DOMAIN ADAPTATION FOR ANTIBIOTIC RESISTANCE CLASSIFICATION WITH MALDI-TOF DATA

Leila Aro-Sati

Student ID: 01716809

Promotor: Dr. Daniel Peralta Cámara

Copromotor: Dr. Xaviera López Cortés

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Master of Science in Statistical Data Analysis.

Academic year: 2023-2024

Gent, September 2, 2024

The promotor,	The author,
Dr. Daniel Peralta Cámara	Leila Aro-Sati

ACKNOWLEDGEMENTS

This thesis is part of a Project FONDECYT Iniciación en Investigación No.11220897 from ANID Chile. The result of the collaboration of UGent and Universidad Católica del Maule in Chile. The integration of Artificial Intelligence in a real-life problem such as antibiotic resistance truly fascinates me.

First and foremost, I would like to thank my supervisor, Dr. Daniel Peralta Cámara, for his continuous guidance, support, and encouragement during my research and studies. Your guidance has been invaluable in the progress of this thesis, and I sincerely appreciate your patience and understanding.

I would like to express my gratitude to my co-supervisor Prof. Xaviera López Cortés from Universidad Católica del Maule for her collaboration, input, and constructive feedback during our meeting in Ghent. It has significantly improved the quality of this work.

My deepest acknowledgments to my family for their continuous support and encouragement throughout this journey. Your belief in me has been a constant source of motivation and inspiration.

Finally, I would like to thank all teachers of MaStat program and fellow students. Your kindness and support are deeply appreciated.

Thank you all.

CONTENTS

ABSTRACT

Antimicrobial Resistance (AMR) is one the biggest health threats and tackling this problem is essential as it is also an economic and security threat that will cost cumulatively about 100 trillion USD by 2050. As an alternative to the standard antibiotic sensitivity test (AST), clinical microbiology laboratories started to apply mass spectrometry (MS), more particularly the MALDI-TOF technique. However, since the increasing prevalence of antibiotic resistance in pathogenic bacteria continues to evolve it demands innovative approaches for the accurate and timely identification of resistant strains. One of the recent approaches is the application of Machine Learning (ML) techniques to assist in the detection of antimicrobial resistance.

In this thesis, we aim to improve and broaden the application of the recently developed AMR prediction model MSDeepAMR. By applying different multi-label approaches, this research aims to enhance the accuracy and reliability of predictions. By integrating data from different species across various antibiotics, we aim to develop a robust model capable of generalizing across various domains of bacterial resistance. While there are examples in the literature of attempts to apply DL methods for AMR, to the best of our knowledge, the subsequent application of multi-label approaches has not yet been explored. In the first step, a multi-label model for predicting resistance across different antibiotics was implemented. It was found that a single model could be utilized for all given antibiotics, offering an experimentally more straightforward approach. The second step involves the implementation of the model that combines all datasets and predicts across multiple bacteria and antibiotics simultaneously. This model demonstrated decent overall performance and comparable results. Subsequently, a transfer learning approach was applied to test the model on an external dataset. Finally, a self-labeling technique has been applied to the final model. This resulted in considerable improvements in predicting AMR.

Keywords: Antimicrobial Resistance (AMR), MALDI-TOF, MSDeepAMR, Deep Learning, CNN, Transfer Learning, Self-Labeling

CHAPTER 1

RESEARCH OBJECTIVES AND OUTLINE

1.1 Introduction

Fleming's discovery of penicillin in 1928 promoted the "golden era" of antibiotic development in the 1950s to combat infectious diseases. However, infectious pathogens can evolve and develop resistance to previously developed antibiotics. For instance, methicillin-resistant *Staphylococcus aureus* causes deaths of nearly 50,000 people every year in the United States and Europe (Conly and Johnston, 2005). Another example of antibiotic-resistant disease is multidrug-resistant tuberculosis (MDR-TB) which developed resistance to the standard and powerful antibiotics.

The problem of antimicrobial resistance (AMR) of bacteria causes hundreds of thousands of deaths annually. Recognizing the severity of this issue, the EU Regulation in 2019/6 has taken measures to limit the use of antibiotics not only in treatment but also in the form of feed additives for livestock (Parlament and the Council of EU, 2019). The World Health Organization (WHO) has also recognized this widespread phenomenon as a major global threat in 2014 (Bengtsson-Palme et al., 2018).

Resistance of bacteria to antibiotics has gained importance not only in clinical conditions but also in agriculture and aquaculture as well. About 70 percent (by weight) of antibiotics are used to prevent infections or promote growth in livestock (O'Neill, 2016). This raises a threat to human health due to the transmission of antibiotic-resistant bacteria (ARBs) from food-producing animals to humans as shown in Figure 1.1 (Kim and Ahn, 2022). Residual antibiotics in food promote low-dose exposure and indirect harm to humans via antibiotic resistance (Chen et al., 2019). Moreover, drug-resistant strains are transmitted from plant- and animal-derived products (meat, eggs, poultry) to humans via environmental media, food products, and direct contact (e.g., agricultural workers) (Smith et al., 2013). For example, chicken broiler samples

from Morocco showed that isolated strains of *S. aureus* were resistant to penicillin (54%) and ciprofloxacin (17%) (Mourabit et al., 2021). *E. coli* strains isolated from poultry in Poland demonstrated high resistance to ampicillin (100%), doxycycline (100%), and ciprofloxacin (81.3%) (Racewicz et al., 2022). The latter complements the findings of Vieira et al. (2011) on the resistant strains of *E. coli* that cause bloodstream infections in humans potentially derived from food sources. Studies have shown that overuse of antibiotics in food-producing animals reduces the effectiveness of those antibiotics in patient treatments (Chen et al., 2019; Ghorbani et al., 2016).

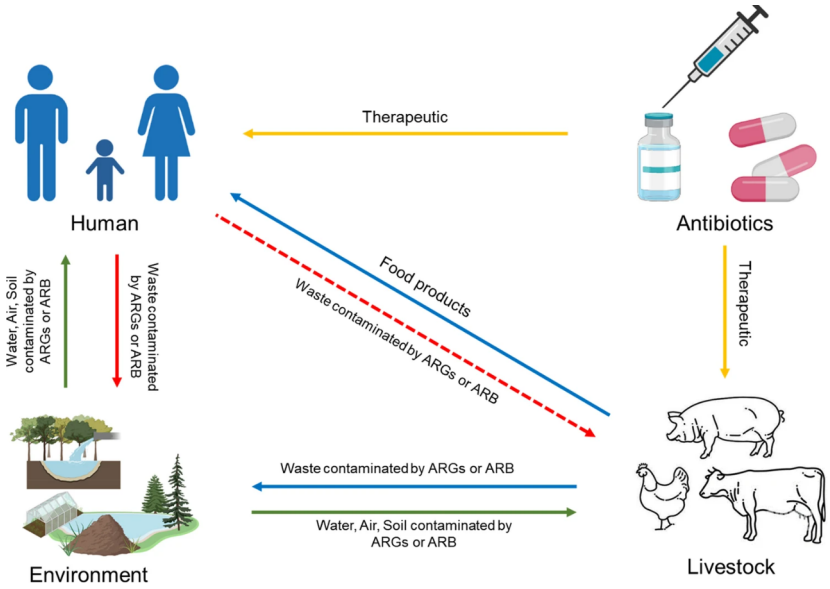

Figure 1.1: Transmission of antibiotic-resistant genes (ARGs) and antibiotic-resistant bacteria (ARBs) (Kim and Ahn, 2022).

Therefore, it is important to detect and monitor the resistance of bacteria towards antibiotics to track the spread of drug resistance globally and take proactive actions. The matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) was introduced in clinical laboratories in the early 2000s for the identification of microorganisms. Nowadays, MALDI-TOF MS is also used to rapidly determine AMR. The commonly used analytical approach using this technique is analysis of the fingerprint spectra in the range of 2000 to 20000 Daltons (Da), reflecting the composite proteome of a bacterial cell. Its ability to create unique spectral fingerprints from bacterial samples enables efficient AMR analysis. The rapid success of MALDI-TOF MS is based on its broad application, accuracy of identification for very diverse groups of microorganisms, robustness, and cost-effectiveness (Florio et al., 2020). MALDI-TOF MS is an effective and reliable monitoring tool that allows rapid and accurate diagnosis of antimicrobial susceptibility. Moreover, it encourages optimal treatment strategies for patients

with infectious diseases which includes appropriate drug selection, reduced hospitalization duration, and overall improved patient outcomes.

1.2 Problem Statement

The reasons for the change in the susceptibility of bacteria are intrinsic resistance (as a result of inherent characteristics) and spontaneous mutation (Blair et al., 2015). In the latter, bacteria defending themselves against antibacterial agents, including antibiotics, evolve and develop different mechanisms to counteract the effects of antibiotics. As a result, bacteria become partially or entirely resistant to a given antibiotic (Acar and Rostel, 2001).

Given the structure of the MALDI-TOF MS fingerprints that display common characteristics such as sharp signal peaks and heteroscedasticity of the base noise level, it is common to use raw mass spectra data for specific data analysis procedures. AI can assist in the lengthy process of antimicrobial susceptibility tests (AST) and promote faster and more efficient patient treatment by applying machine learning techniques. The machine learning techniques that have been commonly used for bacterial species identification mainly exploit artificial neural networks (ANN) including traditional machine learning methods such as random forests (RFs), support vector machines(SVMs), k-nearest neighbors (KNN), logistic regression (LR) (De Bruyne et al., 2011; Mortier et al., 2021; González et al., 2023). While a DT model is traceable (compared to for example an ANN model), a DT model might not perform as well with increasing feature complexity (Ali et al., 2023).

At present, machine learning techniques such as deep learning (DL) are also implemented by researchers for AMR analysis (Popa et al., 2022; Weis et al., 2022; López-Cortés et al., 2024). However, the application of DL in AMR is still underexplored. Current research proposes a tailored approach for multi-label classification problems that allows simultaneous prediction of multiple resistance profiles by exploiting complex MS data as an adaptation to the current research of López-Cortés et al. (2024). Some additional techniques such as self-labeling and custom loss functions were applied. This allowed us to train a model on a larger and more diverse dataset that could potentially improve its ability to generalize better. However, since bacterial resistance profiles can vary significantly across different geographical regions (discussed in Section 2.1) and slight differences in sample collections parameters (López-Cortés

et al., 2024) can limit the generalizability of a DL model. Hence, we employ a transfer learning approach on the DRIAMS-B dataset.

1.3 Objectives

The aim of the current research is to expand research on ML application in AMR by the use of MS data and MSDeepAMR model through the application of the multi-label approach to predict resistance across multiple bacteria species and antibiotics. The step-by-step approach:

1. To develop separate models for each bacterial species that can simultaneously predict resistance across multiple antibiotics.
2. To train a single model per antibiotic across multiple bacterial species.
3. To develop a comprehensive single model that can simultaneously predict resistance across multiple antibiotics and bacterial species.
4. To assess the performance of the model on the external dataset and exploit a self-labeling technique.

The current dissertation is structured as follows:

Chapter 2 gives a basic overview of the current AMR situation and MS application. It also includes a brief explanation of the traditional methods, ML methods, and the recently introduced MSDeepAMR method.

Chapter 3 details the methodology and different approaches used in the current research. It describes the current dataset, deep neural network architecture, model training, and evaluation metrics.

Chapter 4 presents the results of the experiments. This includes tables and figures, a detailed explanation of the results, and a comparison of the different approaches.

Chapter 5 discusses the obtained results, and summarizes the findings. The chapter also provides directions for the future research.

CHAPTER 2

RESEARCH BACKGROUND

Currently, AMR is at dangerously high levels which has a severe global health threat. This increases the importance of applying innovative methods for the accurate and timely identification of resistant bacteria stains. This chapter reviews the biological context of AMR, recent advancements, and current methodologies in the field of AMR detection by focusing on the integration of machine learning techniques.

2.1 AMR and Current Technologies

The causes of AMR are complex and diverse. The AMR develops fast in countries where antibiotics are sold without prescription and used as growth-promoting additives in livestock farms. As such, antibiotic prescription has been inappropriate at least in half of the cases (Tacconelli, 2009).

According to Murray et al. (2022) the six pathogens that are responsible for the death of 929,000 people due to AMR and 3.57 million deaths associated with AMR in 2019 are *Echerichia coli, Staphylococcus aureus, Klebsiella pneumoniae, Streptococcus pneumoniae, Acinetobacter baumannii,* and *Pseudomonas aeruginosa*. These six pathogens have been identified as priority pathogens by WHO (Murray et al., 2022).

However, it is important to note that the share of the AMR burden differed across different geographical locations and socioeconomic factors across the globe (Murray et al., 2022). As such, the study of Murray et al. (2022) on AMR burden in 2019 shows that low-income countries had twice as many deaths attributed to AMR as in high-income countries. The top contributors to the development of AMR in developing countries are (Chokshi et al., 2019):

1. Lack of surveillance of resistance development.

   Indeed, low- and middle-income countries' data on AMR surveillance lack representativeness due to the limited resources and lack of trained and qualified staff (Iskandar et al., 2021).

2. Poor quality control of available antibiotics.

   Several studies have shown that the quality of the drugs is below the standards in developing countries. This also includes relabelled antibiotics whose shelf life is about to expire or even counterfeit drugs (with low or no active ingredients) (Okeke et al., 1999).

3. Clinical misuse.

   The lack of appropriate diagnostic approaches can lead to antibiotic resistance. Many clinicians rely on general symptoms and signs rather than laboratory tests (Achanta et al., 2013).

4. Ease of availability.

   In most developing countries in Asia, Africa, and Latin America antibiotics are available without prescription in local pharmacies, hospitals, and roadside stalls without prescription (Chokshi et al., 2019; Dooling et al., 2014; Satyanarayana et al., 2016). This leads to the overuse of antibiotics due to self-medication. The proportion of patients who self-medicate is probably higher because patients are often reluctant to admit having taken antibiotics before visiting a hospital. According to Chokshi et al. (2019) the ease of availability is potentially the main driver of the AMR.

Whereas in developed countries main factors are (Chokshi et al., 2019):

1. Poor hospital-level regulation.

   Antibiotics are used not only in the treatment of infectious diseases but also prophylactically to reduce the risk of infections during surgeries and clinical procedures. Hence, one of the main sources of antibiotic-resistant infections in developed countries is hospital-acquired infections (nosocomial infections) (Chokshi et al., 2019; Weinstein, 2001). Special hospital programs (e.g. antimicrobial stewardship programs) and hospital regulations were introduced to tackle this problem.

2. Overuse of antibiotics in food-producing animals.

   Consumption of animal-derived products contaminated with antibiotic-resistant pathogens can lead to the transmission of drug-resistant strains to humans as explained in Figure 1.1 (Chen et al., 2019; Smith et al., 2013; Kim and Ahn, 2022).

Traditional methods of antimicrobial susceptibility testing (AST) are effective but time-consuming which can take up to 72 hours to obtain the results (O'Neill, 2016). These methods are widely used in hospital laboratories to guide patient treatments. However, these methods require microbiology facilities and specially trained personnel for accurate implementation and are applicable only for cultivable bacteria (Boolchandani et al., 2019). These drawbacks along with lengthy testing time in detection can hinder timely patient treatment.

Currently, there are several popular methods for AST purposes (Gajic et al., 2022):

- PCR - polymerase chain reaction. This method is both rapid and highly sensitive (Liu et al., 2019). However, the disadvantages are that PCR is prone to errors and that at least some prior sequence data is required.
- qPCR - a quantitative polymerase chain reaction. While this method yields more information than PCR (measurements are performed in real-time), the limitations of PCR still remain.
- NGS - next-generation sequencing. As the whole DNA is sequenced, more information can be gleaned (Behjati and Tarpey, 2013). But as a relatively new technology, it is not yet fully standardized and uptake globally is uneven.
- MALDI-TOF MS - matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. It is a fast and accurate method, but the downside is the high initial cost of the instrumentation (Singhal et al., 2015).

In this research, we focus on the latter technique.

2.2 MS in Microbiology

A mass spectrometer (MS) is a device that separates ions by their mass-to-charge ratio (m/z). However, it does require that the analyte molecules are in a charged state (i.e. ionized). After ionization the ions are separated by their m/z, using for example electric fields, in the mass analyzer. Once the ions are separated, they reach the detector that measures their abundance (Gross, 2019).

In MALDI the sample is mixed with a matrix (that would absorb laser light), which is then excited using a laser beam (Hillenkamp et al., 1991). In comparison with other ionization

techniques (e.g. electron ionization), it is a ”soft” ionization technique – resulting in fewer fragments. This makes it especially suitable for biological samples. The MALDI ionization technique is often coupled with a time of flight (TOF) mass analyzer. The TOF mass analyzer offers a combination of high resolution and the ability to analyze a wide m/z range (Xian et al., 2012).

The initially available ionization techniques (chemical or electrical) were not suitable for the analysis of biomolecules – these techniques use a high amount of energy, which can lead to unpredictable degradation of the biomolecules (Sauer and Kliem, 2010). The later developed electrospray ionization (ESI) and MALDI are softer and thus more suitable for the analysis of large biomolecules. The analysis of bacteria using MS was first described in the 1970s (Anhalt and Fenselau, 1975).

The use of MALDI-TOF MS for the analysis of bacteria was first described in 1990s (Cain et al., 1994). From this point onwards MALDI-TOF MS has revolutionized microbial identification in clinical settings (Singhal et al., 2015). Initially, MALDI-TOF MS was used for rapid and accurate identification of bacteria and yeast species. This technique has been further explored to develop a fast and reliable method for ASTs. Hence, several studies have been conducted to investigate the possibility of MALDI-TOF for speedy and accurate AMR detection in bacterial and fungi pathogens (Florio et al., 2018). Nowadays, it is successfully used in clinical microbiology laboratories to quickly identify pathogens and their resistance profiles. As such, it detects peaks associated with drug resistance, for example, the presence of an enzyme associated with AMR (Yoon and Jeong, 2021).

2.3 Machine Learning and AMR

As the penetration of electronic health records has increased, large amounts of data have become available (Beam and Kohane, 2018). This presents an opportunity to apply machine learning (ML) techniques. The number of publications applying ML to AMR research has increased rapidly since 2012 (Farhat et al., 2023). One application area of ML has been to improve antibiotic selection and clinical decision-making by using information already available from the electric patient record (Sakagianni et al., 2023). While the uptake of ML in the clinical setting might not be very rapid due to safety concerns, it is likely to be more applicable in the laboratory setting (Macesic et al., 2017). Another area where ML techniques have increasingly been

applied to biomedical data is to enhance the accuracy and efficiency of diagnostic procedures. Initial applications in bacterial species identification utilized algorithms such as support vector machines (SVM), random forests (RF), and k-nearest neighbors (KNN) (De Bruyne et al., 2011). The main advantage of the SVM approach is its generalization performance. Furthermore, SVM offers a solution and discriminative power. However, this performance comes at a high computational cost, especially for large sets of data. Because SVM training is most often executed as a quadratic programming problem (Cervantes et al., 2020). The RF approach allows the importance of each feature to be measured with respect to the training data set and the proximity between samples can be measured. However, RF can incur biases depending on the differences in attribute levels and correlation features (Prajwala, 2015). While the KNN approach is comparatively simple to implement, processing large sets of data is slow. Additionally, the KNN approach is sensitive to the use of unnecessary parameters (Cervantes et al., 2020).

Recent advances have seen the integration of more sophisticated models, including artificial neural networks (ANN) and deep learning (DL) architectures, which can handle large datasets and identify complex patterns in mass spectrometry data (Weis et al., 2022). ANN can solve complex problems, evaluate features, and allow for multivariate features. However, they are inherently not traceable and become increasingly complex as more layers and/or nodes are added (Ali et al., 2023). Additionally, gradient boosting machine (GBM) and decision tree (DT) have grown in popularity as well (Sakagianni et al., 2023). While a DT model is traceable (compared to for example an ANN model), a DT model might not perform as well with increasing feature complexity (Ali et al., 2023) compared to DL models.

Multi-label classification (MLC) has been used previously to predict AMR (Ren et al., 2022). However, in the paper of Ren et al. (2022) used whole-genome sequencing data and only the resistance of *E. coli*. Current research, applies MLC to predict AMR of several bacteria across multiple antibiotics. Additionally, in the work presented here MALDI-TOF MS data is used, which is a faster method than whole-genome sequencing (Sakagianni et al., 2023). In the context of AMR, MLC has been applied to MALDI-TOF MS data to predict the resistance of *S. aureus* to two antibiotics (Zhang et al., 2022). There are also research on application of MLC on smaller datasets (PATRIC) which applied similar masked loss function for incomplete labels as in the current research (Tharmakulasingam et al., 2022). CNNs have found different applications in AMR research. For example, identifying novel sites in the genome of a bacteria linked

to AMR in tuberculosis (Green et al., 2022). Closely related to current research is the predicting AMR of bacteria based on their MALDI-TOF MS spectra by utilizing CNNs (López-Cortés et al., 2024). To our knowledge, current research exploits multi-label multi-bacteria classification with custom loss function and self-labelling approach on MALDI-TOF MS data from DRIAMS dataset by utilizing the MSDeepAMR model (CNN) for the first time.

2.4 MSDeepAMR

Deep learning, particularly convolutional neural networks (CNN), has shown promise in enhancing AMR prediction by leveraging the detailed information contained in raw mass spectrometry data (López-Cortés et al., 2024). Studies have demonstrated that CNNs can effectively classify antibiotic resistance across various bacterial species by learning to recognize intricate patterns in spectral data.

The study by López-Cortés et al. (2024) introduced the DeepAMR model, which applies deep neural networks to predict AMR directly from raw MALDI-TOF MS spectra. This model achieved higher classification performance compared to traditional ML approaches (Weis et al., 2022), particularly when combined with transfer learning techniques to adapt pre-trained models to new datasets (DRIAMS-B, DRIAMS-C and DRIAMS-D) for certain bacteria-antibiotic combinations.

The deep learning model applied by López-Cortés et al. (2024) is a convolutional neural network (CNN) designed to classify antibiotic resistance in bacteria. The model consists of the following layers as represented in Figure 2.1:

- Input Layer: accepts the binned mass spectra data (6000,)
- Convolutional Layers:
  - Convolutional layers (4);
  - ReLu Activation function;
  - Average pooling;
  - Batch normalization.
- Fully Connected Layers
- Dropout rate: 0.65

- Output Layer
- Dense layer with 1 neuron

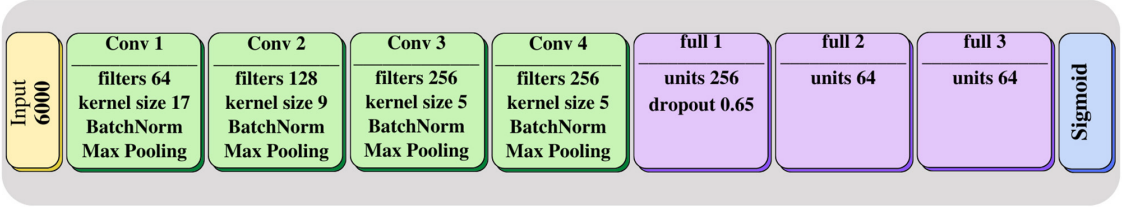

Figure 2.1: DeepAMR architecture (López-Cortés et al., 2024).

The use of multiple convolutional layers allows the model to learn increasingly abstract features from the raw input data, while the fully connected layer integrates these features to make the final classification. The sigmoid activation function indicates the probability of belonging to one of the classes (0 or 1).

In conclusion, the integration of MALDI-TOF MS with machine learning techniques, particularly deep learning model MSDeep AMR, could potentially accelerate the prediction of antimicrobial resistance which could ensure effective and timely treatments.

CHAPTER 3

METHODOLOGY

The current study applies a Deep Learning architecture to detect antibiotic resistance across multiple bacterial species from raw Mass Spectrometry (MS) data. The approach extends the current paper of López-Cortés et al. (2024). The DRIAMS dataset (Drug Resistance in Infectious Agents Modelling and Surveillance) has been chosen to investigate antibiotic resistance patterns in bacterial pathogens due to its high number of samples and public availability (Weis et al., 2022). The size of the database is approximately 300,000 mass spectra with over 750,000 antibiotic resistance profiles which includes initially 803 different types of bacterial and fungal pathogens. The sub-collection of DRIAMS-A has been selected as it contains the largest number of publicly available samples. Hence, the study focuses on the DRIAMS-A dataset for training and cross-validation, with transfer learning applied to the DRIAMS-B dataset. After extraction of the data, binned mass spectra were computed to obtain fixed-length vectors suitable for DL algorithms (López-Cortés et al., 2024). Further, the data was split into training and test sets (80/20). The 10-fold cross-validation was applied to the training set to ensure robust model evaluation. Lastly, the performance metrics are calculated on the test set to assess the model performance. Figure 3.1. visually illustrates the general workflow of this research.

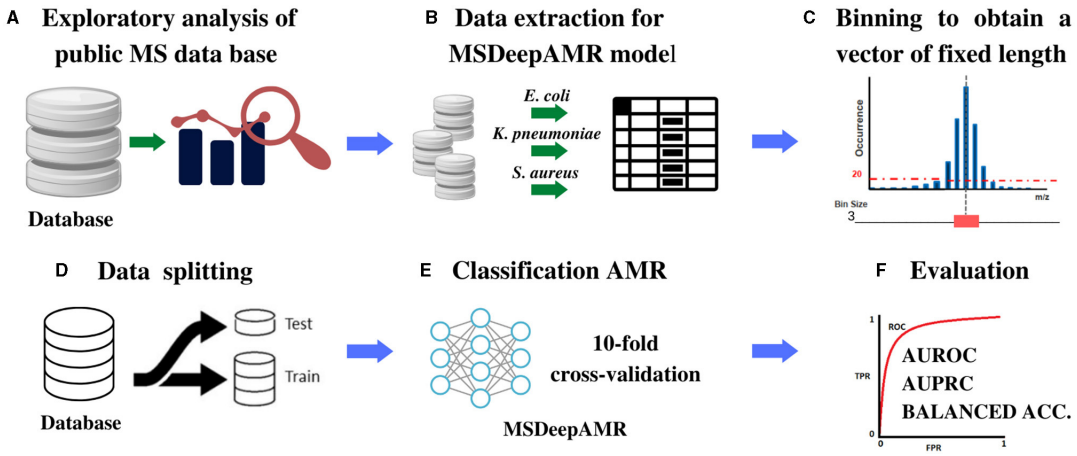

Figure 3.1: Scheme of the methodology proposed for the identification of AMR (López-Cortés et al., 2024).

3.1 Dataset overview

Current research exploits the following priority pathogens (Murray et al., 2022) due to the largest number of samples: *E. coli, K. pnuenomiae, S. aureus* in DRIAMS-A dataset. The extracted dataset consists of 6000 features representing raw MS data represented by binned vectors of size 3 Da in the range of 2000 to 20000 Da and species information. The last columns include the antibiotics with class labels 1 for resistance and 0 for susceptibility. The following bacteria-antibiotics combination were examined:

- ***E. coli***: Ciprofloxacin, Ceftriaxone, Cefepime, Piperacillin-Tazobactam, and Tobramycin.
- ***K. pneumoniae***: Ciprofloxacin, Ceftriaxone, Cefepime, Meropenem, and Tobramycin.
- ***S. aureus***: Ciprofloxacin, Fusidic acid, Oxacillin, Ceftriaxone, and Clindamycin.

This selection of combinations was based on the public availability, clinical relevance, and amount of samples available. Tables 3.1 and 3.2 provide a detailed overview of the class distribution for resistant and susceptible strains for each combination in the DRIAMS-A and DRIAMS-B datasets respectively. These tables highlight the higher number of susceptible strains in each dataset, i.e. class imbalance in each bacteria-antibiotic combination. This could potentially lead to several challenges in the training and evaluation of ML models. For example, biased model predictions towards the majority class, i.e. susceptible samples.

Table 3.1: DRIAMS-A: Susceptible and Resistant Strains Across Antibiotics

Bacteria	Antibiotic	# Susceptible (%)	# Resistant (%)
E. coli	Ciprofloxacin	3445 (70%)	1466 (30%)
	Ceftriaxone	3875 (78%)	1086 (22%)
	Cefepime	4051 (83%)	839 (17%)
	Piperacillin-Tazobactam	4449 (93%)	350 (7%)
	Tobramycin	4240 (87%)	636 (13%)
K. pneumoniae	Ciprofloxacin	2325 (81%)	513 (19%)
	Ceftriaxone	2411 (84%)	449 (16%)
	Cefepime	2477 (87%)	362 (13%)
	Meropenem	2794 (98%)	61 (2%)
	Tobramycin	2527 (89%)	319 (11%)
S. aureus	Ciprofloxacin	3141 (84%)	616 (16%)
	Fusidic acid	3513 (93%)	253 (7%)
	Oxacillin	3064 (81%)	726 (19%)
	Ceftriaxone	2928 (80%)	712 (20%)
	Clindamycin	3056 (85%)	519 (15%)

Complementary to the above-mentioned tables, Figures 3.2 and 3.3 provide a graphical representation of the distribution of resistant and susceptible strains.

Table 3.2: DRIAMS-B: Susceptible and Resistant Strains Across Antibiotics

Bacteria	Antibiotic	# Susceptible (%)	# Resistant (%)
E. coli	Ciprofloxacin	154 (72%)	59 (28%)
	Ceftriaxone	168 (79%)	45 (21%)
	Cefepime	170 (80%)	43 (20%)
	Piperacillin-Tazobactam	164 (77%)	49 (23%)
K. pneumoniae	Ciprofloxacin	130 (85%)	22 (15%)
	Ceftriaxone	134 (88%)	18 (12%)
	Cefepime	134 (88%)	18 (12%)
	Meropenem	146 (99%)	1 (1%)
S. aureus	Oxacillin	325 (94%)	21 (6%)
	Clindamycin	311 (89%)	37 (11%)
	Ceftriaxone	0 (0%)	0 (0%)
	Ciprofloxacin	322 (93%)	26 (8%)
	Fusidic acid	326 (94%)	20 (6%)

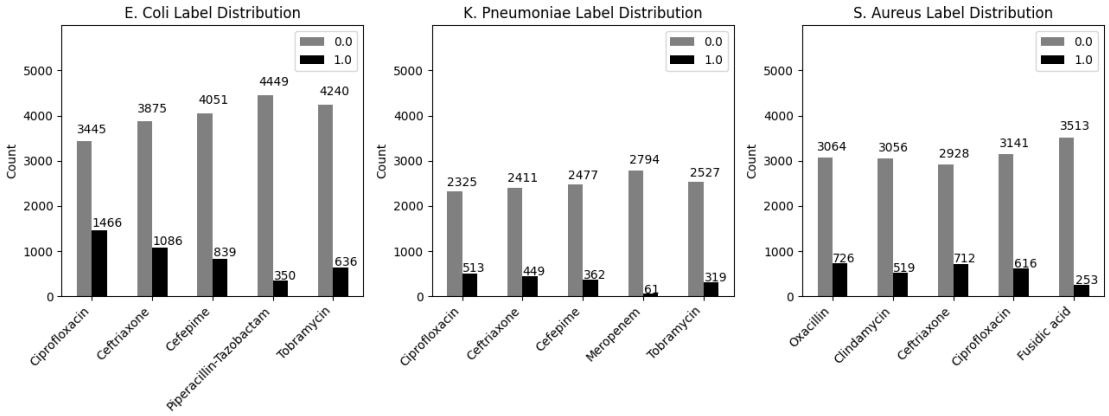

Figure 3.2: Distribution of susceptible (0) and resistant (1) strains across various antibiotics for *E. coli*, *K. pneumoniae*, and *S. aureus* in the DRIAMS-A dataset. These figures correspond to the data shown in Table 3.1.

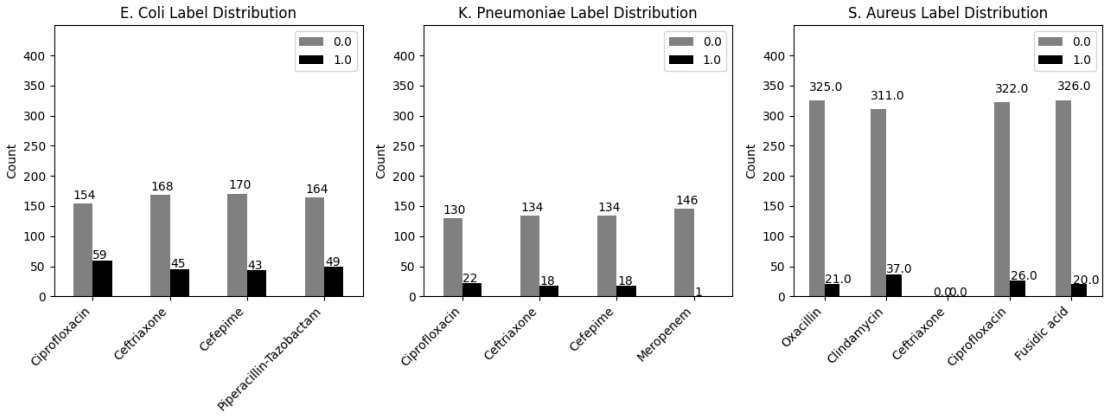

Figure 3.3: Distribution of susceptible (0) and resistant (1) strains across various antibiotics for *E. coli*, *K. pneumoniae*, and *S. aureus* in the DRIAMS-B dataset. These figures correspond to the data shown in Table 3.2.

3.2 Approaches

In this thesis, three different approaches were implemented to predict antibiotic resistance. These approaches were compared with the baseline results of López-Cortés et al. (2024). The approaches are as follows::

3.2.1 Multi-label Classification

Objective: to develop separate models for each bacterial species that can simultaneously predict resistance across multiple antibiotics.

In this approach, a separate model is trained for each bacterial species *E. coli*, *K. pneumoniae*, and *S. aureus*. This is a multi-label classification problem where each label represents a different antibiotic. The advantage of this approach is that it is able to predict resistance to multiple antibiotics for a single species.

3.2.2 Multi Bacteria Classification

Objective: to train a single model per antibiotic across multiple bacterial species.

Contrary to the Multi-label Classification approach, the Multi Bacteria approach focuses on creating individual models for each antibiotic, rather than each bacterial species. Here, a single model is responsible for predicting resistance to one antibiotic, but across multiple species, including *E. coli*, *K. pneumoniae*, and *S. aureus*. This allows us to train a model per antibiotic that can be potentially generalized across species.

Importantly, Table 3.3 below outlines the antibiotics that were used to implement the current approach, as not all antibiotics are present in more than one bacterial species. For example, Piperacillin-Tazobactam is only available in *E. coli* strains, hence we can not implement Multi Bacteria approach with the given antibiotic, this will simply lead to a baseline result with one bacteria-antibiotic approach.

The species information was one-hot encoded, which allows us to exploit species-specific information during a 10-fold cross-validation process, training, and testing. This also promotes the calculation of performance metrics per species thus ensuring a robust evaluation of the model.

Table 3.3: Antibiotics Across Bacterial Species

Bacteria	Applicable Antibiotics	Single Antibiotics
E. coli	Ciprofloxacin, Ceftriaxone, Cefepime, Tobramycin	Piperacillin-Tazobactam
K. pneumoniae	Ciprofloxacin, Ceftriaxone, Cefepime, Tobramycin	Meropenem
S. aureus	Ciprofloxacin, Ceftriaxone	Oxacillin, Fusidic acid, Clindamycin

3.2.3 Multi-label Multi Bacteria

Objective: to develop a comprehensive single model that can simultaneously predict resistance across multiple antibiotics and bacterial species.

The Multi-label Multi Bacteria approach represents a more complex classification task, where a single model is trained to predict antibiotic resistance across multiple species and antibiotics. This approach combines the shared information across species and antibiotics to improve prediction accuracy.

This approach utilizes self-labeling techniques and a masking technique to handle incomplete labels. This is particularly important given that not all antibiotics are present and relevant for every bacterial species (as shown in Table 3.4). Self-labeling involves predicting incomplete labels during training to enhance the model's ability to learn from incomplete data.

Table 3.4: Available Antibiotics Across Bacterial Species

Bacteria	Available Antibiotics	Absent Antibiotics
E. coli	Ciprofloxacin, Ceftriaxone, Cefepime Piperacillin-Tazobactam, Tobramycin	Meropenem, Oxacillin, Fusidic acid, Clindamycin
K. pneumoniae	Ciprofloxacin, Ceftriaxone, Cefepime, Meropenem, Tobramycin	Piperacillin-Tazobactam, Oxacillin, Fusidic acid, Clindamycin
S. aureus	Ciprofloxacin, Ceftriaxone Oxacillin, Fusidic acid, Clindamycin	Cefepime, Piperacillin-Tazobactam, Tobramycin, Meropenem

Species encoding has also been implemented in this approach. Similar to the Multi Bacteria Approach, it allows the model to differentiate between species. Species encoding, masking technique, and self-labeling approach are discussed further in detail in the next section.

This approach combines two previous ones. It develops a single model that can handle multiple tasks. This potentially reduces the computational load and time compared to training separate models for each species or antibiotic. However, the increased complexity of the model requires careful implementation and validation.

3.3 Data Preparation

In order to ensure robust and smooth training of the model, the following steps were implemented during the data preprocessing steps to address specific requirements for each of the approaches. Table 3.5 gives an overview of the data-preprocessing steps applied to each of the approaches.

Table 3.5: Summary of Preprocessing Steps by Approach

Preprocessing Step	**Multilabel Classification**	**Multi Bacteria Problem**	**Multilabel Multi Bacteria**
Unnecessary Columns/Rows	+	+	+
Normalization	+	+	+
Species Encoding	N/A	+	+
Masking-technique	N/A	N/A	+
Self-Labeling	N/A	N/A	+

1. **Unnecessary columns** that do not contribute to prediction were removed in approaches. As such irrelevant information has been removed keeping the features of mass spectra peaks intensity, species information, and labels.

2. **Normalization**

   Features were scaled using normalization to ensure comparability and consistency. The maximum normalization technique is applied due to the nature of the raw mass spectra data (sharp peaks), i.e. each sample's maximum is scaled to 1.

3. **Species encoding**

   Bacterial species information was one-hot encoded as binary columns to allow the model to differentiate between species for the Multi Bacteria Problem and the Multi-label Multi Bacteria Approach. For the Multi-label Classification approach, species encoding was not necessary, since it deals with a single species at a time.

   It is important to note that features were first normalized and then only concatenated with the one-hot encoded species information. This ensures that in our training data, the last three columns are species information distinct from mass spectra peak values. Later,

during the cross-validation, training, and testing processes, it allows us to evaluate the model's performance individually for each species-antibiotic combination.

4. **Incomplete Labels**

   The incomplete labels for antibiotic resistance status is present in all three datasets of *E. coli, K. pneumonia* and *S. aureus*. To deal with incomplete labels, several approaches have been applied:

   - **Removed** for Multi-label Classification and Multi Bacteria Classification.
   - **Masked Binary Crossentropy Loss** function applied on Multilabel Multi Bacteria Approach.

     Since the Multi-label Multi Bacteria approach combines all datasets and not all antibiotics are present in each bacterial species, we can not simply delete incomplete labels, as it leads to an empty dataset. Therefore, we applied the masking technique.

     First, incomplete labels were identified, then they were set to the value of -1 (distinct from 0 and 1). Then by applying the custom function, we create a mask where all -1 values are not considered. The mask is applied to both the true and predicted labels. Therefore, the error between the target and predicted values for the missing value indices will be zero in the masked loss defined below (Tharmakulasingam et al., 2022) and will not affect the masked binary crossentropy value. Finally, the binary crossentropy loss function is calculated only for the masked values, this way we eliminate the model's bias by incomplete entries.

$$Mask(m_{i,j}) = \begin{cases} 0 & if y(i,j) == -1 \\ 1 & if y(i,j) \neq -1 \end{cases}$$

$$Maskedloss = -\frac{1}{n} \sum_{j \in n} \left(my_{i,j} \log\left(\hat{y}_{i,j}\right) + \left(1 - my_{i,j}\right) \log\left(1 - \hat{y}_{i,j}\right) \right)$$

   - **Iterative self-labeling** approach applied complementary to the custom masked loss function on Multi-label Multi Bacteria approach and evaluated separately. This is a semi-supervised learning technique that allows the model to assign labels to unlabelled data points during iterative training.

The training dataset is further split into training and validation sets using stratification. The pre-trained model on Multi-label Multi Bacteria has been loaded and applied in the self-labeling loop. Then, the model predicts the labels for incomplete labels, where we apply a confidence threshold of 0.70. This implies that only if the confidence is above the set threshold, the incomplete label (-1) is replaced by the predicted label (0 or 1). Finally, the labels are updated for the next iteration and the model is retrained.

During the iterative self-training process we exploit early stopping, learning rate reduction, and model checkpoint callbacks. This way we save the best-performed model only (based on validation loss) to employ further on the test set.

3.3.1 Model architecture

The deep learning model applied is a convolutional neural network (CNN) designed to classify antibiotic resistance in bacteria, namely MSDeepAMR López-Cortés et al. (2024). The model consists of the layers discussed in Section 2.4. but with a distinct output layer for different approaches:

- Multi-label Classification: Dense layer with 5 neurons;
- Multi Bacteria Classification: Dense layer with 1 neuron;
- Multi-label Multi Bacteria: Dense layer with 9 neurons.

3.4 Model training

Initially, the model is split into training and test sets with a ratio of 80/20 with stratification to maintain the label distribution.

Next, we apply a 10-fold stratified cross-validation strategy applied to the training set to ensure the robustness and generalizability of the model. In each fold, the data were split into training and validation sets with a ratio of 80/20, applying stratified sampling to maintain the same class distribution across folds. This way we get a training set, validation set, and separate unseen test set.

The model was trained using the Adam optimizer with a learning rate of 0.0001, a batch size of 32, and a binary cross-entropy loss function for Multi-label Classification and Multi Bacteria

Classification. Whereas a custom masked binary crossentropy loss function for the Multi-label Multi Bacteria approach as it was discussed above. The training process also includes the following steps:

- Early stopping - to prevent overfitting, the patience parameter was set to 4. Training is stopped if the validation loss does not improve for four consecutive epochs.
- Learning rate - reduced by 0.1 if the validation loss plateau for two epochs.

3.5 Evalutaion metrics

The performance of ML techniques (when applied in AMR research) is most often measured with the area under receiver operating curve (AUROC), especially in binary classification problems of biological nature class (Chicco, 2017). The calculation of AUROC involves computing the area under the ROC curve, which represents the true positive rate or "recall" versus the false positive rate. This metric measures the model's discriminative ability, where a value of AUROC equal to 1 indicates a perfect model. Whereas a value of 0.5 indicates performance similar to random guessing. Given the nature of the dataset which is imbalanced, the balanced accuracy has also been chosen as a metric. Balanced accuracy is the average between sensitivity and specificity. It is useful when the dataset is imbalanced, i.e. one of the values occurs far less frequently than the other (Brodersen et al., 2010). The area under the precision-recall curve (AUPRC) is sometimes used as an alternatively to AUROC but based on precision and recall, focusing on correctly classified positive values (minority class). AUPRC has been assumed to perform better with imbalanced data for binary classification problem (McDermott et al., 2024). Hence, the model's performance was assessed with the following metrics: balanced accuracy, Area Under the Precision-Recall Curve (AUPRC), and Area Under the Receiver Operating Characteristic Curve (AUROC).

CHAPTER 4

RESULTS

4.1 Introduction

In this chapter, the results of the three different models are discussed. First, the findings of the DRIAMS-A dataset are discussed. Next, the outcomes of the transfer learning on the DRIAMS-B dataset are examined. Finally, the self-labeling approach has been evaluated on DRIAMS-A dataset for the Multi-label Multi Bacteria approach.

The baseline results (single model per bacteria per antibiotic) serve as a reference point, i.e. the models before application of the multi-label classification (MLC).

In all three approaches, we applied 10-fold cross-validation and tested on a separate test set. The combined results of all approaches with 10-fold cross-validation are summarized in Tables 4.4, 4.5, 4.6.

4.2 Multi-label Classification per Bacteria

This approach deploys separate models for each bacteria species across multiple antibiotics. The comparison between the baseline results and the MLC per Bacteria is presented in Table 4.1 across various metrics: Balanced Accuracy (B. Acc), AUROC (Area Under the Receiver Operating Characteristic Curve), and AUPRC (Area Under the Precision-Recall Curve).

The balanced accuracy shows a slight improvement or remains consistent with baseline results. For example, Ceftriaxone for *E.coli* improved from 0.80 to 0.82. The standard deviation is generally similar to the baseline indicating the improvement is consistent across different folds.

The AUROC values are generally higher for the MLC per Bacteria approach, indicating better discriminative ability compared to the baseline. For example, Ceftriaxone for *E. coli* improved from 0.87 to 0.91.

Regarding AUPRC scores in Table 4.1 the model performs better for almost all bacteria-antibiotic combinations except for Ciprofloxacin antibiotic across all bacteria.

Overall, this approach either improves or maintains the performance when compared to the baseline. The improvements are generally consistent, with relatively small increases in variability that might reflect challenges associated with class imbalance.

Bacteria	Antibiotic	Baseline Results			MLC per Bacteria		
		B. Acc	AUROC	AUPRC	B. Acc	AUROC	AUPRC
E. coli	Ciprofloxacin	0.74 ± 0.01	0.85 ± 0.03	0.75 ± 0.03	0.74 ± 0.02	0.84 ± 0.03	0.73 ± 0.03
	Ceftriaxone	0.80 ± 0.01	0.87 ± 0.03	0.79 ± 0.03	**0.82 ± 0.04**	**0.91 ± 0.03**	**0.82 ± 0.04**
	Cefepime	0.78 ± 0.02	0.88 ± 0.02	0.70 ± 0.03	**0.79 ± 0.04**	**0.90 ± 0.03**	**0.73 ± 0.04**
	Piperacillin-T.	0.51 ± 0.04	0.64 ± 0.04	0.14 ± 0.05	**0.52 ± 0.02**	**0.73 ± 0.06**	**0.23 ± 0.09**
	Tobramycin	0.55 ± 0.03	0.76 ± 0.02	0.30 ± 0.04	**0.56 ± 0.02**	0.75 ± 0.04	**0.35 ± 0.07**
K. pneumoniae	Ciprofloxacin	0.59 ± 0.03	0.76 ± 0.02	0.53 ± 0.03	**0.64 ± 0.03**	0.76 ± 0.05	0.50 ± 0.06
	Ceftriaxone	0.76 ± 0.02	0.82 ± 0.01	0.68 ± 0.02	0.76 ± 0.05	**0.87 ± 0.04**	**0.74 ± 0.06**
	Cefepime	0.75 ± 0.01	0.83 ± 0.01	0.60 ± 0.03	0.72 ± 0.05	**0.86 ± 0.04**	**0.64 ± 0.06**
	Meropenem	0.55 ± 0.04	0.83 ± 0.03	0.20 ± 0.05	0.52 ± 0.05	**0.92 ± 0.04**	**0.30 ± 0.14**
	Tobramycin	0.64 ± 0.02	0.83 ± 0.03	0.54 ± 0.02	0.64 ± 0.03	0.81 ± 0.07	0.49 ± 0.06
S. aureus	Ciprofloxacin	0.75 ± 0.01	0.85 ± 0.02	0.70 ± 0.02	0.65 ± 0.05	0.81 ± 0.04	0.59 ± 0.08
	Fusidic acid	0.48 ± 0.04	0.68 ± 0.03	0.10 ± 0.06	**0.51 ± 0.01**	**0.71 ± 0.09**	**0.21 ± 0.06**
	Oxacillin	0.87 ± 0.01	0.93 ± 0.02	0.85 ± 0.01	0.83 ± 0.03	0.92 ± 0.01	0.85 ± 0.02
	Ceftriaxone	-	-	-	0.83 ± 0.03	0.93 ± 0.01	0.86 ± 0.02
	Clindamycin	-	-	-	0.57 ± 0.03	0.75 ± 0.04	0.42 ± 0.06

Table 4.1: Comparison of Baseline Results and MLC per Bacteria

4.2.1 *E. coli*

The balanced accuracy score in 10-fold cross-validation is slightly higher or similar to the baseline results (Table 4.1). Moreover, in most cases, AUROC scores show improved performance compared to baseline results. Specifically, results with Ceftriaxone, Cefepime, and Piperacillin-T. (Figure 4.1). Regarding AUPRC scores, it demonstrates higher results for all antibiotics except for Ciprofloxacin which is slightly lower than baseline. The AUROC curve indicates the model's ability to distinguish between susceptible and resistant classes. AUPRC summarizes the ability to identify the resistant class in the presence of imbalanced data. The confusion matrix results are presented in Figure 4.2. Notably, the model makes almost no correct resistance (1) predictions for Piperacillin-T. and Tobramycin with the lowest number of resistant labels.

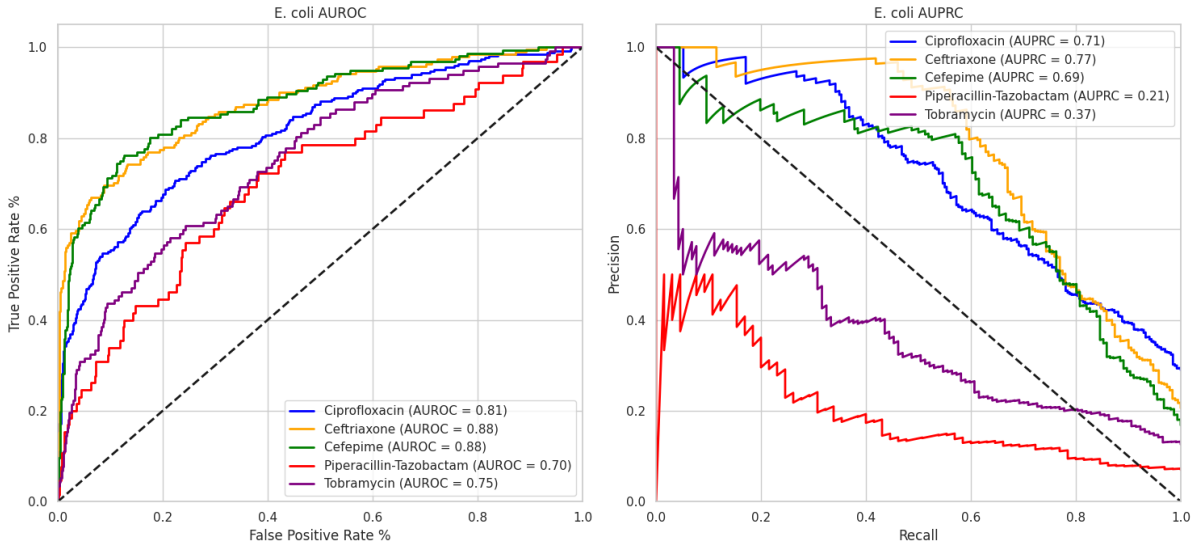

Figure 4.1: AUROC and AUPRC curves for the Test set predictions of *E. coli*.

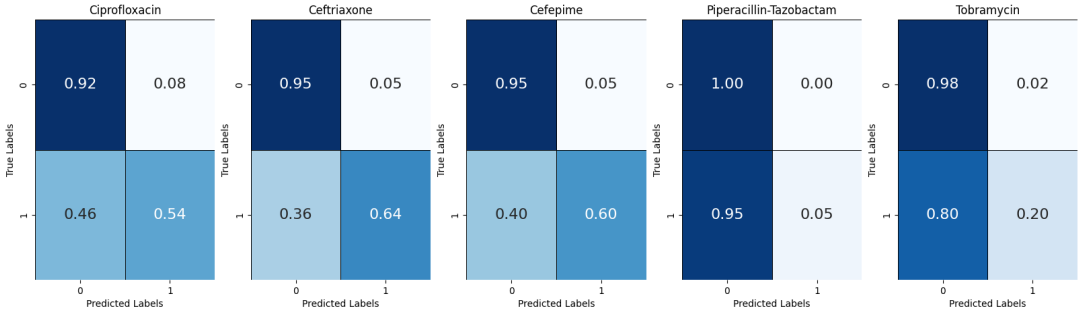

Figure 4.2: Normalized Confusion Matrices for Predicting Antibiotic Resistance in *E. coli* across Ciprofloxacin, Ceftriaxone, Cefepime, Piperacillin-T., Tobramycin. The model struggles with correctly identifying resistance especially Piperacillin-T. and Tobramycin.

4.2.2 *K. pneumoniae*

Similarly to *E. coli*, the balanced accuracy score for *K. pneumoniae* is almost the same as the baseline results (Table 4.1). Likewise, in most cases, AUROC scores show improved performance compared to baseline results. Both AUROC and AUPRC scores for Ciprofloxacin and Cefepime results exceed baseline results. The confusion matrix results are shown in Figure 4.4. Notably, the model makes almost no correct resistance (1) predictions for Meropenem and Tobramycin which have the least amount of resistant samples (low AUPRC).

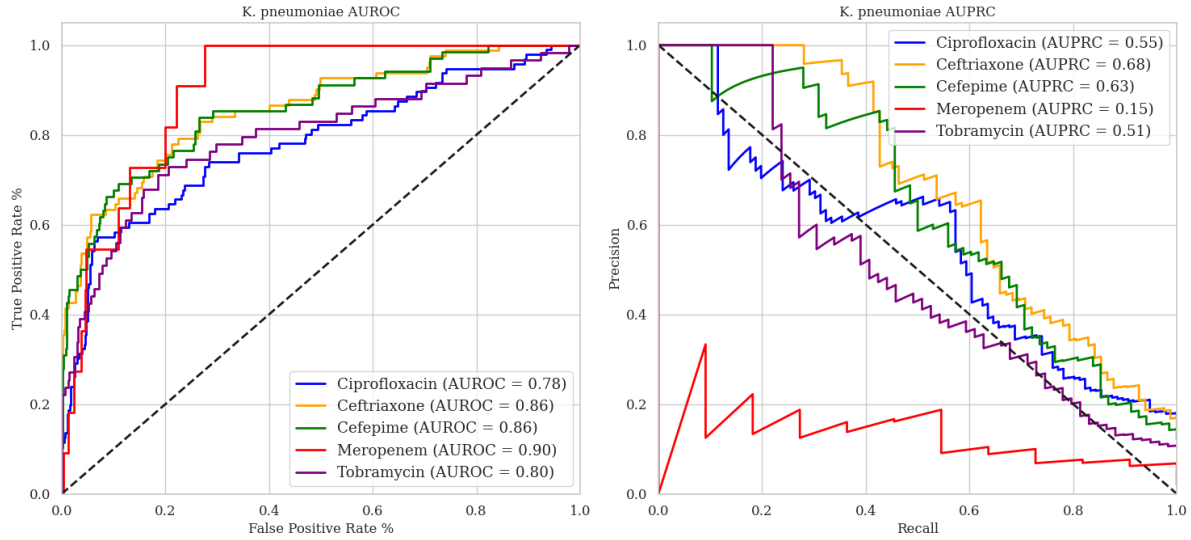

Figure 4.3: AUROC and AUPRC curves for the Test set predictions of *K. pneumoniae*.

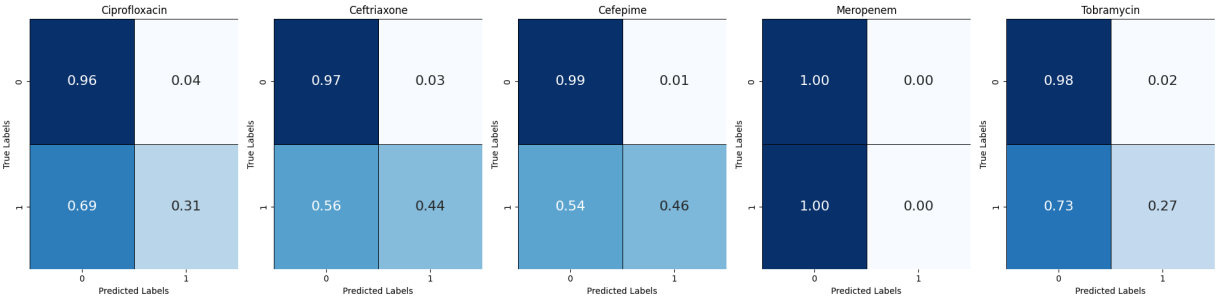

Figure 4.4: Normalized Confusion Matrices for Predicting Antibiotic Resistance in *K. pneumoniae* across Ciprofloxacin, Ceftriaxone, Cefepime, Meropenem, Tobramycin. The model struggles with correctly identifying resistance especially with Meropenem and Tobramycin.

4.2.3 *S. aureus*

The balanced accuracy score for *K. pneumoniae* is slightly lower than the baseline results (Table 4.1). For Ciprofloxacin AUROC and AUPRC scores are considerably lower, for Oxacillin they are similar to baseline results and for Fusidic acid the scores are higher (Table 4.1). The confusion matrix results in Figure 4.6 demonstrate that the model makes almost no correct predictions for Fusidic acid and Clindamycin which have the least amount of resistant samples.

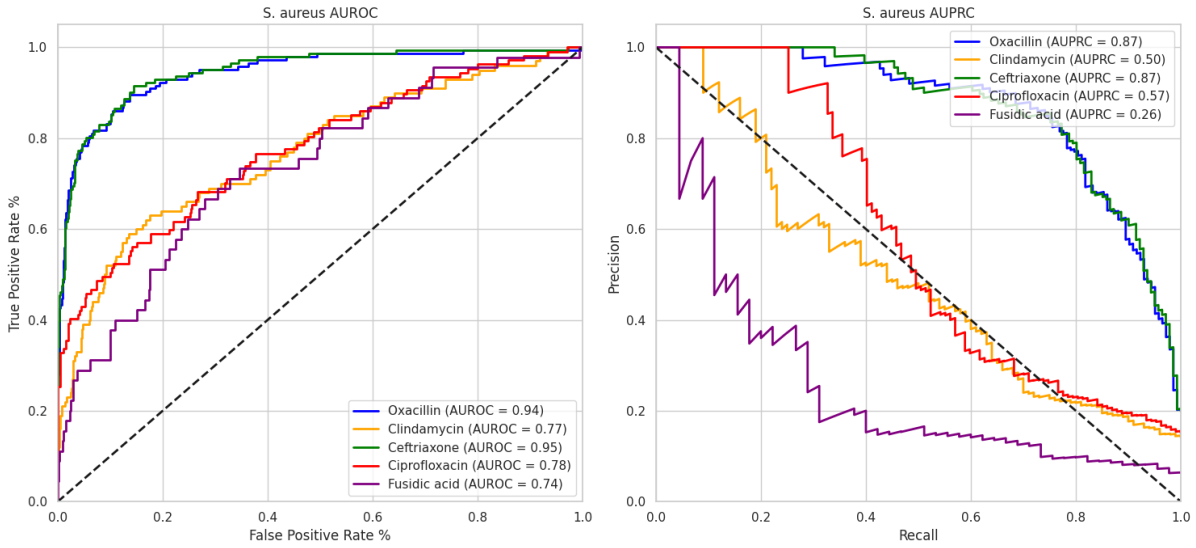

Figure 4.5: AUROC and AUPRC curves for the Test predictions of *S. aureus*.

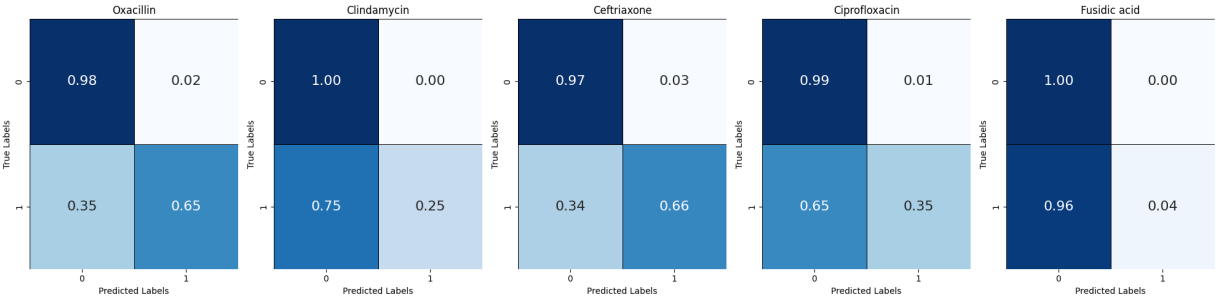

Figure 4.6: Normalized Confusion Matrices for Predicting Antibiotic Resistance in *K. pneumoniae* across Oxacillin, Clindamycin, Ceftriaxone, Ciprofloxacin, Fusidic acid. The model struggles with correctly identifying resistance especially with Fusidic acid and Clindamycin.

4.3 Multi Bacteria Classification per Antibiotic

This approach creates a single model per antibiotic but across different bacteria species. To differentiate between species, species information was one-hot encoded. It shows either similar or slightly lower balanced accuracy compared to the baseline (Table 4.2). For example, in the case of *E. coli* with Ciprofloxacin, the balanced accuracy drops slightly from 0.74 ± 0.01 to 0.71 ± 0.03. Both AUROC and AUPRC generally improve or stay the same. For example, in *K. pneumoniae* with Ceftriaxone, the AUROC increases from 0.82 ± 0.01 to 0.93 ± 0.04 and AUPRC increases from 0.68 ± 0.02 to 0.87 ± 0.06.

Overall, the multi-bacteria approach per antibiotic tends to perform better in terms of AUROC and AUPRC, potentially better handling the class imbalance for certain bacteria-antibiotic com-

binations. The model performs best for Ceftriaxone and Cefepime antibiotics. The best test results obtained from *S. aureus* resistance to Ceftriaxone with AUROC = 0.98 and AUPRC = 0.97 as shown in Figure 4.9.

Antibiotic	Bacteria	Baseline Results			Multi Bacteria per Antibiotic		
		B. Acc	AUROC	AUPRC	B. Acc	AUROC	AUPRC
Ciprofloxacin	*E. coli*	0.74 ± 0.01	0.85 ± 0.03	0.75 ± 0.03	0.71 ± 0.03	0.82 ± 0.01	0.68 ± 0.04
	K. pneumoniae	0.59 ± 0.03	0.76 ± 0.02	0.53 ± 0.03	**0.69 ± 0.03**	**0.83 ± 0.06**	**0.66 ± 0.10**
	S. aureus	0.75 ± 0.01	0.85 ± 0.02	0.70 ± 0.02	0.72 ± 0.03	0.84 ± 0.03	0.69 ± 0.04
Ceftriaxone	*E. coli*	0.80 ± 0.01	0.87 ± 0.03	0.79 ± 0.03	**0.80 ± 0.03**	**0.90 ± 0.01**	**0.81 ± 0.03**
	K. pneumoniae	0.76 ± 0.02	0.82 ± 0.01	0.68 ± 0.02	**0.84 ± 0.05**	**0.93 ± 0.04**	**0.87 ± 0.06**
	S. aureus	-	-	-	0.92 ± 0.01	0.84 ± 0.03	0.80 ± 0.03
Cefepime	*E. coli*	0.78 ± 0.02	0.88 ± 0.02	0.70 ± 0.03	0.76 ± 0.04	0.87 ± 0.03	0.69 ± 0.05
	K. pneumoniae	0.75 ± 0.01	0.83 ± 0.01	0.60 ± 0.03	**0.74 ± 0.03**	**0.86 ± 0.07**	**0.65 ± 0.09**
Tobramycin	*E. coli*	0.55 ± 0.03	0.76 ± 0.02	0.30 ± 0.04	**0.60 ± 0.05**	**0.77 ± 0.03**	**0.40 ± 0.17**
	K. pneumoniae	0.64 ± 0.02	0.83 ± 0.03	0.54 ± 0.02	0.60 ± 0.05	0.76 ± 0.07	0.40 ± 0.17

Table 4.2: Comparison of Baseline Results and Multi Bacteria Approach

4.3.1 Ciprofloxacin

The performance is highest for *Escherichia coli* (AUROC = 0.84, AUPRC = 0.76), followed by *S. aureus* (AUROC = 0.83, AUPRC = 0.63), with the lowest performance observed for *K. pneumoniae* (AUROC = 0.76, AUPRC = 0.47). In contrast to the 10-fold cross-validation results (Table 4.2), AUPRC score in the test for *K. pneumoniae* is considerably lower.

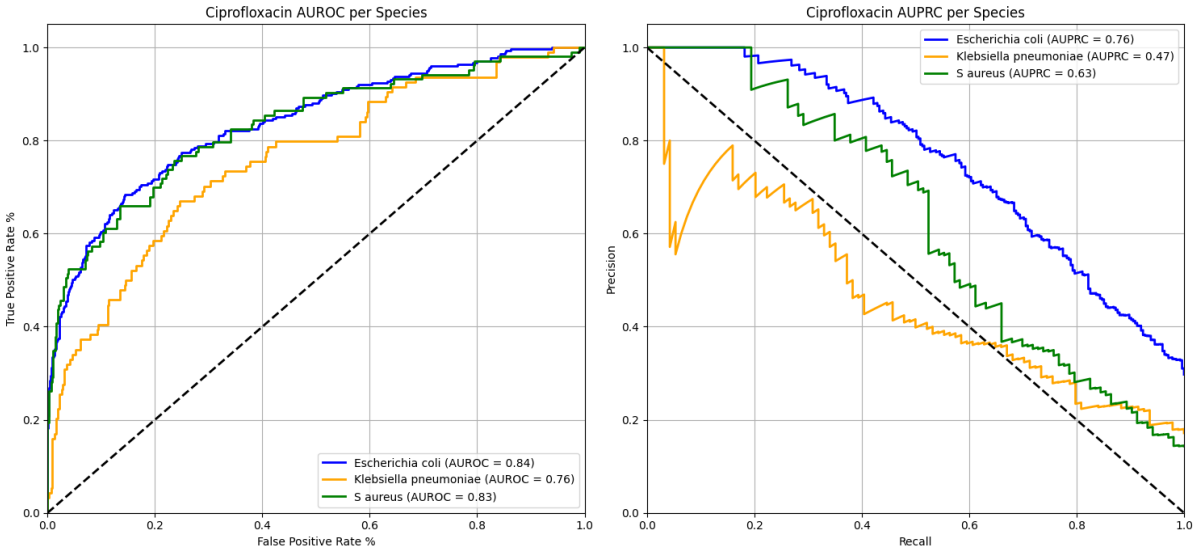

Figure 4.7: Ciprofloxacin AUROC and AUPRC curves for the Test predictions of *E.coli, K. pneumonia, S. aureus*.

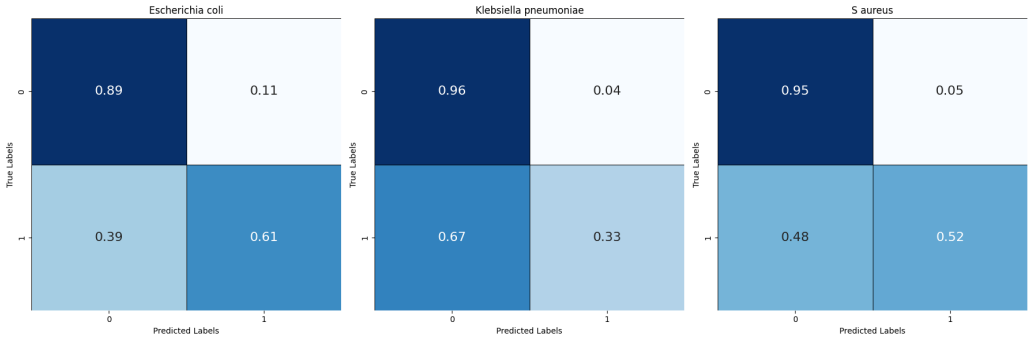

Figure 4.8: Normalized Confusion Matrices for Predicting Antibiotic Resistance across Ciprofloxacin in *E. coli, K. pneumonia, S. aureus*. The model struggles with correctly identifying resistance, especially in K. pneumonia strains.

4.3.2 Ceftriaxone

The performance is highest for *S. aureus* (AUROC = 0.98, AUPRC = 0.97), followed by *K. pneumoniae* (AUROC = 0.92, AUPRC = 0.85), with the lowest performance observed for *E. coli* (AUROC = 0.91, AUPRC = 0.83) as shown in Figure 4.9. Compared to the baseline results, balanced accuracy and both AUROC and AUPRC for all bacteria *E. coli, K. pneumoniae, S. aureus* are considerably higher (Table 4.1). Confusion matrices in Figure 4.10 show considerable improvements in the model's performance in terms of sensitivity (recall) compared to previous models, especially for *S. aureus*.

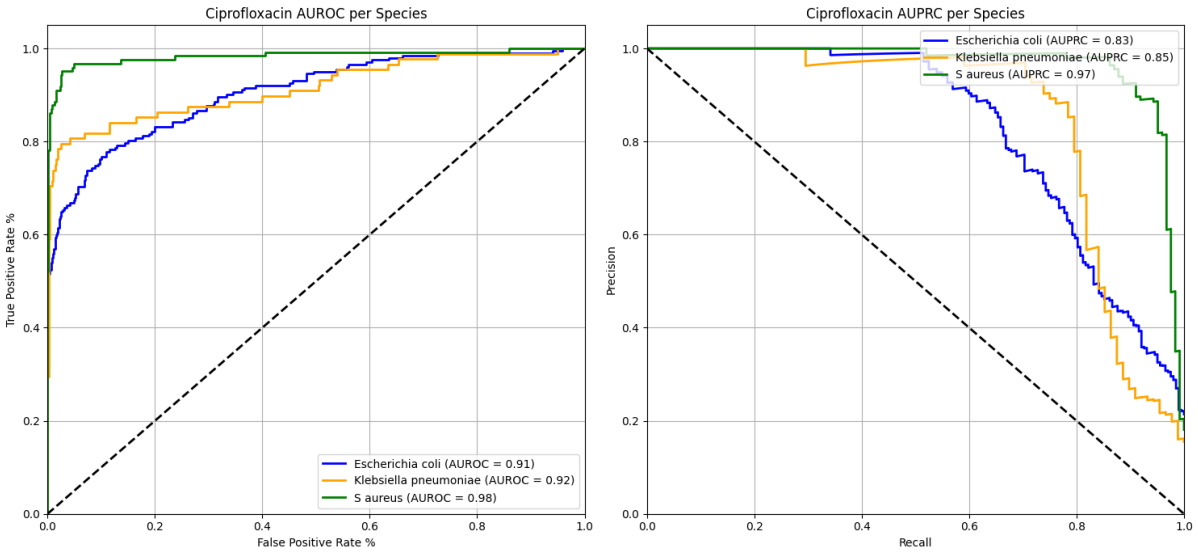

Figure 4.9: Ceftriaxone AUROC and AUPRC curves for the Test predictions of *E.coli, K. pneumonia, S. aureus*.

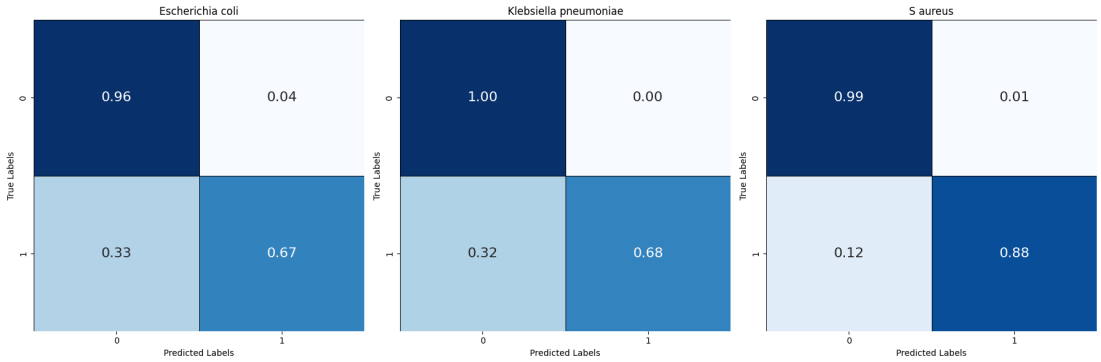

Figure 4.10: Normalized Confusion Matrices for Predicting Antibiotic Resistance across Ceftriaxone in *E. coli, K. pneumonia, S. aureus*. The model struggles the most with correctly identifying resistance in *E. coli, K. pneumoniae* strains

4.3.3 Cefepime

The performance is nearly identical for both *E. coli* and *K. pneumoniae* (AUROC = 0.86, AUPRC = 0.67 and 0.64 respectively). Compared to the baseline results, both AUROC and AUPRC scores for both bacteria *E.coli, K. pneumoaniae* are similar to the baseline. Confusion matrices in Figure 4.14 show the model's difficulty in predicting resistant cases.

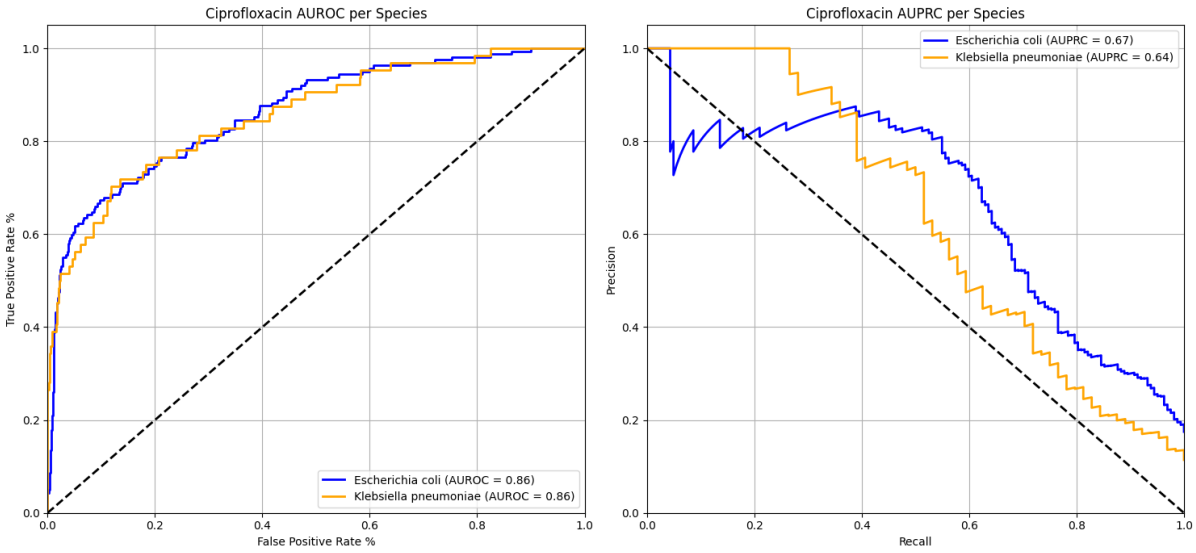

Figure 4.11: Cefepime AUROC and AUPRC curves for the Test predictions of *E.coli, K. pneumonia, S. aureus*.

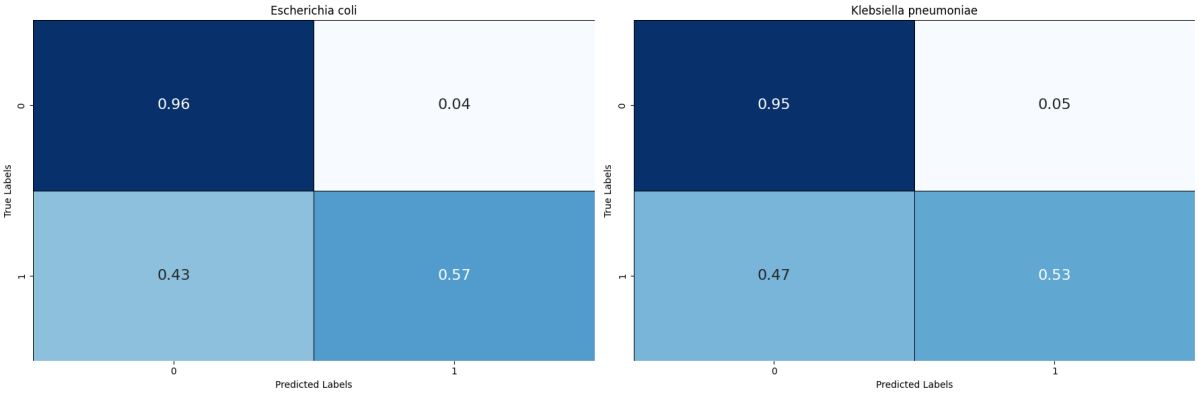

Figure 4.12: Normalized Confusion Matrices for Predicting Antibiotic Resistance across Cefepime in *E. coli, K. pneumonia*. The model struggles with correctly identifying resistance especially in K. pneumonia strains.

4.3.4 Tobramycin

The performance is better for *K. pneumoniae* with AUROC = 0.81 and AUPRC = 0.43, whereas for *E. coli* AUROC = 0.71 and AUPRC = 0.43. Compared to the baseline results, AUROC and AUPRC scores for both species are lower. Confusion matrices in Figure 4.14 show the model's difficulty in predicting resistant cases.

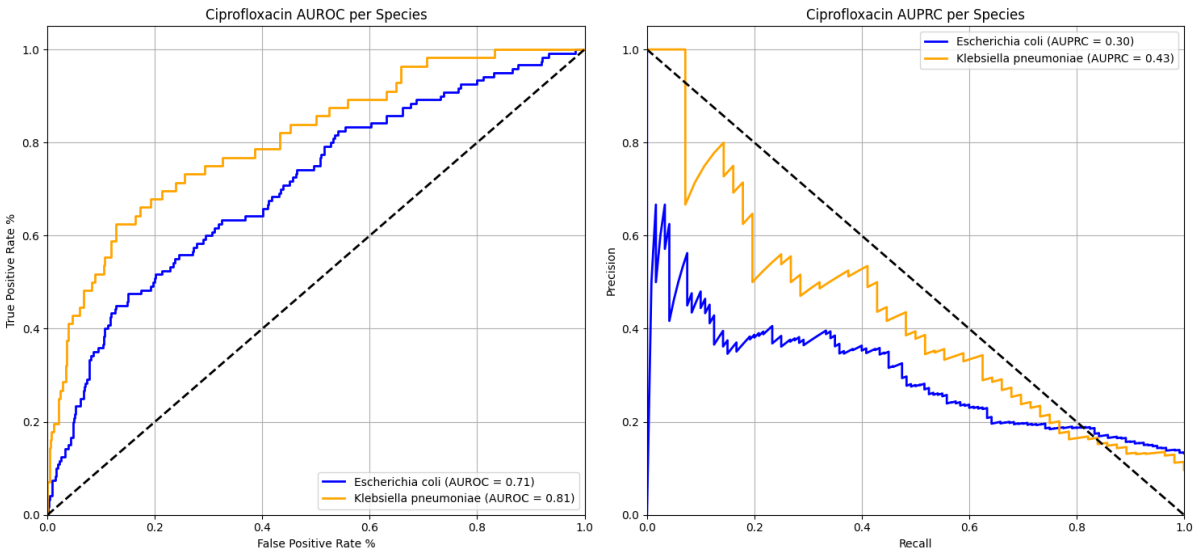

Figure 4.13: Tobramycin AUROC and AUPRC curves for the Test predictions of *E.coli, K. pneumonia, S. aureus*.

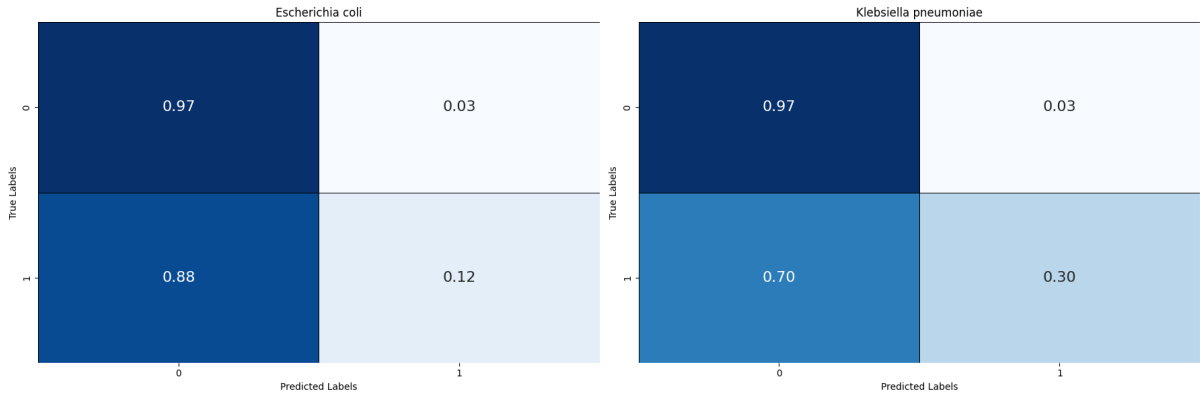

Figure 4.14: Normalized Confusion Matrices for Predicting Antibiotic Resistance across Tobramycin in *E. coli, K. pneumonia*. The model struggles with correctly identifying resistance classes in both bacteria.

4.4 Multi-label Multi Bacteria Classification

This model aims to predict resistance simultaneously across multiple bacteria and antibiotics. Some bacteria-antibiotic combinations demonstrate improved performance, while most of them show similar outcomes as shown in Table 4.3 except for Ciprofloxacin in *S. aureus*. The main challenge that has been encountered along with class imbalance is incomplete labels. Combining all three datasets of bacteria species results in a dataset where every row and column has at least one incomplete label. Hence, simply removing them would have resulted in an empty dataset. To address this problem two main techniques have been applied: masking and self-labeling. In this section results from the masking technique are discussed.

Bacteria	**Antibiotic**	**Baseline Results**			**MLC Multibacteria**		
		B. Acc	**AUROC**	**AUPRC**	**B. Acc**	**AUROC**	**AUPRC**
E. coli	Ciprofloxacin	0.74 ± 0.01	0.85 ± 0.03	0.75 ± 0.03	**0.75 ± 0.03**	0.85 ± 0.02	**0.76 ± 0.03**
	Ceftriaxone	0.80 ± 0.01	0.87 ± 0.03	0.79 ± 0.03	**0.80 ± 0.02**	**0.90 ± 0.02**	**0.82 ± 0.03**
	Cefepime	0.78 ± 0.02	0.88 ± 0.02	0.70 ± 0.03	0.77 ± 0.03	**0.90 ± 0.03**	**0.72 ± 0.07**
	Piperacillin-T.	0.51 ± 0.04	0.64 ± 0.04	0.14 ± 0.05	**0.51 ± 0.01**	**0.68 ± 0.01**	**0.20 ± 0.05**
	Tobramycin	0.55 ± 0.03	0.76 ± 0.02	0.30 ± 0.04	**0.55 ± 0.02**	**0.76 ± 0.02**	**0.35 ± 0.05**
K. pneumoniae	Ciprofloxacin	0.59 ± 0.03	0.76 ± 0.02	0.53 ± 0.03	**0.63 ± 0.05**	0.76 ± 0.05	0.50 ± 0.09
	Ceftriaxone	0.76 ± 0.02	0.82 ± 0.01	0.68 ± 0.02	0.74 ± 0.06	0.83 ± 0.06	0.68 ± 0.13
	Cefepime	0.75 ± 0.01	0.83 ± 0.01	0.60 ± 0.03	0.73 ± 0.05	0.84 ± 0.05	0.61 ± 0.12
	Meropenem	0.55 ± 0.04	0.83 ± 0.03	0.20 ± 0.05	0.52 ± 0.04	**0.88 ± 0.03**	**0.35 ± 0.16**
	Tobramycin	0.64 ± 0.02	0.83 ± 0.03	0.54 ± 0.02	0.61 ± 0.05	0.80 ± 0.05	0.46 ± 0.06
S. aureus	Ciprofloxacin	0.75 ± 0.01	0.85 ± 0.02	0.70 ± 0.02	0.67 ± 0.03	0.81 ± 0.03	0.59 ± 0.04
	Fusidic acid	0.48 ± 0.04	0.68 ± 0.03	0.10 ± 0.06	**0.50 ± 0.01**	**0.71 ± 0.06**	**0.18 ± 0.10**
	Oxacillin	0.87 ± 0.01	0.93 ± 0.02	0.85 ± 0.01	0.82 ± 0.02	0.92 ± 0.02	0.84 ± 0.03
	Ceftriaxone	-	-	-	0.82 ± 0.01	0.92 ± 0.03	0.85 ± 0.03
	Clindamycin	-	-	-	0.56 ± 0.02	0.70 ± 0.02	0.38 ± 0.07

Table 4.3: Comparison of Baseline Results and MLC Multi Bacteria

The AUROC and AUPRC scores visualised in Figure 4.15 provide insights into the model's performance across different bacterial species and antibiotics.

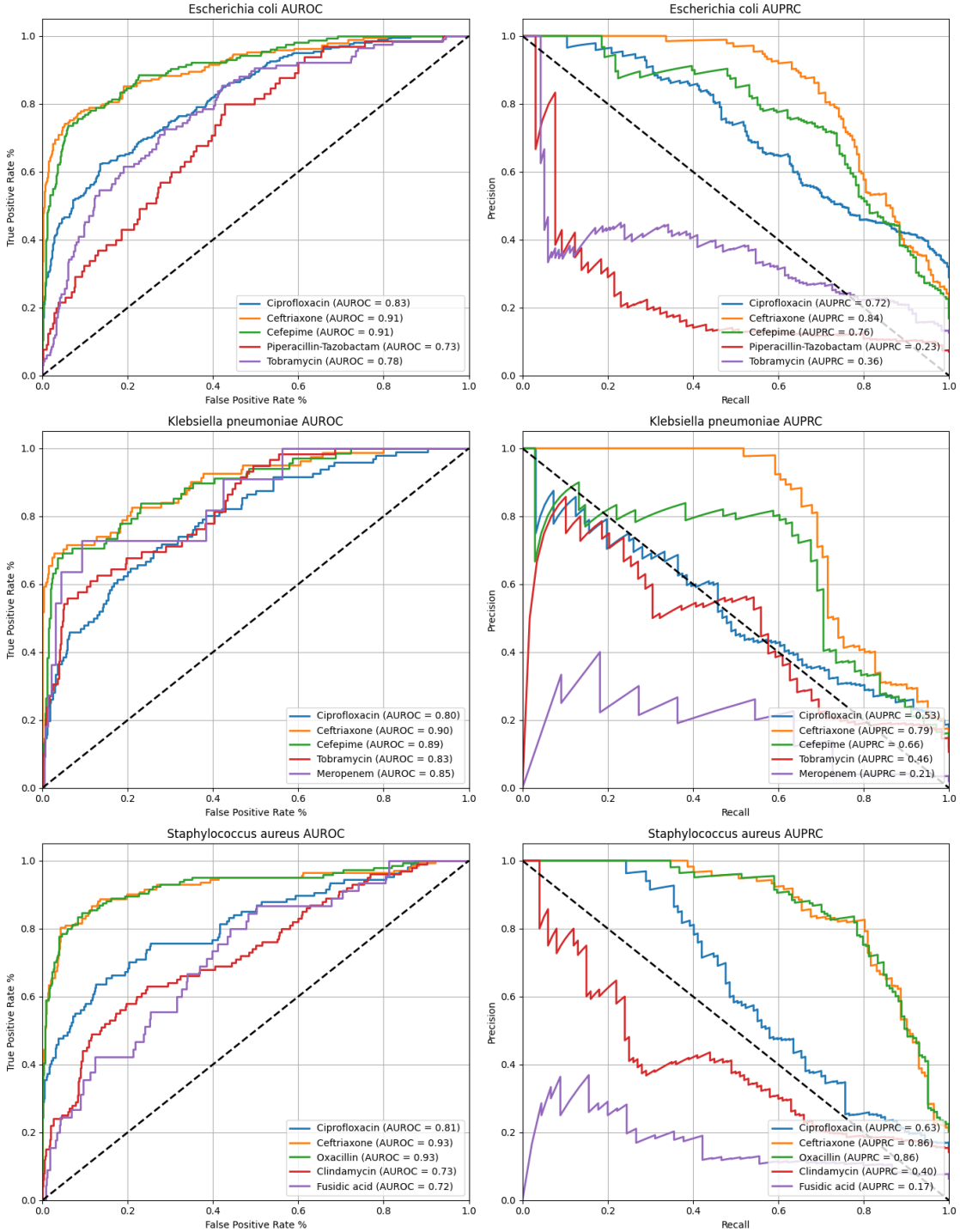

Figure 4.15: Multi-label Multi Bacteria approach: AUROC and AUPRC curves for the Test predictions of *E.coli, K. pneumonia, S. aureus*.

In general, compared to the baseline results, we observe considerable improvement in both AUROC and AUPRC scores. Specifically, for *E. coli* with Ceftriaxone, the AUROC increased

to 0.91 and the AUPRC to 0.84, while for *E. coli* with Cefepime, the AUROC also reached 0.91 and the AUPRC improved to 0.76. The baseline results for these combinations were lower, with AUROCs of 0.87 and 0.88, and AUPRCs of 0.79 and 0.76, respectively.

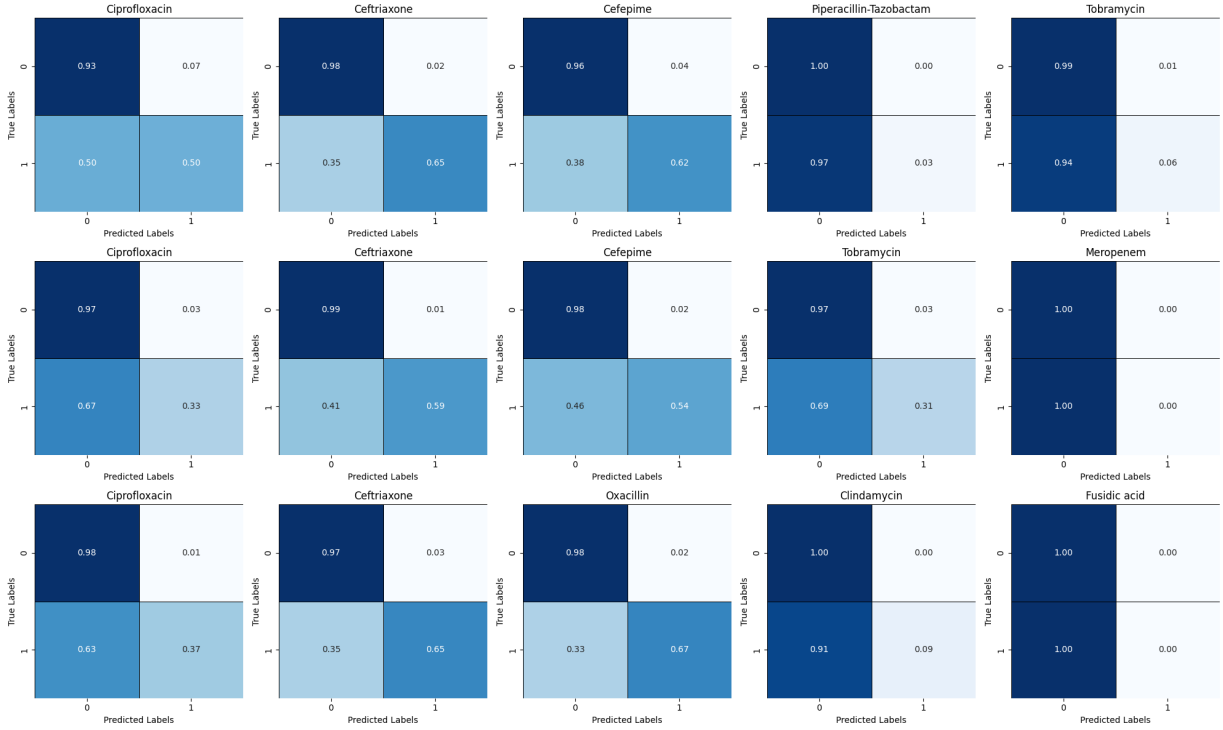

Figure 4.16: Normalized Confusion Matrices for Predicting Antibiotic Resistance across all bacteria-antibiotic combinations. Each row corresponds to species *E. coli*, *K.pneumoniae*, and *S. aureus*

Referring to confusion matrices in Figure 4.16, we see that the model struggles with correctly identifying resistance especially in *E. coli* strains for Piperacillin-T and Tobramycin, in *K. pneumonia* strains for Meropenem and Tobramycin, and in *S. aureus* strains for Clindamycin and Fucidic acid antibiotics.

4.5 Comparative Analysis

First, the 10-fold cross-validation has been applied to all datasets and approaches to provide insights into the model's stability and consistency across different folds. It is important to note that due to the class imbalances (i.e. higher number of susceptible labels than resistant), AUPRC scores are generally lower. Nevertheless, the standard deviation is low which shows that the results are stable. Next, the models were evaluated on the test set. This is a final evaluation of the model's performance.

Overall, it can be seen that there is no a single approach that considerably outperforms other approaches or baseline results as shown in Tables 4.4, 4.5, 4.6.

Nevertheless, it is important that for some species certain approaches perform better. As such:

E. coli: The multi-label classification per bacteria and multi-label multibacteria models tend to outperform others for Ceftriaxone and Cefepime.

K. pneumoniae: The multi bacteria per antibiotic and multi-label multi bacteria approaches show better results for Ceftriaxone.

S. aureus: The multi-label per bacteria and multi-label per antibiotic approaches perform better for Oxacillin.

Bacteria	Antibiotic	Baseline	MLC per Bacteria	MLC Multi Bacteria	MLC per Antibiotic
	Ciprofloxacin	0.74 ± 0.01	0.74 ± 0.02	0.75 ± 0.03	0.71 ± 0.03
	Ceftriaxone	0.80 ± 0.01	**0.82 ± 0.04**	0.80 ± 0.02	0.80 ± 0.03
E. coli	Cefepime	0.78 ± 0.02	**0.79 ± 0.04**	0.77 ± 0.03	0.76 ± 0.04
	Piperacillin-T.	0.51 ± 0.04	**0.52 ± 0.02**	0.51 ± 0.01	
	Tobramycin	0.55 ± 0.03	0.56 ± 0.02	0.55 ± 0.02	**0.60 ± 0.03**
	Ciprofloxacin	0.59 ± 0.03	0.64 ± 0.03	0.63 ± 0.05	**0.69 ± 0.03**
	Ceftriaxone	0.76 ± 0.02	0.76 ± 0.05	0.74 ± 0.06	**0.84 ± 0.05**
K. pneumoniae	Cefepime	**0.75 ± 0.01**	0.72 ± 0.05	0.73 ± 0.05	0.74 ± 0.08
	Meropenem	**0.55 ± 0.04**	0.51 ± 0.03	0.52 ± 0.04	
	Tobramycin	**0.64 ± 0.02**	0.64 ± 0.03	0.61 ± 0.05	0.60 ± 0.05
	Ciprofloxacin	**0.75 ± 0.01**	0.65 ± 0.05	0.67 ± 0.03	0.72 ± 0.03
	Fusidic acid	0.48 ± 0.04	**0.51 ± 0.01**	0.50 ± 0.01	
S. aureus	Oxacillin	**0.87 ± 0.01**	0.83 ± 0.03	0.82 ± 0.02	
	Ceftriaxone	-	**0.83 ± 0.03**	0.82 ± 0.01	0.80 ± 0.02
	Clindamycin	-	**0.57 ± 0.03**	0.56 ± 0.02	

Table 4.4: Balanced Accuracy Score combined for all models applied

Bacteria	Antibiotic	Baseline	MLC per Bacteria	MLC Multi Bacteria	Multi Bacteria
	Ciprofloxacin	0.85 ± 0.03	0.84 ± 0.03	**0.85 ± 0.02**	0.82 ± 0.01
	Ceftriaxone	0.87 ± 0.03	**0.91 ± 0.03**	0.90 ± 0.02	0.90 ± 0.01
E. coli	Cefepime	0.88 ± 0.02	**0.90 ± 0.03**	**0.90 ± 0.03**	0.87 ± 0.03
	Piperacillin-T.	0.64 ± 0.04	**0.73 ± 0.06**	0.68 ± 0.01	
	Tobramycin	0.76 ± 0.02	0.75 ± 0.04	0.76 ± 0.02	**0.77 ± 0.03**
	Ciprofloxacin	0.76 ± 0.02	0.76 ± 0.05	0.76 ± 0.05	**0.83 ± 0.06**
	Ceftriaxone	0.82 ± 0.01	0.87 ± 0.04	0.83 ± 0.06	**0.93 ± 0.04**
K. pneumoniae	Cefepime	0.83 ± 0.01	**0.86 ± 0.04**	0.84 ± 0.05	**0.86 ± 0.07**
	Meropenem	0.83 ± 0.03	**0.92 ± 0.04**	0.88 ± 0.03	
	Tobramycin	**0.83 ± 0.03**	0.81 ± 0.07	0.80 ± 0.05	0.76 ± 0.07
	Ciprofloxacin	**0.85 ± 0.02**	0.81 ± 0.04	0.81 ± 0.03	0.84 ± 0.03
	Fusidic acid	0.68 ± 0.03	**0.70 ± 0.00**	**0.71 ± 0.06**	
S. aureus	Oxacillin	**.93 ± 0.02**	0.92 ± 0.01	0.92 ± 0.02	
	Ceftriaxone	-	0.93 ± 0.01	0.92 ± 0.03	0.92 ± 0.01
	Clindamycin	-	0.75 ± 0.04	0.70 ± 0.02	

Table 4.5: AUROC Score combined for all models applied

Bacteria	Antibiotic	Baseline	MLC per Bacteria	MLC Multi Bacteria	Multi Bacteria
	Ciprofloxacin	0.75 ± 0.03	0.73 ± 0.03	**0.76 ± 0.03**	0.68 ± 0.04
	Ceftriaxone	0.79 ± 0.03	**0.82 ± 0.04**	**0.82 ± 0.03**	0.81 ± 0.03
E. coli	Cefepime	0.70 ± 0.03	**0.73 ± 0.04**	0.72 ± 0.07	0.69 ± 0.05
	Piperacillin-T.	0.14 ± 0.05	**0.23 ± 0.09**	0.20 ± 0.05	
	Tobramycin	0.30 ± 0.04	0.35 ± 0.07	0.35 ± 0.05	**0.40 ± 0.05**
	Ciprofloxacin	0.53 ± 0.03	0.50 ± 0.06	0.50 ± 0.09	**0.66 ± 0.10**
	Ceftriaxone	0.68 ± 0.02	0.74 ± 0.06	0.68 ± 0.13	**0.87 ± 0.06**
K. pneumoniae	Cefepime	0.60 ± 0.03	0.64 ± 0.06	0.61 ± 0.12	**0.65 ± 0.09**
	Meropenem	0.20 ± 0.05	0.30 ± 0.14	**0.35 ± 0.16**	
	Tobramycin	**0.54 ± 0.02**	0.49 ± 0.06	0.46 ± 0.06	0.40 ± 0.17
	Ciprofloxacin	**0.70 ± 0.02**	0.59 ± 0.08	0.59 ± 0.04	0.69 ± 0.04
	Fusidic acid	0.10 ± 0.06	**0.21 ± 0.06**	0.18 ± 0.10	
S. aureus	Oxacillin	**0.85 ± 0.01**	**0.85 ± 0.02**	0.84 ± 0.03	
	Ceftriaxone	-	**0.86 ± 0.02**	0.85 ± 0.03	0.80 ± 0.03
	Clindamycin	-	0.42 ± 0.06	0.38 ± 0.07	

Table 4.6: AUPRC Score combined for all models applied

Bacteria	Antibiotic	Baseline	MLC per Bacteria	MLC Multi Bacteria	Multi Bacteria
	Ciprofloxacin	0.85	0.81	0.83	**0.84**
	Ceftriaxone	0.87	0.88	**0.92**	0.91
E. coli	Cefepime	0.88	0.88	**0.91**	0.86
	Piperacillin-T.	0.64	0.70	**0.73**	
	Tobramycin	0.76	0.75	**0.78**	0.71
	Ciprofloxacin	0.76	0.78	**0.80**	0.76
	Ceftriaxone	0.86	0.86	0.90	**0.92**
K. pneumoniae	Cefepime	0.83	0.86	**0.89**	0.86
	Meropenem	0.83	**0.90**	0.85	
	Tobramycin	**0.83**	0.80	**0.83**	0.81
	Ciprofloxacin	**0.85**	0.75	0.81	0.83
	Fusidic acid	0.68	**0.72**	**0.72**	
S. aureus	Oxacillin	0.93	**0.95**	0.93	
	Ceftriaxone		0.94	0.93	**0.98**
	Clindamycin		**0.78**	0.73	

Table 4.7: AUROC Test Set DRIAMS-A

Bacteria	Antibiotic	Baseline	MLC per Bacteria	MLC Multi Bacteria	Multi Bacteria
	Ciprofloxacin	0.75	0.71	0.72	**0.76**
	Ceftriaxone	0.79	0.77	**0.84**	0.83
E. coli	Cefepime	0.70	0.69	**0.76**	0.67
	Piperacillin-T.	0.14	0.21	**0.23**	
	Tobramycin	0.30	**0.37**	0.36	0.30
	Ciprofloxacin	0.53	**0.55**	0.53	0.47
	Ceftriaxone	0.68	0.68	0.79	**0.85**
K. pneumoniae	Cefepime	0.60	0.63	**0.66**	0.64
	Meropenem	0.20	0.15	**0.21**	
	Tobramycin	**0.54**	0.51	0.46	0.43
	Ciprofloxacin	**0.70**	0.57	0.63	0.63
	Fusidic acid	0.10	**0.22**	0.17	
S. aureus	Oxacillin	0.85	**0.86**	**0.86**	
	Ceftriaxone		0.86	0.86	**0.97**
	Clindamycin		**0.55**	0.40	

Table 4.8: AUPRC Score Test Set DRIAMS-A

4.6 Transfer learning

Transfer learning allows the model to leverage knowledge learned from one domain and apply it to another where domain-specific differences can be significant. Three transfer learning strategies were performed using DRIAMS-B dataset: the first is zero-shot without any training and treating the whole dataset as a test set. Second, with freezing the weights of convolutional layers, and the final one is a warm start (retraining the entire neural network).

The performance of transfer learning models on Multi-label Multi Bacteria approach is summarized in Table 4.10. Whereas the Multi-label classification per Bacteria results are provided in Table 4.9. The baseline results from the paper of (López-Cortés et al., 2024) are also included.

4.6.1 Multi-label per Bacteria

Zero-shot learning provides a strong baseline, outperforming the baseline zero-shot results in all cases. Moreover, the results of training only on local data perform worse compared to other techniques. Further strategies of transfer learning tend to improve performance, namely freezing convolutional layers and warm start. Notably, it performs better for *E.coli* and *K. pneumoniae* and worse for *S. aureus* when compared with zero-shot results. The warm start outcomes vary, being mostly similar to zero-shot results for *E. coli* and *K. pneumoniae* and worse for *S. aureus* strains where zero-shot tends to outperform. The final results are shown in Table 4.9

Bacteria	**Antibiotic**	**Zero-shot**		**Trained on Local Data**		**Freezing Conv Layers**		**Warm start**	
		AUROC	**AUPRC**	**AUROC**	**AUPRC**	**AUROC**	**AUPRC**	**AUROC**	**AUPRC**
E. coli	Ciprofloxacin	0.75	0.68	0.72	0.44	0.74	0.66	**0.78**	**0.73**
	Ceftriaxone	**0.90**	**0.82**	0.55	0.27	0.89	0.80	0.89	0.78
	Ceftriaxone (baseline)	*0.80*	*0.54*	*0.74*	*0.50*	*0.77*	*0.53*	*0.94*	*0.75*
	Cefepime	0.92	0.82	0.48	0.22	**0.93**	**0.83**	0.89	0.74
	Piperacillin-Tazobactam	0.42	0.21	0.61	0.31	0.36	0.20	**0.75**	**0.65**
K. pneumoniae	Ciprofloxacin	0.50	0.32	0.30	0.12	**0.65**	**0.30**	0.49	0.25
	Ceftriaxone	0.49	0.32	0.31	0.10	**0.69**	**0.45**	0.68	0.46
	Ceftriaxone (baseline)	*0.36*	*0.10*	*0.44*	*0.32*	*0.35*	*0.15*	*0.57*	*0.35*
	Cefepime	0.47	0.28	0.42	0.11	**0.68**	**0.44**	0.64	0.53
S. aureus	Ciprofloxacin	**0.70**	**0.26**	0.55	0.19	0.38	0.07	0.41	0.07
	Fusidic acid	**0.61**	**0.12**	0.42	0.05	0.49	0.31	0.49	0.31
	Oxacillin	**0.76**	**0.34**	0.68	0.26	0.66	0.18	0.71	0.22
	Oxacillin (baseline)	*0.72*	*0.18*	*0.68*	*0.39*	*0.72*	*0.18*	*0.79*	*0.27*
	Clindamycin	0.57	0.14	0.41	0.09	0.83	0.29	**0.83**	**0.30**

Table 4.9: Multi-label Classification per Bacteria: performance comparison across different transfer learning strategies on DRIAMS-B dataset.

4.6.2 Multi-label Multi Bacteria

Similar to Multi-label per Bacteria, zero-shot learning provides a strong baseline and outperform other strategies for *E. coli* - Ciprofloxacin, *S. aureus* - Ciprofloxacin, Fusidic acid and Oxacillin. In all cases, zero-shot results outperform the baseline results for zero-shot. Training on local data produce the worst results. Further strategies of transfer learning tend to improve the performance of some bacteria-antibiotic combinations. Freezing convolutional layers shows the best results for *K. pneumoniae* - Ciprofloxain, Ceftriaxone and Cefepime. The warm start outcomes vary, resulting in best outcomes for *E-coli* - Ceftriaxone, Cefepime, Piperacillin-T, and exceptionally outperfoming for *S. aureus* - Clindamycin. The final results are shown in Table 4.10.

Bacteria	**Antibiotic**	**Zero-shot**		**Trained on Local Data**		**Freezing Conv Layers**		**Warm start**	
		AUROC	**AUPRC**	**AUROC**	**AUPRC**	**AUROC**	**AUPRC**	**AUROC**	**AUPRC**
E. coli	Ciprofloxacin	**0.83**	**0.73**	0.76	0.49	0.65	0.55	0.73	0.62
	Ceftriaxone	0.81	0.64	0.60	0.34	0.89	0.82	**0.89**	**0.84**
	Ceftriaxone (*baseline*)	*0.80*	*0.54*	*0.74*	*0.50*	*0.77*	*0.53*	*0.94*	*0.75*
	Cefepime	0.84	0.66	0.62	0.34	0.76	0.70	**0.89**	**0.84**
	Piperacillin-Tazobactam	0.65	0.29	0.58	0.36	0.75	0.64	**0.85**	**0.77**
K. pneumoniae	Ciprofloxacin	0.50	0.32	0.30	0.12	**0.65**	**0.30**	0.49	0.29
	Ceftriaxone	0.49	0.32	0.31	0.10	**0.69**	**0.45**	0.51	0.42
	Ceftriaxone (*baseline*)	*0.36*	*0.10*	*0.44*	*0.32*	*0.35*	*0.15*	*0.57*	*0.35*
	Cefepime	0.47	0.28	0.42	0.11	**0.68**	**0.44**	0.58	0.44
S. aureus	Ciprofloxacin	**0.70**	**0.26**	0.55	0.19	0.38	0.07	0.41	0.07
	Fusidic acid	**0.61**	**0.12**	0.31	0.05	0.49	0.31	0.42	0.08
	Oxacillin	**0.76**	**0.34**	0.38	0.06	0.68	0.19	0.71	0.22
	Oxacillin (baseline)	*0.72*	*0.18*	*0.68*	*0.39*	*0.72*	*0.18*	*0.79*	*0.27*
	Clindamycin	0.57	0.14	0.41	0.09	0.83	0.29	**0.83**	**0.30**

Table 4.10: Multi-label Multi Bacteria: performance comparison across different transfer learning strategies on DRIAMS-B dataset.

4.7 Self-labeling

A self-labeling approach iteratively updates the labels by predicting the labels of initially unlabeled data points. This aims to improve the training process with each iteration. Initially, the dataset is split into train and test sets. The training set is further split into training and validation sets. The self-labeling is applied only on the training set. The pre-trained model of the Multi-label Multi Bacteria approach is loaded. In each iteration, the model predicts probabilities for the unlabeled data. If the predicted probability exceeds the threshold (0.70) then the corresponding label is updated, otherwise, it does not change. Importantly, only unlabeled data

points are updated and considered. Additionally, during the iteration process, the masked cross-entropy loss function is applied to penalize the model only on incorrect prediction of known labels. Due to computational and time limits, the iteration number of 10 has been set. However, we can see from the 4.11 that the number of unlabeled instances does not decrease after the 9th iteration. In the beginning, the dataset had 36,559 labeled and 30,473 unlabeled data points. After applying the interactive self-labeling for 10 iterations the number of labeled instances increased to 64,335 and unlabeled data points decreased to 13,075 as shown in Table 4.11.

Iteration	**# Labeled**	**# Unlabeled (before)**	**# Unlabeled (after)**
Initial	36,559	30,473	30,473
1	38,220	30,473	29,812
2	39,670	29,812	29,102
3	40,939	29,102	28,732
4	42,379	28,732	26,659
5	45,364	26,659	21,668
6	48,968	21,668	18,064
7	52,953	18,064	13,680
8	56,915	13,680	13,079
9	60,722	13,079	13,075
10	64,335	13,075	13,075

Table 4.11: Number of Labeled and Unlabeled Instances Before and After Each Iteration

The most noticeable change in the reduction of the unlabeled labels (-1) appeared for Oxacillin dropped from 5026 to just 1 (Figures 4.17 and 4.18), all of which were assigned to resistance class (1). A similar result is seen for Clindamycin and Tobramycin. Whereas Fusidic acid label distribution has not changed. The distribution appears to become more balanced for Cefepime and Tobramycin as a result of increased resistant (1) labels. A full overview of the label distributions before and after self-labeling is shown in Figures 4.17 and 4.18. Further, the best model of the self-labeling approach is tested on the test set. The results of the test set are presented in tables 4.12 and 4.13.

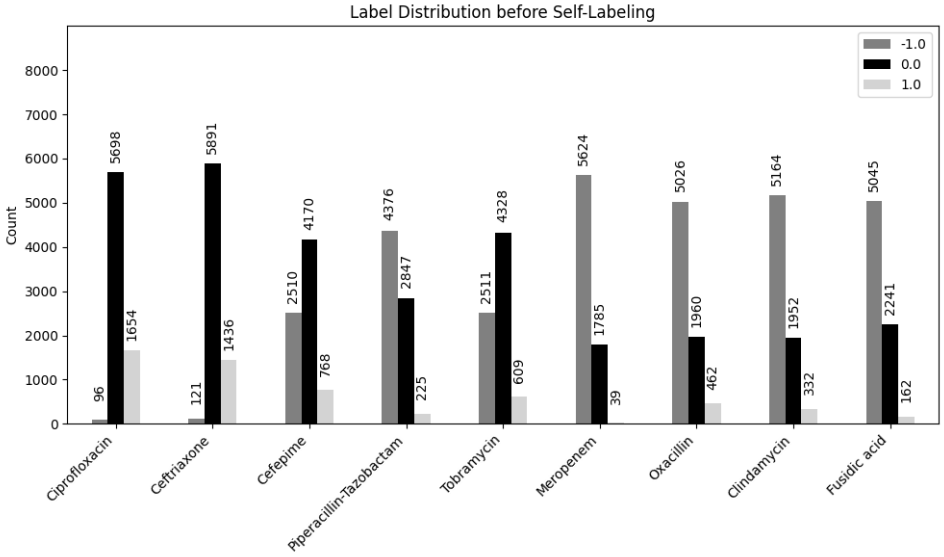

Figure 4.17: DRIAMS-A dataset: label distribution (-1 incomplete, 0 susceptible, 1 resistant) before self-labeling.

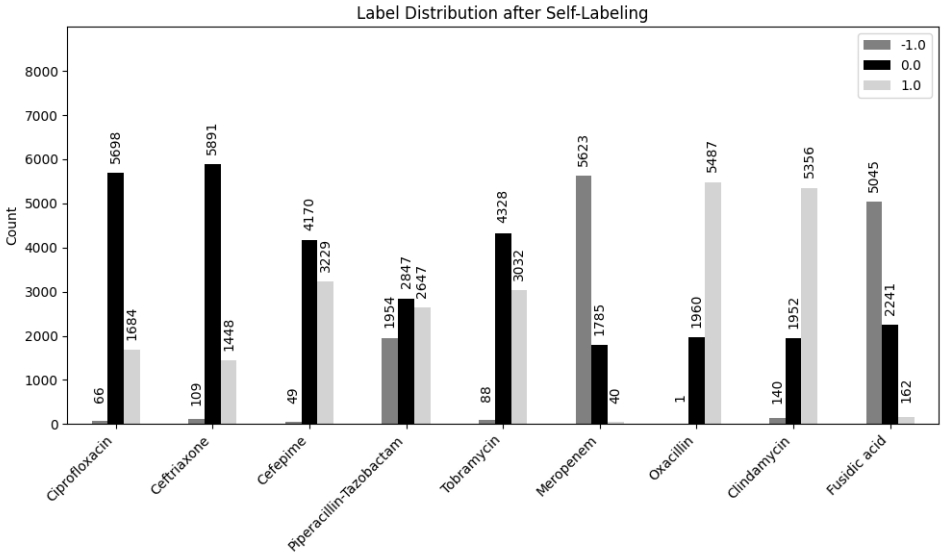

Figure 4.18: DRIAMS-A dataset: label distribution (-1 incomplete, 0 susceptible, 1 resistant) after self-labeling.

As can be seen from the Tables 4.12 and 4.13 self-labeling approach outperforms the baseline and other methods. For comparison purposes, previously obtained results for the Multi-label Multi Bacteria approach are presented as well. As such, for *E.coli* it shows improved performance for Ciptrofloxacin, Ceftriaxone and Piperacillin-T. in terms of both AUROC and AUPRC. For *K. pneumoniae* - Ciptrofloxacin, Meropenem, Cefepime, Tobramycin both AUROC and AUPRC scores outperform previous outcomes. For *S. aureus* - Oxacillin shows slight improvement in AUPRC score.

This approach is straightforward and underlines the potential of a self-labeling approach in improving the models in antibiotic resistance research. Especially in such settings where imbalanced and limited data are common constraints in building an accurate and applicable model in clinical settings.

Bacteria	Antibiotic	Baseline	MLC per Bacteria	MLC Multi Bacteria	Multi Bacteria	MLC Multi Bacteria (self-labeling)
	Ciprofloxacin	0.85	0.81	0.83	0.84	**0.86**
	Ceftriaxone	0.87	0.88	**0.92**	0.91	**0.92**
E. coli	Cefepime	0.88	0.88	**0.91**	0.86	0.90
	Piperacillin-T.	0.64	0.70	0.73		**0.76**
	Tobramycin	0.76	0.75	**0.78**	0.71	0.76
	Ciprofloxacin	0.76	0.78	0.80	0.76	**0.84**
	Ceftriaxone	0.86	0.86	0.90	**0.92**	0.89
K. pneumoniae	Cefepime	0.83	0.86	0.89	0.86	**0.91**
	Meropenem	0.83	0.90	0.85		**0.96**
	Tobramycin	0.83	0.80	0.83	0.81	**0.87**
	Ciprofloxacin	0.85	0.75	0.81	**0.83**	0.79
	Fusidic acid	0.68	0.72	0.72		0.71
S. aureus	Oxacillin	0.93	**0.95**	0.93		0.93
	Ceftriaxone		0.94	0.93	**0.98**	0.93
	Clindamycin		0.78	0.73		0.73

Table 4.12: AUROC Test Set DRIAMS-A

Bacteria	Antibiotic	Baseline	ML per Bacteria	MLC Multi Bacteria	Multi Bacteria	MLC Multi Bacteria (self-labeling)
	Ciprofloxacin	0.75	0.71	0.72	**0.76**	**0.76**
	Ceftriaxone	0.79	0.77	0.84	0.83	**0.85**
E. coli	Cefepime	**0.76**	0.73	**0.76**	0.67	0.71
	Piperacillin-T.	0.14	0.23	0.20		**0.25**
	Tobramycin	0.30	0.37	0.36	0.30	**0.39**
	Ciprofloxacin	0.53	0.55	0.53	0.47	**0.62**
	Ceftriaxone	0.68	0.68	0.79	**0.85**	0.80
K. pneumoniae	Cefepime	0.60	0.63	0.65	0.64	**0.69**
	Meropenem	0.20	0.15	0.21		**0.42**
	Tobramycin	0.54	0.51	0.46	0.48	**0.56**
	Ciprofloxacin	**0.70**	0.57	0.63	0.63	0.56
	Fusidic acid	0.10	**0.22**	0.17		0.17
S. aureus	Oxacillin	0.85	**0.86**	0.84		**0.86**
	Ceftriaxone		0.86	0.86	**0.97**	0.87
	Clindamycin	**0.55**	0.42	0.38		0.40

Table 4.13: AUPRC Test Set DRIAMS-A

CHAPTER 5
CONCLUSION

The objective of this research was to explore and improve the performance of antimicrobial resistance prediction models (MSDeepAMR) by applying different multi-label approaches and domain adaptation techniques. The study aimed to evaluate whether a multi-label approach could enhance model accuracy across various bacteria and antibiotics. For these purposes, mass spectrometry (MS) data from the DRIAMS-A and DRIAMS-B datasets has been used. The initial proposed models included single bacteria-specific models (multi-label classification per bacteria), multi-label models for multiple bacteria, and models focusing on individual antibiotics. Subsequently, transfer learning and self-labeling techniques were incorporated.

Multi-label per Bacteria Approach: The model is trained to predict the resistance to multiple antibiotics but for each bacteria separately. This approach adapts to specific domains (different bacteria). The results indicated that this approach effectively captured the domain-specific characteristics, leading to similar to baseline or improved performance between different bacteria.

Multi Bacteria per Antibiotic: The model is trained to predict the resistance of multiple bacteria to a specific antibiotic. This allows to understand how different bacteria respond to a particular antibiotic. The results showed that this method could adapt to the specific resistance patterns of various bacterial species, achieving competitive performance, particularly for the antibiotic Ceftriaxone.

Multi-label Multi Bacteria Approach: Instead of creating a separate model for each bacteria antibiotic pair, a single model is trained to handle species and antibiotics together. This allowed the model to learn a more diverse set of data that could potentially capture patterns that are not domain-specific (bacteria-antibiotic) but across different bacterial species and antibiotics. The complexity of this model required careful handling of incomplete labels, which was achieved through the implementation of a custom-masked binary cross-entropy loss function and self-labeling techniques. The integration of these methods led to notable improvements in AUROC and AUPRC scores for several bacteria-antibiotic combinations.

Transfer learning: The models for Multi Bacteria per Antibiotic and Multi-label Multi Bacteria were tested on the external DRIAMS-B dataset to evaluate the adaptability and performance of the models. Subsequently, different transfer learning experiments were implemented. The results obtained by warm start (retraining the entire neural network) on Multi-label Classification per Bacteria outperformed the results of baseline *K. pneumoniae* - Ceftriaxone. The results show that zero-shot and warm start performed better compared to the method of freezing convolutional layers.

Challenges with Incomplete Labels

The dataset includes a vast amount of incomplete labels, to address this issue a masked binary cross-entropy loss function was implemented. This loss function specifically masked out the missing labels (denoted as -1) during training. Another method applied was self-labeling, which iteratively predicted and updated the incomplete labels.

The self-labeling technique resulted in improvements in model performance across several bacteria-antibiotic combinations. The AUROC and AUPRC scores generally increased after self-labeling, particularly for antibiotics like Ceftriaxone and Ciprofloxacin. Overall, the integration of masking and self-labeling techniques showed improved results compared to the baseline results.

Recommendations for Future Research

Given the promising results of this study, further research could explore more sophisticated approaches, such as Domain Adversarial Neural Network (DANN) or probability calibration technique on imbalanced data which may offer better performance in complex multi-label classification tasks. Additionally, integrating the self-labeling model with more data sources, such as patient demographics could further improve prediction accuracy.

The results suggest that further development and refinement of these techniques could lead to more accurate and reliable models, contributing significantly to the field of antimicrobial resistance prediction.

The integration of domain adaptation techniques, including multi-label approaches, transfer learning, and self-labeling, has demonstrated promising results in the prediction of antibiotic resistance using MALDI-TOF MS data. Despite the challenges posed by incomplete labels and domain shifts, the study provides a foundation for future work that could lead to more accurate and reliable models.

BIBLIOGRAPHY

Acar, J. and Rostel, B. (2001). Antimicrobial resistance: an overview. *Revue Scientifique et Technique-Office International des Epizooties*, 20(3):797–810.

Achanta, S., Jaju, J., Kumar, A. M., Nagaraja, S. B., Shamrao, S. R. M., Bandi, S. K., Kumar, A., Satyanarayana, S., Harries, A. D., Nair, S. A., et al. (2013). Tuberculosis management practices by private practitioners in andhra pradesh, india. *PloS one*, 8(8):e71119.

Ali, T., Ahmed, S., and Aslam, M. (2023). Artificial intelligence for antimicrobial resistance prediction: challenges and opportunities towards practical implementation. *Antibiotics*, 12(3):523.

Anhalt, J. P. and Fenselau, C. (1975). Identification of bacteria using mass spectrometry. *Analytical chemistry*, 47(2):219–225.

Beam, A. L. and Kohane, I. S. (2018). Big data and machine learning in health care. *Jama*, 319(13):1317–1318.

Behjati, S. and Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood-Education and Practice*, 98(6):236–238.

Bengtsson-Palme, J., Kristiansson, E., and Larsson, D. J. (2018). Environmental factors influencing the development and spread of antibiotic resistance. *FEMS microbiology reviews*, 42(1):fux053.

Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O., and Piddock, L. J. (2015). Molecular mechanisms of antibiotic resistance. *Nature reviews microbiology*, 13(1):42–51.

Boolchandani, M., D'Souza, A. W., and Dantas, G. (2019). Sequencing-based methods and resources to study antimicrobial resistance. *Nature Reviews Genetics*, 20(6):356–370.

Brodersen, K. H., Ong, C. S., Stephan, K. E., and Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In *2010 20th international conference on pattern recognition*, pages 3121–3124. IEEE.

Cain, T. C., Lubman, D. M., and Weber Jr, W. J. (1994). Differentiation of bacteria using protein profiles from matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 8(12):1026–1030.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., and Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408:189–215.

Chen, J., Ying, G.-G., and Deng, W.-J. (2019). Antibiotic residues in food: extraction, analysis, and human health concerns. *Journal of Agricultural and Food Chemistry*, 67(27):7569–7586.

Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData mining*, 10(1):35.

Chokshi, A., Sifri, Z., Cennimo, D., and Horng, H. (2019). Global contributors to antibiotic resistance. *Journal of global infectious diseases*, 11(1):36–42.

Conly, J. and Johnston, B. (2005). Where are all the new antibiotics? the new antibiotic paradox. *The Canadian journal of infectious diseases & medical microbiology*, 16(3):159.

De Bruyne, K., Slabbinck, B., Waegeman, W., Vauterin, P., De Baets, B., and Vandamme, P. (2011). Bacterial species identification from maldi-tof mass spectra through data analysis and machine learning. *Systematic and applied microbiology*, 34(1):20–29.

Dooling, K. L., Kandeel, A., Hicks, L. A., El-Shoubary, W., Fawzi, K., Kandeel, Y., Etman, A., Lohiniva, A. L., and Talaat, M. (2014). Understanding antibiotic use in minya district, egypt: physician and pharmacist prescribing and the factors influencing their practices. *Antibiotics*, 3(2):233–243.

Farhat, F., Athar, M. T., Ahmad, S., Madsen, D. Ø., and Sohail, S. S. (2023). Antimicrobial resistance and machine learning: past, present, and future. *Frontiers in Microbiology*, 14:1179312.

Florio, W., Baldeschi, L., Rizzato, C., Tavanti, A., Ghelardi, E., and Lupetti, A. (2020). Detection of antibiotic-resistance by maldi-tof mass spectrometry: an expanding area. *Frontiers in cellular and infection microbiology*, 10:572909.

Florio, W., Tavanti, A., Ghelardi, E., and Lupetti, A. (2018). Maldi-tof ms applications to the detection of antifungal resistance: state of the art and future perspectives. *Frontiers in Microbiology*, 9:2577.

Gajic, I., Kabic, J., Kekic, D., Jovicevic, M., Milenkovic, M., Mitic Culafic, D., Trudic, A., Ranin, L., and Opavski, N. (2022). Antimicrobial susceptibility testing: a comprehensive review of currently used methods. *Antibiotics*, 11(4):427.

Ghorbani, B., Ghorbani, M., Abedi, M., and Tayebi, M. (2016). Effect of antibiotics overuse in animal food and its link with public health risk. *Int J Sci Res Sci Technol*, 2:46–50.

González, C., Astudillo, C. A., López-Cortés, X. A., and Maldonado, S. (2023). Semi-supervised learning for maldi–tof mass spectrometry data classification: An application in the salmon industry. *Neural Computing and Applications*, 35(13):9381–9391.

Green, A. G., Yoon, C. H., Chen, M. L., Ektefaie, Y., Fina, M., Freschi, L., Gröschel, M. I., Kohane, I., Beam, A., and Farhat, M. (2022). A convolutional neural network highlights mutations relevant to antimicrobial resistance in mycobacterium tuberculosis. *Nature communications*, 13(1):3817.

Gross, J. H. (2019). *Massenspektrometrie*. Springer.

Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Analytical chemistry*, 63(24):1193A–1203A.

Iskandar, K., Molinier, L., Hallit, S., Sartelli, M., Hardcastle, T. C., Haque, M., Lugova, H., Dhingra, S., Sharma, P., Islam, S., et al. (2021). Surveillance of antimicrobial resistance in low-and middle-income countries: a scattered picture. *Antimicrobial Resistance & Infection Control*, 10:1–19.

Kim, J. and Ahn, J. (2022). Emergence and spread of antibiotic-resistant foodborne pathogens from farm to table. *Food Science and Biotechnology*, 31(12):1481–1499.

Liu, H. Y., Hopping, G. C., Vaidyanathan, U., Ronquillo, Y. C., Hoopes, P. C., and Moshirfar, M. (2019). Polymerase chain reaction and its application in the diagnosis of infectious keratitis. *Medical Hypothesis, Discovery and Innovation in Ophthalmology*, 8(3):152.

López-Cortés, X. A., Manríquez-Troncoso, J. M., Hernández-García, R., and Peralta, D. (2024). Msdeepamr: antimicrobial resistance prediction based on deep neural networks and transfer learning. *Frontiers in Microbiology*, 15:1361795.

Macesic, N., Polubriaginof, F., and Tatonetti, N. P. (2017). Machine learning: novel bioinformatics approaches for combating antimicrobial resistance. *Current opinion in infectious diseases*, 30(6):511–517.

McDermott, M., Hansen, L. H., Zhang, H., Angelotti, G., and Gallifant, J. (2024). A closer look at auroc and auprc under class imbalance. *arXiv preprint arXiv:2401.06091*.

Mortier, T., Wieme, A. D., Vandamme, P., and Waegeman, W. (2021). Bacterial species identification using maldi-tof mass spectrometry and machine learning techniques: a large-scale benchmarking study. *Computational and Structural Biotechnology Journal*, 19:6157–6168.

Mourabit, N., Arakrak, A., Bakkali, M., Zian, Z., Bakkach, J., and Laglaoui, A. (2021). Antimicrobial resistance trends in staphylococcus aureus strains carried by poultry in north of morocco: A preliminary analysis. *Journal of Food Quality*, 2021(1):8856004.

Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Aguilar, G. R., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., et al. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The lancet*, 399(10325):629–655.

Okeke, I. N., Lamikanra, A., and Edelman, R. (1999). Socioeconomic and behavioral factors leading to acquired bacterial resistance to antibiotics in developing countries. *Emerging infectious diseases*, 5(1):18.

O'Neill, J. (2016). Tackling drug-resistant infections globally: final report and recommendations.

Parlament, E. and the Council of EU (2019). Regulation (eu) 2019/6 of the european parliament and of the council of 11 december 2018 on veterinary medicinal products and repealing directive 2001/82/ec. *Off. J. Eur. Union*, 4.

Popa, S. L., Pop, C., Dita, M. O., Brata, V. D., Bolchis, R., Czako, Z., Saadani, M. M., Ismaiel, A., Dumitrascu, D. I., Grad, S., et al. (2022). Deep learning and antibiotic resistance. *Antibiotics*, 11(11):1674.

Prajwala, T. (2015). A comparative study on decision tree and random forest using r tool. *International journal of advanced research in computer and communication engineering*, 4(1):196–199.

Racewicz, P., Majewski, M., Biesiada, H., Nowaczewski, S., Wilczyński, J., Wystalska, D., Kubiak, M., Pszczoła, M., and Madeja, Z. E. (2022). Prevalence and characterisation of antimicrobial resistance genes and class 1 and 2 integrons in multiresistant escherichia coli isolated from poultry production. *Scientific Reports*, 12(1):6062.

Ren, Y., Chakraborty, T., Doijad, S., Falgenhauer, L., Falgenhauer, J., Goesmann, A., Schwengers, O., and Heider, D. (2022). Multi-label classification for multi-drug resistance prediction of escherichia coli. *Computational and Structural Biotechnology Journal*, 20:1264–1270.

Sakagianni, A., Koufopoulou, C., Feretzakis, G., Kalles, D., Verykios, V. S., and Myrianthefs, P. (2023). Using machine learning to predict antimicrobial resistance—a literature review. *Antibiotics*, 12(3):452.

Satyanarayana, S., Kwan, A., Daniels, B., Subbaraman, R., McDowell, A., Bergkvist, S., Das, R. K., Das, V., Das, J., and Pai, M. (2016). Use of standardised patients to assess antibiotic dispensing for tuberculosis by pharmacies in urban india: a cross-sectional study. *The Lancet Infectious Diseases*, 16(11):1261–1268.

Sauer, S. and Kliem, M. (2010). Mass spectrometry tools for the classification and identification of bacteria. *Nature reviews microbiology*, 8(1):74–82.

Singhal, N., Kumar, M., Kanaujia, P. K., and Virdi, J. S. (2015). Maldi-tof mass spectrometry: an emerging technology for microbial identification and diagnosis. *Frontiers in microbiology*, 6:791.

Smith, T. C., Gebreyes, W. A., Abley, M. J., Harper, A. L., Forshey, B. M., Male, M. J., Martin, H. W., Molla, B. Z., Sreevatsan, S., Thakur, S., et al. (2013). Methicillin-resistant staphylococcus aureus in pigs and farm workers on conventional and antibiotic-free swine farms in the usa. *PloS one*, 8(5):e63704.

Tacconelli, E. (2009). Antimicrobial use: risk driver of multidrug resistant microorganisms in healthcare settings. *Current opinion in infectious diseases*, 22(4):352–358.

Tharmakulasingam, M., Gardner, B., La Ragione, R., and Fernando, A. (2022). Explainable deep learning approach for multilabel classification of antimicrobial resistance with missing labels. *IEEE Access*, 10:113073–113085.

Vieira, A. R., Collignon, P., Aarestrup, F. M., McEwen, S. A., Hendriksen, R. S., Hald, T., and Wegener, H. C. (2011). Association between antimicrobial resistance in escherichia coli isolates from food animals and blood stream isolates from humans in europe: an ecological study. *Foodborne pathogens and disease*, 8(12):1295–1301.

Weinstein, R. A. (2001). Controlling antimicrobial resistance in hospitals: infection control and use of antibiotics. *Emerging infectious diseases*, 7(2):188.

Weis, C., Cuénod, A., Rieck, B., Dubuis, O., Graf, S., Lang, C., Oberle, M., Brackmann, M., Søgaard, K. K., Osthoff, M., et al. (2022). Direct antimicrobial resistance prediction from clinical maldi-tof mass spectra using machine learning. *Nature Medicine*, 28(1):164–174.

Xian, F., Hendrickson, C. L., and Marshall, A. G. (2012). High resolution mass spectrometry. *Analytical chemistry*, 84(2):708–719.

Yoon, E.-J. and Jeong, S. H. (2021). Maldi-tof mass spectrometry technology as a tool for the rapid diagnosis of antimicrobial resistance in bacteria. *Antibiotics*, 10(8):982.

Zhang, J., Wang, Z., Wang, H.-Y., Chung, C.-R., Horng, J.-T., Lu, J.-J., and Lee, T.-Y. (2022). Rapid antibiotic resistance serial prediction in staphylococcus aureus based on large-scale maldi-tof data by applying xgboost in multi-label learning. *Frontiers in Microbiology*, 13:853775.