Preprint submitted to bioRxiv, 2025, pp. 1–14

doi:
Advance Access Publication Date: Day Month Year
Paper

PAPER

Species-agnostic and Salmonella-specific Models for Antimicrobial Resistance Prediction Using FCGR and ResNet-18

Saifelden M. Ismail[1,*] and Salma H. Fayed[2]

[1]Independent Researcher, Giza, Egypt and [2]Independent Researcher, Giza, Egypt
[*]Corresponding author. saifeldenmi@gmail.com
FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

Antimicrobial resistance (AMR) prediction from bacterial genomes remains a major challenge for clinical microbiology and surveillance. We developed deep learning models based on Frequency Chaos Game Representation (FCGR) and a ResNet-18 architecture to classify resistance phenotypes directly from whole-genome assemblies. Using homology-aware clustering to prevent genomic data leakage, we compared a species-specific model for *Salmonella enterica* with a species-agnostic model spanning *S. enterica*, *Escherichia coli*, and *Staphylococcus aureus*. The *Salmonella enterica* model achieved high predictive accuracy, particularly for cephalosporins, while performance was lower for tetracycline and ampicillin. The species-agnostic model generalized less effectively overall but improved prediction for underrepresented species such as *E. coli*. Benchmarking against the gene-based tool ResFinder confirmed that FCGR-based models can approach curated predictions while capturing distributed genomic features not restricted to known resistance genes. This suggests that deep learning can highlight genomic patterns with potential biological relevance beyond established markers. Overall, species-focused models give the most reliable performance, while species-agnostic approaches can boost prediction for poorly sampled pathogens and aid AMR surveillance.

Key words: Antimicrobial resistance, Deep learning, Frequency Chaos Game Representation (FCGR), ResNet-18, Species-agnostic modeling, *Salmonella enterica*, Genomic prediction

Introduction

Antimicrobial resistance (AMR) is one of the greatest health threats of the 21st century [1]. It causes around 1.2 million deaths each year, and this number is rising [2]. The way bacteria respond to antibiotics shows how much they have evolved. They have considerable genetic flexibility, which allows them to resist nearly all available antibiotics by various methods. These methods include changing their target sites, altering their efflux pumps, and inactivating enzymes [1]. This ability to change, combined with the slow development of new antibiotics, has created a pressing need for better ways to predict and monitor AMR.

Traditional antimicrobial susceptibility testing (AST) methods take a long time, usually 18 to 24 hours, to identify bacteria and determine their susceptibility [2]. These delays can lead to incorrect initial treatment, resulting in treatment failures and increased resistance. About 50% of antibiotic treatments begin without proper identification of the pathogen [3]. We urgently need faster and more accurate predictions of antimicrobial resistance.

Recent advances in artificial intelligence (AI) and machine learning (ML) have opened new opportunities to address these challenges. Deep learning methods have proven very effective in predicting antimicrobial resistance (AMR). Green et al. achieved 91.9% sensitivity for predicting resistance in *Mycobacterium tuberculosis* using convolutional neural networks (CNNs), which reduced diagnostic times from weeks to hours [4]. However, current ML-based AMR prediction methods have several drawbacks. A recent meta-analysis showed that while ML methods reach pooled area under the curve (AUC) scores around 0.82, they experience challenges with retrospective methodology, inconsistent data processing, and limited clinical validation [5]. Additionally, most studies are still in the early stages of development with little real-world application [6].

Current genomic methods for predicting antimicrobial resistance primarily depend on databases that identify known resistance genes by matching patterns against established sources like ResFinder, NCBI, and CARD. Tools such as ABRicate [7] are the current standard for quick genotypic resistance prediction. However, they are limited because they rely on previously characterized resistance mechanisms. These methods cannot find new resistance patterns or reflect the complex, multi-gene nature of resistance that arises through various evolutionary pathways [8].

© The Author 2025. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

One promising way to tackle these limitations is by using alignment-free genomic representations. Chaos Game Representation (CGR) and its extension, Frequency Chaos Game Representation (FCGR), have become effective tools for transforming DNA sequences into high-density, information-rich 2D images. These images keep biological meaning while standardizing input sizes [9]. FCGR converts sequences of different lengths into equal-sized images based on k-mer frequencies [10]. This makes them ideal for deep learning applications. Recently, combining FCGR with CNNs for viral classification achieved 96.29% accuracy. This shows the potential of this method for genomic analysis [11]. Ren et al. demonstrated that SNP encodings, including FCGR, along with machine learning methods, can provide strong predictive performance for antimicrobial resistance in *E. coli* [12]. However, their study only focused on predictions for one species. It also used limited evaluation metrics and did not compare results against standard gene-based tools like ResFinder. Additionally, their data splitting method did not control for similarities between training and test isolates, raising concerns about genomic leakage.

Recent reviews show that AI and machine learning are playing a bigger role in managing antimicrobial resistance. Their uses include identifying pathogens, recognizing resistance patterns, and predicting treatment outcomes [13]. However, even though machine learning applications for antimicrobial resistance are growing in areas like genomic prediction, understanding mechanisms, and clinical guidance, this field is still young. There are many chances for research that can lead to practical applications [14].

Despite these advances, a significant gap still exists in creating species-agnostic AMR prediction models. Most current AI-based AMR research focuses on predicting resistance in single species. This limits how useful the models are in clinical settings, where it is crucial to quickly identify resistance patterns across various pathogens. The consistent nature of FCGR representation provides an opportunity to develop models that can generalize across different bacterial species by learning resistance patterns that go beyond individual species boundaries.

Building on these foundations, this study improves alignment-free genomic AMR prediction by applying ResNet-18 architectures to Frequency Chaos Game Representations (FCGR) of whole genome sequences. To ensure a thorough evaluation, we use homology-aware data splitting and compare our predictions against the gene-database-based tool ResFinder [15] as a relevant clinical baseline. Besides species-specific models, we explore the possibility of species-agnostic modeling to see if conserved genomic signals of resistance can be detected across pathogens. Although other image-based ways to encode biological sequences have shown promise [16], this study focuses only on FCGR. By working independently of predefined resistance gene databases, our approach could uncover new resistance mechanisms and set a strong foundation for AMR surveillance across different species.

Methods

Dataset description

Table 1 summarizes the primary sources of datasets used for this study, including species, source journals, project numbers (if available) and their respective publications ([17], [18], [19]). These datasets provided the foundation from which antibiotics of interest were selected with their corresponding samples. Only samples with labels for the chosen drugs were retained, while irrelevant samples were removed. For the species-agnostic model, antibiotics were selected based off commonality and label quality between species, elaborated on in Section 2.1.2.

Fig. 1. Correlation matrix of antibiotic resistance labels in the *Salmonella enterica* dataset. The values represent pairwise correlations between resistance profiles for the antibiotics included in this study.

Having described the general dataset sources, the logical next step is to first ground the analysis in a single well-characterized species. This ensures that resistance patterns can be examined in a controlled setting before testing whether they generalize across pathogens.

Salmonella enterica-specific model

Figure 1 displays the correlation matrix of antibiotic resistance labels in the *Salmonella enterica* dataset. It shows that some antibiotics, such as cefoxitin, ceftiofur, and ceftriaxone, are strongly correlated with each other. In contrast, tetracycline and gentamicin have weaker correlations.

This outcome is expected. The antibiotics mentioned are all β-lactams (cephalosporins) with similar chemical structures and resistance mechanisms, such as β-lactamase activity. Tetracycline and gentamicin, on the other hand, show weak correlations with other drugs. This likely occurs because they belong to different classes and have different mechanisms of action.

While this species-specific exploration provides clarity on resistance structures within *Salmonella*, our broader goal is to test whether such models can capture cross-species genomic signals. This motivates a shift toward constructing a species-agnostic dataset that integrates multiple pathogens into a single framework.

Species-agnostic model

Table 2 contains a summary of the curated species-agnostic (general) dataset used in this study. Panel (a) has the species distribution, with *Salmonella enterica* comprising roughly half of the samples, followed by *Staphylococcus aureus* and *Escherichia coli* making up roughly a quarter of the dataset each. Panel (b) reports the overall proportion of NULL values

Table 1. Primary sources of datasets used in this study. Species, source journals, project numbers, and associated publications are listed.

Species	Primary Source	Project Number(s)	Publication
Staphylococcus aureus	mSystems Journal	Multiple datasets	DOI: 10.1128/msystems.01185-20
Salmonella enterica	JCM Journal	PRJNA292661, PRJNA292666	DOI: 10.1128/jcm.01260-18
Escherichia coli	mSystems Journal	Multiple datasets	DOI: 10.1128/msystems.00913-20

Table 2. Summary statistics of the species-agnostic model (10,079 samples). Values are rounded to two decimal places. The table shows (a) species distribution, (b) proportion of NULL values per drug, (c) ratio of resistant+intermediate (R+I) to susceptible (S) samples per drug and species, and (d) percentage of NULL values per drug and species.

(a) Species distribution

Species	Count	%
Salmonella enterica	5084	50.44
Staphylococcus aureus	2563	25.43
Escherichia coli	2432	24.13

(b) NULL values per drug (%)

Drug	NULL %
tetracycline	28.42
gentamicin	7.24

(c) (R+I)/S ratio per drug and species

Species	tetra	tenta
S. aureus	0.76	0.64
S. enterica	1.24	0.16
E. coli	2.16	0.22

(d) NULL percentage per drug and species

Species	tetra	genta
S. aureus	30.16	27.27
S. enterica	0.00	0.00
E. coli	85.98	1.27

per label, showing that tetracycline contains a high fraction of missing labels when directly compared with gentamicin. Panel (c) presents the ratio of resistant+intermediate (R+I) to susceptible (S) samples per species, showing a skew towards resistance in *E. coli*, while *S. aureus* and *S. enterica* are closer to balance. Finally, panel (d) details the percentage of NULLs per label for each species, showing that the proportion of NULLs is uneven across species, with *E. coli* having a particularly high fraction of NULLs for tetracycline.

By comparing these panels, it becomes clear that the species-agnostic setup introduces both new opportunities and new challenges. On one hand, the diversity of samples raises the prospect of detecting truly cross-species genomic markers; on the other, the presence of missing labels and class imbalance complicates evaluation. These observations frame the analytical strategies adopted in later sections, where NULL masking and imbalance-aware metrics are employed to ensure fair comparisons.

Dataset splits and homology considerations

Building on the description of dataset composition in the previous subsection, we next outline how the data were partitioned for training, validation, and testing. Proper splitting was given significant focus on in this study to prevent information leakage between homologous genomes and to ensure a fair evaluation of both species-specific and species-agnostic models.

Because of the large dataset size and computational constraints, utilizing standard homology-based clustering tools such as MMseqs2 [20] and CD-HIT [21] was not feasible. Instead, we used a homology-aware clustering strategy using *sourmash* [22], which relies on MinHash sketching to approximate k-mer similarity. For each dataset, genome assemblies were initially sketched with *sourmash* (version 4.8.6) using a k-mer size of 31 and a scaling factor of 1000, producing signatures that capture the k-mer content of each genome. Pairwise similarities between all genomes were then calculated with *sourmash compare* and converted into a distance matrix ($d = 1 - s$, where s is the Jaccard similarity).

A graph was constructed in which genomes were represented as nodes, and edges were drawn between genomes with distances below a predefined threshold ($d \leq 0.05$). Connected components of this graph were then extracted using the *networkx* library, with each component representing a homology cluster. Each genome was assigned to a cluster, and these clusters were used as the unit of splitting: all genomes within a cluster were assigned entirely to either the training or the test set. This process ensured that highly similar genomes (differing by at most 5% in their MinHash distance) did not appear in both sets.

To build the independent test set, clusters were sorted by size and grouped into bins to avoid biasing the split toward either very small or very large clusters. From each bin, a fixed proportion of clusters (20%) was randomly selected and assigned to the test set, while the remaining clusters were reserved for training. This ensured that the test set contained a representative distribution of cluster sizes and label coverage, while maintaining separation of homologous genomes. To ensure diversity, the largest bin was excluded from test selection so that the test set would not be dominated by a single large cluster.

We ensured that cross-validation was preformed within the training clusters to further divide the data into training and validation folds. For the species-specific models, we used *GroupKFold* with clusters as the grouping variable, ensuring that all genomes from a given cluster were assigned to the same fold.

For the general (species-agnostic) model, the splitting procedure was applied separately within each species, and the resulting training and test clusters were then aggregated across species. Additionally, we employed *StratifiedGroupKFold*, which stratifies by species labels while grouping by clusters. This ensured that validation folds were balanced across species while still preventing homologous overlap. This two-stage procedure (cluster-based test set construction followed by

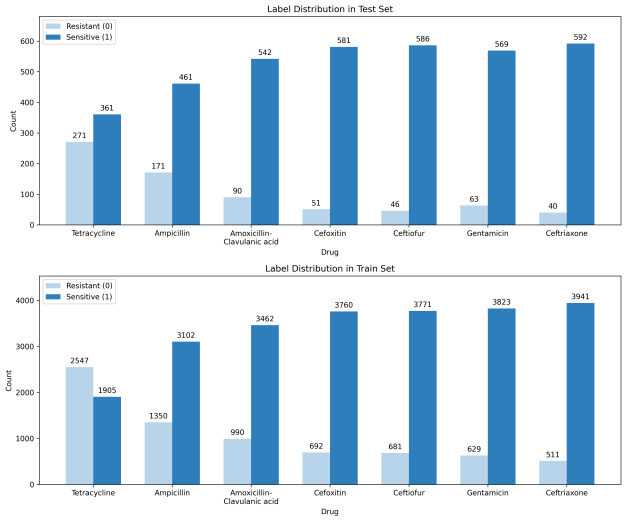

Fig. 2. Label distribution for the *Salmonella enterica*-specific model training and test sets. The dataset was split into 4452 training samples (87.57%, 154 clusters) and 632 test samples (12.43%, 36 clusters).

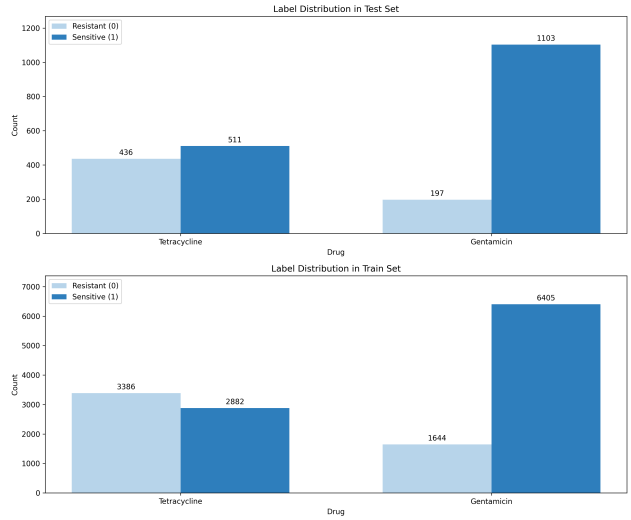

Fig. 3. Label distribution for the species-agnostic (general) model training and test sets. The dataset was split into 8639 training samples (85.71%, 1447 clusters) and 1440 test samples (14.29%, 359 clusters).

cluster-aware cross-validation) provided a thorough evaluation framework that avoided data leakage and ensured fair comparisons between species-specific and species-agnostic models.

Having described the overall strategy for dataset partitioning, we now turn to the direct application of this process in the *Salmonella*- specific model.

Salmonella enterica-specific model

Figure 2 displays the class distribution across the 7 drug labels in both the training and test sets. As described in the figure, the dataset was split into 4452 training samples (87.57%, 154 clusters) and 632 test samples (12.43%, 36 clusters) using the previously mentioned homology aware splitting algorithm.

This figure notably highlights the dataset class imbalance across labels, with some antibiotics (e.g., cefoxitin, ceftiofur, ceftriaxone) having few resistant samples when compared to others (e.g., tetracycline, ampicillin). This imbalance is addressed by utilizing positive weights while training the model, as well as imbalance-aware metrics (e.g., balanced accuracy, MCC, Jaccard) during model evaluation on the independently curated test set.

While the *Salmonella*-specific split illustrates the challenges of class imbalance within a single pathogen, the next subsection considers the broader task of balancing resistance profiles across multiple species.

Species-agnostic model

Figure 3 shows the distribution of resistant and susceptible samples across the species-agnostic (general) model sets. This dataset was split into 8639 training samples (85.71%, 1447 clusters) and 1440 test samples (14.29%, 359 clusters) using a more species-aware, homology-based splitting algorithm.

While tetracycline is somewhat balanced, gentamicin shows a skew towards susceptible samples. This imbalance is present across both the test and training datasets and is different per species, as is presented in Figure 2. To improve optimization for imbalanced classes across species, species-specific positive weights are applied during optimization, as described in Section 2.4.

Frequency Chaos Game Representation (FCGR)

Having established how the datasets were split and controlled for homology, the next step was to transform the raw genomic sequences into a format suitable for machine learning. To this end, we applied the Frequency Chaos Game Representation (FCGR) approach, which converts nucleotide sequences into structured numerical arrays.

Each genome assembly was converted into a Frequency Chaos Game Representation (FCGR) matrix, which encodes the frequency of all possible k-mers in a two-dimensional grid [10]. We used the open-source *fcgr* Python library [23] to generate these matrices. This library provides an efficient and accurate implementation of FCGR, supporting direct conversion from FASTA/FNA files into k-mer frequency matrices.

We evaluated multiple k-mer sizes (6, 7, and 8) during development of the *Salmonella* model, and based on empirical performance (see Table 3), we selected $k = 8$ for all subsequent experiments. For each genome, all contigs were concatenated into a single sequence prior to FCGR computation. The resulting $2^k \times 2^k$ matrices were stored as *.npy* files to avoid repeated conversion during training, and subsequently compressed into archives for efficient transfer to local memory in the training environment.

During training, each FCGR matrix was normalized using min–max scaling to the range $[0, 1]$. This normalization ensured that the relatively weak signals characteristic of FCGR representations (with many low-intensity values and few high-intensity peaks) were preserved and made comparable across genomes.

Figure 4 shows sample FCGR images for the three species included in this study. Upon inspection, the FCGRs of *E. coli* and *Salmonella* seem to strongly resemble each other, though *E. coli* demonstrates darker regions in some areas, likely indicating subtle differences in k-mer composition. *Staphylococcus*, on the other hand, exhibits a visually distinct

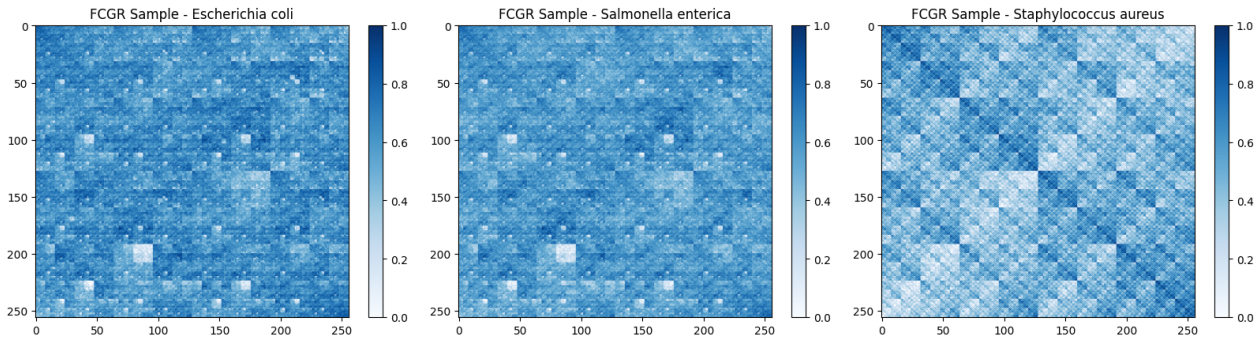

Fig. 4. Frequency Chaos Game Representation (FCGR) images of *Escherichia coli*, *Salmonella enterica*, and *Staphylococcus aureus*. The 8-mer FCGR matrices in this figure were normalized and log-transformed to enhance contrast and make genomic patterns more visually distinct merely for display purposes.

pattern with a blockier structure. These visual differences imply that a model trained across species may benefit from the shared genomic structure of *E. coli* and *Salmonella*, while *Staphylococcus* may need a more species-specific feature set. This observation motivates our later decision to train species-specific models on select drugs and a subset of samples from the Species-agnostic model's dataset.

With the genomic sequences now transformed into structured numerical inputs, the next step was to prepare the phenotypic labels in a way that was both biologically meaningful and computationally stable for model training.

Label encoding and preprocessing

Resistance phenotypes were binarized by encoding resistant (R) and intermediate (I) isolates as `0`, and susceptible (S) isolates as `1`. Although this representation is inverted compared to common conventions (where resistance is taken as the positive class), we adopted it for two reasons. First, preliminary experiments indicated that the models converged more stably and achieved superior predictive performance when susceptibility was treated as the positive outcome, despite resistance being the rarer label. Second, intermediate isolates were merged with resistant due to their low frequency and biological plausibility: many intermediate cases carry at least some genomic features of resistance and cluster more closely with resistant than susceptible isolates. To ensure that this label choice did not bias evaluation, we report only symmetric, polarity-invariant metrics (Balanced Accuracy, MCC, and Jaccard), and handle missing labels (NULLs) by masking them on a per-drug basis during training and evaluation.

Deep convolutional neural network architecture

With the genomic sequences transformed into numerical FCGR representations and the resistance phenotypes encoded into stable, binarized labels, the next step was to design and train a predictive model capable of learning relevant resistance-associated patterns. To achieve this, we employed a deep convolutional neural network (CNN) architecture tailored to the specific properties of the data.

We employed a ResNet-18 architecture [24] as the backbone for all models. ResNet-18 was selected due to its relatively small size, which is well-suited to the dataset scale, and because its residual connections help preserve weak but informative signals in the FCGR representations. As noted earlier, FCGR images contain many low-intensity features and relatively few high values; the skip connections in ResNet mitigate the risk of these weaker signals being diminished through successive convolutional layers. The first convolutional layer was modified to accept single-channel input, and the final fully connected layer was replaced with a linear classifier matching the number of antibiotic resistance classes. All models were trained from scratch without pretrained weights.

Training parameters differed slightly between the general and species-specific models. The general model was trained for 60 epochs with a batch size of 64, while all species-specific models (including *Salmonella*) were trained for 19 epochs with a batch size of 32. Optimization was performed using the Adam optimizer with the one-cycle learning rate schedule. For species-specific models, MixUp augmentation [**?**] was applied as a regularization strategy.

Loss functions were designed to account for both class imbalance and missing labels. For species-specific models, we used a masked, weighted binary cross-entropy loss with class-specific positive weights computed from the training set. For the general model, this was extended to a species-aware variant, in which a species-by-class weight matrix was applied dynamically according to the species identity of each batch.

Evaluation was carried out using cluster-aware cross-validation. For species-specific models, we used *GroupKFold* with clusters as the grouping variable. For the general model, we used *StratifiedGroupKFold*, which stratifies by species while grouping by clusters, ensuring balanced folds across species while preventing homologous overlap. Model checkpoints were saved based on the Matthews correlation coefficient (MCC) on the validation set.

Together, these design decisions established a neural architecture and training regimen optimized to preserve subtle genomic signals, handle class imbalance, and fairly evaluate performance across species, laying the foundation for the subsequent results presented in Section 4.

Implementation

The complete implementation was developed in Python within Jupyter notebooks, chosen to facilitate transparency, collaboration, and reproducibility. Due to the computationally intensive nature of the pipeline, the workflow was divided into separate modules, with each major step implemented in a dedicated notebook (data acquisition, table generation, drug selection, encoding, and training). This modular design was

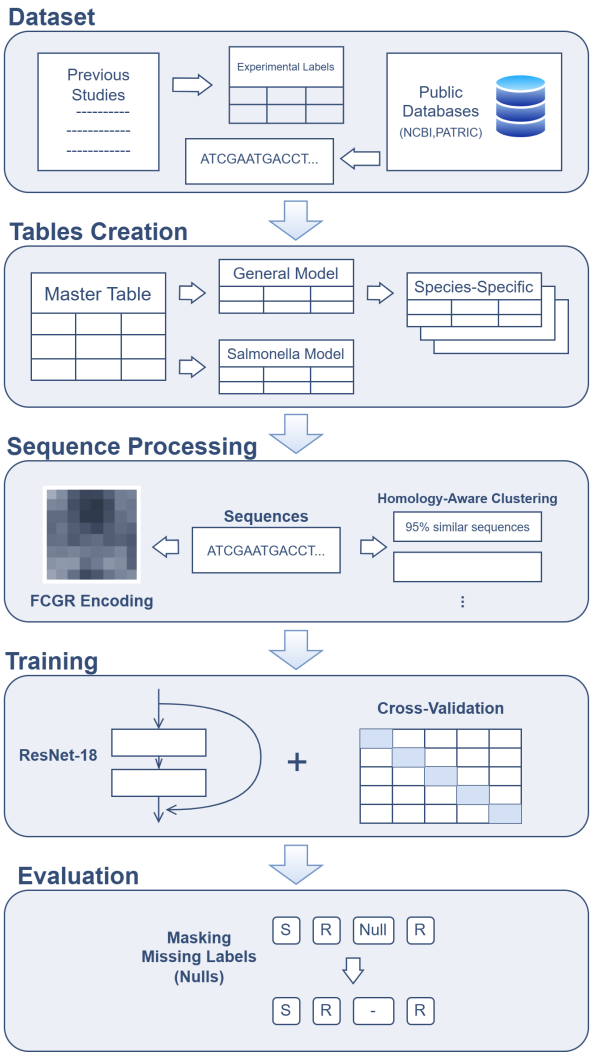

Fig. 5. Overview of the workflow pipeline

necessary as executing the entire workflow in a single notebook would often exceed the memory constraints of the Google Colab environment and lead to crashes. Separating time-consuming steps also minimized unnecessary recomputation, enabling experiments to be conducted with greater efficiency and reliability. The complete workflow is illustrated in Figure 5.

Software dependencies

All analyses were performed using Python 3, with the primary libraries and versions listed in Table S1. Each notebook in the repository specifies its imports explicitly, and, for reproducibility, a *environment.yml* file will be included in the GitHub repository to allow for direct installation of all dependencies through *Conda*. This ensures that results can be replicated in other Python environments without manual package resolution.

Hardware environment

All experiments, including data acquisition, preprocessing, and model training, were executed using the Google Colab platform. The standard configuration consisted of a single NVIDIA Tesla T4 GPU paired with 2 vCPUs @ 2.2 GHz, 13 GB of RAM, and 100 GB of available storage. This setup was selected primarily for its ease of accessibility and collaborative features, while also reflecting a common deep learning environment available to researchers free of charge. Training times per model varied by experiment: the species-agnostic (general) model required approximately one hour to complete training across 5-fold cross-validation, whereas species-specific models, such as the *Salmonella*-only pipeline, required approximately 30–40 minutes under the same evaluation strategy.

Code and data availability

All Jupyter notebooks used in this study—including those for preprocessing, model training, and results generation—are openly available at GitHub. The repository further includes automatically generated tables (stored without rounding to preserve full precision) and figure-production scripts. Raw sequencing data are available from the public repositories cited in Section 2, while all processed data required to reproduce analyses are provided in the GitHub repository. A DOI (e.g., via Zenodo) will be generated for the repository at the time of publication to ensure long-term archival access.

Results and Discussion

Having established the dataset preparation strategy, preprocessing steps, and model architecture in the previous section, we now present the results of our experiments. To interpret these results meaningfully, it is first necessary to define the evaluation metrics used throughout this study, with special attention to issues of class imbalance and missing labels that arise in antimicrobial resistance prediction.

Evaluation metrics

For binary classification, let TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively. Predictions $\hat{y}$ were obtained by thresholding model outputs (logits at 0.0, probabilities at 0.5). For the general model, samples with missing labels (NULLs) were ignored when computing metrics (masking). Metrics were computed per drug and averaged across drugs.

Balanced Accuracy accounts for class imbalance by averaging recall across both classes:

$$\text{Balanced Accuracy} = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right). \quad (1)$$

Precision measures the proportion of predicted resistant samples that are correct:

$$\text{Precision} = \frac{TP}{TP+FP}. \quad (2)$$

Recall (sensitivity) measures the proportion of true resistant samples that are correctly identified:

$$\text{Recall} = \frac{TP}{TP+FN}. \quad (3)$$

The F1 score is the harmonic mean of precision and recall:

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (4)$$

Table 3. Comparison of overall performance of FCGR models with different k-mer sizes (6-mer, 7-mer, 8-mer) on the independent test set. Values are rounded to two decimal places.

Model (k-mer)	Balanced Accuracy	Precision	Recall	F1	F_β	ROC AUC	Jaccard	MCC
6-mer	0.80	0.92	0.90	0.91	0.90	0.86	0.84	0.51
7-mer	0.83	0.94	0.91	0.91	0.92	0.90	0.87	0.56
8-mer	0.82	0.94	0.94	0.93	0.93	0.89	0.89	0.66

The F_β score generalizes F1 by weighting recall more heavily (here $\beta = 2$):

$$F_\beta = \frac{(1+\beta^2)\cdot \text{Precision}\cdot \text{Recall}}{\beta^2\cdot \text{Precision} + \text{Recall}}. \tag{5}$$

The Matthews Correlation Coefficient (MCC) is a balanced correlation measure even for imbalanced data:

$$\text{MCC} = \frac{TP\cdot TN - FP\cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \tag{6}$$

The Jaccard index measures the overlap between predicted and true resistant sets:

$$\text{Jaccard} = \frac{TP}{TP+FP+FN}. \tag{7}$$

The ROC AUC is the area under the receiver operating characteristic curve:

$$\text{ROC AUC} = \int_0^1 TPR(FPR^{-1}(x))\,dx, \tag{8}$$

where

$$TPR = \frac{TP}{TP+FN}, \quad FPR = \frac{FP}{FP+TN}. \tag{9}$$

Together, these metrics provide a comprehensive framework for evaluation: some emphasize balance under class skew (Balanced Accuracy, MCC), others capture precision–recall trade-offs (Precision, Recall, F_1, F_β), while ROC AUC and Jaccard quantify discriminative ability and overlap. Using this multi-metric evaluation approach allows us to assess model performance robustly, avoiding the pitfalls of relying on a single measure. In the next subsection, we apply these metrics to compare results across both species-specific and species-agnostic models.

Performance on independent test set

Having defined the evaluation framework in the previous subsection, we now apply these metrics to assess model performance on the independent test set. We begin by examining the *Salmonella*-specific model, which serves as a focused case study, before moving to the broader species-agnostic model.

Salmonella enterica-specific model

Table 3 compares the overall performance of the FCGR models when trained across several k-mer sizes (6-mer, 7-mer, and 8-mer) on the independent test set. Across all metrics, the 7-mer and 8-mer models outperformed the 6-mer model, with higher balanced accuracy, F1, and MCC values.

The 7-mer model achieved the highest overall balanced accuracy at 0.83 and solid performance across other metrics (F_β = 0.92, MCC = 0.55), while the 8-mer model achieved the best overall discriminative performance (MCC = 0.66, Jaccard = 0.89). In comparison, the 6-mer model lagged behind across all metrics, suggesting that 6-mers provide insufficient resolution to capture genomic patterns.

In all, these results suggest that increasing k-mer size improves model performance, though possibly stagnating as we increase beyond. All subsequent experiments were performed on 8-mer representations, though it is likely that 7-mer may be sufficient.

Having established 8-mer FCGRs as the representation of choice, we next evaluate the final *Salmonella enterica*-specific model across drugs on the independent test set.

Table 4 reports the final performance of the *Salmonella enterica*-specific model on the independent test set using 8-mer FCGR images. Metrics are reported overall and across drugs, with balanced accuracy serving as an imbalance-sensitive replacement for accuracy.

Overall, the model demonstrated strong performance across metrics (Balanced Accuracy = 0.82, F1 = 0.93, MCC = 0.66), indicating reliable predictive ability on a held-out test set. Performance significantly varied across drugs, however, achieving near-perfect predictions for cefoxitin, ceftiofur, and ceftriaxone (Balanced Accuracy $\geq$ 0.94, F1 $\geq$ 0.99, MCC $\geq$ 0.84), indicating a relative ease of classification for these antibiotics. gentamicin also showed high performance (Balanced Accuracy = 0.86, MCC = 0.80).

In contrast, the model struggled to predict tetracycline, ampicillin, and amoxicillin-clavulanic acid, with lower balanced accuracy (0.71–0.79) and MCC values (0.55–0.58).

It is likely that earlier observations made about Figure 1 serve as a direct explanation for the relatively stellar performance of cefoxitin, ceftiofur, and ceftriaxone. During training, the model can leverage the shared drug patterns across these antibiotics, effectively tripling the available training signal for each drug.

Taken together, these test results show that the *Salmonella*-specific model performs robustly overall, but that its performance appears drug-dependent.

Figure 6 shows the confusion matrices for the *Salmonella enterica*-specific model on the independent test set. These matrices support the performance trends reported in Table 4 with good predictive ability for cefoxitin, ceftiofur, and ceftriaxone and more frequent misclassification in the resistant class for tetracycline, ampicillin, and amoxicillin-clavulanic acid.

Building on these findings from a single species, we next turn to the general, species-agnostic model to assess whether shared genomic signals across species can improve prediction, particularly for underrepresented taxa.

Species-agnostic model

Table 5 compares the performance of the species-agnostic (general) model with species-specific models across species and drugs on the independent test set. Metrics are reported per

Table 4. Final performance of the *Salmonella enterica*-specific model on the independent test set for 8-mer fcgr images. Metrics are reported overall and per drug, rounded to two decimal places. Balanced accuracy is shown instead of standard accuracy.

Class	Balanced Accuracy	Precision	Recall	F1	F_β	ROC AUC	Jaccard	MCC
overall	0.82	0.94	0.94	0.93	0.93	0.89	0.89	0.66
tetracycline	0.79	0.92	0.66	0.77	0.70	0.90	0.62	0.58
ampicillin	0.71	0.83	0.98	0.90	0.94	0.93	0.81	0.55
amoxicillin-clavulanic acid	0.74	0.92	0.97	0.95	0.96	0.87	0.90	0.57
cefoxitin	0.94	0.99	0.99	0.99	0.99	0.95	0.98	0.86
ceftiofur	0.98	1.00	0.99	0.99	0.99	0.99	0.99	0.92
gentamicin	0.86	0.97	0.99	0.98	0.99	0.97	0.96	0.80
ceftriaxone	0.96	1.00	0.98	0.99	0.98	0.98	0.98	0.84

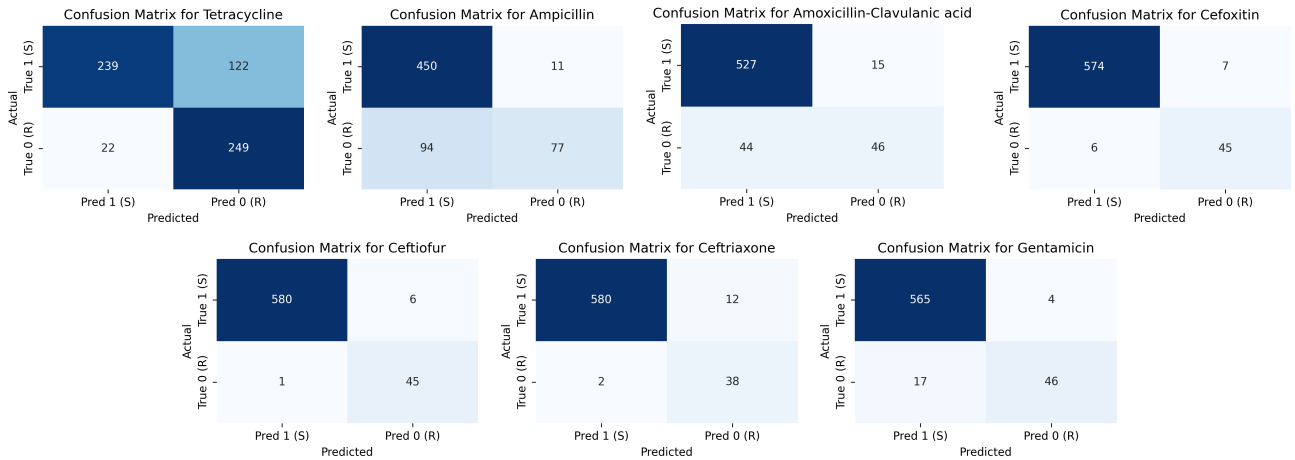

Fig. 6. Confusion matrices for the *Salmonella enterica*-specific 8-mer model on the independent test set, shown for all eight antibiotics. Each matrix is normalized and labeled by drug.

drug, per species, and overall. The species-specific models shown here were trained separately for the purpose of direct comparison, and are distinct from one another. Notably, the *Salmonella enterica* model included in this table is not the same as the main *Salmonella*-specific model presented throughout the paper, but rather a separate model trained on the two drugs tested in the general model in order to limit any possible interference.

For *E. coli*, the general model consistently outperformed the species-specific model, particularly on tetracycline and gentamicin, where balanced accuracy and MCC were substantially higher. This supports the previous hypothesis that the general model can leverage shared genomic patterns from similar species to improve performance on underrepresented species.

In contrast, for *Salmonella enterica*, the species-specific model achieved superior performance across most metrics, reflecting the abundant and high-quality labels in the *Salmonella* dataset. Its performance seemingly held back from optimizing for three species. For *Staphylococcus aureus*, results were more mixed, with the species-specific model performing somewhat better, though differences are not as pronounced.

Taken together, these results suggest that a species-agnostic model can perform comparably to models trained on individual species, and may compensate for weaker species datasets when similarities between species exist.

Cross-validation results

While the independent test set provides a strong indication of generalization ability, it represents only a single held-out split. To further assess the stability and robustness of model performance, we also conducted cross-validation. This allows performance to be evaluated across multiple folds of the training set, providing insight into consistency and variance across different partitions of the data. We first present results for the *Salmonella*-specific model, followed by those for the more challenging species-agnostic model.

Salmonella enterica-specific model

Table 6 reports the cross-validation performance of the *Salmonella enterica*-specific model. This model achieved a mean balanced accuracy of 0.81 ± 0.06 and an MCC of 0.58 ± 0.10, with precision, recall, and F1 scores of around 0.90. These results demonstrate that the *Salmonella*-specific model is not only accurate on a single test set but also stable across multiple folds, with low variance providing confidence in the robustness of its predictive patterns.

Having established consistency for the single-species case, we next turn to the species-agnostic model, where training data are more heterogeneous and label quality varies across species.

Species-agnostic model

Table 7 reports the cross-validation performance of the species-agnostic (general) model. The model achieved a mean balanced accuracy of 0.76 ± 0.04 and an MCC of 0.49 ± 0.09, with

Table 5. Comparison of the species-agnostic (general) model and species-specific models on the independent test set. Metrics are reported per species and drug, rounded to two decimals.

Species	Drug	Model	Bal. Acc	Precision	Recall	F1	ROC AUC	Jaccard	MCC
E. coli	tetracycline	General	0.66	0.50	0.46	0.48	0.64	0.32	0.33
		Specific	0.49	0.00	0.00	0.00	0.52	0.00	-0.07
	gentamicin	General	0.70	0.92	0.92	0.92	0.82	0.86	0.40
		Specific	0.72	0.92	0.94	0.93	0.84	0.87	0.46
Salmonella enterica	tetracycline	General	0.78	0.76	0.98	0.86	0.91	0.75	0.64
		Specific	0.84	0.84	0.92	0.88	0.91	0.78	0.70
	gentamicin	General	0.83	0.96	0.99	0.98	0.96	0.96	0.77
		Specific	0.93	0.99	0.99	0.99	0.98	0.98	0.88
Staphylococcus aureus	tetracycline	General	0.68	0.73	0.62	0.67	0.77	0.51	0.37
		Specific	0.64	0.67	0.63	0.65	0.65	0.48	0.28
	gentamicin	General	0.69	0.81	0.72	0.76	0.74	0.61	0.36
		Specific	0.68	0.77	0.89	0.82	0.77	0.70	0.40
Overall (General)			0.73	0.83	0.88	0.84	0.84	0.76	0.47

Table 6. Cross-validation performance of the *Salmonella enterica*-specific model. Values are reported as mean ± standard deviation across folds, rounded to two decimals.

	Val. Loss	Bal. Acc	Prec.	Recall	F1	F_β	ROC AUC	Jaccard	MCC
Mean	0.20	0.81	0.92	0.90	0.90	0.89	0.87	0.83	0.58
SD	0.04	0.06	0.03	0.05	0.05	0.05	0.05	0.07	0.10

Table 7. Cross-validation performance of the species-agnostic (general) model. Values are reported as mean ± standard deviation across 5 folds, rounded to two decimals.

	Val. Loss	Bal. Acc	Prec.	Recall	F1	F_β	ROC AUC	Jaccard	MCC
Mean	0.50	0.76	0.84	0.84	0.82	0.83	0.84	0.74	0.49
SD	0.15	0.04	0.06	0.02	0.04	0.03	0.04	0.05	0.09

precision, recall, and F1 scores consistently around 0.82–0.84. Variance across folds remained moderate, indicating stable convergence even under heterogeneous data conditions.

In comparison to the *Salmonella*-specific model, however, the general model shows weaker performance, which is likely attributable to the noisier or incomplete labels found in *E. coli* and *Staphylococcus aureus*. Moreover, the masking of missing values across drugs may further reduce effective training signal. These observations reinforce the theme already noted in the test set evaluation: the general model is more sensitive to dataset quality, but compensates by leveraging shared genomic patterns for underrepresented species.

In summary, cross-validation results complement the independent test set findings by confirming that both species-specific and general models remain stable across multiple folds, albeit with consistent performance advantages for the *Salmonella*-specific setup.

Comparison to baselines

To better contextualize the performance of our deep learning approach, we next compare it against widely used gene-based and classical machine learning baselines. This allows us to evaluate whether sequence-based CNN models provide tangible advantages over conventional predictors of antimicrobial resistance.

To establish a gene-based baseline with already existing tools, we evaluated predictions from *ABRicate* [7] using the *ResFinder* resistance gene database [15]. For each genome assembly, *ABRicate* was run across all samples, and the resulting resistance gene labels were mapped to antibiotic classes using a curated drug–class dictionary (details on the mapping available in Section 3). This mapping translated detected resistance genes into binary resistance/susceptibility (R/S) labels for each antibiotic of interest. If a genome contained at least one resistance gene corresponding to the drug's class, the isolate was labeled as resistant (R); otherwise, it was labeled as susceptible (S). Intermediate (I) labels were merged with resistant, consistent with previous decisions.

These predictions were then treated as results from a pretrained model, and evaluated using the same metrics to serve as a baseline for comparison. When comparing the ResFinder results in Table 8 with the *Salmonella enterica*-specific model in Table 4, several important contrasts emerge.

Differences are particularly evident in Balanced Accuracy (BA) and MCC. For tetracycline, ResFinder reports a high BA of 0.98 and MCC of 0.96, whereas the CNN model achieves only 0.79 BA and 0.58 MCC, indicating reduced balanced detection performance. For ampicillin, ResFinder again shows strong results (BA = 0.96, MCC = 0.94) in contrast to the CNN model (0.71 BA, 0.55 MCC), suggesting the database method is more balanced overall. In amoxicillin-clavulanic acid, BA (0.96 vs. 0.74) and MCC (0.95 vs. 0.57) similarly favor ResFinder. In contrast, for cefoxitin, ResFinder achieves a BA of 0.97 and MCC of 0.88, but the CNN model performs more competitively with a BA of 0.94 and MCC of 0.86, showing improved class balance under test conditions. Finally, for gentamicin, ResFinder reports 0.95 BA and 0.94 MCC, while the CNN model, though slightly weaker (BA = 0.86, MCC = 0.80), still demonstrated strong predictive consistency.

Taken together, ResFinder consistently reports higher balanced accuracy across a majority of drugs, while the CNN model sometimes narrows the gap, particularly for cefoxitin and related cephalosporins. Notably, the species-agnostic CNN model, despite being trained without direct resistance gene annotations and only across a selected set of pathogens, approaches curated database-based performance for some drugs in *Salmonella*. This suggests that sequence-derived, species-independent features can capture underlying resistance signals, highlighting the model's potential for generalization beyond curated gene lists.

To further benchmark against non-deep learning methods, we also trained a Random Forest classifier on the same *Salmonella* dataset using gene presence/absence features. Performance was heterogeneous: while recall, F1, and AUROC values were high with respect to the majority class (e.g., F1 >0.95 for ceftriaxone, gentamicin, and ceftiofur), the balanced accuracy remained modest (0.56–0.65 across drugs), indicating poor discrimination of resistant isolates. MCC values ranged from 0.14 (tetracycline) up to 0.51 (ceftriaxone), reflecting only moderate capture of phenotype variance.

In summary, the Random Forest provided slightly more informative predictions than ABRicate with ResFinder, but still underperformed compared to our deep learning approach. These comparisons reinforce the need for more expressive models—like CNNs with FCGR representations—for drug-specific AMR prediction, particularly in cases where gene-based databases are incomplete or limited in scope.

Biological and practical implications

Having compared our models against established baselines, we next turn to the biological significance of the learned features. Beyond raw predictive metrics, an important question is whether our models recognize sequence-level signals of resistance that are clinically relevant.

Table 8. Performance metrics of ResFinder predictions for the species-agnostic model evaluated on *Salmonella enterica*, *Staphylococcus aureus*, and *Escherichia coli*. Values are rounded to two decimals. Note: Cephalosporins were streamlined in *Salmonella* by removing Ceftiofur (lacks direct mapping) and Ceftriaxone (redundant with other cephalosporins).

Species	Drug	Bal. Acc	Prec.	Recall	F1	$F_{\beta=0.5}$	ROC AUC	Jaccard	MCC
Salmonella enterica	tetracycline	0.98	0.99	0.98	0.98	0.99	0.98	0.97	0.96
	ampicillin	0.96	0.99	0.93	0.96	0.98	0.96	0.92	0.94
	amoxicillin-clavulanic acid	0.96	0.98	0.93	0.96	0.97	0.96	0.92	0.95
	cefoxitin	0.97	0.83	0.96	0.89	0.85	0.97	0.80	0.88
	gentamicin	0.95	0.98	0.91	0.94	0.96	0.95	0.89	0.94
Staphylococcus aureus	tetracycline	0.86	0.68	0.90	0.78	0.72	0.86	0.64	0.68
	gentamicin	0.89	0.79	0.86	0.82	0.80	0.89	0.70	0.76
Escherichia coli	tetracycline	0.70	0.21	0.96	0.34	0.24	0.70	0.20	0.28
	gentamicin	0.96	0.96	0.93	0.95	0.96	0.96	0.90	0.93

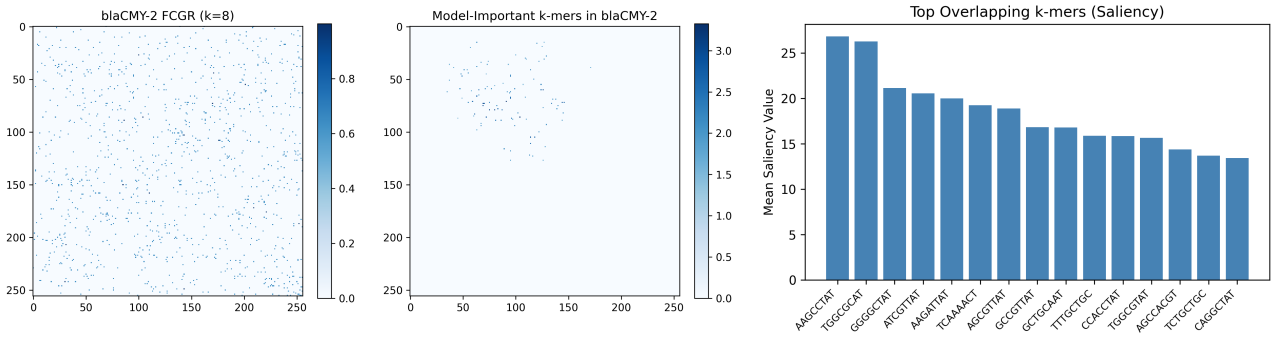

Fig. 7. In total, 102 out of 2814 model-important k-mers (3.6%) were present in blaCMY-2, accounting for 5.0% of the overall weighted saliency mass.

To explore sequence features underlying resistance for the *Salmonella* model, saliency maps were generated for 16 resistant genomes and used to rank k-mers by contribution to the resistance label. Saliency analysis focused on Ceftiofur, which showed strong predictive performance within the trained model. For each genome, the top 3000 salient k-mers were selected based on the elbow of the ranked saliency distribution, and the union across genomes was analyzed. This allowed us to identify k-mers that consistently appeared among the most important features across multiple resistant genomes, which we defined as "common" resistance-associated motifs. Their mean saliency values provided a global ranking of sequence elements most predictive of resistance.

Building on this, we next asked whether the model-important k-mers were specifically enriched within *blaCMY-2*, a clinically relevant β-lactamase associated with ceftiofur resistance. The *blaCMY-2* sequence [25] was visualized as an 8-mer frequency chaos game representation (FCGR) and overlaid with the common salient k-mers. In total, 102 of the 2814 globally important k-mers (3.6%) were found in *blaCMY-2*, corresponding to 5.0% of the weighted saliency mass, defined as

$$\text{Weighted Saliency (\%)} = \frac{\sum_{k \in \mathcal{K}_{\text{gene}}} S(k)}{\sum_{k \in \mathcal{K}_{\text{total}}} S(k)} \times 100.$$

This relatively modest overlap suggests that the model does not treat *blaCMY-2* as the sole driver of its resistance prediction. Nevertheless, several of the overlapping k-mers ranked highly by saliency (Fig. 7), indicating that the model may still register *blaCMY-2* through a small set of informative motifs. At the same time, the results are consistent with the possibility that the model relies more heavily on additional motifs distributed elsewhere in the genome, rather than exclusively on specific genes like *blaCMY-2*.

Taken together, this interpretation underscores a key practical implication: deep learning models using FCGR inputs can identify distributed genomic features that extend beyond known resistance genes. This hints at their potential to reveal novel or polygenic contributions to antimicrobial resistance, while still retaining sensitivity to established markers when they are present.

Limitations and future work

While our results demonstrate the feasibility and competitive performance of FCGR-based deep learning models for predicting antimicrobial resistance, several important limitations must be acknowledged. Recognizing these constraints helps place our findings in context and guides directions for future research.

This study has several important limitations. First, the datasets used for *E. coli* and *S. aureus* were of relatively low quality, with a large proportion of missing or incomplete antibiotic resistance labels. This limited the effective sample size and introduced a species imbalance in the agnostic model as well as reducing the overall scope of the models that could reasonably be trained. Second, while the computational requirements for training the current ResNet-18 models were manageable, the preprocessing pipeline for generating FCGR matrices was not. The training of larger, more expressive models will require substantially

greater computational resources. These constraints restricted the exploration of deeper architectures and more extensive hyperparameter searches.

Future work will focus on addressing these limitations. A key priority is the collection and curation of more comprehensive, high-quality, and fully labeled datasets across multiple bacterial species. With such datasets, stronger and more generalizable models can be trained, allowing for improved prediction performance and broader applicability. In addition to this, generalized models may be leveraged to identify species-agnostic biomarkers of resistance. By analyzing saliency maps from trained models, it may be possible to extract recurrent k-mer patterns that act as "fingerprints" of resistance across species. Since each position in an FCGR corresponds to a specific k-mer, these saliency maps could be reverse-mapped to genomic regions, potentially revealing conserved resistance-associated motifs. Combining this approach with signal processing techniques may provide a novel framework for biomarker discovery, enabling predictive modeling with biological interpretability.

Taken together, these limitations highlight both the challenges and opportunities of applying deep learning to AMR prediction. Addressing them will not only enhance technical performance but also deepen the biological insights that can be drawn from such models, providing a clear path for future work at the intersection of computational modeling and infectious disease genomics.

Conclusion

In this work, we showed how effective Frequency Chaos Game Representation (FCGR) is when paired with ResNet-18 for predicting antimicrobial resistance directly from bacterial genomes, without needing predefined resistance gene databases. The model specific to *Salmonella enterica* achieved high predictive accuracy, especially for cephalosporins where resistance mechanisms are well captured through shared genomic patterns. Meanwhile, the species-agnostic model was able to generalize across various pathogens, improving predictions for species with limited or imbalanced data, such as *E. coli*. These findings suggest that genomic signals of resistance are conserved enough to be used by deep learning methods.

When we compared our models with traditional gene-based methods like ResFinder, they reached similar performance levels in some cases. However, they also offered the benefit of possibly identifying new, database-independent resistance signatures. We ensured model robustness by using homology-aware clustering to avoid data leakage. This strengthened the credibility of both our experimental results and overall conclusions.

Still, there are challenges to address. Curated datasets are often small and lack complete labels, and preprocessing pipelines for FCGR require a lot of computing power. To overcome these issues, we need to create larger, high-quality multi-species collections and use more advanced deep learning architectures. This will be essential for closing the performance gap with curated gene-based methods and for moving toward systems that can be used in clinical settings.

Overall, this study demonstrates the potential of FCGR-based deep learning for rapid, scalable, and species-agnostic AMR prediction. By merging computational efficiency with biological relevance, this framework lays the groundwork for next-generation genomic surveillance tools that can respond to an evolving resistance landscape and improve antimicrobial stewardship efforts.

Acknowledgments

We thank Sama S. Eltaher (Zewail City University of Science and Technology) for providing biological insights and guidance on bioinformatics tools, which greatly supported this study. We also acknowledge our colleague Mariam I. Mohamed for valuable assistance with data collection.

References

1. Jose M. Munita and Cesar A. Arias. Mechanisms of antibiotic resistance. *Microbiology Spectrum*, 4(2):10.1128/microbiolspec.vmbf–0016–2015, 2016.
2. Alexander Zagajewski, Piers Turner, Conor Feehily, Hafez El Sayyed, Monique Andersson, Lucinda Barrett, Sarah Oakley, Mathew Stracy, Derrick Crook, Christoffer Nellåker, Nicole Stoesser, and Achillefs N. Kapanidis. Deep learning and single-cell phenotyping for rapid antimicrobial susceptibility detection in escherichia coli. *Communications Biology*, 6(1):1164, Nov 2023.
3. Antti Vasala, Vesa P. Hytönen, and Olli H. Laitinen. Modern tools for rapid diagnostics of antimicrobial resistance. *Frontiers in Cellular and Infection Microbiology*, Volume 10 - 2020, 2020.
4. Anna G. Green, Chang Ho Yoon, Michael L. Chen, Yasha Ektefaie, Mack Fina, Luca Freschi, Matthias I. Gröschel, Isaac Kohane, Andrew Beam, and Maha Farhat. A convolutional neural network highlights mutations relevant to antimicrobial resistance in mycobacterium tuberculosis. *Nature Communications*, 13(1):3817, Jul 2022.
5. Rui Tang, Rui Luo, Shiwei Tang, Haoxin Song, and Xiujuan Chen. Machine learning in predicting antimicrobial resistance: a systematic review and meta-analysis. *International Journal of Antimicrobial Agents*, 60(5):106684, 2022.
6. Tabish Ali, Sarfaraz Ahmed, and Muhammad Aslam. Artificial intelligence for antimicrobial resistance prediction: Challenges and opportunities towards practical implementation. *Antibiotics*, 12(3), 2023.
7. Torsten Seemann. Abricate: Mass screening of contigs for antimicrobial and virulence genes. `https://github.com/tseemann/abricate`, 2018. Accessed: YYYY-MM-DD.
8. Nenad Macesic, Oliver J. Bear Don't Walk, Itsik Pe'er, Nicholas P. Tatonetti, Anton Y. Peleg, and Anne-Catrin Uhlemann. Predicting phenotypic polymyxin resistance in klebsiella pneumoniae through machine learning analysis of genomic data. *mSystems*, 5(3):10.1128/msystems.00656–19, 2020.
9. Daniel Lichtblau. Alignment-free genomic sequence comparison using fcgr and signal processing. *BMC Bioinformatics*, 20(1):742, Dec 2019.
10. Hannah Franziska Löchel and Dominik Heider. Chaos game representation and its applications in bioinformatics. *Comput Struct Biotechnol J*, 19:6263–6271, nov 2021.
11. Jorge Avila Cartes, Santosh Anand, Simone Ciccolella, Paola Bonizzoni, and Gianluca Della Vedova. Accurate and fast clade assignment via deep learning and frequency chaos game representation. *GigaScience*, 12:giac119, 12 2022.

12. Y. Ren, T. Chakraborty, S. Doijad, L. Falgenhauer, J. Falgenhauer, A. Goesmann, A. C. Hauschild, O. Schwengers, and D. Heider. Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning. *Bioinformatics*, 38(2):325–334, Jan 2022.
13. Hazrat Bilal, Muhammad Nadeem Khan, Sabir Khan, Muhammad Shafiq, Wenjie Fang, Rahat Ullah Khan, Mujeeb Ur Rahman, Xiaohui Li, Qiao-Li Lv, and Bin Xu. The role of artificial intelligence and machine learning in predicting and combating antimicrobial resistance. *Computational and Structural Biotechnology Journal*, 27:423–439, 2025.
14. Melis N. Anahtar, Jason H. Yang, and Sanjat Kanjilal. Applications of machine learning to the problem of antimicrobial resistance: an emerging model for translational research. *Journal of Clinical Microbiology*, 59(7):10.1128/jcm.01260–20, 2021.
15. E. Zankari, H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F. M. Aarestrup, and M. V. Larsen. Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*, 67(11):2640–2644, Nov 2012.
16. Saeedeh Akbari Rokn Abadi, Amirhossein Mohammadi, and Somayyeh Koohi. WalkIm: Compact image-based encoding for high-performance classification of biological sequences using simple tuning-free CNNs. *PLoS One*, 17(4):e0267106, April 2022.
17. Wei Wang, Michelle Baker, Yue Hu, Jin Xu, Dajin Yang, Alexandre Maciel-Guerra, Ning Xue, Hui Li, Shaofei Yan, Menghan Li, Yao Bai, Yinping Dong, Zixin Peng, Jinjing Ma, Fengqin Li, and Tania Dottorini. Whole-genome sequencing and machine learning analysis of staphylococcus aureus from multiple heterogeneous sources in china reveals common genetic traits of antimicrobial resistance. *mSystems*, 6(3):10.1128/msystems.01185–20, 2021.
18. Finlay Maguire, Muhammad Attiq Rehman, Catherine Carrillo, Moussa S. Diarra, and Robert G. Beiko. Identification of primary antimicrobial resistance drivers in agricultural nontyphoidal salmonella enterica serovars by using machine learning. *mSystems*, 4(4):10.1128/msystems.00211–19, 2019.
19. Nicole Pearcy, Yue Hu, Michelle Baker, Alexandre Maciel-Guerra, Ning Xue, Wei Wang, Jasmeet Kaler, Zixin Peng, Fengqin Li, and Tania Dottorini. Genome-scale metabolic models and machine learning reveal genetic determinants of antibiotic resistance in escherichia coli and unravel the underlying metabolic adaptation mechanisms. *mSystems*, 6(4):10.1128/msystems.00913–20, 2021.
20. Martin Steinegger and Johannes Söding. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028, 2017.
21. Li Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152, December 2012.
22. Luiz Irber, N. Tessa Pierce-Ward, Mohamed Abuelanin, Harriet Alexander, Abhishek Anant, Keya Barve, Colton Baumler, Olga Botvinnik, Phillip Brooks, Daniel Dsouza, Laurent Gautier, Mahmudur Rahman Hera, Hannah Eve Houts, Lisa K. Johnson, Fabian Klötzl, David Koslicki, Marisa Lim, Ricky Lim, Bradley Nelson, Ivan Ogasawara, Taylor Reiter, Camille Scott, Andreas Sjödin, Daniel Standage, S. Joshua Swamidass, Connor Tiffany, Pranathi Vemuri, Erik Young, and C. Titus Brown. sourmash v4: A multitool to quickly search, compare, and analyze genomic and metagenomic data sets. *Journal of Open Source Software*, 9(98):6830, 2024.
23. Abhishek Halder, Piyush, Bernadette Mathew, and Debarka Sengupta. fcgr: A python library for accurate frequency chaos game representation of dna sequences. `https://github.com/abhi1238/FCGR/tree/main`, 2023. Accessed: 2025-09-07.
24. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
25. NCBI GenBank. blacmy-2 gene, genbank accession jn714983.1. `https://www.ncbi.nlm.nih.gov/nuccore/JN714983.1`, 2025. Accessed: 10 September 2025.

Saifelden M. Ismail

Saifelden M. Ismail is a senior-year Communications and Information Engineering student with a broad interest in computer science. His academic background includes systems programming and software development, alongside experience in bioinformatics.

Salma H. Fayed .

Salma H. Fayed is a senior-year Communications and Information Engineering student.

Supplementary Material

This section provides supplementary information related to implementation and reproducibility. Tables S1–S2 document the software libraries and code resources used in this study.

Table S1. Python libraries used in the pipeline.

Core Libraries
PyTorch (v2.8.0+cu126)
FastAI (v2.8.4)
Scikit-learn (v1.6.1)
Pandas (v2.2.2)
NumPy (v2.0.2)
Bioinformatics Libraries
Sourmash (v4.9.4)
BioPython (v1.85)
ABRicate (v1.0.1)
Visualization Libraries
Matplotlib (v3.10.0)
Seaborn (v0.13.2)
Plotly (v5.24.1)

Table S2. Code access on GitHub.

Code resources and access links
Dataset Download: Automated scripts for downloading WGS data from NCBI/PATRIC [GitHub]
ABRicate Prediction Generation: Script for generating ABRicate predictions and storing them [GitHub]
ABRicate Metrics Generation: Script for evaluating ABRicate results using the same metrics [GitHub]
Master Table Creation: Integration of phenotypic data with genomic metadata [GitHub]
FCGR Encoding: Frequency Chaos Game Representation image generation [GitHub]
Drug Selection: Filtering & quality-based drug selection and dataset filtering [GitHub]
Homology-Aware Splitting: Sequence similarity clustering for data leakage prevention [GitHub]
Model Training Pipeline: PyTorch/FastAI implementation with cross-validation [GitHub]
Interpretability Analysis: Saliency map and feature interpretation pipeline (used in Biological and practical implications section) [GitHub]
Baseline Random Forest: Gene presence/absence–based Random Forest baseline model for comparison [GitHub]