resistance—a literature review. Antibiotics, 12(3), 452.
22. Salman Ibne Eunus. (2025). antibiotic_resistance-geissen-dataset: Predicting antibiotic resistance genes [Data set]. Kaggle. https://www.kaggle.com/datasets/salmaneunus/antibiotic-resistance-geissen-dataset
23. Van Camp, P. J., Haslam, D. B., & Porollo, A. (2020). Prediction of antimicrobial resistance in Gram-negative bacteria from whole-genome sequencing data. Frontiers in microbiology, 11, 1013.
24. Vanalli, C., Mari, L., Righetto, L., Casagrandi, R., Gatto, M., & Cattadori, I. M. (2020). Within-host mechanisms of immune regulation explain the contrasting dynamics of two helminth species in both single and dual infections. PLoS computational biology, 16(11), e1008438.
25. Waddington, C., Carey, M. E., Boinett, C. J., Higginson, E., Veeraraghavan, B., & Baker, S. (2022). Exploiting genomics to mitigate the public health impact of antimicrobial resistance. Genome medicine, 14(1), 15.