Predicting Antimicrobial MICs and Associated Genomic Features for Salmonella: A Multilayer Perspective

Ramyasri Veerapaneni, Participant, MSU/USDA Graduate Summer Research Experience Program

Mentor: Dr. Zhiqian Chen

Introduction

- A bacterium causing illness in humans and animals
- Commonly found in animal intestines and can contaminate various food types
- the second leading cause of foodborne illness in the U.S
- Salmonella is a genus of bacteria that includes different species. Differentiating between its species and serotypes is crucial for disease prevention, as each variation poses unique health risks.
- Different types of Salmonella have diverse characteristics.
- K-mers is a k-base pairs from the DNA, short genome sequences used to compare and understand different Salmonella types.

Objective

- **Primary aim:** Predict Minimum Inhibitory Concentration (MIC) values
- **Distinct Approach:** Regression-based approach for continuous MIC values, different from previous classification task
- **Methodology:** Use of Graph Convolutional Networks and Feature Extraction
- **Data:** MIC values and whole genome sequences of Salmonella from NARMS

Model Architecture

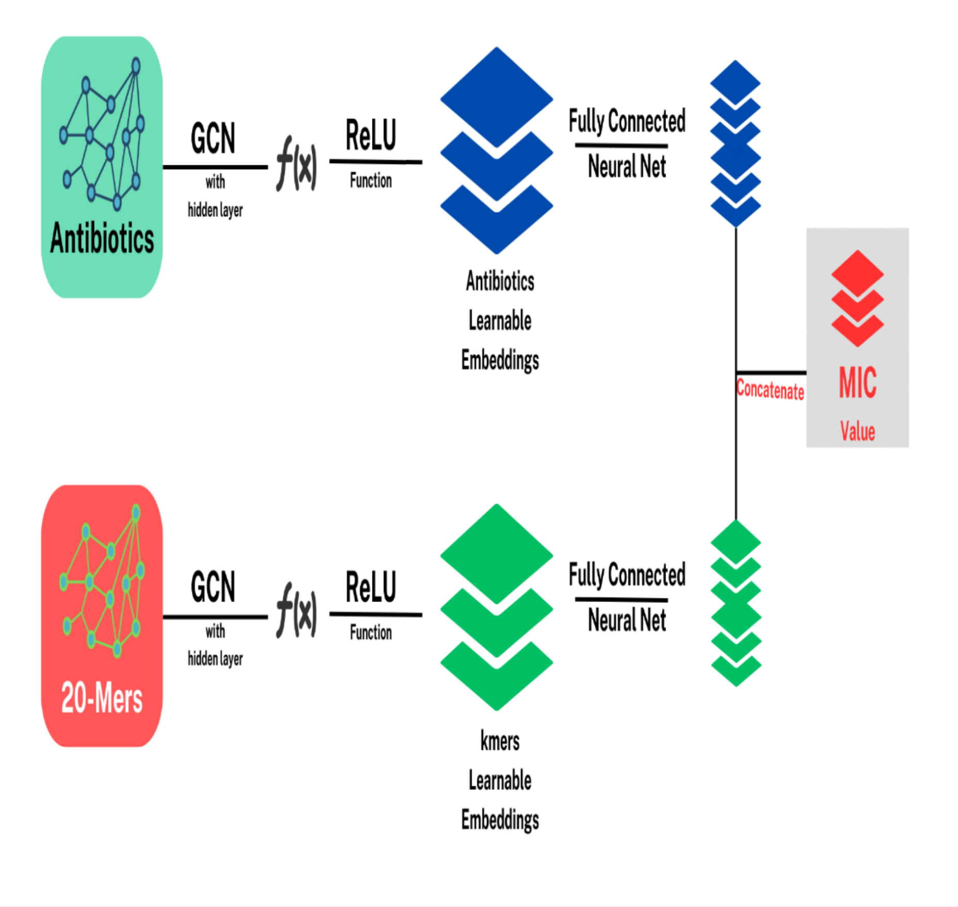

Figure 1: GCN model architecture

- Throughout this process, the model is essentially learning how to map the original features of antibiotics and k-mers and their relationships to the MIC values.
- This is how it acquires the ability to predict MIC values, helping us understand antibiotic resistance.

Multilayer Network

- **Nodes**: Elements (antibiotics/20-mers).
- **Edges**: Connections based on shared attributes.
- **Interlayer edges**: Represent influence of k-mers on MIC of antibiotics
- **MIC:** Lowest concentration of an antibiotic that stops bacterial growth

Results

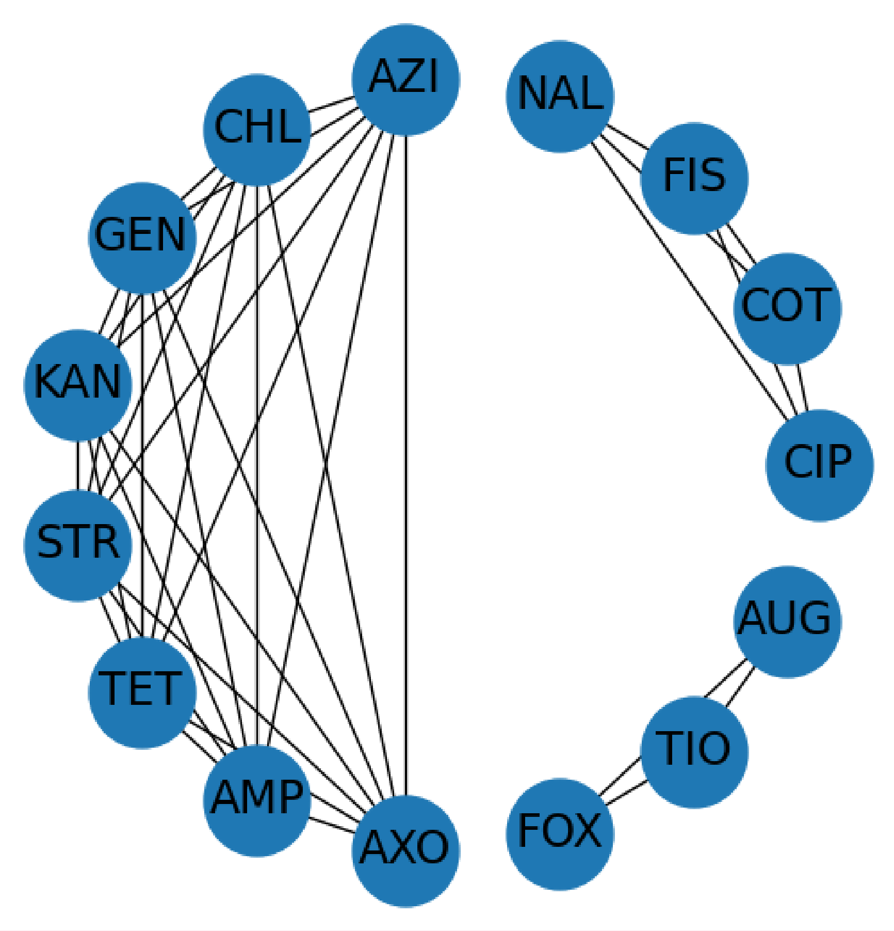

Figure 2: Antibiotic Similarity

- The consistent downward trend in both MSE and MAPE indicates that the model's predictions are becoming progressively more accurate, with lower error values.
- These results highlight the effectiveness of the GCN approach for predicting the MIC values and pave the way for further improvements and optimization.

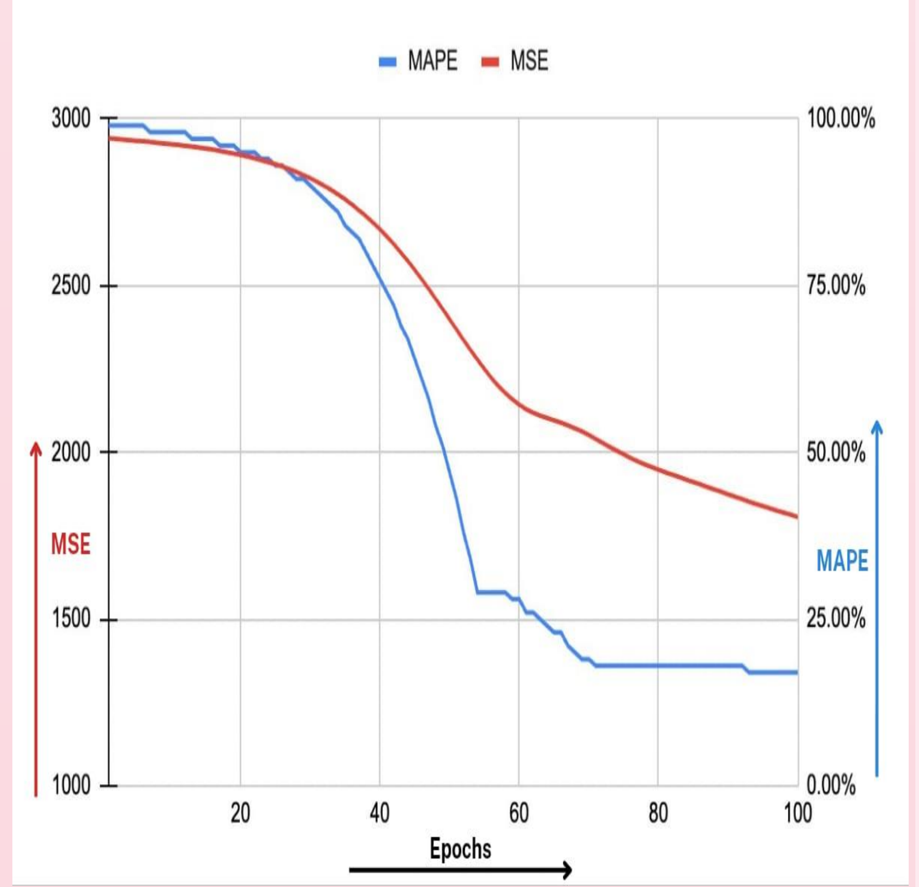

Figure 3: Result for MIC values

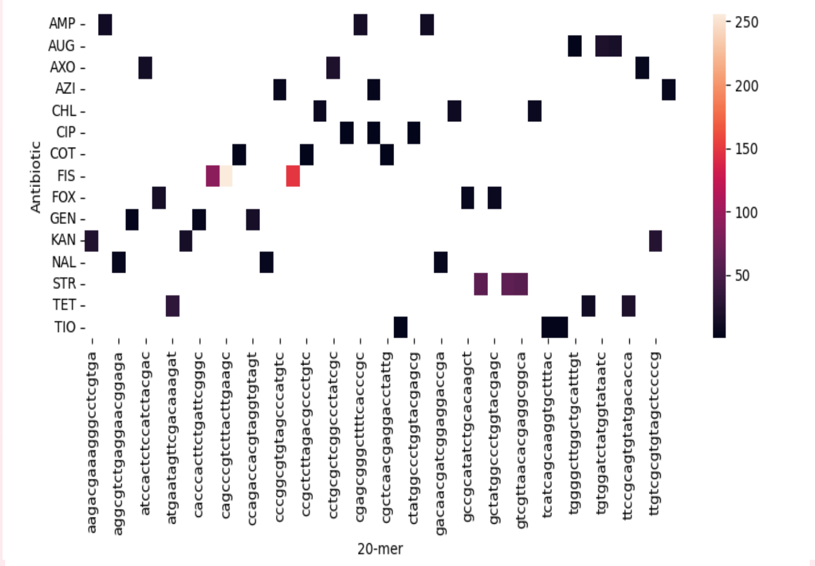

Figure 4: Antibiotic-20-mer Interactions: An MIC Heatmap

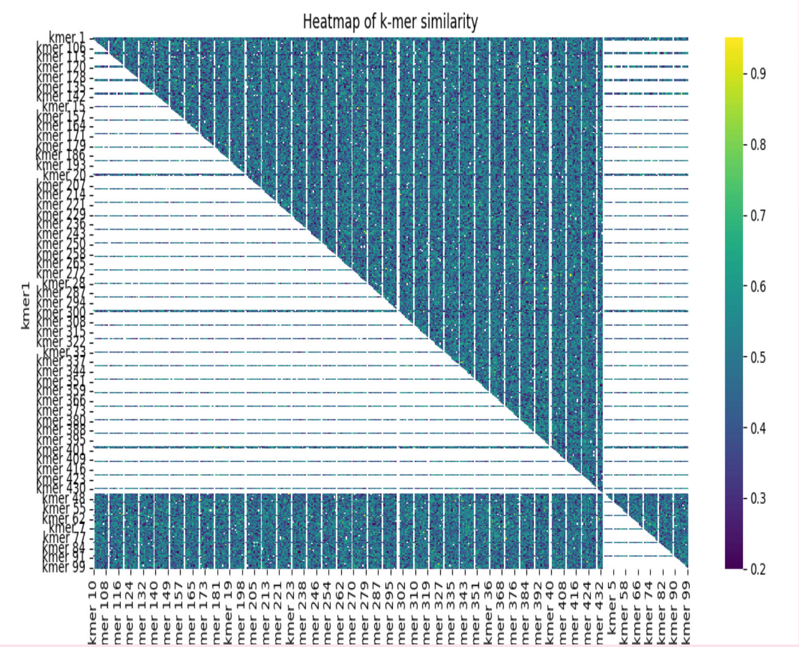

Figure 5: k-mer Similarity

Conclusion

- Utilizes genomic features to capture intricate k-mer, antibiotic, and MIC relationships
- Regression-based approach for continuous MIC values, different from previous classification task
- Demonstrated learning capability shown by decreasing MSE and MAPE over training epochs

Future Work : Explore diverse network architectures and incorporate additional data sources to improve predictions