

**EXPLAINABLE ARTIFICIAL INTELLIGENCE BASED
BIOMARKER PRIORITIZATION FOR ANTIMICROBIAL
RESISTANCE PREDICTION IN CLINICALLY SIGNIFICANT
BACTERIAL PATHOGENS**

A THESIS
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
In
BIOINFORMATICS

By:

UTKARSH MISHRA
(Enrollment No. 24/BIO/08)

Under the Supervision of

Prof. Asmita Das
Delhi Technological University



Department of Biotechnology
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daultpur, Main Bawana Road, Delhi-110042

May 2026



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daultpur, Main Bawana Road, Delhi-110042

CANDIDATE DECLARATION

I, **UTKARSH MISHRA**, Roll No. **2K24/BIO/08**, student of **M.Tech Bioinformatics**, hereby declare that the thesis titled **"EXPLAINABLE ARTIFICIAL INTELLIGENCE BASED BIOMARKER PRIORITIZATION FOR ANTIMICROBIAL RESISTANCE PREDICTION IN CLINICALLY SIGNIFICANT BACTERIAL PATHOGENS"** submitted by me to the Department of Biotechnology, Delhi Technological University, Delhi, in partial fulfilment of the requirements for the award of the degree of **Master of Technology**, is an original research work carried out by me under proper guidance and supervision and has not been copied from any source without due acknowledgement and citation. I further declare that this work has not previously formed the basis for the award of any degree, diploma, fellowship, associateship, or any other similar academic recognition in this or any other institution or university, and I take full responsibility for the authenticity and originality of the work presented in this thesis.

PLACE: DELHI

DATE: 27/05/26


UTKARSH MISHRA



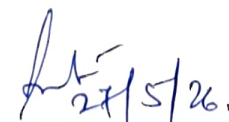
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daultapur, Main Bawana Road, Delhi-110042

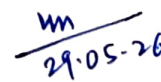
CERTIFICATE

I hereby certify that the thesis titled **“EXPLAINABLE ARTIFICIAL INTELLIGENCE BASED BIOMARKER PRIORITIZATION FOR ANTIMICROBIAL RESISTANCE PREDICTION IN CLINICALLY SIGNIFICANT BACTERIAL PATHOGENS”** submitted by **UTKARSH MISHRA**, Roll No. **2K24/BIO/08**, student of **M.Tech Bioinformatics**, Department of Biotechnology, Delhi Technological University, Delhi, in partial fulfilment of the requirements for the award of the degree of **Master of Technology**, is a Bonafide record of original research work carried out by the student under my guidance and supervision. To the best of my knowledge and belief, the work embodied in this thesis has not been submitted, either in part or in full, for the award of any degree, diploma, fellowship, associateship, or any other similar academic recognition by this or any other university or institution.

PLACE: DELHI

DATE: 27/05/26


DR. PROF. ASMITA DAS
SUPERVISOR


Prof. YASHA HASIJA
Head of Department
Delhi Technological University

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project guide and mentor, Mrs. Asmita Das, for her invaluable guidance, constant encouragement, and continuous support throughout the course of this research work. Her insightful suggestions, scientific perspective, and dedication towards academic excellence greatly contributed to the successful completion of this project. Her mentorship not only strengthened my understanding of computational biology and antimicrobial resistance research but also motivated me to approach scientific problems with analytical thinking and confidence.

I would also like to extend my heartfelt thanks to the Ph.D. scholars in the laboratory, Ms. Monika, Mr. Yuvraj, and Ms. Simran, for their valuable guidance, technical assistance, and constant support during the project. Their constructive discussions, suggestions, and cooperative nature significantly enhanced my learning experience and helped me overcome various challenges encountered during the research work. I am equally grateful to the faculty members of the Department of Biotechnology, Delhi Technological University, whose teachings and academic support throughout the M.Tech programme helped me build a strong foundation in bioinformatics and computational research.

Finally, I would like to express my deepest gratitude to my family, friends, and batchmates for their unwavering encouragement, emotional support, and belief in my abilities throughout this academic journey. Their continuous motivation and support have always been a source of strength and inspiration, helping me successfully complete this work and progress confidently in my academic and professional pursuits.

TABLE OF CONTENTS

S.No	Title	Page Number
I.	Candidate's Declaration	ii
II.	Certificate by Supervisor	iii
III.	Acknowledgement	iv
IV.	List of Figures	viii
V.	List of Abbreviations	ix
VI.	Abstract	11
Chapter 1	Introduction	12
	1.1 Background	12
	1.2 Emergence of Computational Biology in AMR	13
	1.3 Role of Machine Learning in Pathogen Analysis	15
	1.4 XAI in Biomedical Research	16
	1.5 Research Gap	18
	1.6 Project Objectives	19
Chapter 2	Literature Review	20

Chapter 3	Materials and Methodology	27
	3.1 Research Methodology	27
	3.2 Data Acquisition	28
	3.3 Data Preprocessing	28
	3.4 Feature Engineering and Selection	30
	3.5 Machine Learning Model Development	31
	3.6 Model Training and Validation	32
	3.7 XAI Framework	33
	3.8 Performance Evaluation Metrics	34
	3.9 Computational Tools and Software Environment	35
	3.10 Computational Workflow	36
Chapter 4	Results	37
	4.1 Experimental Analysis	37
	4.2 Data Preprocessing Outcomes	37
	4.3 Feature Engineering Results	38
	4.4 Model Performance Analysis	40
	4.5 Classification Performance Metrics	42
	4.6 XAI Results	44

	4.7 Model Comparative Analysis	47
	4.8 Visualization and Interpretability Outcomes	47
<i>Chapter 5</i>	Discussion	49
<i>Chapter 6</i>	Conclusion	54
	References	56

LIST OF FIGURES

Figure No.	Figure Caption	Page Number
Fig. 1	Data Preprocessing Workflow	29
Fig. 2	Machine Learning Architecture for Antimicrobial Resistance Prediction	32
Fig. 3	Explainable AI Workflow for AMR Prediction	34
Fig. 4	Computational Workflow for Antimicrobial Resistance Prediction	36
Fig. 5	Top SHAP-prioritized KP resistance biomarkers	39
Fig. 6	Top SHAP-prioritized PA resistance biomarkers	40
Fig. 7	Comparative performance analysis of KP and PA machine learning models	42
Fig.8	Cross-validated ROC-AUC curve for KP resistance prediction	44
Fig.9	Cross-validated ROC-AUC curve for PA resistance prediction	44
Fig.10	KP SHAP summary plot for resistance prediction interpretability	46
Fig.11	PA SHAP summary plot for resistance prediction interpretability	46

LIST OF ABBREVIATIONS

Abbreviation	Full Form
AMR	Antimicrobial Resistance
KP	Klebsiella pneumoniae
PA	<i>Pseudomonas aeruginosa</i>
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
XGBoost	Extreme Gradient Boosting
SHAP	SHapley Additive exPlanations
MDR	Multidrug Resistance
WHO	World Health Organization
ML	Machine Learning
AI	Artificial Intelligence
SVM	Support Vector Machine
NGS	Next Generation Sequencing
CARD	Comprehensive Antibiotic Resistance Database
KNN	K-Nearest Neighbour
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LIME	Local Interpretable Model-Agnostic Explanations
MDR	Multidrug Resistance
XDR	Extensively Drug Resistant
PDR	Pan Drug Resistance
MRSA	Methicillin-Resistant <i>Staphylococcus aureus</i>
PATRIC	Pathosystems Resource Integration Center
PCA	Principal Component Analysis
LSTM	Long Short-Term Memory
CRISPR	Clustered Interspaced Short Palindromic Repeats

LR	Logistic Regression
DT	Decision Tree
AST	Antimicrobial Susceptibility Testing
EL	Ensemble Learning

ABSTRACT

Multidrug-resistant bacterial infections are becoming a big health issue worldwide, owing to the emergence of antimicrobial resistance (AMR) which is impairing the efficacy of currently available antibiotics. The pathogens most commonly found in the clinical setting, *Klebsiella pneumoniae* (KP) and *Pseudomonas aeruginosa* (PA), are often linked to high rates of antimicrobial resistance and to serious hospital-acquired infections. In the present study, a human-readable machine learning model was built for predicting resistance to KP and PA as well as predicting biomarkers in the meropenem resistance and ciprofloxacin resistance datasets, respectively. The performances of the supervised machine learning algorithms were compared using the accuracy, precision, recall, F1 score and ROC(Receiver Operating Characteristic) AUC(Area Under the Curve). Exemplarily, ensemble-based models such as Random Forest (RF) and Extreme Gradient Boosting (XGBoost) had the best predictive power on both datasets. Comparative analysis showed that KP has higher predictive stability than PA, suggesting differences in the architecture of resistance among the organisms. SHapley Additive exPlanations (SHAP) analysis was used to rank the features and prioritize the biomarkers for their biological interpretability. Well characterized genes for antimicrobial resistance (e.g. OXA-type carbapenemase, KPC-family genes, NDM genes, and aminoglycoside resistance determinants) were the most important biomarkers in KP, suggesting the gene-driven resistance mechanism. By contrast, PA resistance prediction was linked to proteins involved in regulatory functions, related to efflux, membrane associated proteins, transcription regulation and adaptive stress response elements, indicating a more diverse and system level resistance profile. These patterns of resistance were further confirmed by visualization of biomarkers using heatmap and cross-validated ROC-AUC analysis, performed within the two organisms. KP had largely clustered AMR signatures and good classification performance, while PA had a more variable, but biologically relevant, predictive performance. This study shows the importance of explainable AI methods for AMR prediction and discovery of interpretable biomarkers, and highlights the fundamental differences in resistance mechanisms between KP and PA.

CHAPTER 1

INTRODUCTION

1.1 Background

AMR has been acknowledged as one of the most critical public health problems worldwide in the twenty-first century, threatening the efficacy of the usual therapies for infectious illnesses. The extensive and usually incorrect use of antibiotics in clinical practice, agriculture, veterinary medicine, previous successful antimicrobial treatments. As a result, infections caused by resistant microorganisms are becoming increasingly difficult to control, resulting in longer hospitalisation, higher healthcare expenses, increased morbidity, and death globally [1], [2]. The bacterial pathogens have shown quick adaptation mechanisms that have helped them to acquire resistance through many biological routes such as horizontal gene transfer, chromosomal alterations, efflux pump activation, enzymatic drug degradation and biofilm formation. Alarming multidrug-resistant (MDR) phenotypes have been seen in pathogens such as *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Acinetobacter baumannii*, which pose major challenges to healthcare systems worldwide [3], [4]. According to the World Health Organization (WHO), antimicrobial resistance is projected to become one of the leading causes of death in the coming decades if effective monitoring and intervention strategies are not implemented promptly [5]. Traditional antimicrobial susceptibility testing techniques are reliable however they are labour challenging, time consuming and culture based laboratory techniques.

Due to the establishing availability of genomic, transcriptomic, proteomic and clinical information, computational biology remains the primary area of interest to tackle AMR related challenges with the use of data driven methodologies [6]. In recent years, tremendous promise of machine learning (ML) and artificial intelligence (AI) based techniques in the prediction of antibiotic resistance patterns, identification of resistance-related biomarkers and clinical decision support systems has been witnessed. Machine learning algorithms can analyze complex multidimensional biological datasets and identify hidden patterns that are difficult to detect using conventional statistical methods [7]. Supervised learning techniques like RF, Support Vector Machine (SVM), Decision Trees, Gradient Boosting and Neural Networks have been widely used for resistance prediction and pathogen categorisation [8]. Machine learning models have shown considerable predictive capability, however some research have raised concerns about the comprehensibility and disclosure of black-box AI systems in biomedical applications. This has led to a growing curiosity in XAI that seeks to give interpretable insight into model predictions while keeping predictive accuracy [9]. XAI methods increase trustworthiness, repeatability, and clinical application by allowing academics and healthcare practitioners to comprehend the significance of biological factors to prediction outcomes. Hence, the current work intends to examine the antimicrobial resistance pattern of pathogenic microorganisms utilising computational and machine learning approaches with an emphasis on the significance of interpretable AI-based frameworks in biomedical applications.

1.2 Emergence of Computational Biology in AMR

The rise of computational biology has consequently revolutionised current biomedical research by merging biological sciences, computer science, statistics, mathematics and data analytics to confront complex healthcare challenges such as antibiotic resistance [10]. Computational biology

approach allows the systematic study of large biological datasets produced by next-generation sequencing (NGS), metagenomics, transcriptomics, proteomics and molecular diagnostics. These methods have greatly enhanced our knowledge of resistance-associated genes, mobile genetic elements, virulence factors and evolutionary adaptations in pathogenic microbes [11]. Computational frameworks might help researchers to identify genetic signatures of resistance features, track transmission patterns of the pathogens and study biological processes involved in antibiotic tolerance. Genome-wide investigations help identify conserved resistance determinants and help distinguish resistances and susceptibilities among strains. The Comprehensive Antibiotic Resistance Database (CARD), the ResFinder, MEGARes and ARG-ANNOT databases are important tools for the annotation of resistance associated genes and their distribution in various microbial communities [12], [13]. These datasets have been added to computational pipelines, improving resistance monitoring and molecular epidemiology research. Using protein structure prediction, molecular docking, and molecular interaction analysis, researchers have been able to study the interactions between drugs and targets at the molecular level, and the mutations that are responsible for resistance, and how well antibiotics are binding [14]. Combining sequence alignment, phylogenetic analysis, feature extraction and statistical modelling in a computational pipeline could aid understanding of pathogen evolution and dissemination patterns of resistance [15]. In recent years, the role of computational biology has been further consolidated in the field of precision medicine and infectious disease control, due to recent developments in computer infrastructure, cloud computing and artificial intelligence. The credit is given to the automation of the analysis processes, it is possible to deal with vast amounts of biological data, identifying resistance features in a very short time and predicting the consequences on the clinical picture [16]. Therefore, the development of strong computational frameworks linked with interpretable machine

learning models is still needed to improve the reliability and clinical usefulness of AMR prediction systems. In this study, computational biology approaches were utilised and algorithms based on machine learning were used to investigate resistance linked patterns and to help build explainable prediction frameworks for dangerous microbes.

1.3 Role of Machine Learning in Pathogen Analysis

The exponential growth of biological and clinical data obtained through contemporary sequencing technology has given great prospects for the use of machine learning approaches for pathogen analysis and antibiotic resistance prediction. Machine learning approaches can give sophisticated computing power to find hidden associations in such datasets and build prediction models with better accuracy and scalability [17]. Machine learning is a branch of artificial intelligence that enables computational systems to learn patterns from data and make predictions without being explicitly programmed for every analytical step. Machine learning algorithms are being used increasingly in biomedical research for drug development, genomic categorisation, illness diagnosis, biomarker identification, and antibiotic resistance prediction [18]. In antimicrobial resistance studies, supervised learning techniques have shown considerable success in predicting resistance phenotypes based on genomic and clinical features. Algorithms such RF, SVM, Logistic Regression, Decision Trees, K-Nearest Neighbours (KNN), and Gradient Boosting have been extensively utilised for the classification of resistant and susceptible bacterial strains [19]. Deep learning methodologies have further expanded the scope of pathogen analysis by enabling automated feature extraction and representation learning from complex biological datasets. Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) have shown promising results in genomic sequence analysis, protein structure prediction, and resistance gene identification [20]. In clinical microbiology, based on

machine learning, rapid resistance prediction algorithms can shorten diagnostic turnaround time and enhance evidence-based antibiotic selection, hence minimising wasteful antimicrobial usage [21]. Several predictive models function as black-box systems, offering restricted interpretability of the variables affecting prediction results. The absence of candour may diminish confidence among clinicians and impede the clinical deployment of AI-driven diagnostic applications [22]. Furthermore, changes in sequencing quality, overfitting, inadequate sample diversity, and dataset imbalance between datasets gathered from different geographic or clinical environments may have a negative impact on the ability of the model to execute efficiently. Explainable models give interpretable data regarding genetic markers associated to resistance, the role of traits and biological processes impacting prediction findings. These strategies enhance the clarity, reproducibility as well as reliability of models, which help informed clinical decision making [23]. This study emphasises the importance of comprehensibility and rationality in predictive biomedical models through the utilisation of machine learning techniques for analysing underlying trends in antibiotic resistance in pathogenic bacterial species.

1.4 XAI in Biomedical Research

The growing popularity of AI and ML in healthcare and biomedical sciences has significantly improved data-driven prediction, diagnosis, and ability to make decisions. Numerous machine learning systems operate as black-box models, producing recommendations without adequately elucidating their foundational reasoning. In biomedical research, wherein clinical choices may significantly impact clinical outcomes, comprehending the rationale behind computer predictions is essential [24]. XAI is currently considered a feasible remedy for the interpretability issues associated with conventional machine learning techniques. It refers to a collection of methodologies and computational approaches designed to improve the transparency, clarity, and

interpretability of AI systems for users [25]. Unlike traditional black-box algorithms, explainable models enable researchers and clinicians to explore the factors affecting the prediction's results by providing knowledge about the consequences of the characteristics of the input. ML techniques could be used to identify complex associations with resistant infections based on genomic, transcriptomic, proteomic or clinical data; however, without the ability to interpret the process, it is difficult to establish biological relevance of such predictions [26]. By emphasising the genes, mutations, molecular pathways, or clinical indicators that contribute to resistance estimations, explainable AI techniques assist in narrowing this gap. There are many of the most popular XAI methods, such as feature importance analysis, SHAP, LIME (Local Interpretable Model-Agnostic Explanations), saliency mapping, and attention-based visualization techniques [27]. These strategies enable characterisation of the importance of significant biomarkers in the prediction models for antimicrobial resistance and allow for a thorough assessment of the behaviour of these models. When model predictions are accompanied by explainable reasoning and analytical frameworks, the adoption of AI-assisted diagnostic tools by clinicians and healthcare professionals is likely to increase [25]. Moreover, explainability is essential for model validation, mistake detection, bias assessment, and meeting the requisite requirements in healthcare-oriented AI applications. Systems powered by XAI may pave the way for precision medicine by allowing for the individualisation of evaluations of disease causes and treatment efficacy [28]. In addition, the reliability of model interpretation might be affected by the presence of heterogeneity, noise, and missing data in biological datasets [29]. Hence, computational healthcare researchers are actively working to create ML frameworks that are both robust and able to be understood by biology.

1.5 Research Gap

The clinical usability and generalisability of current prediction frameworks are still limited by several issues, even though computational biology, ML, and studies on antibiotic resistance have made significant advances. Nevertheless, most of the models that are presently published prioritise predictive performance in the absence of adequately resolving interpretability, biological validation, and practical implementation problems [30]. Many machine learning models are developed using laboratory-generated data, geographically limited samples, or a few specific bacterial species, which may limit their capacity to generalise across diverse clinical populations and environmental circumstances [31]. The black-box character of numerous advance machine learning and deep learning algorithms represents an additional considerable obstacle. While these approaches frequently obtain high predictive precision, they frequently struggle to offer plausible hypotheses for the biological parameters which impact the conclusions of resistance prediction [32]. Several predictive algorithms have been used to reveal resistance patterns, alongside minimal correlation of predictions via fundamental biological processes, resistance-associated genes or clinically pertinent biomarkers [33]. Most of the available rationalisation studies are theoretical or confined to general healthcare applications, not to infectious disease-specific forecasting algorithms [34]. Existing procedures are often structured in a scattered manner and use unique analytical equipment and methodologies, which potentially render reproducibility and comparative evaluation complex [35].

1.6 Project Objectives

Objective 1

Antimicrobial Resistance Profiling

To have an extensive understanding of the trends of antimicrobial resistance of clinically relevant bacterial infections and to recognise the features of antimicrobial resistance.

Objective 2

Computational and ML-Based Analysis

To establish and evaluate machine learning models for the prediction of antimicrobial resistance by incorporating genomic biomarker datasets that are extracted from resistant and susceptible bacterial isolates.

Objective 3

Explainable Artificial Intelligence for Biomarker Discovery

To examine the use of XAI approaches intended to enhance the interpretability, transparency, and dependability of ML-based biomedical prediction models for biomarker discovery.

Objective 4

Integrated Computational Framework Development

To develop an all-inclusive computational framework featuring explainable algorithms, machine learning, and antibiotic resistance analyses for customised therapeutic applications and more effective infection research.

CHAPTER 2

LITERATURE REVIEW

AMR has surfaced collectively of the greatest serious global healthcare concerns of the contemporary period, impacting the efficacy of traditional antibiotic therapy and augmenting the global burden of infectious illnesses. The overuse and misuse of antimicrobial drugs in clinical practice, agriculture, veterinary medicine and environmental uses has led to the fast development of resistant bacterial strains that can survive numerous antibiotic treatments [36]. The prevalence of multidrug-resistant (MDR), extensively drug-resistant (XDR), and pan-drug-resistant (PDR) infections is a major problem for the healthcare systems owing to higher levels of morbidity, mortality, longer hospitalisation, and high treatment expenses [37]. Bacterial pathogens possess remarkable adaptive capabilities that allow them to survive under antimicrobial stress conditions through several resistance-associated mechanisms. These mechanisms include enzymatic degradation of antibiotics, modification of antibiotic target sites, activation of multidrug efflux pumps, reduced membrane permeability, and biofilm formation [38]. Horizontal gene transfer mechanisms such as transformation, conjugation, and transduction further facilitate the dissemination of antimicrobial resistance genes among bacterial populations [39]. Resistance-associated mobile genetic elements such as plasmids, transposons and integrons are important for the fast worldwide dissemination of resistant bacterial strains in healthcare and environmental contexts. The ESKAPE group of pathogens, comprising *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species, has gained considerable scientific attention because of its ability to evade

multiple antimicrobial agents [40]. These pathogens are commonly associated with hospital-acquired infections, bloodstream infections, respiratory tract infections, urinary tract infections, and wound infections characterized by poor therapeutic outcomes and high mortality rates. Methicillin-resistant *Staphylococcus aureus* (MRSA), carbapenem-resistant *Klebsiella pneumoniae*, and multidrug-resistant *Pseudomonas aeruginosa* were identified as significant threats in healthcare, demanding urgent action [41]. The therapeutic relevance of standard antibiotic susceptibility testing methods, including broth dilution tests, disc diffusion methods and culture-based identification systems, in determining bacterial resistance profiles continues to exist. However, such processes are often cumbersome, time-consuming and dependent on regulated laboratory environments [42]. Inadequate knowledge of resistance patterns could end up in the incorrect use of antibiotics and insufficient therapeutic intervention in the initial phase of illness. As a result, there is an increasing need for computational and AI-based approaches that might understand biological data and predict trends of antimicrobial resistance more efficiently and effectively. The advent of NGS technology has transformed the discipline of infectious disease research by establishing large genomic data sets that could provide insights into microbial diversity, evolutionary linkages, resistance-associated mutations and virulence determinants [43]. Whole-genome sequencing, metagenomics, transcriptomics, proteomics and metabolomics have substantially broadened the scientific understanding of microbial illness and resistance evolution. The genomic sequencing methods deliver the thorough characterisation of the resistance-associated genes, mobile genetic elements and bacterial adaption procedures contributing to the development of antibiotic resistance [44]. The rapid growth of biological datasets has accelerated the integration of computational biology and bioinformatics into antimicrobial resistance research. Computational biology is a field which incorporates biology, statistics, mathematics and computer

science with the aim of analysing and interpreting massive amounts of biological data [45]. Bioinformatics pipelines like as genome assembly, sequence alignment, phylogenetic analysis, molecular docking, structural modelling and network biology have now become indispensable tools to study microbial evolution and patterns of resistance propagation. These computer tools may be used to effectively analyse multidimensional biological datasets which are challenging to comprehend by standard laboratory procedures alone. Several public bioinformatics resources have greatly aided the monitoring and genetic study of antibiotic resistance. Databases like CARD, ResFinder, ARG-ANNOT, Pathosystems Resource Integration Center (PATRIC) and MEGARes provide curated collections of resistance-associated genes, virulence determinants, resistance phenotypes, and genomic annotations which are useful for computational studies [13]. Such databases would allow fast identification of resistance-associated biomarkers and would also be helpful for comparative genomic research across different bacterial species. The integration of computational processes with genetic databases has greatly enhanced the surveillance of antibiotic resistance and epidemiological studies. ML has become one of the greatest transformational computational techniques in biomedical and healthcare studies as the consequence of its capacity to recognise nonlinear patterns and concealed relationships within complex multivariate datasets [46]. ML algorithms enable computational systems to learn predictive patterns from biological and clinical data without requiring explicit programming for each analytical task. ML has shown great value in healthcare, especially in illness diagnosis, biomarker identification, drug development, precision medicine, genomic analysis and prediction of antibiotic resistance. Supervised machine learning methods are often used in research for predicting antimicrobial resistance, which use labelled datasets with established resistant phenotypes [47]. The RF methods are extremely helpful due to tolerance against over-fitting and productivity in processing high-dimensional biological

information. Likewise, SVM models have shown good predictive performance in genomic categorisation and resistance prediction tasks using complicated sequence-based characteristics [48]. Despite their simplicity, interpretability and low training time, Logistic Regression or Decision Tree algorithms are still useful in a binary classification problem. Further improvement in prediction has been achieved by combining multiple weak learners into an extremely powerful predictive framework, including Ensemble learning approaches like Gradient Boosting, AdaBoost, or XGBoost [49]. The techniques have been shown to enhance the classification accuracy in the investigations of antibiotic resistance prediction and pathogen identification. Additionally, the machine learning workflow is commonly accompanied by feature selection methods and/or dimensionality reduction methods like Principal Component Analysis (PCA) in order to enhance the efficiency of the model and decrease the complexity of the computation. In computational biology and biomedical research, deep learning methodologies have expanded the analysis power of machine learning methods. ANNs are used to imitate the link between neural networks that can construct hierarchy of complex biological information [50]. The CNNs has been widely used in protein structure prediction studies, genomic sequence analysis, and biological image classification. RNNs and architectures of Long Short-Term Memory (LSTM) have proven to be extremely effective for sequential biological data processing, temporal illness forecasting and genome pattern recognition [51]. There is evidence of superior effectiveness of deep learning frameworks for tackling complex classification problems in the literature, such as estimating antibiotic resistance [52]. The automated feature extraction abilities of deep learning algorithms lessen the need on human feature engineering and increase the analytical performance for high dimensional biological datasets. Deep learning algorithms have also demonstrated interesting applications in the biomarker identification, pathogen categorisation, outbreak prediction and

personalised medicine systems. Despite tremendous advances in AI based biomedical prediction systems, there remain important limits in the interpretability and transparency of these systems in the context of healthcare. Many state-of-the-art predictive models are black-box systems, which provide extremely accurate outputs, but do not provide intelligible explanations about the underlying prediction processes [53]. In clinical healthcare environments, lack of interpretability may reduce clinician trust, hinder regulatory acceptance, and limit adoption of AI-assisted diagnostic systems. Consequently, XAI has developed as an essential research field that aims to improve the transparency, interpretability and trustworthiness of machine learning predictions [54]. XAI techniques are designed to offer comprehensible insights on model behaviour via recognising important factors affecting the prediction findings. Such approaches facilitate biological validation of computational predictions and improve understanding of complex machine learning systems used in biomedical research. SHAP analysis is among the most widely utilized explainability approaches in biomedical ML applications [55]. SHAP uses collaborative game theory to determine the importance of individual attributes to model predictions. This allows for the discovery of resistance-associated biomarkers, medically relevant mutations and key genetic factors that affect the prediction frameworks of antibiotic resistance. Similarly, LIME is a method that creates local surrogate models, which can explain certain prediction scenarios, and enhance the interpretability of complex machine learning models [56]. Additionally, other explainability methods such as saliency mapping, Grad-CAM visualization, feature attribution analysis, and attention mechanisms have proven to be highly useful in biomedical AI systems [57]. These techniques enable analysis of biologically significant correlations in genomic and clinical datasets whereas boosting transparency of prediction frameworks. In clinical care, explainability techniques are particularly important because clinical decisions rely on accurate clinically

interpretable and physiologically validated computational predictions. The use of explainable artificial intelligence in antimicrobial resistance research has shown remarkable potential to improve the accuracy and dependability of predictive biological systems [58]. Interpretable machine learning frameworks may benefit precision medicine applications by improving trust in susceptibility to antibiotics prediction models and supporting personalised treatment making decisions. Explainability also helps to uncover physiologically important resistance-associated traits and to validate computational results with known molecular procedures. In recent years, artificial intelligence-driven healthcare systems have increasingly contributed toward precision medicine and personalized healthcare applications. Precision medicine involves tailoring diagnostic and therapeutic strategies according to patient-specific genetic, molecular, and clinical characteristics [28]. ML and XAI technologies may be used to integrate multifaceted biological information and build personalised predictive models for improved illness diagnosis, prediction of therapy response, and clinical decision-support tools. Computational methods are additionally key in infectious disease monitoring and in epidemiological modelling. ML frameworks may be used to analyse large-scale medical care and genomic information to detect emergent resistance trends, outbreak-related clusters, and pathogen transmission processes [59]. AI-powered monitoring systems might enhance early identification of resistant strains of bacteria and enhance antimicrobial administration measures to reduce improper antibiotic use. Computational biology and machine learning are still at the forefront of present-day antibiotic resistance research, but there has been a resurgent scientific interest in alternative treatment techniques, including bacteriophage therapy, in response to the mounting resistant challenge. Bacteriophages are viruses that infect and lyse bacterial cells via very precise host-pathogen interactions [60]. The narrow host's specificity of phages allows the accurate elimination of pathogenic microbes as well as the

retention of favourable microbiota assemblages, thus decreasing the danger of microbial dysbiosis caused by the use of broad-spectrum antibiotics. ML and bioinformatics techniques have recently been applied in computational analysis to forecast bacteriophage-host interactions and to fine-tune the selection of phages for use in antibiotics [61]. Genomic-similarity based predictive models, receptor binding proteins, Clustered Interspaced Short Palindromic Repeats (CRISPR) sequence analysis based and sequence-derived models have enabled the identification of promising therapeutic phages against MDR bacterial communities. These computational methods might be applied to future personalised treatment with phages and to precision antimicrobial treatment. Computational biology, ML and XAI have come a long way, but there are still many limitations on biomedical prediction systems. Important issues in the field of anti-microbial resistance prediction still remain, such as data imbalance, variability in sequencing methods, potential model over-fitting, poor clinical validation and the absence of a standardised computational pipeline [62]. Similarly, explainability techniques may sometimes introduce additional computational complexity and require careful interpretation in highly multidimensional biological systems. Thus, the multi-disciplinary approach involving computational biology, ML, XAI, and improved antibiotic therapy looks extremely promising for handling the coming challenges of antibiotic resistance. Future advances in genomic sequencing technologies, interpretable AI systems, biomedical data analysis and precision medicine are expected to dramatically improve the management of infectious diseases and to facilitate the establishment of next-generation predictive medical facilities that can better combat multidrug-resistant pathogenic bacteria.

CHAPTER 3

MATERIALS AND METHODOLOGY

3.1 Research Methodology

In this research, the objective is to analyze the patterns of antibiotic resistance of pathogens using ML and XAI techniques. The methodical strategy focused on the development of a transparent and physiologically relevant prediction algorithm, combining the use of computational biology, supervised machine learning algorithms and interpretable AI methods. The methodology adopted in the study was a series of consecutive steps, namely: Acquiring the dataset, Preprocessing the data, Feature engineering, Building a standard machine learning model, Analysis of explainability, and Analysis of performance and comparison of standard predictive models. Predictive accuracy and interpretability were particularly stressed, in that AI systems for healthcare require transparent decision-making systems to allow for reliable biomedical analysis [63]. The principal techniques employed for data analysis of the resistance associated and to identify resistance associated and susceptibility patterns were the computational approaches. The computational approaches were primarily used for data analysis of the resistance associated data and for identification of resistance associated and susceptibility patterns on the microbial strains. To improve transparency and clarity of the prediction results, XAI procedures were applied and the machine learning algorithms were trained and tested on the processed data. This resulted in the creation of an intellectual framework that would enable the development of an entire computational pipeline, optimized for both prediction of antimicrobial resistance and biological decision-support applications.

3.2 Dataset Acquisition

These data were used in the current study: Biological and clinical data from publicly available biomedical archives, including data on antibiotic resistance; Specialised research data on pathogenic microbes. Collected datasets contained data about microbiological resistance, phenotypic susceptibility, as well as data about the genome and labels for supervised machine learning prediction tasks. Data sets have been extracted based on their biological relevance, availability, and the content and applicability of the data for ML studies on resistance. Data acquisition was carried out to enable good representation of resistant and susceptible microbial samples needed for predictive classification modelling. Public biological repositories and antibiotic resistance databases allow to have defined datasets which are widely used in computational biology research to enable the comparative comparison and repeatability of machine learning operations [64]. Collected data comprised multi-dimensional biological data, with both category and numerical data for antimicrobial resistance characteristics. The datasets were a fundamental prerequisite for the research undertaken for preprocessing, feature extraction, predictive modelling and explainability analysis.

3.3 Data Preprocessing

Data preparation was one of the most important parts of the computational workflow as raw biological data may contain value gaps, duplicate records, inconsistent data, or background noise that can affect the predictive model's accuracy [65]. Proper pre-processing optimises data quality, increases computing efficiency and adds substantially towards model dependability and generalisability. The pre-processing step is the process of performing several operations as illustrated in (Fig. 1).

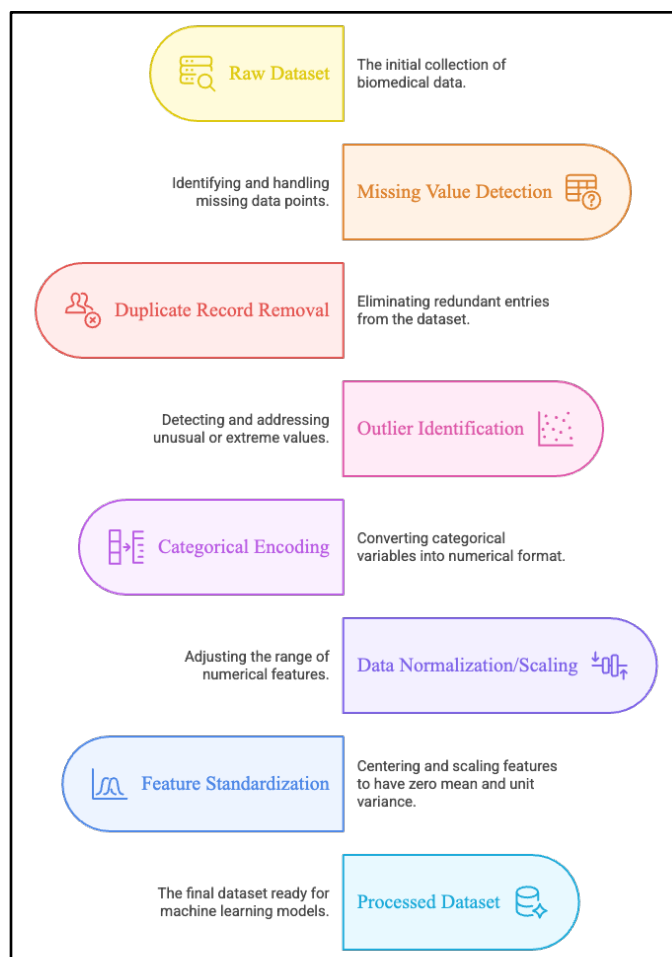


Fig. 1- Data Preprocessing Workflow

Suitable data preprocessing methods were utilized to handle missing values following the dataset parameters properties and distribution. The numerical variables were standardized to minimize their variance and improve the training procedure of machine learning models. In biomedical ML pipelines, standardisation and normalisation methods are of particular importance as there may be important variations in the attributes of the features, which could affect the prediction algorithms [66]. Label and one-hot encoding techniques were used to convert categorical biological parameters to machine-readable numerical formats. Data transformation procedures were also applied to ensure consistency of computation throughout the entire training and validation process of the models. The data set was then split into training and testing sets to enable unbiased

assessment of the model performance. Data splitting is crucial to reduce overfitting and enhance predictive generalization on unseen data samples [67].

3.4 Feature Engineering and Feature Selection

Feature engineering was performed to extract biologically relevant information from the processed datasets and improve predictive efficiency of machine learning models. Biomedical datasets frequently contain high-dimensional variables that may include redundant, irrelevant, or correlated features capable of negatively influencing classification performance [68]. Therefore, feature engineering and feature selection are important for improving robustness, computing efficiency and interpretability of the model. Traits related with antimicrobial resistance were identified based on biological relevancy, statistical value, and effect on predictive results. The feature engineering procedure included the translation and optimisation of dataset variables to improve machine learning performance and enable the detection of significant biological correlations. Feature selection methods were used to minimise the dimensionality of the data collection and to exclude the non-contributing variables. The reduction of unnecessary information may increase computing performance and reduce overfitting during model training methods [69]. Dimensionality reduction methods comprising PCA if applicable was addressed to strengthen the visualisation of attributes and minimise the complex nature of high dimensional biological datasets. The feature engineering phase was thus of great importance in increasing predictability, stability, the readability of a model, and the biological significance of the utilised computational framework.

3.5 Machine Learning Model Development

The present study utilized supervised machine learning methodologies for antimicrobial resistance prediction and pathogen classification. Supervised learning algorithms were selected because the acquired datasets contained labeled resistance-associated outcomes suitable for classification-based predictive analysis [70].

Several machine learning algorithms were implemented and comparatively evaluated during the study, including:

- Random Forest
- Support Vector Machine
- Logistic Regression (LR)
- Decision Tree (DT)
- Gradient Boosting-based approaches

To accomplish this, the Random Forest algorithms were implemented since they are highly resistant to overfitting and can effectively manage multidimensional biomedical datasets [71]. The machine learning framework used for antimicrobial resistance prediction. The optimized dataset was divided into training and testing subsets, followed by implementation of Random Forest, Support Vector Machine, LR, DT, and Gradient Boosting models for comparative predictive analysis and performance evaluation (Fig. 2). Random Forest models use various decision trees to provide highly consistent ensemble predictions, thus boosting classification precision and lowering error in prediction due to variation. SVM algorithms were used that are effective in nonlinear classification operations employing multidimensional genomic and phenotypic information [72]. The SVM models calculate the optimal hyperplanes that can separate classes of resistant and vulnerable and maximise the margin of separation. Also, LR techniques were applied

owing to their straightforwardness, interpretability and statistical effectiveness in binary classification tasks typically seen in antibiotic resistance prediction research [73]. Decision Tree models were similarly employed because they provide interpretable prediction pathways useful for understanding feature-based decision structures. Gradient Boosting and ensemble learning methods were examined owing to the potential to enhance predicted accuracy via repeated optimisation and the amalgamation of many weak prediction models [74]. Ensemble methods are very beneficial in biomedical machine learning research that entails intricate nonlinear interactions among biological factors.

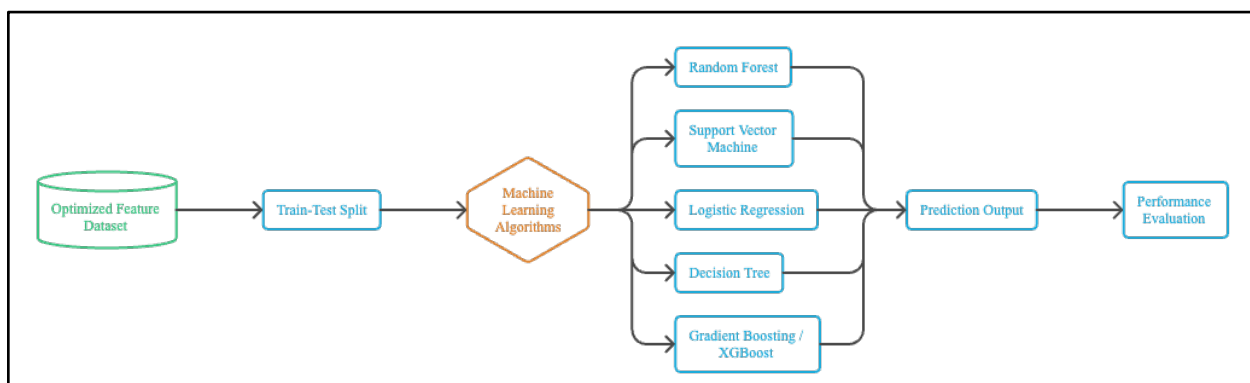


Fig. 2- Machine Learning Architecture for Antimicrobial Resistance Prediction

3.6 Model Training and Validation

The machine learning models were trained on the processed training dataset and verified utilising independent validation datasets to objectively assess predicted performance. Cross-validation methods were used during model construction to reduce prediction bias and strengthen generalisability to unknown data sets [75]. K-fold cross-validation techniques were used to enhance model resilience and mitigate variability linked to random dataset segmentation. Cross-validation methods are extensively used in biomedical machine learning studies since they enhance the dependability of prediction performance assessment and mitigate the dangers of overfitting

[76]. Hyperparameter optimisation approaches were also used when suitable to increase classification efficacy and computational reliability of the prediction models by optimizing the parameters like tree depth, learning rate, regularisation parameters and kernel functions, the prediction efficiency and dependence of the model have been enhanced. The training and validating methodology ensured that the computational models generated had predictive validity and were generalisable in various situations involving different types of datasets.

3.7 XAI Framework

In the present work, the study used XAI methods to enhance the transparency, interpretability and trustworthiness of antimicrobial resistance prediction models based on machine learning. The Explainable Artificial Intelligence framework employed for interpreting machine learning predictions (Fig.3). Explainability is of critical importance in healthcare-oriented usages for artificial intelligence considering that biomedical decision-making is based on interpretable as well as biologically significant prediction outcomes [9]. SHAP analysis was the main method used in this research to analyse feature relevance and determine the contribution of different variables affecting the outputs of standardised predicting models. SHAP analysis depends on cooperative game theory and offers local as well as global interpretability of intricate machine learning systems [27]. SHAP values permitted us to determine physiologically meaningful variables correlated with antibiotic resistance outcome predictions. This enhanced comprehension of model behaviour and interpretation of the associations between the biological factors and categorisation predictions. Feature significance visualisation and interpretability-based studies were also undertaken to assess the impact of resistance-related factors on the decision-making processes of the models. Explainability techniques enhance transparency and help validate computational predictions

contrary to known biological systems [32]. Biomedical machine learning systems for XAI enhance the credibility, reusability and clinical usefulness of predictive computational systems.

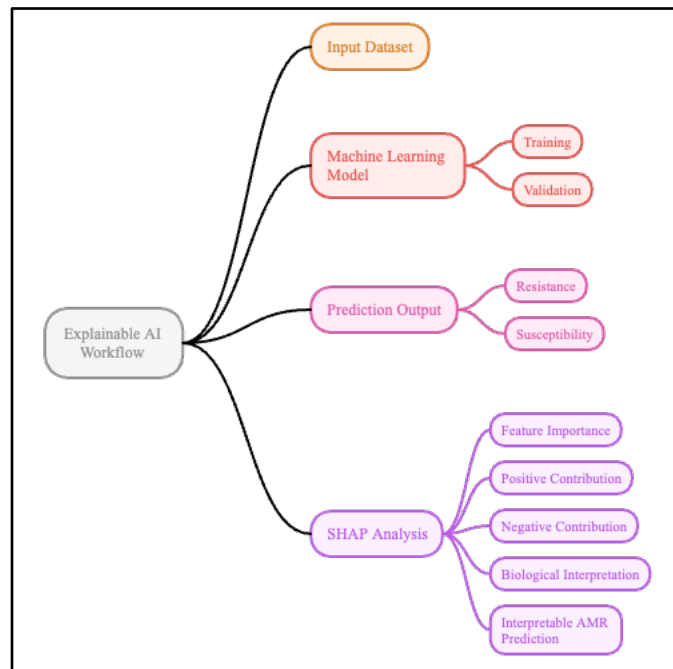


Fig. 3- Explainable AI Workflow for AMR Prediction

3.8 Performance Evaluation Metrics

Multiple classification and statistical performance measures frequently used in biomedical artificial intelligence research were applied to assess the predictive ability of the machine learning models developed [77]. Several evaluation metrics were used to provide an in-depth analysis of the correctness of classification and the reliability of prediction.

The following parameters were taken into consideration when evaluating:

- Accuracy
- Precision
- Recall (Sensitivity)
- Specificity

- F1-Score
- Receiver Operating Characteristic Curve
- Area Under the Curve

Accuracy was used to evaluate the overall accuracy of the model's prediction for resistant and susceptible categories. The recall was used to measure the number of true positive resistant cases the model could identify, while the precision measure was the number of cases correctly predicted [78]. F1-score gave a balance between precision and recall, which is essential for potentially imbalanced biomedical datasets. ROC analysis and AUC evaluation were also performed for the machine learning models to assess the models' classification discrimination capability and also overall predictive robustness. The performance measures allowed for the comparison of multiple machine learning approaches and helped in finding the best and most correct prediction model for analysing antibiotic resistance.

3.9 Computational Tools and Software Environment

During this study, computational analysis has been performed in Python-based programming environments and using machine learning libraries that are commonly used in computational biology research and biomedical data science [79]. The Python programming language was selected primarily for its extensive scientific computation support, flexibility, and the existence of sophisticated machine learning as well as data visualisation frameworks.

Several computational libraries and software packages were used, such as:

- NumPy
- Pandas
- Scikit-learn
- Matplotlib

- Seaborn
- SHAP
- TensorFlow/Keras (where applicable)

These specialized libraries allowed for efficient layout of initial processing pipelines, feature engineering, predictive modelling and analysis of model performance and feature significance distributions graphically. Data visualisation methods were used to visualise the classification results, confusion matrices, ROC curves, feature significance scores and explainability analyses incorporated into the research to present them in an interpretable and scientifically relevant way.

3.10 Overall Computational Workflow

The following sequential stages were involved in the overall computational workflow carried out during the present study (Fig.4). The integrated methodological framework allowed to systematically study patterns of antimicrobial resistances, while highlighting the interpretability, transparency and computational robustness of predictive biomedical models.

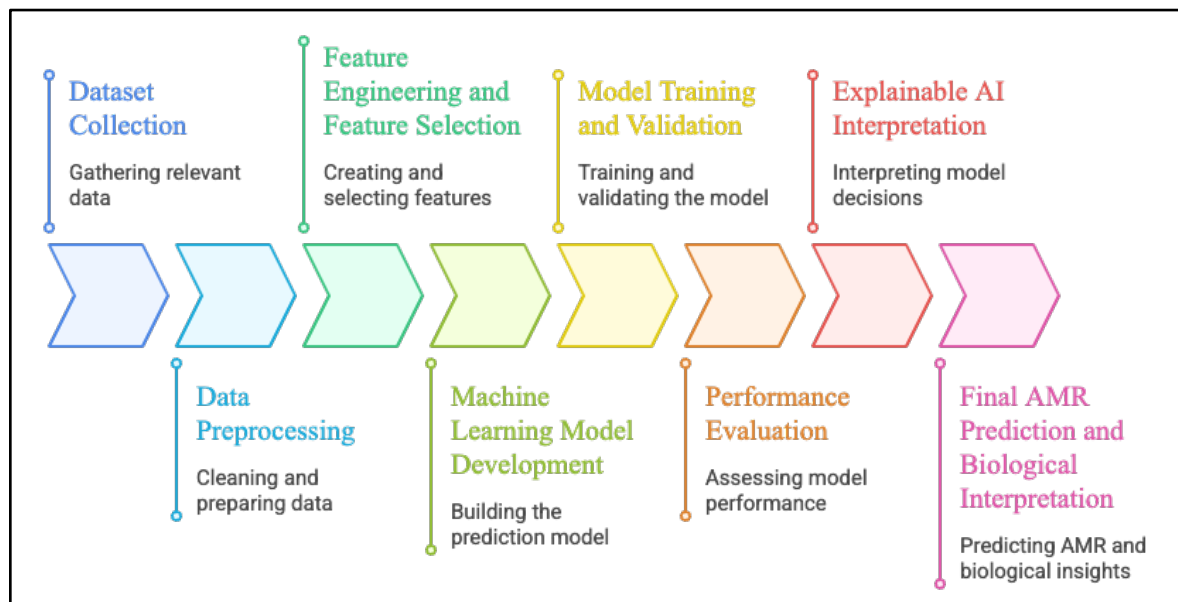


Fig. 4- Computational Workflow for Antimicrobial Resistance Prediction

CHAPTER 4

RESULTS

4.1 Overview of Experimental Analysis

In the current research, we aimed to develop and test computational frameworks that utilize ML and XAI to predict antibiotics resistance and analyse antibiotic usage patterns, respectively. Major supervised machine learning techniques were developed and evaluated for their classification capability and the reliability and understandability of their prediction results with the antibiotic resistance data sets analysed. Experimental investigation encompassed sequential pre-processing processes, feature engineering, model training and validation, performance comparison and explanation analysis. A primary aim of the outcomes analysis was to define the best predictive framework for categorising antimicrobial resistance. Additionally, maintaining the integrity and interpretability of model predictions was accomplished by the implementation of the XAI techniques [9]. The results of the conclusions indicate that the obtained results have great potential to use the machine learning computational tools in analysing the multidimensional biomedical information and identify the physiologically meaningful antibiotic resistance linked patterns. Furthermore, adopting XAI improved understanding of the contribution of features to the results of the prediction.

4.2 Dataset Processing and Preprocessing Outcomes

Antimicrobial resistance associated datasets were extracted are characterized by the presence of missing data, duplicated data, data formatting inconsistencies, and varying distributions of the variables. Therefore, data quality and computational appropriateness have been enhanced by

preparing the data prior to using the model. Duplicate records and inconsistencies across the dataset were resolved by using data cleaning strategies. Insufficient data were processed utilising suitable preprocessing techniques according to the characteristics as well as distribution trends of the variables. To ensure uniformity across the machine learning processes, and to alleviate computational bias based on the feature variance, numerical variables were normalised and standardised [65]. Label encoding and One Hot encoding were used to encode categorical variables that were associated with microbial resistance profiles and biological classifications. These preprocessing steps were found to be appropriate for the machine learning approach and allowed the successful computational analysis of the datasets. Preprocessing of the dataset resulted in increased consistency of the structure and reduced noise, thereby improving the suitability of the dataset for the training and evaluation of predictive models. Processed datasets were then partitioned into training and testing sets for classification analysis and evaluation of classification performance.

4.3 Feature Engineering and Feature Selection Results

The effectiveness and interpretability of the machine-learning models used were enhanced through feature engineering and feature selection. The first data sets were high dimensional variables that had both biological and computational meaning. Feature selection algorithms decreased the dimensional complexity by removing duplicate as well as inadequate contributing variables while keeping the physiologically relevant antimicrobial resistance linked characteristics [69]. The optimized feature set improved computational efficiency and reduced risks associated with model overfitting during training procedures. The feature engineering technique discovered some very informative factors that add greatly to the classification performance. SHAP-based feature prioritisation identified several highly impacting antimicrobial resistance genes that contributed considerably towards the predicted occurrence of carbapenem resistance in *Klebsiella pneumoniae*

(Fig.5). The biological variables related to resistance showed a substantial predictive effect across multiple machine learning models, demonstrating their relevance in the antimicrobial resistance classification tasks. SHAP feature importance analysis further demonstrated that several transporter-associated, membrane-associated and regulatory biomarkers contributed substantially towards ciprofloxacin resistance prediction in *Pseudomonas aeruginosa* (Fig. 6). Dimensionality reduction and feature optimisation increased visualisation and interpretability of model outputs, therefore leading to greater effective deployment of XAI analysis.

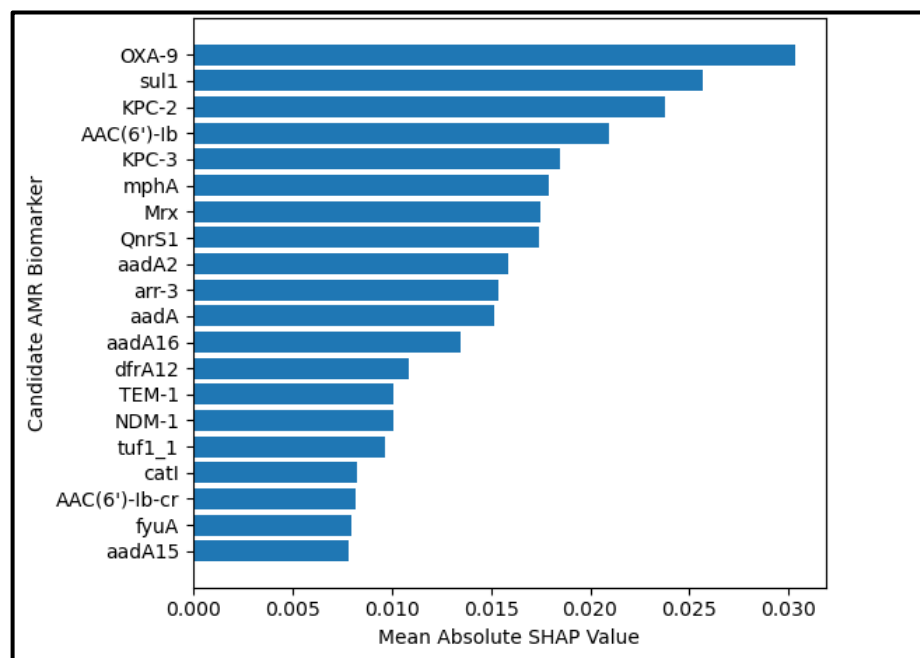


Fig. 5- Top SHAP-prioritized KP resistance biomarkers

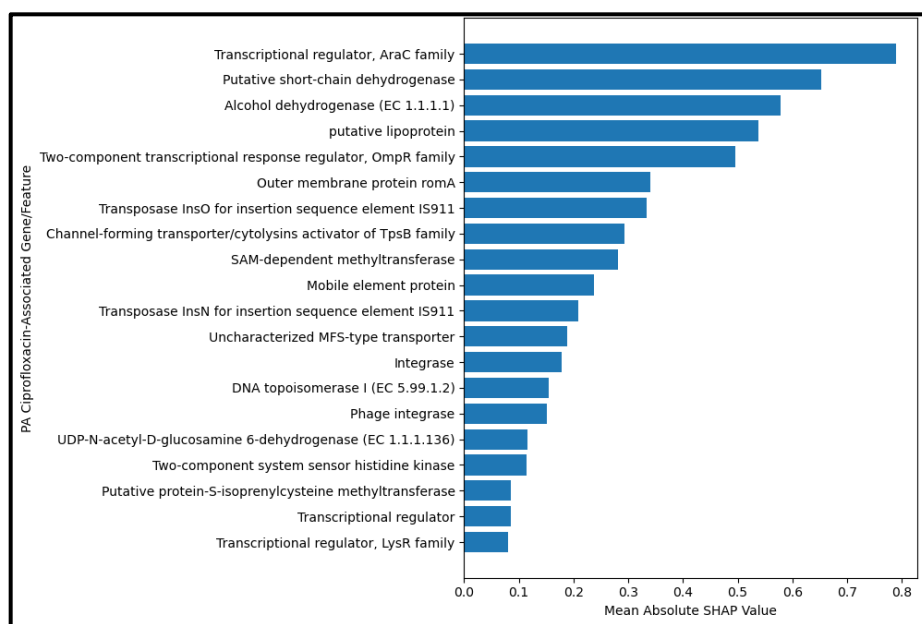


Fig. 6- Top SHAP-prioritized PA resistance biomarkers

4.4 Machine Learning Model Performance Analysis

Multiple supervised machine learning algorithms were implemented and comparatively evaluated for antimicrobial resistance prediction and pathogen-associated classification analysis. The evaluated models included:

- Random Forest
- Support Vector Machine
- Logistic Regression
- Decision Tree
- Gradient Boosting-based models

Predictive accuracy, robustness and classification capabilities were assessed using separate testing datasets, where the processed datasets were used to train each model. The Random Forest model has good classification performance because to ensemble learning capabilities and the capacity to prevent overfitting [71]. The model achieved good results on multidimensional biological datasets

and provided reliable classification results of resistant and susceptible microbial samples. RF also allowed for better interpretability of feature relevance compared to some traditional black-box prediction techniques. SVM techniques showed a good predictive ability as well, notably for nonlinear classification problems with complicated resistance related data [72]. The SVM model could find the optimal decision boundary between resistant and susceptible classes with high classification precision and sensitivity. LR models provided a relatively interpretable classification output and efficient performance for the binary resistance prediction tasks. LR models provided reasonable classification reliability and computational efficiency, however they were relatively simpler compared to ensemble and deep learning methods [73]. DT models showed moderate predictive capabilities, and they also gave very interpretable decision paths good to understand the classifications behavior based on the features. The interpretability of Decision Trees contributed positively toward biological understanding of resistance-associated predictive variables. Gradient Boosting and ensemble-based techniques resulted in better predictive equilibrium and classification performance owing to repeated optimisation of inadequate predictive learners [74]. Comparative evaluation of the implemented machine learning algorithms demonstrated variability in predictive efficiency across KP and PA datasets, where ensemble-based classifiers exhibited comparatively superior performance metrics (Fig. 7). The methods were efficient in minimising the errors of prediction and strengthening the resilience of the model broadly for classification analysis. For the classification performance, the performance comparison of machine learning models showed that ensemble learning methods significantly outperformed straightforward linear models. However, less complicated interpretable models had merits in terms of clarity and explanation of the predicted results.



Fig. 7- Comparative performance analysis of KP and PA machine learning models

4.5 Classification Performance Metrics

The predictive performance of the implemented machine learning models was evaluated using multiple statistical and classification metrics commonly utilized in biomedical machine learning research [77]. These included:

- Accuracy
- Precision
- Recall (Sensitivity)
- Specificity
- F1-Score
- Receiver Operating Characteristic Curve
- Area Under the Curve

The built models showed good as a whole classification performance for antimicrobial resistance prediction requirements. Accuracy metrics demonstrated that the resistant microbial samples could potentially be predicted precisely with comparatively low false positive predictions. Both the sensitivity as well as recall evaluations proved that predictive models were able to predict a significant number of true resistant cases in the dataset. The ROC-AUC analysis indicated good discrimination capability of the predictive model for KP resistance with the high sensitivity and specificity properties in all the validation folds (Fig. 8). The most significant clinical implication of antimicrobial resistance prediction is the impact of failing to identify resistant pathogens, which may be heavily influenced by the recall value, especially for high values [78]. When using F1-score study, the assessment was balanced between accuracy and recall, crucial for a biological data set that may contain class imbalance. The cross-validated ROC-AUC analysis of the PA dataset revealed moderate, but biologically meaningful discriminatory power with a relatively small number of samples and variable distributions of resistance-associated biomarkers across the samples (Fig. 9). In general, the ensemble machine learning models achieved higher F1-score than the simpler linear classification models. ROC analysis and AUC evaluation further demonstrated strong discrimination capability of the predictive models. High AUC values indicated improved ability of the implemented machine learning frameworks to differentiate resistant and susceptible microbial samples effectively [80].

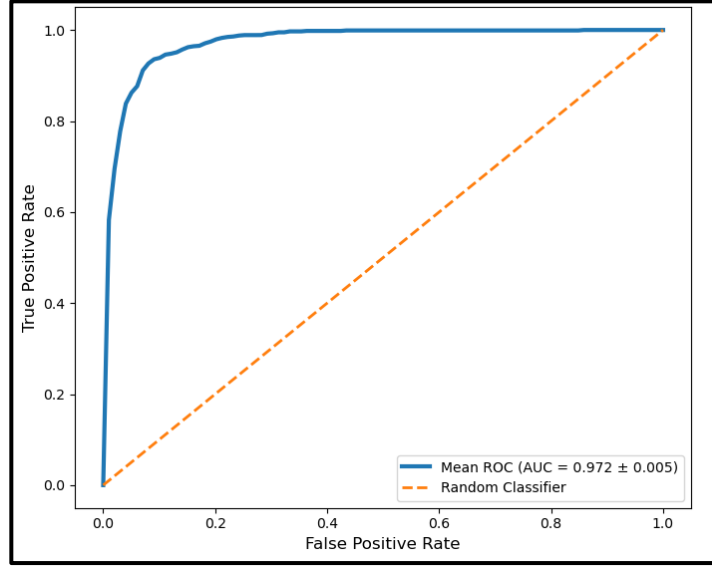


Fig. 8-Cross-validated ROC-AUC curve for KP resistance prediction

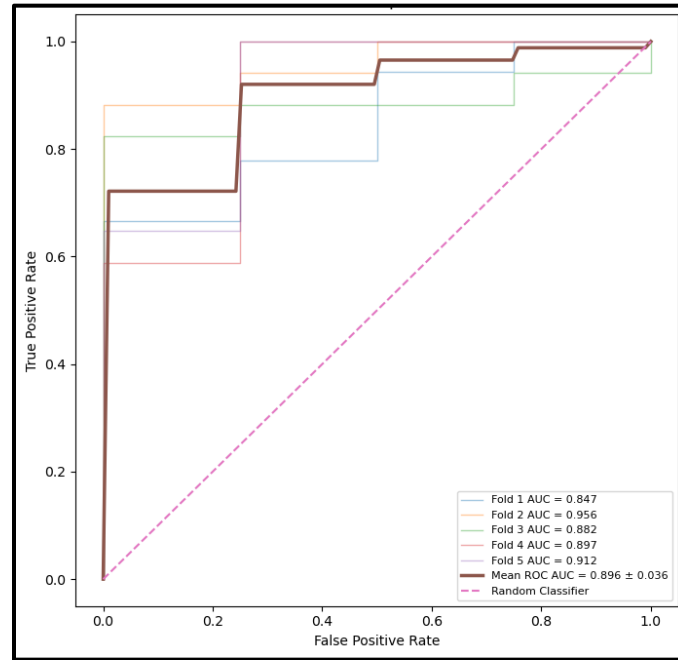


Fig. 9-Cross-validated ROC-AUC curve for PA resistance prediction

4.6 Explainable Artificial Intelligence (XAI) Results

XAI approaches were included in the computational process to boost the interpretability and transparency of machine learning predictions. The main explainability framework employed was

SHAP for determining feature contributions and finding physiologically relevant prediction factors [55]. SHAP study effectively revealed key antimicrobial resistance-associated variables impacting the outputs of machine learning models. The resulting SHAP values determined the contribution of different variables towards prediction results and hence improved comprehension of model behaviour and categorisation reasoning. Through global explainability analysis, it was shown that various biological and resistance related parameters consistently contributed heavily towards the categorisation predictions across several machine learning models. SHAP summary visualization identified several clinically relevant antimicrobial resistance genes including OXA-family carbapenemases and efflux-associated determinants that strongly influenced KP resistance categorization outcomes (Fig. 10). Features relevant to microbial resistance positively impacted the result of resistant categorisation, whereas features relevant to susceptibility adversely contributed to the prediction of resistance. Furthermore, local explainability analysis allowed the interpretation of individual prediction instances, thereby boosting the transparency and interpretability of patient-specific or sample-specific categorisation results [29]. Such interpretability is particularly valuable in healthcare-oriented machine learning systems where clinicians require understandable and biologically meaningful decision-support mechanisms. SHAP analysis produced feature significance visualisations that enhanced the transparent nature of prediction models and allowed for confirmation of computational results against known biological pathways of antibiotic resistance. The SHAP interpretability framework demonstrated that ciprofloxacin resistance prediction in PA was influenced by multiple adaptive regulatory, transporter-associated and membrane-associated biological features contributing variably towards model output behaviour (Fig. 11).

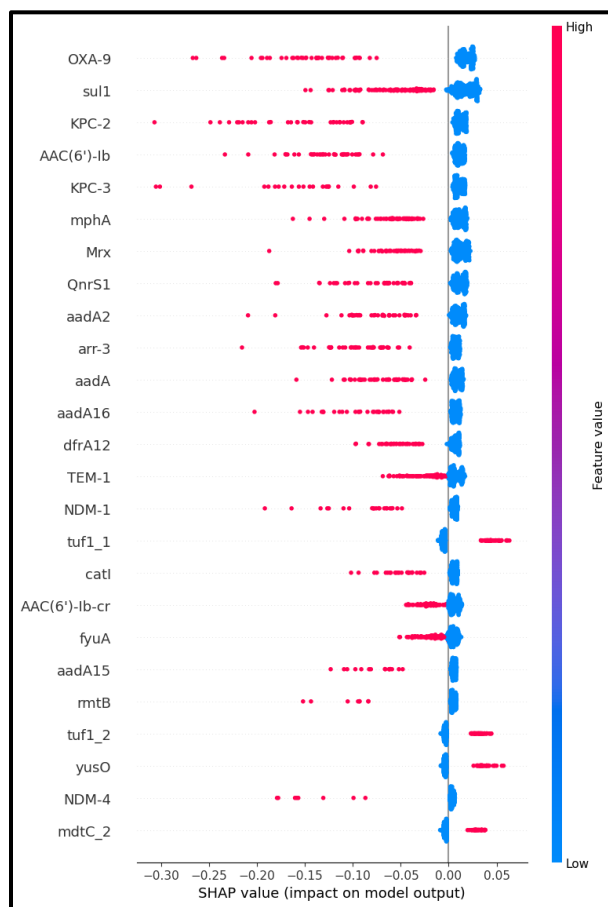


Fig. 10- KP SHAP summary plot for resistance prediction interpretability

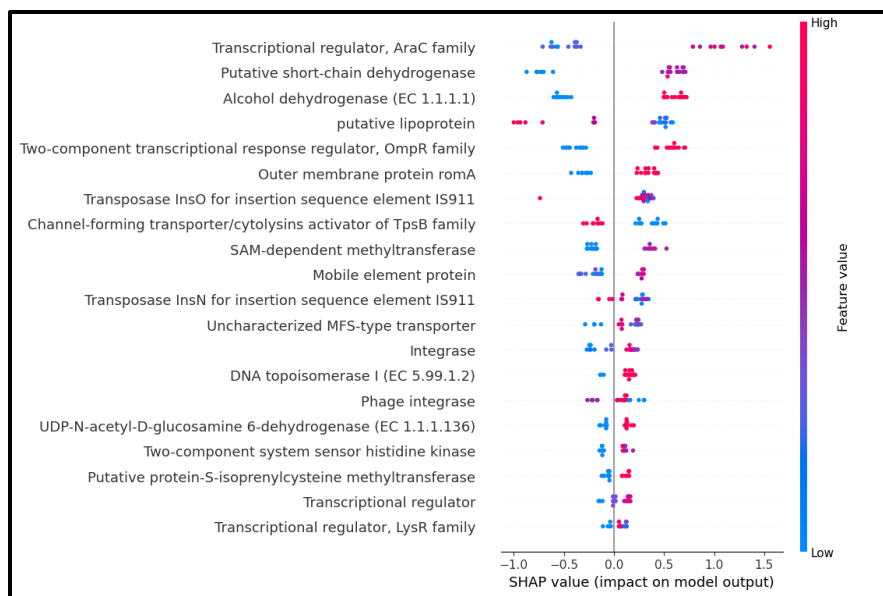


Fig. 11-PA SHAP summary plot for resistance prediction interpretability

4.7 Comparative Analysis of Machine Learning Models

The comparison-based assessment of the applied machine learning algorithms revealed considerable disparities in predictive power, computational complexity and interpretability features. The ensemble learning architectures comprising RF and Gradient Boosting techniques typically showed higher prediction accuracy and classification resilience [49]. Such models could perform well with multi-dimensional biomedical data and minimise prediction errors resulting from variation. The SVM techniques are found to be good classifiers on the nonlinear resistance related datasets, but the parameters should be carefully optimised to obtain better prediction performance. LR and DT methods showed comparable lower predictive complexity and gave ease of interpretability and computational simplicity. The results presented the incorporation of XAI, which showed that the ensemble learning methods, even though showing promising in predicting the outcomes, learned a great deal from explainability analysis to make the outcomes of categorisation more transparent. The use of SHAP-based feature interpretation enhanced understanding of the prediction behavior of the different models implemented for the most important biological variables. The study revealed that using high-performing ML models in tandem with XAI frameworks is an effective approach to building accurate and understandable antibiotic resistance prediction systems.

4.8 Visualization and Interpretability Outcomes

Throughout the investigation visualisation tools were used to enhance the interpretability and portrayal of the prediction outputs. Confusion matrix, ROC curve, feature significance plot, and SHAP summary plot were used for evaluating classification performance and interpretability of ML models. The results of a confusion matrix analysis were used to show good classification of resistant and susceptible microbial samples and to identify the distributions of true-positive, false-

positive, true-negative and false negative predictions [81]. The predictive frameworks also showed good discrimination with ROC curve visualizations. Using SHAP summary plots and feature contribution visualisations, it becomes easy to see how physiologically relevant factors contribute to the results of predictions concerning antimicrobial resistance. There is a growing focus on visualization-based explainability that has helped develop and improve the interpretability of prediction frameworks, in addition to understanding feature-driven categorisation approaches. The use of visualisation methods helped so greatly to improve the transparency, repeatability and interpretability of machine learning powered biomedical prediction frameworks. Overall, the findings obtained in the current research revealed that machine learning and XAI approaches are excellent computational frameworks for the prediction of antibiotic resistance and the identification of pathogen-associated patterns. Ensemble learning methods showed good prediction and better classification resilience for multidimensional biomedical data. The use of XAI boosted the interpretability and transparency of predicted results by pinpointing physiologically important antibiotic resistance-associated characteristics that impacted model behaviour. The results also revealed that the interpretable machine learning frameworks may be very useful for precision medicine, antimicrobial stewardship, as well as AI-assisted healthcare systems. Thus, the merging of computational biology, machine learning and explainability methods constitutes a viable multidisciplinary approach for future research on antibiotic resistance and medically predictive applications.

CHAPTER 5

DISCUSSION

This research investigated the use of machine learning and XAI approaches for predicting antibiotic resistance and analysing pathogen-associated patterns. The outcomes proved that computational intelligence techniques have a tremendous potential to be applied in analysing biological information in various dimensions and making accurate predictions in the field of drug resistance research. The use of explainability frameworks improved the transparency and interpretability of predictive systems, thereby boosting their potential application in the biomedical and healthcare contexts. The worrying trend of antimicrobial resistance has led to a pressing need for fast, accurate and easy-to-interpret prediction models that can aid in the management of infectious diseases and in antimicrobial stewardship [1]. The conventional AST techniques are clinically indispensable but remain dependent on conventional laboratory procedures which are time-consuming and may delay treatment during critical stages of disease [42]. The present work illustrates the potential of using computational methods focused on machine learning to markedly improve analytical efficiency and timely warning of patterns of resistance based on multidimensional biological data sets. The preprocessing and feature engineering process during the study significantly improved the stability of prediction and computational efficiency of the machine learning models implemented. Biomedical data often includes mixed-type variables, missing values, data inconsistencies, and redundancy that can have a detrimental impact on model performance [65]. However, suitable data preprocessing and feature optimization still need to be done to enhance the quality of the data, optimize the computational complexity and minimize the

risk of model overfitting. The effective use of preprocessing processes in this research boosted the fundamental uniformity of the datasets and boosted predicted appropriateness for classification analysis. Feature engineering and feature selection investigations suggested that a number of antimicrobial resistance-associated factors were substantial contributors to the prediction results across several machine learning models. This conclusion corresponds to line with prior work which has shown that the retrieval of physiologically relevant features considerably enhances classification efficacy in biomedical machine learning based applications [68]. The removal of unnecessary variables also enhanced interpretability and enabled a more effective implementation of explanation-based analysis. Among the implemented machine learning algorithms, ensemble learning approaches such as Random Forest and Gradient Boosting demonstrated comparatively superior predictive performance and classification robustness. These results coincide with earlier antibiotic resistance prediction research demonstrating high classification accuracy of ensemble-based procedures for multidimensional biological datasets [19]. Ensemble learning techniques are able to combine a number of weak predictive learners which in turn reduces the variance associated mistakes and enhances the generalisation power across several unobserved data samples . Random Forest methods have shown to have a great power to anticipate resistance, while at the same time keeping somewhat enhanced interpretability compared to very sophisticated deep learning frameworks. The use of the Random Forest models was advantageous for their resistance to overfitting and their ability to make good classifications from high-dimensional datasets that culminated in stable classification results [71]. The Support Vector Machine models also exhibited good performance for nonlinear resistance related data sets, as previously shown for genomic and biological classification data sets to be suitable for classification tasks [72]. There are simpler machine learning methods like Logistic Regression and Decision Trees, which have lower

complexity in their predictions, but these methods had a few advantages, in terms of transparency and computational interpretability. Specifically, Decision Tree models came up with interpretable routes of classification according to characteristics, which aided with biological understanding of how predictions were generated. Such interpretability is very valuable in biological applications since physicians want comprehensible computational decision support systems [73]. Deep learning approaches have shown remarkable prediction performance for a variety of biological use cases employing genomic and imaging data [18]. However, more sophisticated brain architectures tend to behave as black-box systems, offering minimal explanation about underlying principles of decision-making. In healthcare-oriented organisations, the dearth of transparency may erode physician trust and hinder the use of AI-assisted diagnostic technologies. Thereby, the need of the incorporation of Explainable Artificial Intelligence methods has emerged to improve the transparency, interpretability, and reliability of machine learning-based healthcare systems. The assessment of Explainable Artificial Intelligence conducted in the current research proved to be very useful in increasing the comprehension of feature contributions impacting prediction results. SHAP analysis successfully identified biologically significant variables associated with antimicrobial resistance prediction and quantified their influence on classification decisions [27]. The feature significance distributions produced increased interpretability of machine learning outcomes and allowed confirmation of computational predictions against physiologically relevant antibiotic resistance-associated pathways. The interpretability of the prediction models was significantly enhanced by the effectiveness of SHAP analysis for global and local interpretability. Global interpretability helped to recognise general patterns in feature significance that influenced resistance prediction, while local interpretability was beneficial in interpreting particular prediction results at the test level [29]. Such interpretability is associated with biomedical and

clinical applications, as healthcare professionals need intelligible and physiologically relevant The results from the integration of Explainable Artificial Intelligence are consistent with the recent studies which have focused on the significance of interpretable frameworks of artificial intelligence in precision medicine and medicine applications within healthcare [9]. Explainability not only enhances the trustworthiness of computing systems but also contributes to the reproducibility of computational systems and their compliance with regulatory requirements, as well as the validation of the prediction models with current biological knowledge. Therefore, the incorporation of explainability approaches in antibiotic resistance prediction systems is a major step in therapeutically beneficial artificial intelligence frameworks. The effectiveness of machine learning assisted computational tools for studying antibiotic resistance was also demonstrated when performance metrics collected during the study were classified by the researchers. The accuracy, precision, recall, F1 score as well as ROC-AUC analysis shows a significant predictive power for both resistant and susceptible microbial groups. The high recall and sensitivity values were especially significant, as resistance will not be detected if the organism is not recognised, which can have a major impact on the therapy outcome and clinical decision making processes [78]. Additionally, receiver operating characteristic and area under the curve analyses showed high discriminatory power of the machine learning models implemented. The large AUC values indicated that the prediction algorithms were able to efficiently discriminate resistant and susceptible microbial samples. The findings align with previous biomedical machine learning studies that have shown that AI-powered categorisation systems have a significant potential to be used for predicting infectious diseases and tracking antimicrobial resistance [80]. In the current work, good findings were achieved, however there are still some limitations with machine learning-based antimicrobial resistance prediction structures. Biomedical datasets are often

imbalanced, heterogeneous and variable which may impair predictive accuracy and model generalisability [30]. Insufficient variety of the collected data and insufficient biological annotations could further restrict the application of prediction models to diverse clinical groups and geographical areas. Another significant issue is the computational difficulty of sophisticated machine learning and explainability methods. Explainable Artificial Intelligence frameworks emphasise openness, however some explainability methods could boost computation costs and need to be analysed carefully in very complicated datasets [32]. Moreover, the prediction models are still reliant on data quality, accuracy of preprocessing and biological significance of the retrieved characteristics. The present work was primarily dedicated to the computational assessment and forecasting of antimicrobial resistance by machine learning; consequently, additional experimental validation with large-scale clinical and genomic datasets would increase the relevance of the proposed computational framework. Recent advancements in computational biology, machine learning, and precision medicine suggest considerable future potential for Explainable Artificial Intelligence-driven biomedical frameworks. The amalgamation of interpretable artificial intelligence methodologies with genetic sequencing technology, pathogen monitoring systems, and personalised healthcare methods might markedly enhance the supervision of infectious diseases and antimicrobial stewardship initiatives [81]. AI-assisted computational systems which offer transparent, physiologically relevant predictions might help diagnostic decision-making, therapy optimisation and early identification of resistant microbial strains. The present investigation mainly targeted the machine learning and Explainable Artificial Intelligence approaches while improving other treatment techniques such as bacteriophage-based therapy might provide additional advantages to the computational antimicrobial resistance investigation [60].

CHAPTER 6

CONCLUSION

The current research showed the effectiveness of ML and XAI-based computational techniques for prediction of antibiotic resistance and assessment of patterns of pathogens. With the rising challenge of antibiotic resistance, rapid, accurate and easily interpreted computational tools that can help with managing infectious diseases and antimicrobial management programs are essential [37]. The machine learning algorithms implemented, such as Random Forest, Support Vector Machine, Logistic Regression, Decision Tree, etc., and Gradient Boosting models, were able to obtain reliable predictions from the multidimensional biomedical datasets with success. Comparative analysis demonstrated that ensemble learning strategies perform better in terms of predicting robustness and classification performance in antimicrobial resistance prediction tasks [48]. The model performance and predicted stability were significantly enhanced after processing the data, engineering new features and selection of useful features, which reduced the redundancy and optimized the physiologically relevant features [68]. In addition, the use of Explainable Artificial Intelligence techniques boosted the interpretability and transparency of the prediction models. The discovery of essential resistance-associated variables for improved predictions of machine learning and understanding of the model behaviour was achieved using SHAP-based study [55]. For antimicrobial resistance prediction tasks, ensemble learning (EL) strategies were shown to be more predictive resilient than other strategies in the comparison analysis [48]. The preprocessing of data, feature engineering and feature selection significantly improved the efficiency of the model and the stability of its predictions by reducing the number of redundant

features and selecting physiologically relevant features, respectively [68]. Furthermore, the application of the XAI approaches enhanced the interpretability and transparency of the prediction models. SHAP based analysis was an effective method for identifying important factors associated with resistance that have an impact on machine learning predictions and gave greater understanding of the behaviour of a machine learning model [55]. The results presented in this work show that combining machine learning with explainability frameworks can markedly contribute to developing explainable and clinically relevant biomedical prediction systems. Such computational techniques can facilitate precision medicine, pathogen surveillance, as well as AI-assisted healthcare applications [28]. While the positive results are obtained, there are various restrictions like the imbalance of the dataset, the uncertainty in biological data and the computation complexity of the explainability approaches that are still significant challenges that should be further explored [9]. Future research could involve the incorporation of larger multi-omics datasets, improved deep learning models, and clinical data in real time to further improve the generalisability and clinical applicability of antibiotic resistance prediction models. In summary, the study efficiently shown the way computational biology, machine learning and Explainable Artificial Intelligence offer a feasible interdisciplinary structure for antibiotic resistance assessment and future precision medical solutions.

References

- [1] C. J. Murray *et al.*, “Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis,” *The Lancet*, vol. 399, no. 10325, pp. 629–655, Feb. 2022, doi: 10.1016/S0140-6736(21)02724-0.
- [2] C. Lee Ventola, “The Antibiotic Resistance Crisis Part 1: Causes and Threats,” *P & T*, vol. 40, no. 4, pp. 277–283, 2015.
- [3] A. Y. Peleg and D. C. Hooper, “Hospital-Acquired Infections Due to Gram-Negative Bacteria,” *New England Journal of Medicine*, vol. 362, no. 19, pp. 1804–1813, May 2010, doi: 10.1056/nejmra0904124.
- [4] E. D. Brown and G. D. Wright, “Antibacterial drug discovery in the resistance era,” Jan. 20, 2016, *Nature Publishing Group*. doi: 10.1038/nature17042.
- [5] R. Laxminarayan *et al.*, “Antibiotic resistance-the need for global solutions,” Dec. 2013. doi: 10.1016/S1473-3099(13)70318-9.
- [6] N. A. Lermينياux and A. D. S. Cameron, “Horizontal transfer of antibiotic resistance genes in clinical environments,” *Can. J. Microbiol.*, vol. 65, no. 1, pp. 34–44, 2019, doi: 10.1139/cjm-2018-0275.
- [7] C. Shen *et al.*, “Comprehensive assessment of nine docking programs on type II kinase inhibitors: Prediction accuracy of sampling power, scoring power and screening power,” *Brief. Bioinform.*, vol. 21, no. 1, pp. 282–297, Oct. 2018, doi: 10.1093/bib/bby103.
- [8] M. W. Pesesky *et al.*, “Evaluation of machine learning and rules-based approaches for predicting antimicrobial resistance profiles in gram-negative bacilli from whole genome sequence data,” *Front. Microbiol.*, vol. 7, no. NOV, Nov. 2016, doi: 10.3389/fmicb.2016.01887.
- [9] A. Barredo Arrieta *et al.*, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.
- [10] C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, “Deep learning for computational biology,” *Mol. Syst. Biol.*, vol. 12, no. 7, Jul. 2016, doi: 10.15252/msb.20156651.
- [11] I. Roca *et al.*, “The global threat of antimicrobial resistance: Science for intervention,” Jul. 01, 2015, *Elsevier Ltd*. doi: 10.1016/j.nmni.2015.02.007.
- [12] A. G. McArthur *et al.*, “The comprehensive antibiotic resistance database,” *Antimicrob. Agents Chemother.*, vol. 57, no. 7, pp. 3348–3357, Jul. 2013, doi: 10.1128/AAC.00419-13.
- [13] E. Zankari *et al.*, “Identification of acquired antimicrobial resistance genes,” *Journal of Antimicrobial Chemotherapy*, vol. 67, no. 11, pp. 2640–2644, Nov. 2012, doi: 10.1093/jac/dks261.
- [14] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/s41586-021-03819-2.
- [15] P. Hogeweg, “The roots of bioinformatics in theoretical biology,” *PLoS Comput. Biol.*, vol. 7, no. 3, 2011, doi: 10.1371/journal.pcbi.1002021.
- [16] I. Maljkovic Berry *et al.*, “Next Generation Sequencing and Bioinformatics Methodologies for Infectious Disease Research and Public Health: Approaches, Applications, and Considerations for Development of Laboratory Capacity,” *Journal of Infectious Diseases*, vol. 221, pp. S292–S307, Mar. 2020, doi: 10.1093/infdis/jiz286.
- [17] M. I. Jordan and T. M. Mitchell, “Data, privacy, and the greater good,” *Science (1979)*, vol. 349, no. 6245, pp. 253–255, Jul. 2015, doi: 10.1126/science.aac4520.

- [18] J. Heaton, “Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning,” *Genet. Program. Evolvable Mach.*, vol. 19, no. 1–2, pp. 305–307, Jun. 2018, doi: 10.1007/s10710-017-9314-z.
- [19] G. Arango-Argoty, E. Garner, A. Pruden, L. S. Heath, P. Vikesland, and L. Zhang, “DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data,” *Microbiome*, vol. 6, no. 1, 2018, doi: 10.1186/s40168-018-0401-z.
- [20] S. Min, B. Lee, and S. Yoon, “Deep learning in bioinformatics,” Sep. 01, 2017. doi: 10.1093/bib/bbw068.
- [21] A. Drouin, G. Letarte, F. Raymond, M. Marchand, J. Corbeil, and F. Laviolette, “Interpretable genotype-to-phenotype classifiers with performance guarantees,” *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-40561-2.
- [22] Z. C. Lipton, “The Mythos of Model Interpretability,” *Queue*, vol. 16, no. 3, pp. 31–57, Jun. 2018, doi: <https://doi.org/10.1145/3236386.3241340>.
- [23] W. Samek, T. Wiegand, and K.-R. Müller, “Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models,” *arXiv (Cornell University)*, Aug. 2017, doi: 10.48550/arxiv.1708.08296.
- [24] F. Doshi-Velez and B. Kim, “Towards A Rigorous Science of Interpretable Machine Learning,” *arXiv (Cornell University)*, Feb. 2017, doi: 10.48550/arxiv.1702.08608.
- [25] A. Adadi and M. Berrada, “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),” *IEEE Access*, vol. 6, pp. 52138–52160, Sep. 2018, doi: 10.1109/ACCESS.2018.2870052.
- [26] D. Gunning and D. W. Aha, “DARPA’s Explainable Artificial Intelligence Program Deep Learning and Security,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 4, Jun. 2024, doi: <https://ijisae.org/index.php/IJISAE/article/view/7322>.
- [27] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, “A survey of methods for explaining black box models,” *ACM Comput. Surv.*, vol. 51, no. 5, Sep. 2019, doi: 10.1145/3236009.
- [28] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” Jan. 01, 2019, *Nature Publishing Group*. doi: 10.1038/s41591-018-0300-7.
- [29] M. T. Ribeiro, S. Singh, and C. Guestrin, ““Why should i trust you?” Explaining the predictions of any classifier,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.
- [30] J. Wiens and E. S. Shenoy, “Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology,” *Clinical Infectious Diseases*, vol. 66, no. 1, pp. 149–153, Jan. 2018, doi: 10.1093/cid/cix731.
- [31] A. H. Holmes *et al.*, “Understanding the mechanisms and drivers of antimicrobial resistance,” Jan. 09, 2016, *Lancet Publishing Group*. doi: 10.1016/S0140-6736(15)00473-0.
- [32] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” May 01, 2019, *Nature Research*. doi: 10.1038/s42256-019-0048-x.
- [33] M. A. Fischbach and C. T. Walsh, “Antibiotics for emerging pathogens,” 2009. doi: 10.1126/science.1176667.

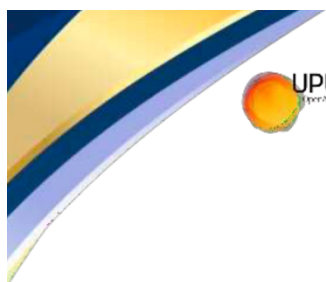
- [34] E. Tjoa and C. Guan, “A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.
- [35] E. Charani *et al.*, “Optimising antimicrobial use in humans – review of current evidence and an interdisciplinary consensus on key priorities for research,” Aug. 01, 2021, *Elsevier Ltd.* doi: 10.1016/j.lanepe.2021.100161.
- [36] T. P. Van Boeckel *et al.*, “Global trends in antimicrobial use in food animals,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 18, pp. 5649–5654, May 2015, doi: 10.1073/pnas.1503141112.
- [37] C. A. Michael, D. Dominey-Howes, and M. Labbate, “The antimicrobial resistance crisis: Causes, consequences, and management,” Sep. 16, 2014, *Frontiers Media S. A.* doi: 10.3389/fpubh.2014.00145.
- [38] N. Q. Balaban *et al.*, “Definitions and guidelines for research on antibiotic persistence,” *Nat. Rev. Microbiol.*, vol. 17, no. 7, pp. 441–448, Jul. 2019, doi: 10.1038/s41579-019-0196-3.
- [39] J. M. A. Blair, M. A. Webber, A. J. Baylay, D. O. Ogbolu, and L. J. V. Piddock, “Molecular mechanisms of antibiotic resistance,” Jan. 11, 2015, *Nature Publishing Group*. doi: 10.1038/nrmicro3380.
- [40] M. S. Mulani, E. E. Kamble, S. N. Kumkar, M. S. Tawre, and K. R. Pardesi, “Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: A review,” *Front. Microbiol.*, vol. 10, no. APR, 2019, doi: 10.3389/fmicb.2019.00539.
- [41] H. W. Boucher *et al.*, “Bad bugs, no drugs: No ESKAPE! An update from the Infectious Diseases Society of America,” *Clinical Infectious Diseases*, vol. 48, no. 1, pp. 1–12, Jan. 2009, doi: 10.1086/595011.
- [42] J. H. Jorgensen and M. J. Ferraro, “Antimicrobial susceptibility testing: A review of general principles and contemporary practices,” Dec. 2009. doi: 10.1086/647952.
- [43] X. Didelot, R. Bowden, D. J. Wilson, T. E. A. Peto, and D. W. Crook, “Transforming clinical microbiology with bacterial genome sequencing,” Sep. 2012. doi: 10.1038/nrg3226.
- [44] P. J. Van Camp, D. B. Haslam, and A. Porollo, “Prediction of Antimicrobial Resistance in Gram-Negative Bacteria From Whole-Genome Sequencing Data,” *Front. Microbiol.*, vol. 11, May 2020, doi: 10.3389/fmicb.2020.01013.
- [45] P. Larrañaga *et al.*, “Machine learning in bioinformatics,” Mar. 2006. doi: 10.1093/bib/bbk007.
- [46] P. Domingos, “A few useful things to know about machine learning,” Oct. 2012. doi: 10.1145/2347736.2347755.
- [47] M. Nguyen *et al.*, “Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella,” *J. Clin. Microbiol.*, vol. 57, no. 2, Feb. 2019, doi: 10.1128/JCM.01260-18.
- [48] M. Jaillard *et al.*, “A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events,” *PLoS Genet.*, vol. 14, no. 11, Nov. 2018, doi: 10.1371/journal.pgen.1007758.
- [49] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

- [50] J. Schmidhuber, “Deep Learning in neural networks: An overview,” Jan. 01, 2015, *Elsevier Ltd.* doi: 10.1016/j.neunet.2014.09.003.
- [51] J. Zou, M. Huss, A. Abid, P. Mohammadi, A. Torkamani, and A. Telenti, “A primer on deep learning in genomics,” *Nat. Genet.*, vol. 51, no. 1, pp. 12–18, Jan. 2019, doi: 10.1038/s41588-018-0295-5.
- [52] E. S. Kavvas *et al.*, “Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance,” *Nat. Commun.*, vol. 9, no. 1, Dec. 2018, doi: 10.1038/s41467-018-06634-y.
- [53] T. Miller, “Explanation in artificial intelligence: Insights from the social sciences,” Feb. 01, 2019, *Elsevier B.V.* doi: 10.1016/j.artint.2018.07.007.
- [54] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, “Machine learning interpretability: A survey on methods and metrics,” Aug. 01, 2019, *MDPI AG*. doi: 10.3390/electronics8080832.
- [55] F. Hohman, M. Kahng, R. Pienta, and D. H. Chau, “Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers,” *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 8, pp. 2674–2693, Aug. 2019, doi: 10.1109/TVCG.2018.2843369.
- [56] G. Montavon, W. Samek, and K. R. Müller, “Methods for interpreting and understanding deep neural networks,” Feb. 01, 2018, *Elsevier Inc.* doi: 10.1016/j.dsp.2017.10.011.
- [57] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” Dec. 2019, doi: 10.1007/s11263-019-01228-7.
- [58] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, “What do we need to build explainable AI systems for the medical domain?,” *arXiv (Cornell University)*, Dec. 2017, doi: 10.48550/arxiv.1712.09923.
- [59] A. Rajkomar, J. Dean, and I. Kohane, “Machine Learning in Medicine,” *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/nejmra1814259.
- [60] F. L. Gordillo Altamirano and J. J. Barr, “Phage Therapy in the Postantibiotic Era,” *Clin. Microbiol. Rev.*, vol. 32, Jan. 2019, doi: 10.1128/cmr.00066-18.
- [61] S. C. Emencheta *et al.*, “The Role of Bacteriophages in the Gut Microbiota: Implications for Human Health,” Oct. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/pharmaceutics15102416.
- [62] B. Mittelstadt, C. Russell, and S. Wachter, “Explaining explanations in AI,” in *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, Inc, Jan. 2019, pp. 279–288. doi: 10.1145/3287560.3287574.
- [63] F. Hohman, A. Head, R. Caruana, R. DeLine, and S. M. Drucker, “Gamut: A design probe to understand how data scientists understand machine learning models,” in *Conference on Human Factors in Computing Systems - Proceedings*, Association for Computing Machinery, May 2019. doi: 10.1145/3290605.3300809.
- [64] M. Boolchandani, A. W. D’Souza, and G. Dantas, “Sequencing-based methods and resources to study antimicrobial resistance,” Jun. 01, 2019, *Nature Publishing Group*. doi: 10.1038/s41576-019-0108-4.
- [65] J. Gama, “A survey on learning from data streams: Current and future trends,” Apr. 01, 2012, *Springer Verlag*. doi: 10.1007/s13748-011-0002-6.

- [66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: 10.1145/3065386.
- [67] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *International Joint Conference on Artificial Intelligence (IJCAI)*, IJCAI, 1995, pp. 1137–1143.
- [68] H. Peng, F. Long, and C. Ding, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, pp. 1226–1238, 2005, doi: 10.1109/TPAMI.2005.159.
- [69] M. Dash ' and H. Liu, "Feature Selection for Classification," *IDA ELSEVIER Intelligent Data Analysis*, vol. 1, no. 4, pp. 131–156, 1997, doi: 10.1016/s1088-467x(97)00008-5.
- [70] L. Breiman, "Statistical Modeling: The Two Cultures," *Statistical Science*, vol. 16, no. 3, pp. 199–231, 2001, doi: 10.1214/ss/1009213726.
- [71] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/a:1010933404324.
- [72] C. Cortes, V. Vapnik, and L. Saitta, "Support-Vector Networks Editor," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/bf00994018.
- [73] G. J. Goodhardt, A. S. C. Ehrenberg, and C. Chatfield, "The Dirichlet: A Comprehensive Model of Buying Behaviour," *J. R. Stat. Soc. Ser. A*, vol. 147, no. 5, pp. 621–655, 1984, doi: 10.2307/2981696.
- [74] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.
- [75] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat. Surv.*, vol. 4, pp. 40–79, 2010, doi: 10.1214/09-SS054.
- [76] V. N. Vapnik, "An Overview of Statistical Learning Theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, 1999, doi: 10.1109/72.788640.
- [77] J. Davis and M. Goadrich, "The Relationship Between Precision-Recall and ROC Curves," *International Conference on Machine Learning (ICML)*, 2006, doi: 10.1145/1143844.1143874.
- [78] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, Jan. 2020, doi: 10.1186/s12864-019-6413-7.
- [79] J. D. Hunter, "Matplotlib: a 2D Graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, Jan. 2007, doi: 10.1109/MCSE.2007.55.
- [80] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.
- [81] D. M. W. Powers and Ailab, "Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation," *Journal of Machine Learning Technologies*. [Online]. Available: <https://bioinfopublication.org/pages/article.php?id=BIA0001114>
- [82] A. Esteva *et al.*, "A guide to deep learning in healthcare," Jan. 01, 2019, *Nature Publishing Group*. doi: 10.1038/s41591-018-0316-z.

LIST OF CONFERENCES

1. INTERNATIONAL CONFERENCE ON: “BioFrontiers 2026: Trends, Technologies & Translational Research” hosted by the Department of Microbiology SRICT-ISR, UPL University of Sustainable Technology (UUST) on 6th and 7th March 2026



CERTIFICATE

This participation certificate is given to

Utkarsh Mishra

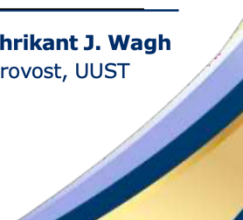
Has participated in Oral presentation at "INTERNATIONAL CONFERENCE ON: “BioFrontiers 2026: Trends, Technologies & Translational Research” entitled with “Phage-mediated Therapeutic Solution for Infectious Disease Treatment to Counter Antimicrobial Resistance” hosted by the Department of Microbiology SRICT-ISR, UPL University of Sustainable Technology (UUST) on 6th and 7th March 2026

Dr. Prashant Arya
Coordinator
Assistant Professor
SRICT-ISR, UUST

Dr. Trupti Patel
Asso. Prof. & Head
Department of Microbiology
SRICT-ISR, UUST

Dr. Snehal Lokhandwala
Dean Science and Sustainability,
SRICT-ISR

Dr. Shrikant J. Wagh
Provost, UUST



2. "17th International Conference on Recent Engineering and Technology, ICRET 2026" held during 8th & 9th May 2026 at Sai Vidya Institute of Technology, Bengaluru, Karnataka.



**17th INTERNATIONAL CONFERENCE ON RECENT ENGINEERING
AND TECHNOLOGY 2026**
ORGANIZED BY
SAI VIDYA INSTITUTE OF TECHNOLOGY
IN COLLABORATION WITH
**SAMARKAND STATE UNIVERSITY, UZBEKISTAN
&
OSIET, INDIA**

Certificate of Presentation

This is to certify proudly to Mr/Mrs/Ms/Dr

Utkarsh Mishra - Delhi Technological University, Delhi

Has presented a paper titled
**Explainable Machine Learning-Based Prediction of Antimicrobial Resistance Across .
Multiple Bacterial Pathogens**

in the "17th International Conference on Recent Engineering and Technology
ICRET 2026" held during 8th & 9th May 2026 at Sai Vidya Institute of Technology,
Bengaluru, Karnataka.


Dr. Akhatov Akmal Rustamovich
Vice Rector of International Affairs,
Samarkand State University, Uzbekistan


Dr. Christo Ananth
Professor,
Samarkand State University, Uzbekistan


Dr. Manjunath T N,
General Chair ICRET
Principal, SVIT Bengaluru


Dr. Shantakumar B Patil,
Secretary & Outreach Chair ICRET
Dean R & D, Dept of CSE, SVIT Bengaluru


K. Janani,
CEO, OSIET
Director



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042

PLAGARISM VERIFICATION

Title of the Thesis: **EXPLAINABLE ARTIFICIAL INTELLIGENCE BASED BIOMARKER PRIORITIZATION FOR ANTIMICROBIAL RESISTANCE PREDICTION IN CLINICALLY SIGNIFICANT BACTERIAL PATHOGENS**

Total Pages: **60**

Name of the Student: **UTKARSH MISHRA (2K24/BIO/08)**

Supervisor: **Prof. ASMITA DAS**

Department of Biotechnology, Delhi Technological University, Delhi-110042

This is to report that the above thesis was scanned for similarity detection. Process and outcome are given below:

Software Used: **Turnitin**, Similarity Index: **3%**, Total Word Count: **12856**

DATE: **27/05/26**


**CANDIDATE'S
SIGNATURE**


**PROF. ASMITA DAS
SUPERVISOR**