

ANTIMICROBIAL RESISTANCE PREDICTION VIA A SOFT VOTING CLASSIFIER MODEL FOR MALDI-TOF MASS SPECTRA

Rocchi Ettore¹, Faccin Mauro¹, Sala Claudia¹, Lombardo Donatella², Ambretti Simone², Castellani Gastone¹

¹ Department of Medical and Surgical Sciences, Università di Bologna, 40138 Bologna, Italy

² Microbiology Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, 40138 Bologna, Italy

Introduction

Antimicrobial resistance (AMR), due to a massive and often uncontrolled use of antibiotics, is nowadays a huge problem with massive implications on both economic and public health terms. In this contribution, we describe a machine-learning-based model capable of predicting the resistance or susceptibility of *Escherichia coli*'s strains to Ampicillin, starting from the MALDI-TOF mass spectra of those strains. This work has been funded by the EU - NextGenerationEU with funds made available by the National Recovery and Resilience Plan (NRRP) - Partenariati Estesi (PE13 - INF-ACT) - CUP J33C22002870005.

Objectives

The main objective of this study is to present a machine learning algorithm which may be of crucial help in clinical contexts, since it allows to specifically tailor an antibiotic therapy based on resistance and susceptibility of the considered strain, providing relevant information to guide clinicians' choices, hence saving precious time.

Methods

Our work will exploit a publicly available dataset: a part of the DRIAMS database, supplied by Weis et al. (2022) [1]. The considered dataset is composed of 213 *Escherichia coli*'s MALDI-TOF mass spectra. Among them 136 are resistant and 77 susceptible to Ampicillin. The raw mass spectra, whose m/z range covers the interval 2k-20k Dalton, have been cleaned up with pre-processing procedures, including denoising and smoothing filters (Savitzky-Golay), a detrending process, the ION count normalisation, and, finally, the data binning (with bin width of 3 Da) to reduce the complexity of the spectra to 6000 features. The pre-processed spectra have then been divided into a training set and a testing set (with proportions 75/25%). The used machine-learning models are a logistic regression classifier, a gradient boosting decision tree, a multi-layer perceptron, a k-nearest neighbours algorithm, a naive Bayes classifier, and a support vector machine. Their hyper-parameters have been optimised via a 5-fold cross validation during the training step. Finally, their predictions have been subjected to the soft voting algorithm to retrieve a unified prediction.

Results

The final voting classifier have been evaluated through the most-known metrics, such as the area under the receiver operating characteristic curve (AUROC), the accuracy score and the very major error (VME). In particular, the voting classifier performed on the testing set (composed of 54 mass spectra) produces a prediction accuracy of 72%, an AUROC value equal to 0.74 and a VME equal to 7, meaning that 13% of resistant spectra have been misclassified as susceptible. The receiver operating characteristic (ROC) curve is shown in Figure 1A, while Figure 1B and 1C show the Precision-Recall curve and the per-class ROC curves, respectively.

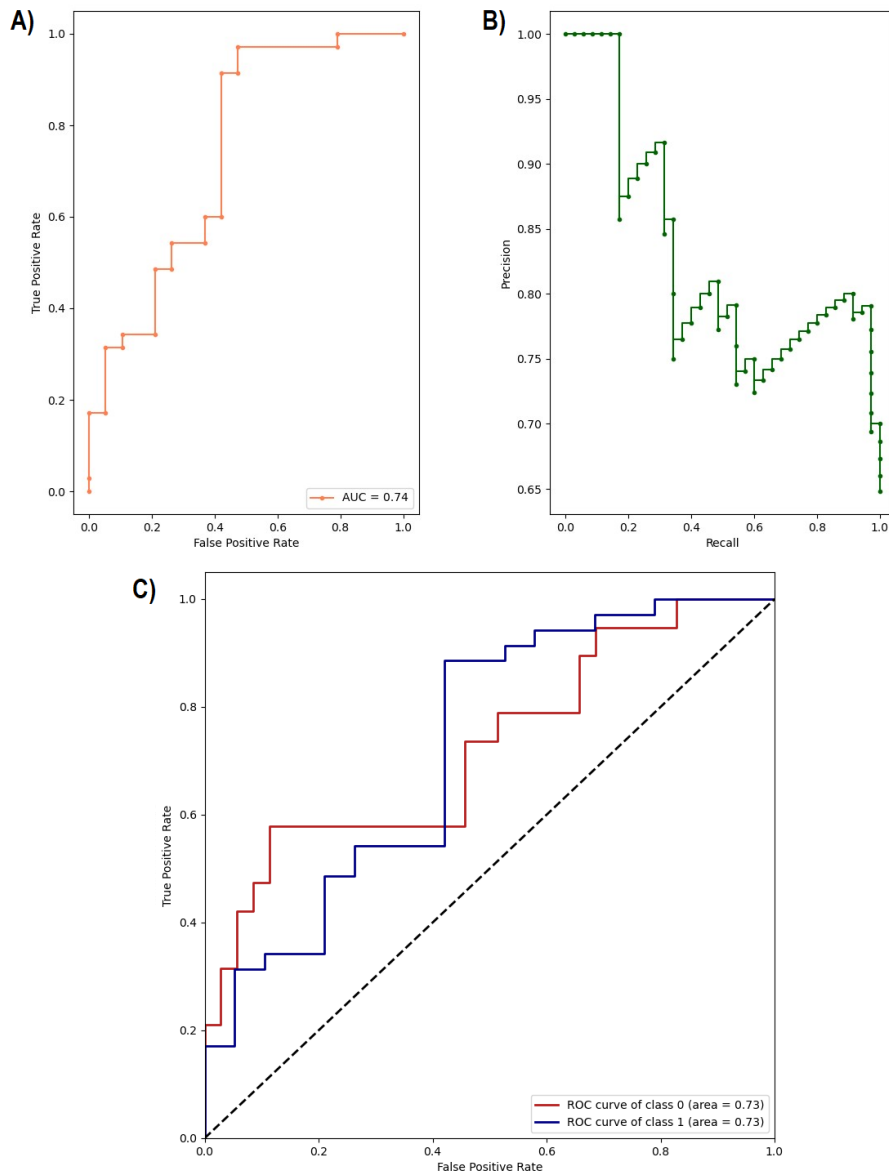


Figure 1: Some metrics' representations for the soft voting classifier: A) ROC curve; B) Precision-Recall curve; C) per-class ROC curve.

Conclusions

This work shows that the ensemble learning method of soft voting classification, and more generally the machine-learning models applied to MALDI-TOF mass spectra, may be a useful and fast tool to predict antimicrobial resistance, thus helping clinicians in the choice of the proper and more effective therapeutical strategy, which in many cases could be of vital importance.

Bibliography

[1] Weis, C., Cuénod, A., Rieck, B. et al., *Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning*. *Nat Med*, 2022 Jan; 28:164–174