CERTIFICATE

It is certified that the work contained in the thesis titled ***"Combating Antimicrobial Resistance with Artificial Intelligence"*** by ***Ritesh Sharma*** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all requirements of Comprehensive Examination, Candidacy, and SOTA for the award of Ph.D. Degree.

Prof. Sanjay Kumar Singh
Professor and HOD
Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi

DECLARATION BY THE CANDIDATE

I, ***Ritesh Sharma***, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of ***Prof. Sanjay Kumar Singh*** from ***July 2019*** to ***April 2023*** at ***Department of Computer Science and Engineering***, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.* reported in journals, books, magazines, reports, dissertations, theses, *etc.*, or available at websites and have not included them in this thesis and have not cited as my own work.

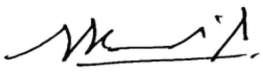
Date: 18/09/2023

Place: Varanasi

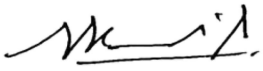
Ritesh Sharma

CERTIFICATE BY THE SUPERVISOR

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.



Prof. Sanjay Kumar Singh
Professor
Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi



Signature of Head of the Department

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Combating Antimicrobial Resistance with Artificial Intelligence

Name of the Student: Ritesh Sharma

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the *Doctor of Philosophy*.

Date:18/09/2023

Place: Varanasi

Ritesh Sharma

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

This thesis is dedicated to GOD and my beloved family.

ACKNOWLEDGEMENT

I bow down to the Lord Almighty, whose grace guided me from inception to the completion of this research work. Though only my name appears on the cover of this dissertation, a great many people have contributed to its production. I am grateful to everyone who helped make this dissertation possible. Firstly, I would like to thank my esteemed supervisor, Prof. Sanjay Kumar Singh, for his invaluable supervision, support, and tutelage during my Ph.D. His immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. I hope that one day I would become as good an advisor to my students as Prof. Sanjay Kumar Singh has been to me. I am thankful to my Doctoral Committee (Prof. Neeraj Sharma and Dr. Ravi Shankar Singh) for reviewing my work and helping me with invaluable suggestions. I am grateful to my collaborators, who helped me a lot. I am also grateful to all the technical and non-technical staff members of the department (Dr. Ram Prasad Meena, Ravi Kumar Bharti, Shubham Pandey, and Prakhar Kumar) for their help and precious assistance. I also want to thank my friends (Vishakha, Jai Shankar, Priyatosh, Sushant, Sandeep, Vipin) for their continous support throughout this journey. This acknowledgment would not be complete without mentioning the painstaking efforts and patience of my families. No words are adequate to express my indebtedness to my parents for their support, blessings, and good wishes. I owe this thesis to my parents, my in-laws, and other family members who always stood by me and provided strength in pursuing this work. I would like to express my heartfelt gratitude to my father, Capt. Y. R. Sharma & mother, Mrs. Anjana Sharma, without their unconditional support, this task would not have been accomplished. I am also grateful to my father-in-law Mr. R. P Duvedi & mother in law Mrs. Lata Duvedi, for their continuous motivation. I would like to say specially thanks to my brother (Nirdesh Kumar Sharma), sister-in-law (Shanu), brother-in-law (Abhinav) for their continuous support.

I am also very much thankful to my wife Dr. Ekta Sharma. She has been the pillar of support during this journey and without her sacrifices this thesis would not have

been completed.

Lastly, I express my hearty thanks to those I missed mentioning by name, who helped directly or indirectly and co-operated with me a lot in completing this Ph.D. research work.

Ritesh Sharma

Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
AMR	Antimicrobial Resistance
AMPs	Antimicrobial Peptides
AVPs	Antiviral Peptides
AFPs	Antifungal Peptides
ABPs	Antibacterial Peptides
MIC	Minimum Inhibitory Concentration
WHO	World Health Organization
RBCs	Red Blood Cells
AUC	Area Under Curve
ELMo	Embeddings from Language Models
HCF	Handcrafted Features
PESTV	Pretrained Embeddings from Seq2vec
OHE	One-Hot Encoding
XGBoost	Extreme Gradient Boosting
SVM	Support Vector Machine
RF	Random Forest
LR	Logistics Regression
NB	Naive Bayes

Abbreviation	**Description**
KNN	K-Nearest Neighbour
ANN	Artificial Neural Network
CM	Classification Module
ERT	Extremely Randomized Trees
AB	AdaBoost
GB	Gradient Boosting
ReLU	Rectified Linear Unit
ICL	Independent Component Layer
Adam	Adaptive Moment Estimation
FDA	Food and Drug Administration
Bi-GRU	Bidirectional Gated Recurrent Unit
Bi-LSTM	Bidirectional Long Short-Term Memory
BiTCN	Bidirectional Temporal Convolutional Network
CNN	Convolutional Neural Network
LIME	Local Interpretable Model-agnostic Explanations
IC	Independent Component
GF	Genomic Features

List of Symbols

Symbol	Description
z_t	update gate
r_t	reset gate
h_t	candidate state
i_t	input gate
o_t	output gate
f_t	forget gate
L	loss function
η	learning rate
σ	activation function
x_t	embedding vector of the amino acid

Preface

The discovery of antibiotics revolutionized modern healthcare and increased the lifespan of humans by more than twenty years, but their unregulated use has resulted in antimicrobial resistance (AMR). AMR kills at least 7 lakh people in poor and middle-income countries each year, and the WHO ranks it as one of the top ten public health hazards. To counteract this condition, novel antimicrobial compounds are required, and antimicrobial peptides (AMPs) show promise in this regard. AMPs are proteins produced by diverse organisms naturally and can be classified into multiple groups, such as antiviral (AVPs), antifungal (AFPs), and antibacterial peptides (ABPs).

To avoid the high costs and time associated with identifying novel AMPs in the lab, researchers use *in-silico* tools for preliminary screening of natural sources. However, the existing tools available for this purpose have poor performance, which limits their applicability for wet-lab researchers. Thus, we have proposed AI-based frameworks for identifying AFPs, AVPs, and ABPs from natural sources in different studies.

All amino acids are not equally important in classifying a peptide. Existing AI-based tools do not provide information about the essential amino acids responsible for classifying a peptide. Therefore, we developed an explainable framework that not only classifies the peptides but also provides information about the essential amino acids responsible for classifying a peptide.

The WHO categorizes bacteria into three categories (critical, high, and medium), and the ESKAPEE pathogens are a major threat as they range from high to critical

WHO-priority pathogens. Earlier studies can identify ABPs from natural resources but do not provide minimum inhibitory concentration (MIC) values against the ESKAPEE pathogens. Thus, for identifying optimal ABPs (which work at low MIC), wet-lab researchers have to test the identified ABPs against ESKAPEE pathogens at different concentrations, leading to a loss of time and money. To address this issue, we proposed a framework that predicts the MIC values for ABPs against ESKAPEE pathogens. The proposed framework can help identify optimal ABPs that can work at low MICs against the entire ESKAPEE group of bacteria.

Peptide toxicity is a major hurdle in the development of therapeutic peptides. Hemolytic activity against red blood cells is one of the key factors to consider when evaluating peptide toxicity. High hemolytic activity can lead to anemia and other blood disorders, making highly hemolytic peptides unsuitable for pharmaceutical use. However, discovering low hemolytic peptides is a labor-intensive and time-consuming process that involves testing on mammalian red blood cells. To address this, we proposed a framework to identify low hemolytic therapeutic peptides.

All the frameworks we developed as part of different studies have also been made available as web applications. The wet-lab researchers can use these tools to narrow the search space while discovering low hemolytic AMPs active against different pathogens.