

Application of Decision Tree Based Machine Learning Algorithms for Prediction of Antimicrobial Resistance

Muhammad Yasir^{1, 2†,*}, Asad Mustafa Karim^{3, †,*}, Sumera Kausar Malik⁴, Amal A. Bajaffer¹, Esam I. Azhar¹

¹ Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah 21589, Saudi Arabia

² Department of Medical Laboratory Sciences, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia

³ Graduate School of Biotechnology, College of Life Sciences, Kyung Hee University, Yongin 17104, South Korea;

⁴ Department of Bioscience and Biotechnology, The University of Suwon, Hwaseong City, Gyeonggi-do 18323, South Korea

* Correspondence: yasirkhattak.mrl@gmail.com (M.Y.); asadmustafa8@gmail.com (A.M.K.)

Figure S1. Bar charts showing importance of top 30 features out of 6,000 features for the Ceftazidime, Meropenem and Ciprofloxacin.

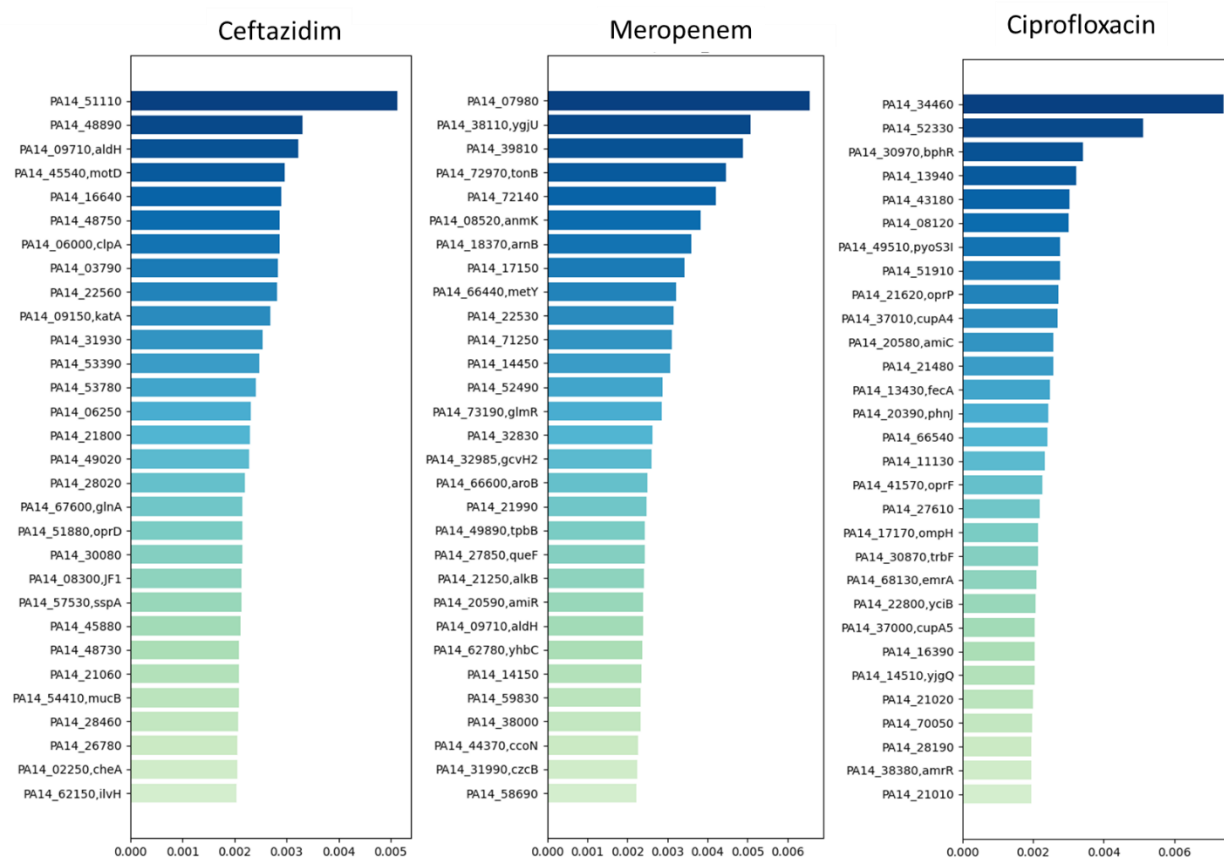


Figure S2. Confusion matrix for the gene expression values of *P. aeruginosa* isolates with a threshold set at 0.5. Rows represent the true transcriptomic values and columns represent the predictions.

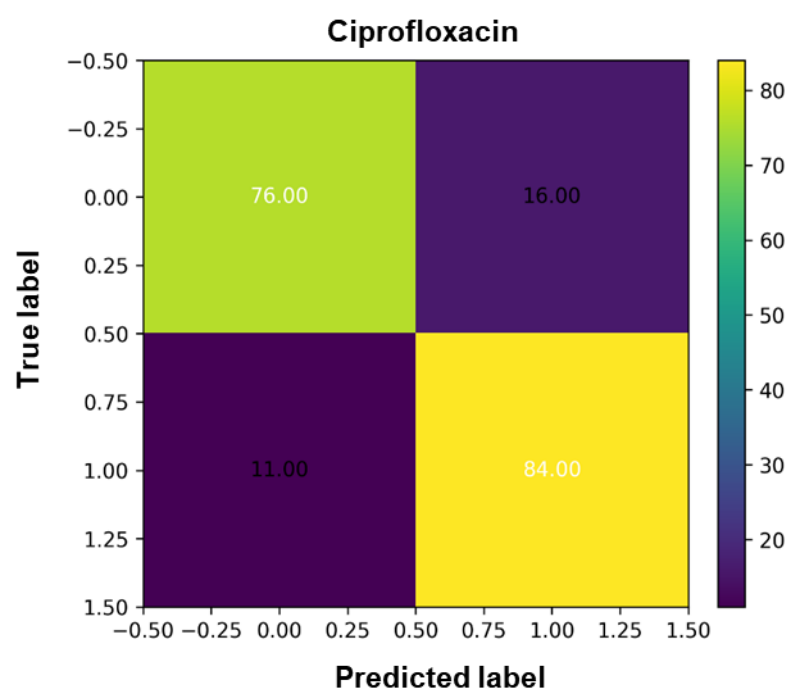


Figure S3. The comparison of specificity of ten classification models used for gene expression prediction for meropenem.

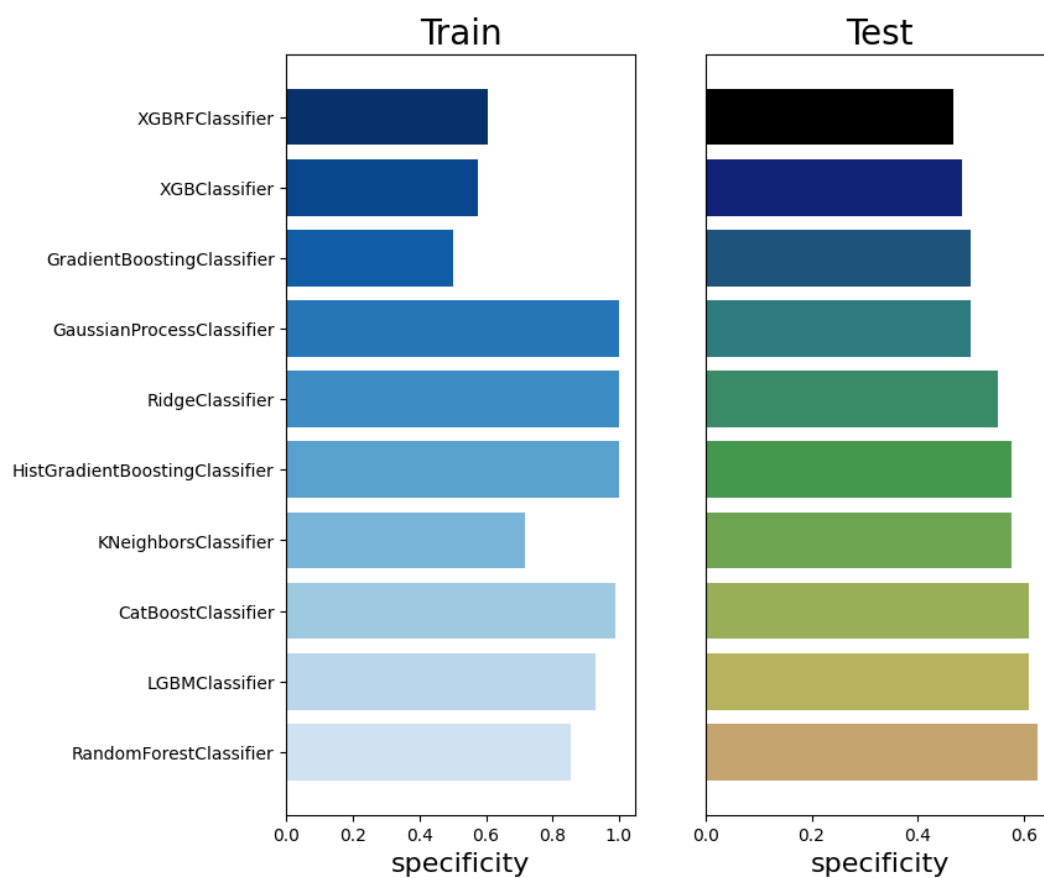


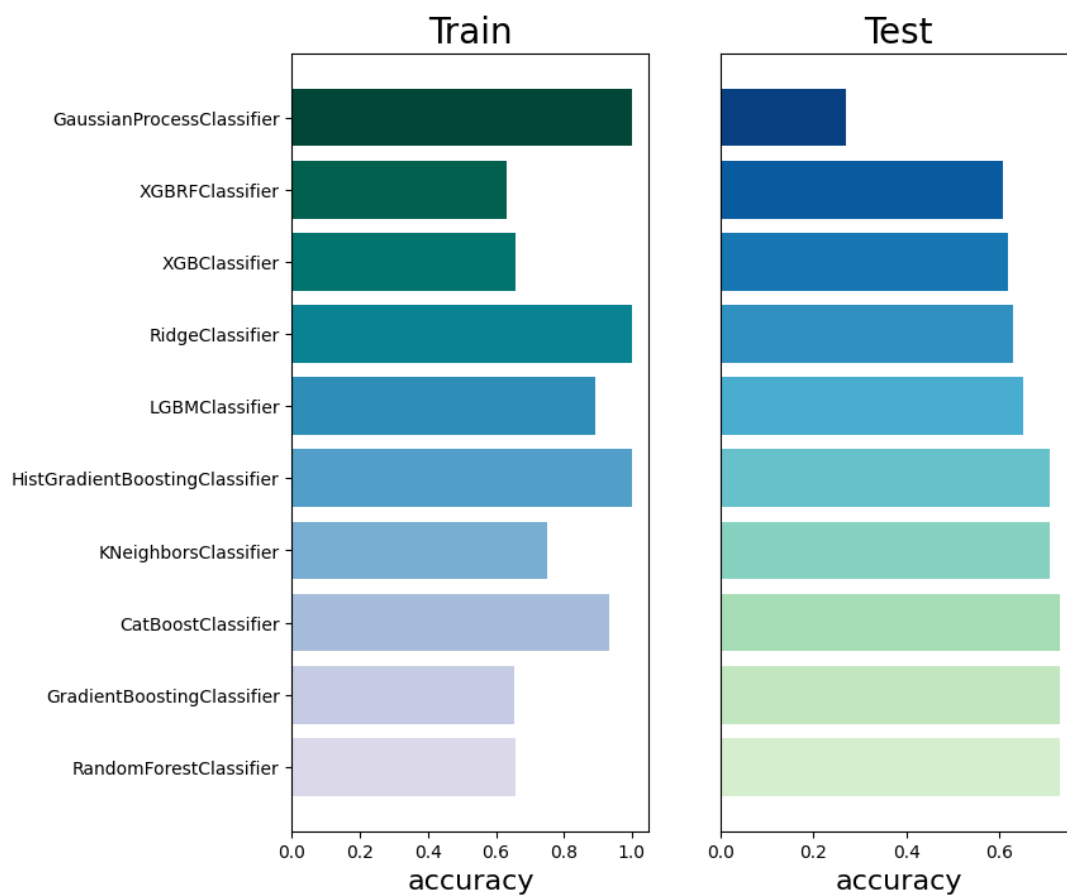
Figure S4. The comparison of accuracy of ten classification models used for gene expression prediction for meropenem.

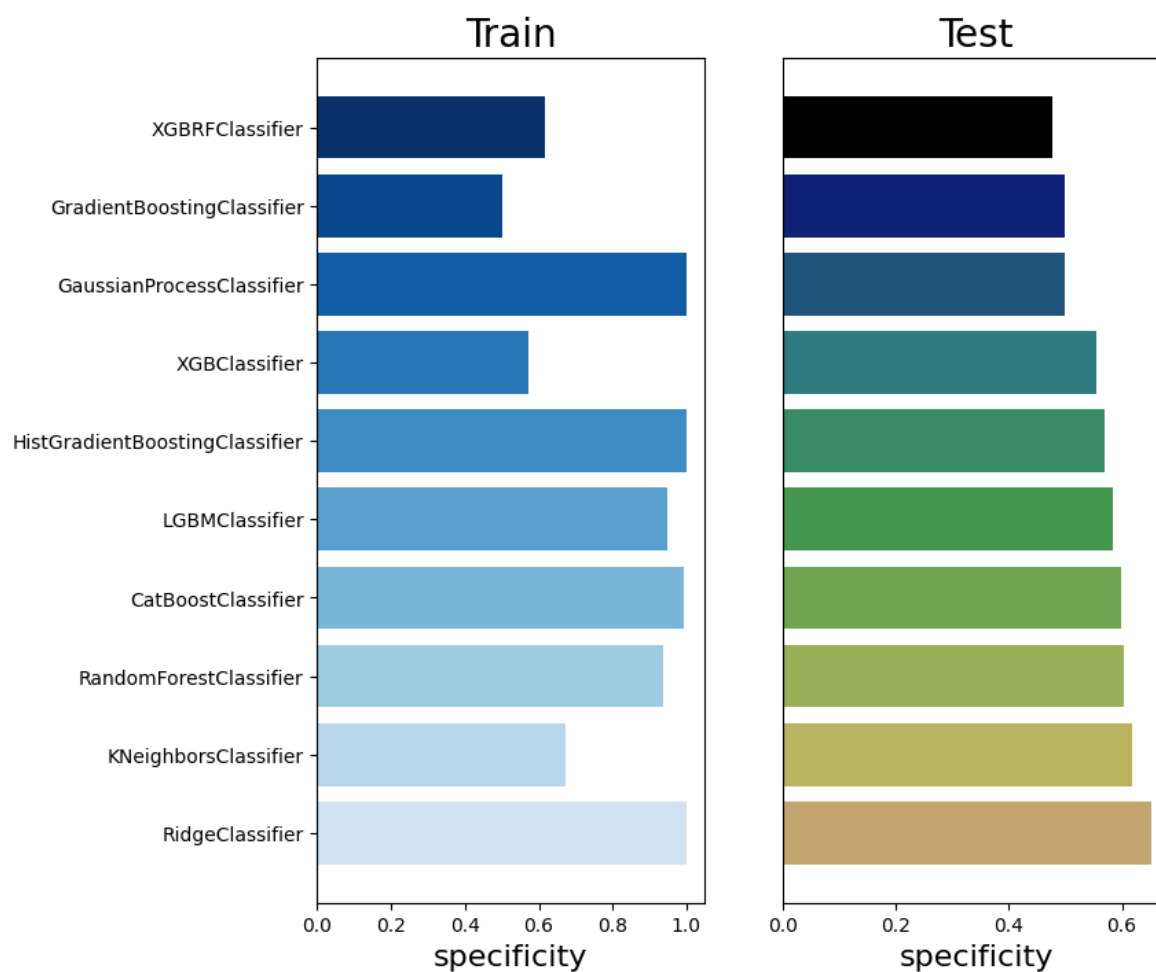
Figure S5. The comparison of specificity of ten classification models used for gene expression prediction for ceftazidime.

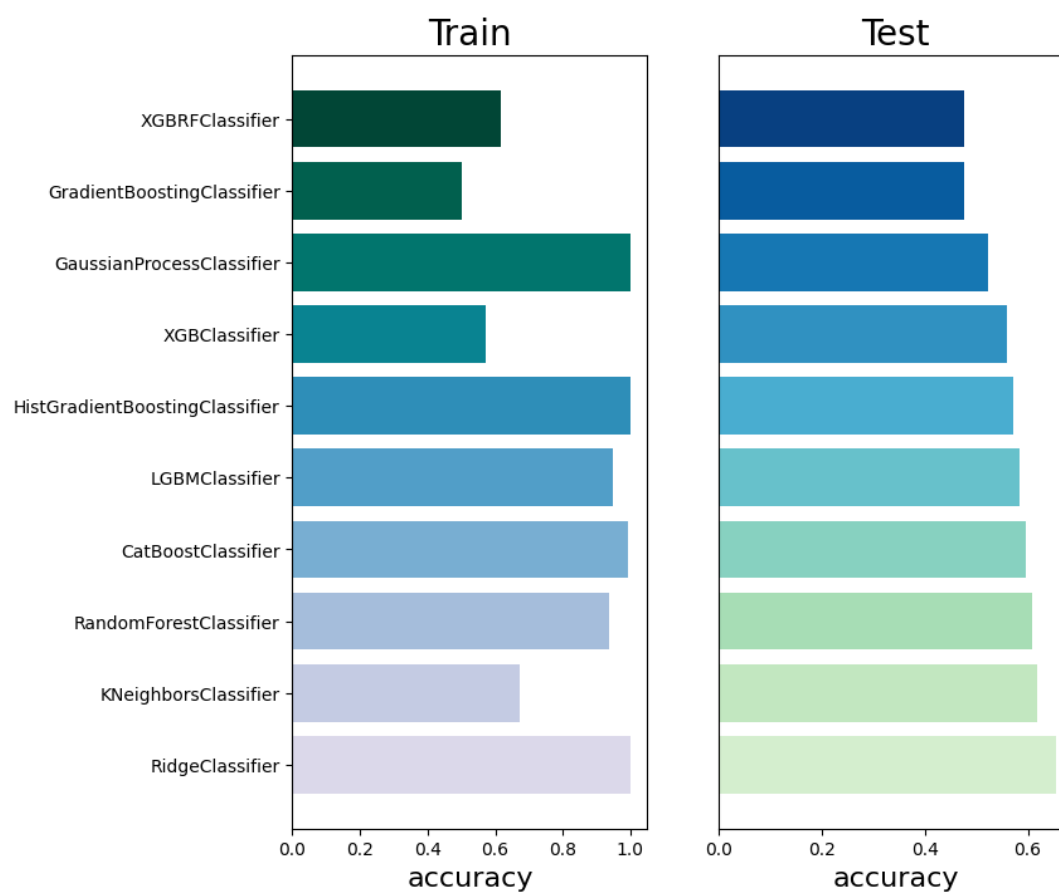
Figure S6. The comparison of accuracy of ten classification models used for gene expression prediction for ceftazidime.

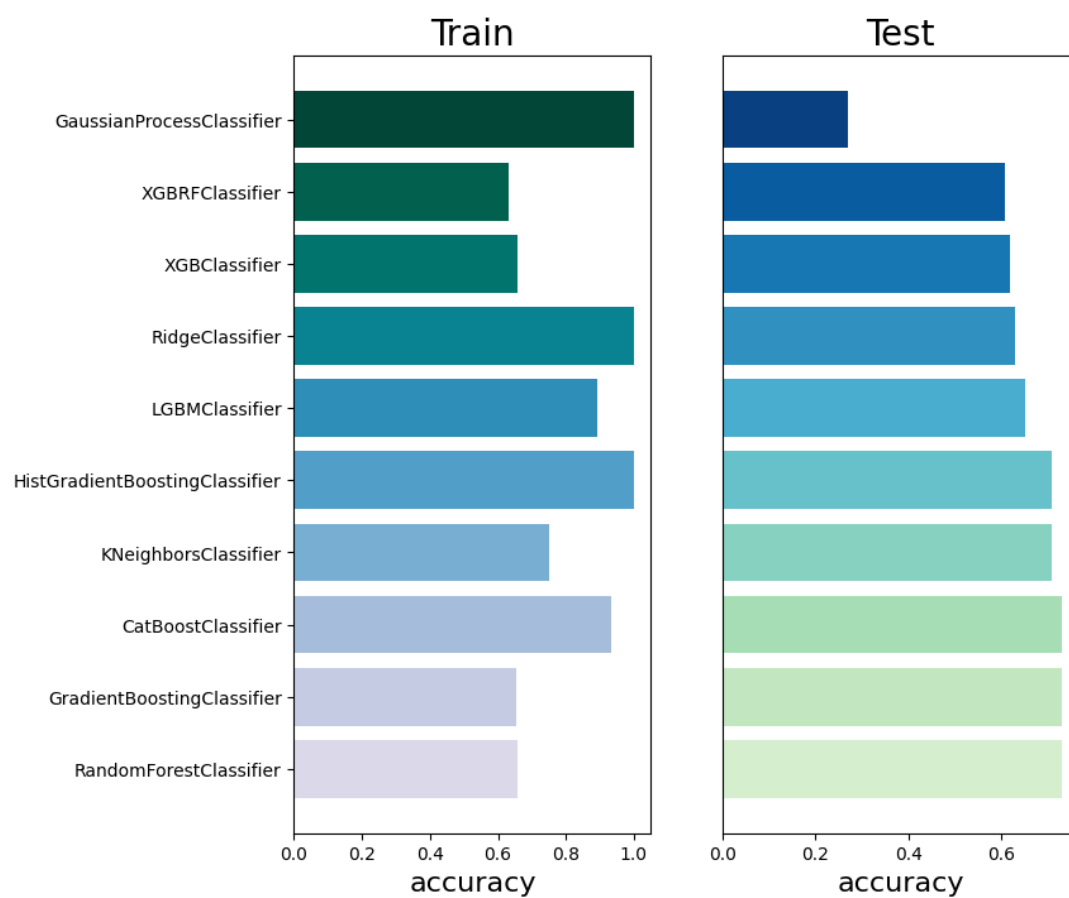
Figure S7. The comparison of accuracy of ten classification models used for gene expression prediction for Ciprofloxacin.

Figure S8. The comparison of specificity of ten classification models used for gene expression prediction for Ciprofloxacin.

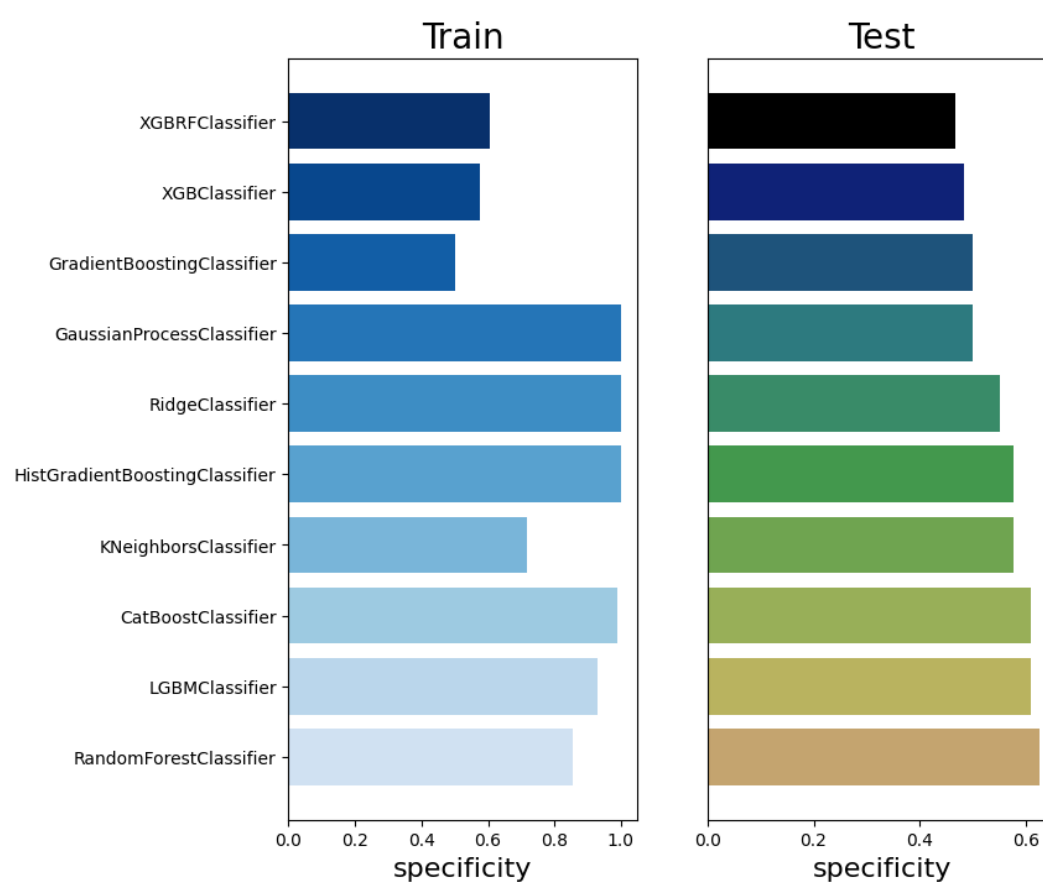
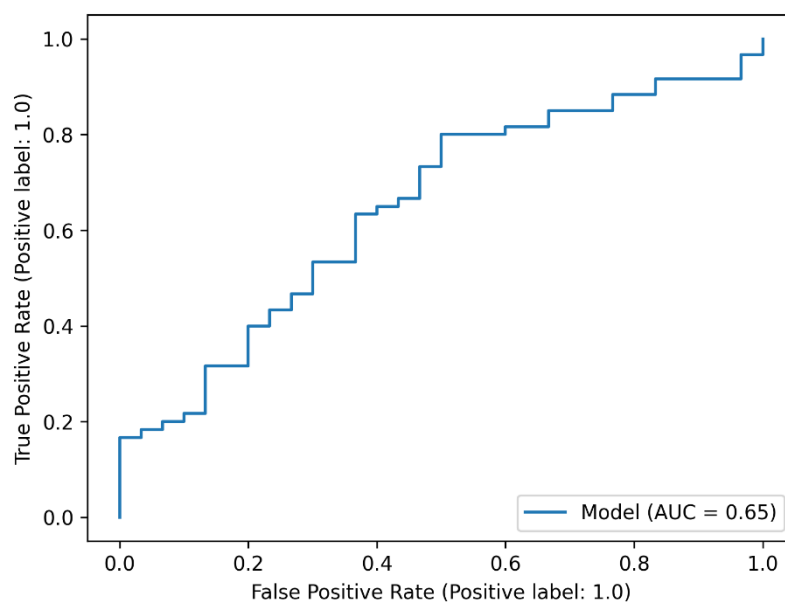
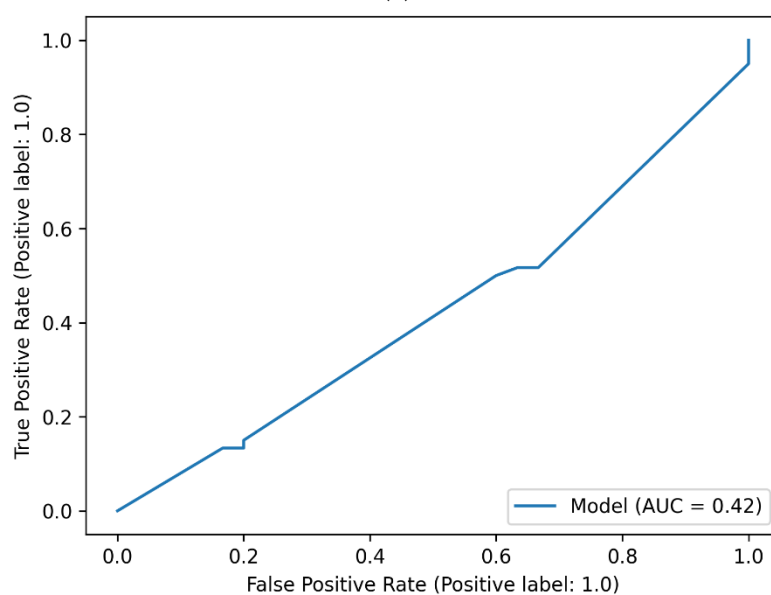


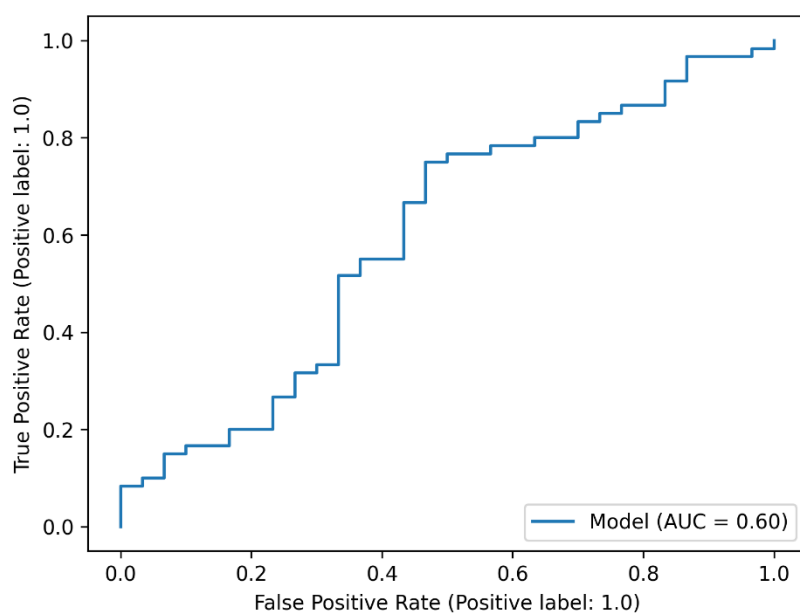
Figure S9. (a). Receiver operating characteristic (ROC) curves for Catboost Classifier in case of ciprofloxacin. AUC value is also shown in the figure. (b). Receiver operating characteristic (ROC) curves for GradientBoosting Classifier in case of ciprofloxacin. AUC value is also shown in the figure. (c). Receiver operating characteristic (ROC) curves for HitsGradientBoosting Classifier in case of ciprofloxacin. AUC value is also shown in the figure. (d) Receiver operating characteristic (ROC) curves for RandomForest Classifier in case of ciprofloxacin. AUC value is also shown in the figure.



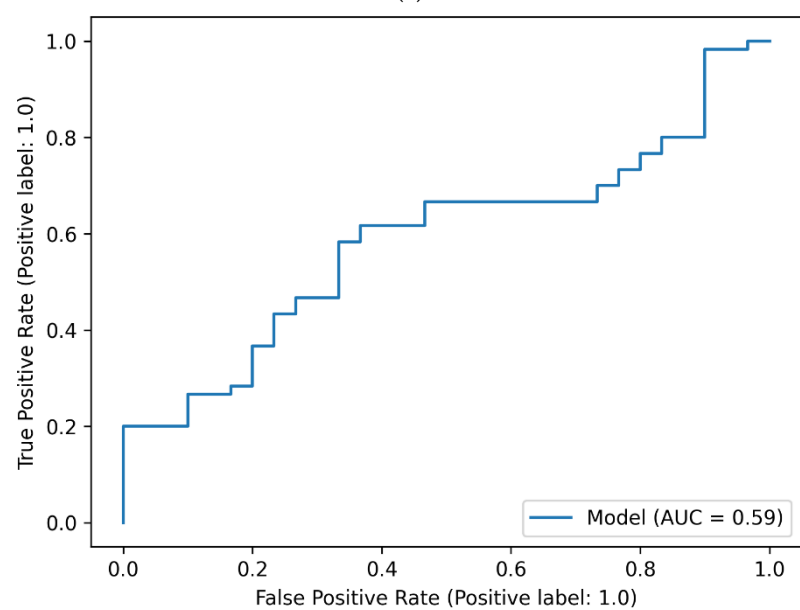
(a)



(b)

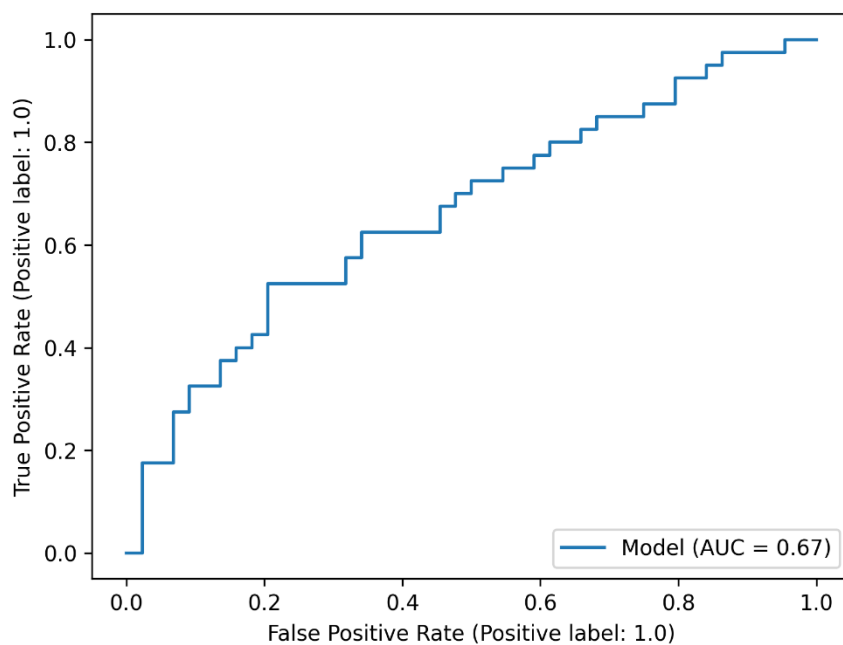


(c)

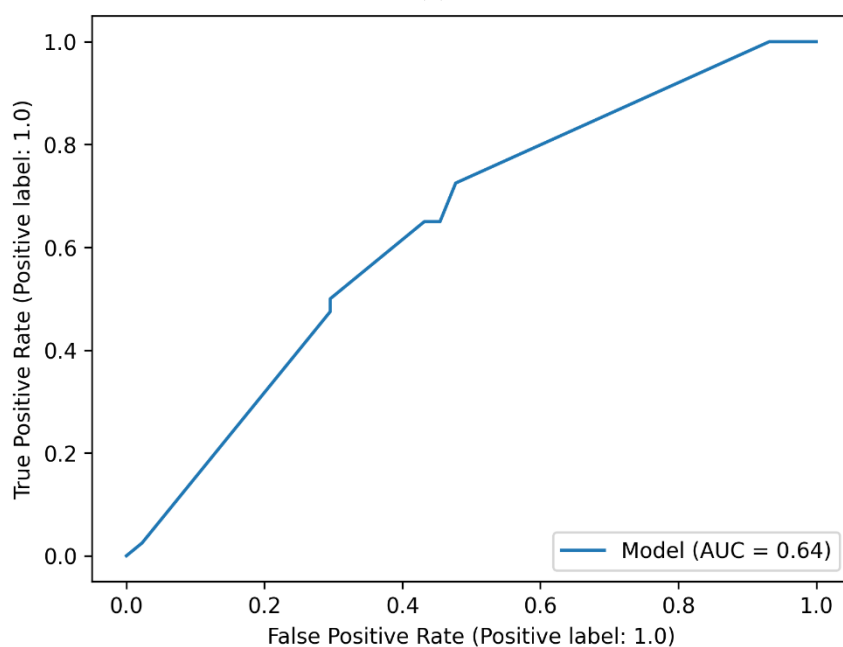


(d)

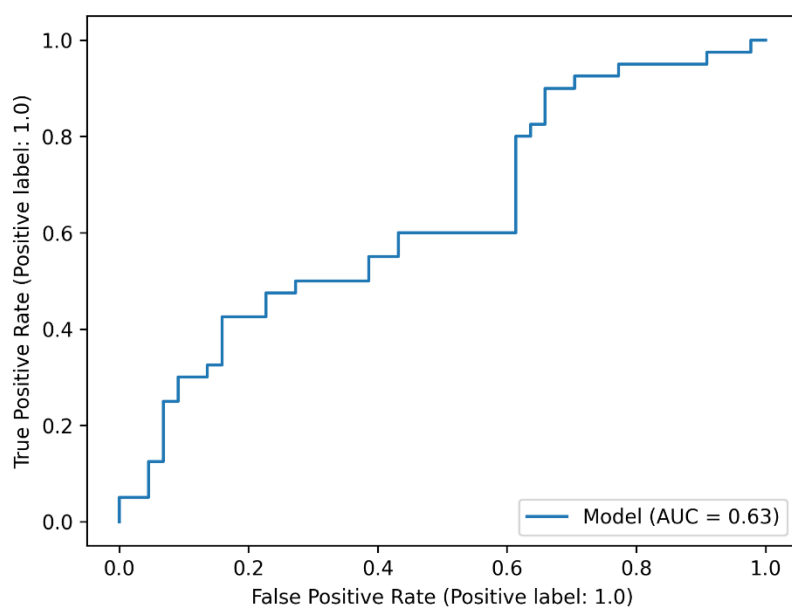
Figure S10. (a). Receiver operating characteristic (ROC) curves for Catboost Classifier in case of Ceftazidime. AUC value is also shown in the figure. (b). Receiver operating characteristic (ROC) curves for GradientBoosting Classifier in case of Ceftazidime. AUC value is also shown in the figure. (c). Receiver operating characteristic (ROC) curves for HitsGradientBoosting Classifier in case of Ceftazidime. AUC value is also shown in the figure. (d). Receiver operating characteristic (ROC) curves for Randomforest Classifier in case of Ceftazidime. AUC value is also shown in the figure.



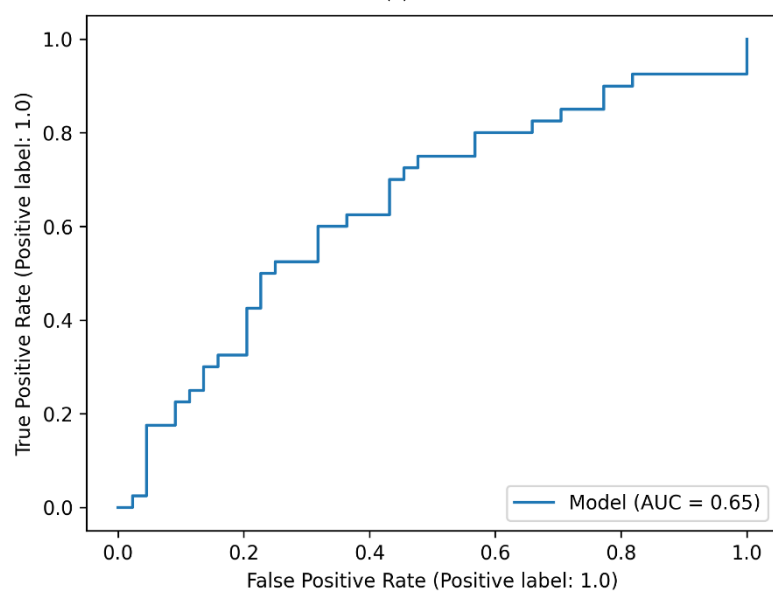
(a)



(b)

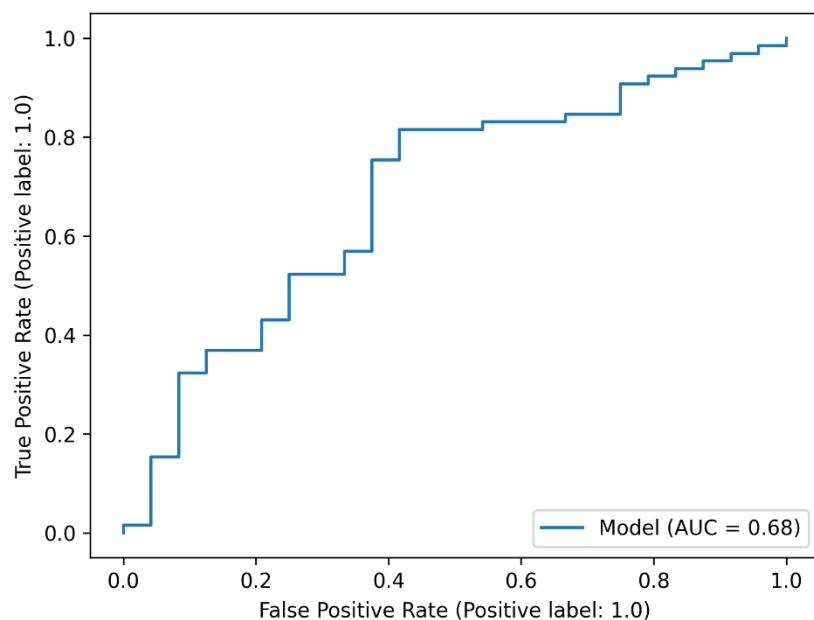


(c)

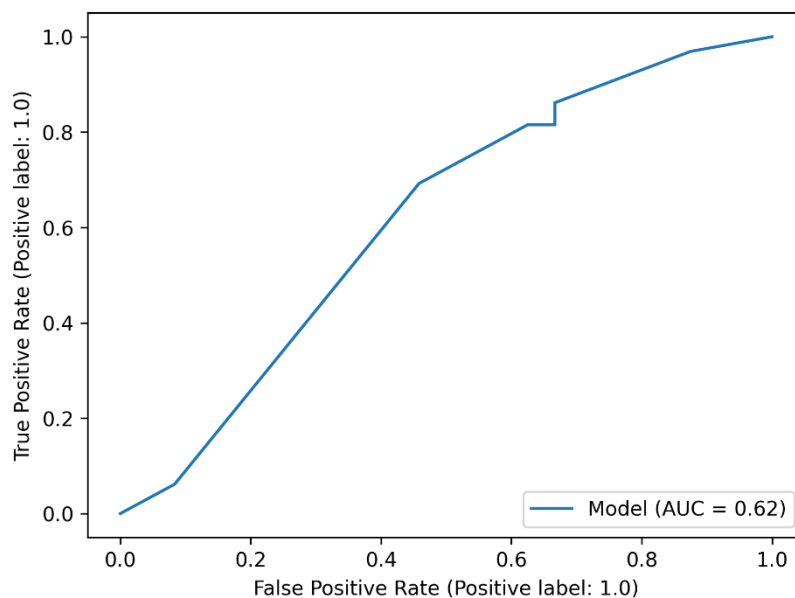


(d)

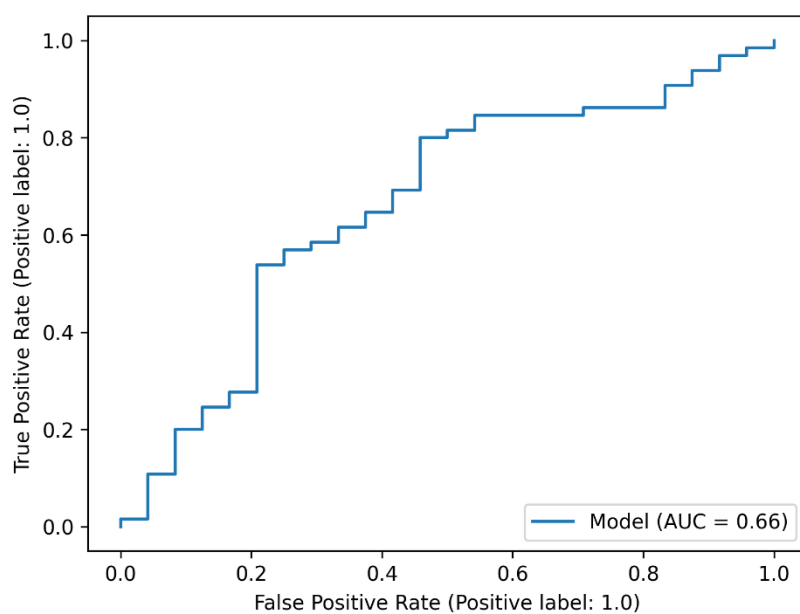
Figure S11 (a). Receiver operating characteristic (ROC) curves for Catboost Classifier in case of meropenem. AUC value is also shown in the figure. **(b).** Receiver operating characteristic (ROC) curves for Gradientboosting Classifier in case of meropenem. AUC value is also shown in the figure. **(c).** Receiver operating characteristic (ROC) curves for HITSGradientboosting Classifier in case of meropenem. AUC value is also shown in the figure. **(d)** Receiver operating characteristic (ROC) curves for Randomforest Classifier in case of meropenem. AUC value is also shown in the figure.



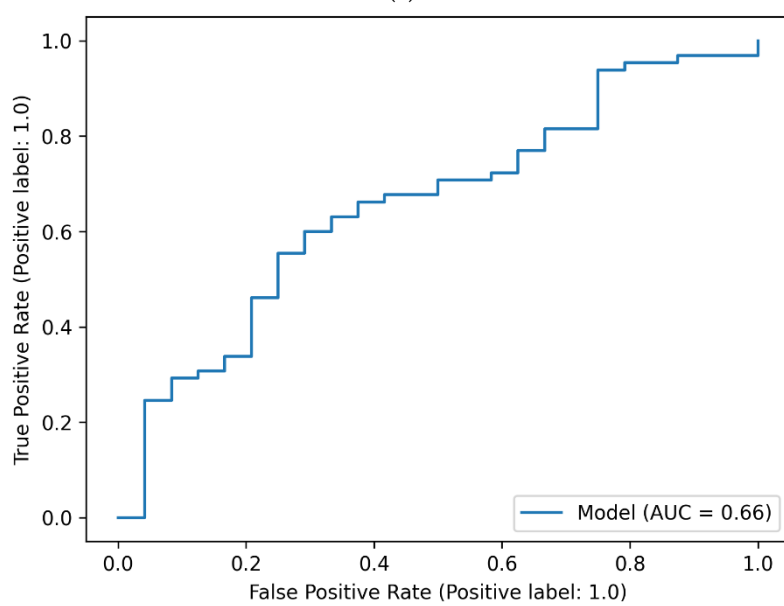
(a)



(b)



(c)



(d)

Table S1. Metrics comparison of ten different ML classification models. Models were compared on the basis of their performance to predict gene expression values of ciprofloxacin on the test datasets.

	F1-score	precision	specificity
CatBoostClassifier	0.586	0.596	0.608
GaussianProcessClassifier	0.25	0.16666	0.5
GradientBoostingClassifier	0.4	0.3333	0.5
HistGradientBoostingClassifier	0.554	0.566	0.575
KNeighborsClassifier	0.539	0.567	0.575
LGBMClassifier	0.598	0.598	0.608
RandomForestClassifier	0.6	0.611	0.625
RidgeClassifier	0.537	0.5446	0.55
XGBClassifier	0.395	0.4799	0.4833
XGBRFClassifier	0.4650	0.4694	0.4666