AMRITA VISHWA VIDYAPEETHAM

Computational Framework for Understanding Microbial Pathogenesis and Antimicrobial Resistance (AMR)

Molecular Biology and Basic Cellular Physiology (24AIM112)
Ethics, Innovative Research, Businesses & IPR (24AIM115)

RAGUL U. - CB.AI.U4AIM24036
RAMKUMAR R. - CB.AI.U4AIM24033
SHWETHA P. - CB.AI.U4AIM24042
PRAGALYA M. - CB.AI.U4AIM24032

Introduction

Antimicrobial Resistance (AMR):

- AMR is when microbes resist drugs, making infections harder to treat.
- Understanding the genetic mechanisms behind AMR is crucial for developing effective prevention and treatment strategies.
- At the same time, understanding protein virulence is crucial for identifying pathogenic factors that contribute to disease progression

AMRITA VISHWA VIDYAPEETHAM

Objective

- Predict Antimicrobial Resistance (AMR)
- Prediction of Virulence factor in Protein
- Apply Machine Learning for Predictions
- UI for easy use
- Provide a Scalable and Cost-Effective Approach

PROGRESS

Data collected

Collection of the data :

- Gene sequence

```
Sequence ID,Description,DNA Sequence,Class
```

- Protein sequence

```
Protein_ID,Sequence,Label
```

Sources:

- NCBI (National Centre for Biotechnology Information) for gene sequences
- UniProt for protein sequences

AMR Prediction using CNN (K-mer + TF-IDF)

In [6]:
```
new_sequence_kmers = " ".join(get_kmers("ATGCGTACGTAGCTAGC"))
new_sequence_tfidf = vectorizer.transform([new_sequence_kmers]).toarray().reshape(1, -1, 1)
cnn_prediction = model_cnn.predict(new_sequence_tfidf)

print("CNN Prediction:", cnn_prediction)
```

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 126ms/step
CNN Prediction: [[0.95005983]]
```

In [10]:
```
# Example new gene sequence
new_sequence = "GGGGGCCGCCTTCGCCACCGGTATTCCTCCAGATCTCTACGCATTTCACCGCTACACCTGGAATTCTACCCC

# Make prediction
predicted_class, confidence = predict_amr_cnn(new_sequence, model_cnn, vectorizer)

# Print results
print(f"Predicted Class: {'AMR (1)' if predicted_class == 1 else 'Non-AMR (0)'}")
print(f"Prediction Confidence: {confidence:.4f}")
```

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 66ms/step
Predicted Class: AMR (1)
Prediction Confidence: 0.9501
```

Protein virulence prediction

MODEL USED	ACCURACY	Feature of the model
ProtBert + Random Forest	92%	Learns relationship between sequences
CNN + BiLSTM	93%	Captures the structural patterns

```
In [51]:
# Test on a new sequence
new_seq = "MGGRWSKSSIVGWPAIRERIRRTDPAADGVGAVSRDLEKHGAITSSNTRGTNADCAWLEAQEESEEVGFPVRPQVPLRPMTYKGALDLSHFLKEKGG"
new_seq_encoded = one_hot_encode(new_seq).reshape(1, 200, 20)

# Predict
prob = model.predict(new_seq_encoded)[0][0]
pred_class = 1 if prob > 0.5 else 0

print(f"Predicted Class: {'Virulent (1)' if pred_class == 1 else 'Non-Virulent (0)'}")
print(f"Prediction Confidence: {prob:.4f}")

1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 71ms/step
Predicted Class: Virulent (1)
Prediction Confidence: 0.8128
```

Fig. Output obtained from the final model (CNN+BiLSTM)

Problems faced

Class imbalance in dataset

- Extracting gene and protein sequences from NCBI (AMR) & UniProt (Virulence) was time-consuming and complex .The dataset had a skewed distribution (more AMR/Virulent sequences than non-AMR/non-virulent ones)

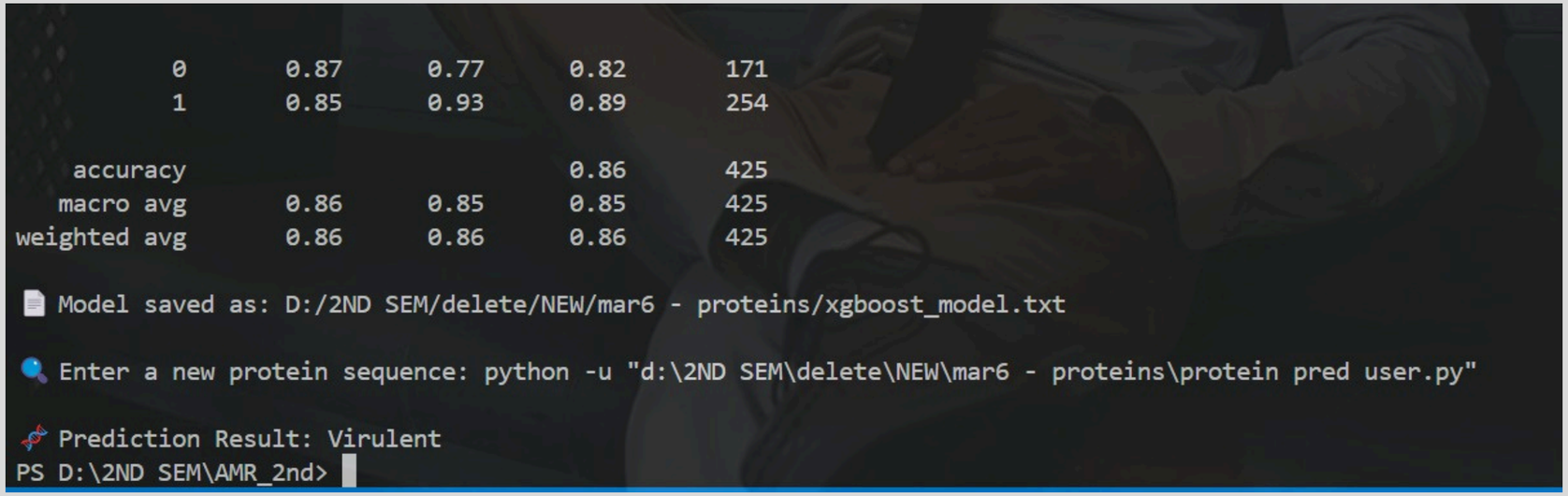
1319 -Virulence protein sequence, 480 -AMR gene sequence,
807-Non Virulence protein sequences 124- Non AMR gene sequences

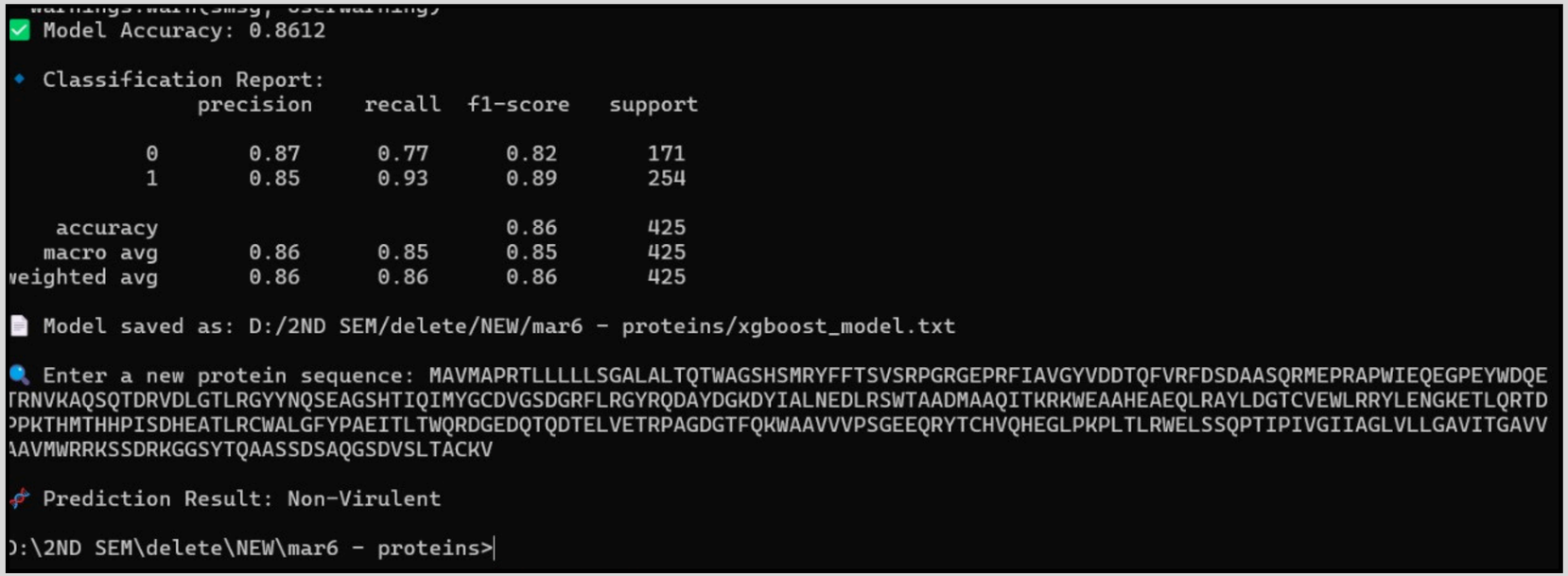
Overfitting in CNN Models

- Initial models trained on gene and protein sequences using CNNs showed high accuracy but suffered from overfitting due to class imbalance.
- Switched from raw sequence input to Word2Vec embeddings to generalize better

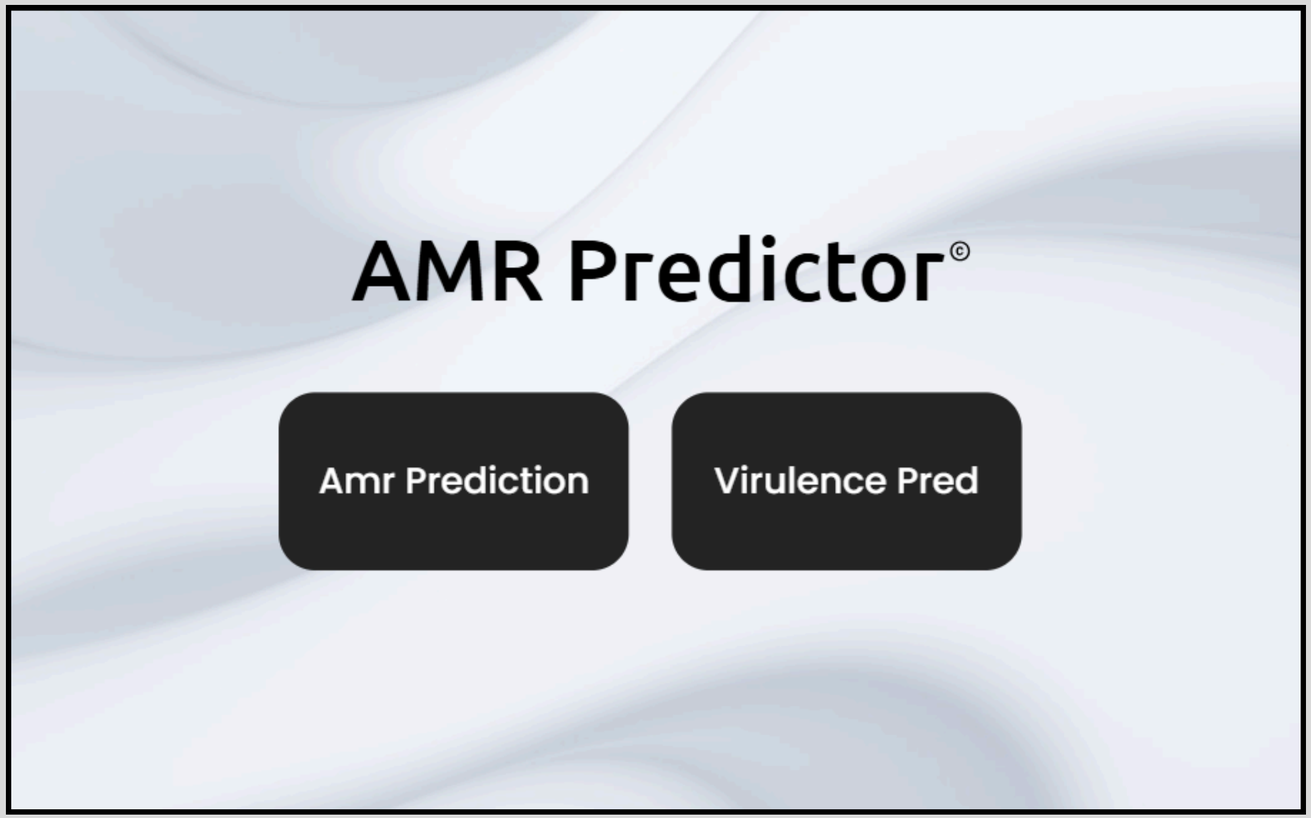
	A	B	C	D	E	F
1	Sequence	Descriptio	DNA Seque	Class	Kmers	Embedding
2	JNOG0100	Escherichia	TTTTTTCC	0	['TTTTTT', 'TTTTTC',	[-0.35838243
3	NZ_JACCN	Escherichia	CTGCTCGA	0	['CTGCTC', 'TGCTCG	[-0.41395122

XG-BOOST model trained for Virulent Prediction





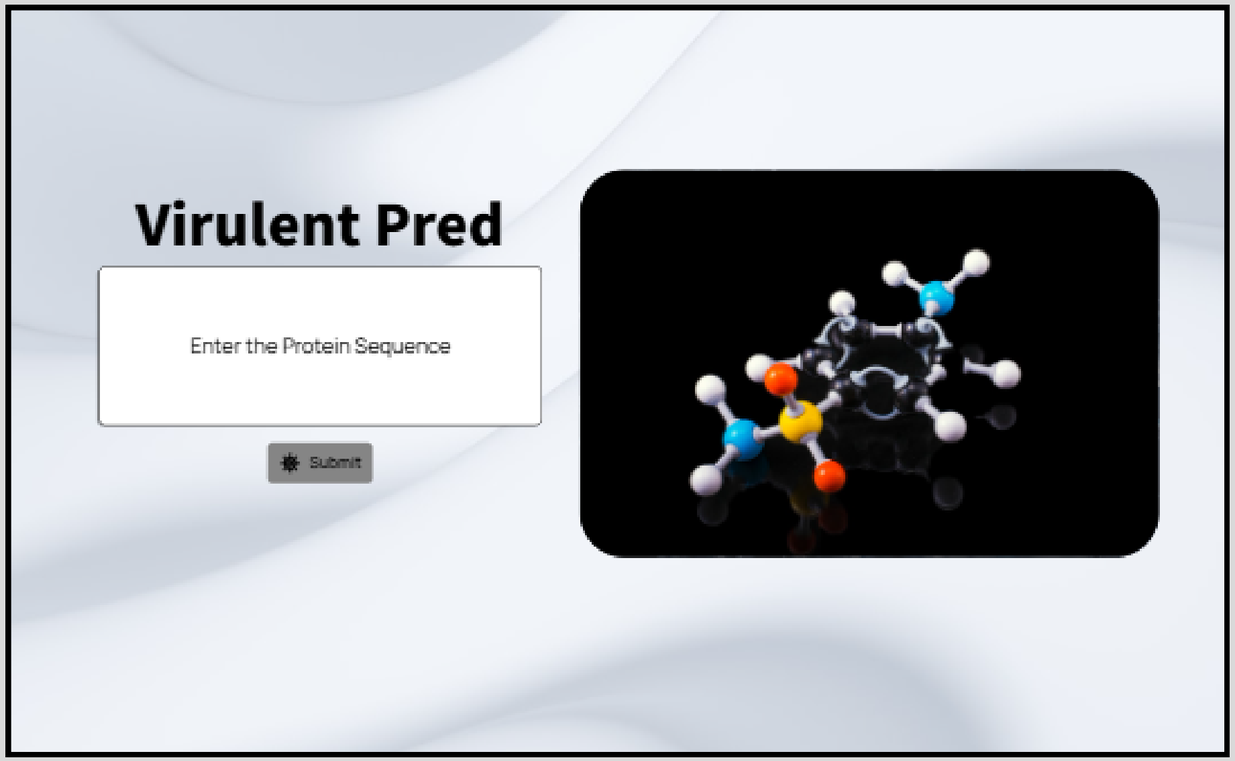
User - Interface



Features :

- AMR prediction
- Virulence Pred
- KEGG
- PDB - Search

User - Interface



FUTURE WORK:

- Further classes of resistance gene will be added for better prediction and understanding.
- Incorporating larger datasets to improve model generalizability.
- Including more bacterial strains, viral genomes, and fungal pathogens.
- Using ML to forecast the emergence of new resistance genes.
- UI interface for integration of amr prediction with virulence prediction for better access.

PAPER 1 - Ethics and antibiotic resistance

- The ethical challenges posed by antibiotic resistance, emphasizing the severe and unevenly distributed health consequences. It critiques common frameworks like patient responsibility, the tragedy of the commons, and antibiotic stewardship, highlighting their limitations

Responsibility and Accountability

Areas of agreement: Ethical analyses have focused on the moral responsibilities of patients to complete antibiotic courses, resistance as a tragedy of the commons and attempts to limit use through antibiotic stewardship.

to the problem.[3] Making and implementing policy to address drug resistance involves balancing multiple ethical values, as well as multiple types of harms and benefits.[2,3]

problems related to resistant bacteria involve multiple species and sectors, and at this broader level, ABR may be best characterized as a One Health problem.[9,10]

Public Health vs. Individual Freedom

The recognition that antibiotic use increases acquired ABR, and thereby may reduce the availability of effective antibiotic therapy for all, led to the formulation that antibiotic use is analogous to one type of collective action problem known as the tragedy of the commons.[23–30] Standard commons

infection[32,33]; (ii) therapeutic uncertainty, as not all patients with a bacterial infection benefit from antibiotic therapy[34]; (ii) microbiological uncertainty, microbiological uncertainty, as pathogens are not uniformly susceptible to all available antibiotic options (and this information may not be available

Conflicts of Interest in Healthcare

ations between individuals' antibiotic consumption and colonization or infection with resistant isolates can be identified even when associations at the ecological ('commons') level are absent.[41,42] The implication of this observation is that the 'commons' necessarily places in opposition. When attempting to resolve this assumed conflict, and in concordance with dominant professional norms, clinicians often prioritize their immediate patient over the interests of other, distant and/or future patients.[32,33,44] In

involves risk, these risks are concentrated among users), and patients who use disproportionately large amounts of antibiotics are not 'free riding' on Although inescapable within the 'commons' formulation, this conflict of interest can disappear when the limitations of this formulation are considered.

Moral hazard arises when prescribers' incentives do not align with those of the other two parties, and qualitative research has identified these adverse drivers of antibiotic prescribing as active in multiple medical contexts.[50] Chief among these appear to need for other policies to address inappropriate antibiotic use.

Moreover, stewardship resources are often concentrated in hospitals, where some of the more overt harms from resistant bacterial infections often

Autonomy and Individual Rights

risks. Unlike infections with respiratory viruses that last days or weeks, resistant bacteria are often carried asymptomatically for years, meaning that cumulative infringements on carriers' lives may be even more significant. Being identified as a carrier can result in restricted access to healthcare and mental health issues due to stigmatization,[60,61] but also infringements on privacy, freedom of movement and free choice of occupation.[12] Determining the con-

promote the spread of resistant organisms (and resistance traits between organisms) as much of the environment in poor communities is contaminated with resistant bacteria.[9] Without addressing these factors, stewardship efforts to reduce antibiotic use among the global poor (and, for that matter, in livestock and agriculture) will remain largely futile.

Above and beyond altruistic motivations for high-income countries to help address these social deter-

Justice and Fairness

the global human population. Even before questions of whether people from such communities can access resistant organisms. The recent COVID-19 pandemic has highlighted the ethical salience of public health measures for asymptomatic infection, which some-

'courses' of antibiotics appear excessive); (ii) the minimally effective antibiotic spectrum for specific antibiotics and/or a healthcare provider to diagnose the disease and supervise access to antibiotics comes the need for basic public health measures (see Fig. 1).

of 'access'.[62] While high-income countries struggle with the development of appropriate policies to curb inappropriate use of antibiotics available in abundance, hundreds of thousands of people die in low-income countries every year for want of access to antibiotics.[63] Yet this contrast hides more

and health policy. From an ethical perspective, policy and clinical decisions should be based on value judgements informed by sound evidence. Where this evidence is lacking there is an ethical imperative for more relevant scientific research and public health surveillance. Policy should also focus on harm reduc-

Much of this disease burden is concentrated in low- and middle-income countries (LMICs) where surveillance for resistance is often incomplete.[2]

and freedom (or liberty).[8] Before embarking on a tour of the ethical issues related to acquired ABR, it is important to review the biology and

PAPER 2 - Machine learning for predicting antimicrobial resistance in critical and high-priority pathogens (Carlos, Daniel, Sergio)

Published 25 Feb 2025

They found that Gradient Boosted Decision Trees (GBDT), Random Forest, and XGBoost were the top-performing ML models.

assessing 688,107 patients and 1,710,867 antimicrobial susceptibility tests. GBDT, Random Forest, and XGBoost were the top-performing ML models for predicting antibiotic resistance in CHPP infections. GBDT exhibited the highest AuROC values compared to Logistic Regression (LR), with a mean value of 0.80 (range 0.77–0.90) and 0.68 (range 0.50–0.83), respectively. Similarly, Random Forest generally showed better AuROC values compared to LR (mean value 0.75, range 0.58–0.98 versus mean value 0.71, range 0.61–0.83). However, some predictors selected by these algorithms align with those suggested by LR.

ETHICS

BIAS

available data in a scenario that closely mirrors real-world cases. Consequently, research endeavors should construct prediction models utilizing clinically significant variables, encompassing all predictors delineated by clinical guidelines, and draw conclusions that align with practical clinical considerations [10]. However, many studies have constructed prediction

Data Privacy and Security

web of factors influencing multidrug resistance [10,14,15]. By leveraging large-scale datasets encompassing clinical, microbiological, and antimicrobial susceptibility tests, and epidemiological variables, ML models can discern subtle patterns and relationships that traditional statistical methods may overlook. These models hold the potential to identify novel risk factors,

Responsible Use of Technology

susceptibility tests in CHPP in real-world healthcare settings. However, limitations such as retrospective methodology for model development, nonstandard data processing, and lack of validation in randomized controlled trials must be considered before applying these models in clinical practice.

PAPER 3 - Artificial intelligence in predicting pathogenic microorganisms' antimicrobial resistance: challenges, progress, and prospects

Author : Yan Li, Xiaoyan Cui, Xiaoyan Yang, Guangqia Liu ,Juan Zhang
Published : 01 November 2024

Objective :

1. This article reviews the latest advancements in AI and ML for predicting antimicrobial resistance in pathogenic microorganisms.
2. highlights the main AI and ML models used in resistance prediction,
3. Finally the paper discuss about the new perspectives and solutions for research into microbial resistance through algorithm optimization, dataset expansion, and interdisciplinary collaboration.

Case Study-3

PATENT: GLYCOMIMETICS TO INHIBIT PATHOGEN-HOST INTERACTIONS (US 9,605,014 B2)

Date of Patent: Mar. 28, 2017

- The patent US9605014B2 describes glycomimetic compounds designed to inhibit pathogen-host interactions
- A glycomimetic as described herein may be used to impregnate filters, masks, and clothing or any combination thereof in prophylactic strategies to reduce transmission and inhibit the binding of any of a variety of pathogens to their target.