School of Computing, Engineering and Digital Technologies Middlesbrough TS1 3BA

Predicting Antimicrobial Resistance (AMR) in Microbial Genomes and Cost Burden Analysis Using Machine Learning

Submitted partial requirements for the degree of MSc Computer Science.

7th January 2025

Shubham Satish Kumar Mishra

W9641416

Supervisor: Dr. Zia Ush Shamszaman

Acknowledgment

The completion of this Master's project would not have been possible without the invaluable support and guidance of several key individuals.

First, I extend my deepest gratitude to my supervisor, Dr. Zia Ush Shamszaman, for his unwavering mentorship throughout this project. Dr. Shamszaman's expertise in machine learning, combined with his dedication to rigorous scholarship, have been instrumental in shaping the direction and quality of this work. His guidance has been essential to my growth as a student.

I also wish to sincerely thank Dr. Bruno Silvester Lopes from the Department of Microbiology for his invaluable contributions and support. Dr. Lopes' profound understanding of the dataset and his willingness to share his expertise have been crucial in helping me navigate the complexities of this topic. His insightful discussions have greatly enhanced my understanding of the biological underpinnings of antimicrobial resistance.

Finally, I acknowledge the invaluable assistance provided by the university staff, my fellow researchers, and my family and friends. Their support and encouragement have been a constant source of motivation throughout this challenging and rewarding journey.

To all those who have contributed to this project, directly or indirectly, I extend my heartfelt appreciation. This accomplishment is a testament to the power of collaboration and the generosity of the scientific community.

Acronyms

Acronym	Full Form	Acronym	Full Form
AMR	Antimicrobial Resistance	ML	Machine Learning
CDC	Centers for Disease Control and Prevention	MLST	Multi-Locus Sequence Typing
CNN	Convolutional Neural Network	OECD	Organisation for Economic Co-operation and Development
DNA	Deoxyribonucleic Acid	PubMLST	Public Multi-Locus Sequence Typing Database
FSA	Food Standards Agency	QALY	Quality-Adjusted Life Year
GBM	Gradient Boosting Machines	QRDR	Quinolone Resistance-Determining Region
GP	General Practitioner	RNN	Random Neural Network
iCaMPS	Campylobacter Attribution Study	SARIMA	Seasonal Autoregressive Integrated Moving Average
LMICs	Low and Middle-Income Countries	SARIMAX	Seasonal Autoregressive Integrated Moving Average with eXogenous variables
SIR	Susceptible-Infectious-Recovered	WHO	World Health Organization
SNP	Single Nucleotide Polymorphism	VSL	Value of Statistical Life
ST	Sequence Type	UK	United Kingdom
US	United States		

Predicting Antimicrobial Resistance (AMR) in Microbial Genomes and Cost Burden Analysis Using Machine Learning

An academic research paper for possible submission to

Bioinformatics

Abstract

Antimicrobial resistance (AMR) in *Campylobacter* has emerged as a major public health threat, making infections increasingly difficult and costly to treat. Predicting and monitoring AMR trends is crucial for guiding interventions and policy to combat resistance. This project aimed to analyze patterns of AMR in *Campylobacter jejuni* and *C. coli* isolates from Scotland and England, develop machine learning models to predict AMR phenotypes from genomic and epidemiological data, and forecast the future burden of AMR in the UK.

The project utilized a comprehensive dataset of 6,683 clinical and animal *Campylobacter* isolates collected between 2001-2017. Whole-genome sequencing data, including multilocus sequence types (MLST), AMR genes, and point mutations, was integrated with epidemiological and temporal information. Exploratory data analysis revealed key AMR drivers, such as mutations in the *gyrA* gene conferring fluoroquinolone resistance and the presence of the *tet(O)* gene associated with tetracycline resistance. A random forest model was trained to predict AMR phenotypes, achieving an accuracy of 74% on a held-out test set and identifying the most informative predictors.

To forecast future AMR trends, time series models (SIR, SARIMA, Prophet) were fit to historical campylobacteriosis incidence data for the UK. The models projected a concerning increase in incidence over the next three decades, with the SARIMA model estimating rates could exceed 130 cases per 100,000 by 2050. A population projection model predicted the UK population will reach 71.3 million by 2050. Combining these projections with current economic cost estimates suggested the total cost of campylobacteriosis, adjusted for inflation, could exceed £1.9 billion annually by 2050. These estimates underscore the substantial and growing public health and economic burden of AMR in *Campylobacter*.

The results of this project highlight the power of machine learning and genomic analysis to elucidate AMR dynamics and predict future trends. The concerning projections emphasize the urgent need for action to combat AMR, including reducing antibiotic misuse, strengthening surveillance, and investing in research and development of new antibiotics and diagnostic tools. The methods and insights from this project can help inform targeted interventions and policies to mitigate the impact of AMR and protect population health. Ultimately, confronting the complex challenge of AMR will require sustained commitment and coordination across sectors, guided by data-driven approaches such as those demonstrated in this work.

Contents

1. Introduction

1.1 Background

Antimicrobial resistance (AMR) emerges as one of the most pressing public health challenges of the 21st century. AMR occurs when bacteria, viruses, fungi, and parasites evolve to become resistant to the drugs designed to kill them, rendering common treatments ineffective. The overuse and misuse of antibiotics, both in healthcare settings and in agriculture, accelerates the development and spread of resistant pathogens. This has profound consequences for human health, leading to prolonged illnesses, higher healthcare costs, and increased mortality rates (World Health Organization, 2020).

Campylobacter, a genus of Gram-negative bacteria, represents a major contributor to this growing crisis. *Campylobacter* species, particularly *C. jejuni* and *C. coli*, stand as the leading cause of bacterial gastroenteritis worldwide (Igwaran & Okoh, 2019). In the UK alone, estimates suggest that *Campylobacter* causes over 500,000 cases of food poisoning each year (Food Standards Agency, 2021). Infection typically occurs through consumption of contaminated poultry, water, or unpasteurized milk, resulting in symptoms such as diarrhea, fever, and abdominal cramps. While most cases prove self-limiting, severe or prolonged infections may require antibiotic treatment.

However, the effectiveness of these treatments faces increasing threats from AMR. Many *Campylobacter* strains develop resistance to clinically important antibiotics, such as fluoroquinolones, macrolides, and tetracyclines (Sproston et al., 2018). This resistance arises through various mechanisms, including target modification, efflux pumps, and enzymatic inactivation (Luangtongkum et al., 2009). Of particular concern is the rapid spread of fluoroquinolone resistance, driven by mutations in the DNA gyrase gene *gyrA* and the overuse of these drugs in poultry production (Bolinger & Kathariou, 2017). Macrolide resistance, while less common, also emerges, posing a serious threat given the importance of these drugs as first-line treatments for severe campylobacteriosis (Florez-Cuadrado et al., 2018).

The consequences of AMR in *Campylobacter* prove far-reaching. Resistant infections associate with more severe illness, longer hospital stays, and higher treatment costs (Wassenaar et al., 2016). At a societal level, AMR hampers efforts to control foodborne illness, strains healthcare systems, and inflicts substantial economic losses due to reduced productivity and increased healthcare expenditure. In the United States alone, estimates indicate that antibiotic-resistant *Campylobacter* infections cost $1.2 billion annually (CDC, 2019). As resistance continues to spread, these individual and societal burdens are only expected to grow.

Confronting the crisis of AMR in *Campylobacter* requires a multifaceted approach guided by a deep understanding of resistance mechanisms, transmission dynamics, and future trends. Mathematical modeling and machine learning offer valuable tools to dissect these complex

factors and guide interventions. Advances in whole-genome sequencing open new avenues to identify and track AMR determinants in *Campylobacter* populations (Whitehouse et al., 2018). Integrating this genomic data with epidemiological and clinical information powers predictive models to forecast resistance patterns and inform surveillance efforts (Nguyen et al., 2019). Furthermore, time series analysis and economic modeling project the future burden of AMR and motivate investment in prevention and mitigation strategies.

Recognizing this potential, the present study aims to harness machine learning and statistical modeling to analyze AMR patterns in *Campylobacter* isolates from the UK and predict the future impact of resistance. By leveraging a unique dataset of over 6,000 sequenced isolates, I seek to identify the key genetic and epidemiological drivers of AMR, build predictive models of resistance phenotypes, and forecast the clinical and economic consequences of AMR in the coming decades. The ultimate goal is to produce actionable insights that inform public health policy and guide efforts to combat the AMR crisis.

In the following sections, I describe the data and methods used in this study, present the results of my analyses, and discuss the implications of my findings for science, policy, and public health practice. I also consider the limitations of my approach and highlight areas for future research. Ultimately, I hope this work demonstrates the power of data-driven approaches to confront the complex challenge of AMR and inspires further efforts to apply these tools to safeguard our most precious medicines.

1.2 Project Objectives

This project aims to leverage advanced data analytics and machine learning techniques to elucidate the patterns, determinants, and future trajectory of antimicrobial resistance in *Campylobacter* isolates from the UK. The specific objectives are threefold:

1. Identify the key genetic and epidemiological factors driving AMR in *C. jejuni* and *C. coli* through comprehensive analysis of whole-genome sequencing and associated metadata. By pinpointing the most influential AMR genes, mutations, and host factors, this work will enhance our mechanistic understanding of resistance development and inform targeted interventions.

2. Develop and validate predictive models to forecast AMR phenotypes from genotypic, epidemiological, and temporal features. Harnessing the power of machine learning, these models will enable proactive identification of emerging resistance threats and guide risk assessment and surveillance efforts.

3. Project the future clinical and economic burden of AMR in *Campylobacter* in the UK over the coming decades. By integrating AMR predictions with population forecasts and cost estimates, this analysis will quantify the potential health and financial consequences of unchecked resistance, informing policy decisions and resource allocation.

1.3 Project Scope

The scope of this project encompasses the analysis of AMR in *Campylobacter* isolates collected from clinical and animal sources across Scotland and England between 2001-2017. The core dataset includes whole-genome sequencing data and accompanying metadata for 6,683 isolates, capturing both *C. jejuni* (the predominant cause of human infections) and *C. coli* (a significant reservoir in livestock).

The genomic data scope spans multi-locus sequence typing (MLST) profiles, AMR gene presence/absence, and point mutations in key resistance loci such as *gyrA*. The epidemiological and temporal variables include host species, geographic region, isolation source, and sampling year. This rich, multi-dimensional dataset enables a comprehensive examination of AMR patterns and drivers over space and time.

The analytical scope of the project includes:

- Descriptive and inferential statistical analyses to characterize AMR trends and associations
- Feature engineering and selection to construct informative predictors for modeling
- Machine learning model development, training, and validation for AMR phenotype prediction
- Time series modeling and forecasting to project future AMR prevalence and incidence rates
- Economic modeling to estimate the healthcare costs and societal burden of AMR

The project deliverables will include a comprehensive research report detailing the methods, results, and implications of the analyses, as well as a suite of predictive models and visualizations to support AMR surveillance and decision-making. While the immediate scope is focused on *Campylobacter* in the UK, the approaches and insights generated here could inform AMR research and policy in other pathogens and settings.

1.4 Research Questions

This project will address the following key research questions:

- What are the dominant genetic mechanisms of AMR in UK *Campylobacter* isolates, and how do they vary across species, regions, and time?
- What are the most important epidemiological and temporal predictors of AMR phenotypes in *Campylobacter*?
- How accurately can machine learning models predict AMR phenotypes from genotypic and epidemiological data, and what are the key drivers of model performance?

- How is the prevalence and incidence of AMR in UK *Campylobacter* projected to evolve over the next three decades, based on current trends and population dynamics?
- What are the potential economic burden of AMR in *Campylobacter* in the UK over the coming decades, and how can this inform policy and interventions?

By systematically investigating these questions, this project will paint a comprehensive picture of the current state, future trajectory, and potential impacts of AMR in UK *Campylobacter*. The answers will directly inform efforts to combat resistance in this key pathogen, while also providing a roadmap for data-driven AMR research and policy across diverse microbial threats. Ultimately, this work aspires to help safeguard the effectiveness of life-saving antibiotics and protect public health in the face of evolving resistance.

2. Literature Review

2.1 Antimicrobial Resistance: A Global Challenge

At its core, AMR is a simple evolutionary process. When a population of bacteria is exposed to an antibiotic, those individuals that happen to carry genes enabling them to survive will multiply, passing on their resistance to future generations. Over time and repeated exposure, the proportion of resistant bacteria increases, until the antibiotic is no longer effective. It's survival of the fittest on a microscopic scale, with profound consequences for human and animal health.

The Scope of the Problem: The AMR crisis is truly global in nature, affecting countries in every region and at every stage of economic development. Resistant pathogens do not respect borders, and can spread rapidly through international travel and trade, healthcare settings, and the food chain. The World Health Organization (WHO) has declared AMR one of the top 10 global public health threats facing humanity, with the potential to undo decades of medical progress (WHO, 2020).

The numbers are staggering. Currently, AMR is estimated to cause at least 700,000 deaths globally each year, including 230,000 deaths from multidrug-resistant tuberculosis alone (O'Neill, 2016). In the European Union, AMR is responsible for over 33,000 deaths and costs the economy €1.5 billion annually in healthcare expenses and productivity losses (Cassini et al., 2019). In the United States, antibiotic-resistant infections affect 2.8 million people and claim more than 35,000 lives each year (CDC, 2019).

But these figures only capture the direct impact. AMR also threatens many of the foundations of modern medicine. Procedures like organ transplants, cancer chemotherapy, and surgeries rely on effective antibiotics to prevent and treat infections. Without these drugs, the risk of complications would be too high, making these life-saving interventions impossible.

The Drivers of Resistance: AMR is a complex problem with multiple, interconnected drivers. At the most basic level, the use of antibiotics creates selection pressure for resistant strains to emerge and spread. The more we use these drugs, the faster resistance develops. Overuse and misuse of antibiotics, in both human medicine and agriculture, are thus major contributors to AMR.

In many countries, antibiotics are still widely available without a prescription, leading to self-medication and inappropriate use for conditions like viral infections that don't respond to these drugs. Even when prescribed correctly, patients may not adhere to the full course of treatment, allowing surviving bacteria to develop resistance.

The use of antibiotics in livestock production is another key driver. Worldwide, more antibiotics are used in animals than in humans, often to promote growth or prevent disease in conditions of poor hygiene and overcrowding (Watts et al., 2017). This creates a reservoir of

resistant bacteria that can spread to humans through direct contact, environmental contamination, and the food chain.

The Need for Action: Tackling AMR requires urgent, coordinated action across multiple sectors, guided by a One Health approach that recognizes the interconnections between human, animal, and environmental health. This includes:

- Improving antibiotic stewardship in healthcare settings to ensure these drugs are used appropriately and only when necessary. This involves implementing prescribing guidelines, monitoring antibiotic use, and educating healthcare workers and the public about AMR.
- Reducing antibiotic use in food production by improving animal husbandry practices, vaccinating livestock, and restricting the use of critically important antibiotics. Many countries have already banned the use of antibiotics for growth promotion, but more needs to be done to ensure responsible use.
- Investing in research and development of new antibiotics, diagnostic tools, vaccines, and alternative therapies. No new classes of antibiotics have been discovered since the 1980s, and the pipeline of new drugs in development is insufficient to keep pace with rising resistance (WHO, 2020).
- Strengthening surveillance and monitoring of AMR at local, national, and global levels to track resistance trends, detect emerging threats, and guide interventions. This requires standardized data collection, sharing, and analysis across sectors and countries.
- Promoting infection prevention and control measures, such as improved hygiene, sanitation, and vaccination coverage, to reduce the need for antibiotics in the first place.

2.2 *Campylobacter* and Public Health

Picture your last barbecue: sizzling burgers, juicy chicken skewers, a perfect summer evening. But lurking in that delicious grilled chicken could be a microscopic threat: *Campylobacter*. This spiral-shaped bacterium is one of the most common causes of foodborne illness worldwide, responsible for an estimated 166 million cases of diarrheal disease each year (Havelaar et al., 2015). In the UK alone, *Campylobacter* causes over 500,000 infections, 100 deaths, and costs the economy around £900 million annually (FSA, 2021).

But what makes *Campylobacter* such a formidable foe? For one, it's a master of survival. The bacterium can persist in a variety of environments, from animal intestines to food processing plants, and only a small number of cells are needed to cause infection (Bolton, 2015). It's particularly associated with poultry, with many chickens carrying the bacterium asymptomatically in their gut. During slaughter and processing, *Campylobacter* can

contaminate the meat, and if this meat is undercooked or cross-contaminates other foods, it can lead to human infection.

When ingested, *Campylobacter* invades the lining of the small intestine, causing inflammation and the hallmark symptoms of campylobacteriosis: diarrhea, abdominal cramps, and fever. While most cases are self-limiting, some can develop into more serious conditions like septicemia, reactive arthritis, or the rare but severe Guillain-Barré syndrome, a form of paralysis (Nachamkin et al., 1998). Certain populations, such as young children, the elderly, and immunocompromised individuals, are at higher risk of severe complications.

As if this wasn't concerning enough, many *Campylobacter* strains are becoming increasingly resistant to antibiotics. Fluoroquinolones and macrolides, two key classes of drugs used to treat severe cases, are becoming less effective (Sproston et al., 2018). This resistance is driven by a combination of factors, including antibiotic overuse in human medicine and agriculture, particularly in poultry production. As resistant strains become more prevalent, the risk of treatment failures and prolonged illness increases.

So what can be done to address this public health threat? Prevention is key. At the individual level, this means practicing safe food handling: thoroughly cooking poultry, avoiding cross-contamination, and washing hands regularly. At the industry level, interventions like improved biosecurity on farms, competitive exclusion products, and decontamination treatments during processing can help reduce *Campylobacter* levels in the food chain (Wagenaar et al., 2015). Reducing antibiotic use in livestock, especially those critical for human medicine, is also crucial for preserving the effectiveness of these drugs.

From a public health perspective, strengthening surveillance is vital. Monitoring *Campylobacter* prevalence and AMR profiles in humans, animals, and food can help track the spread of resistant strains, identify emerging threats, and guide risk management strategies. Advances in genomic sequencing are providing new insights into the transmission dynamics and evolution of AMR in *Campylobacter*, informing more targeted interventions (Cody et al., 2013).

In the face of rising AMR, we need to use every tool in our arsenal to combat this public health challenge. This is where predictive models come in - by leveraging the power of machine learning and big data, we can identify risk factors, forecast resistance trends, and optimize interventions. It's an exciting frontier in the fight against AMR, one that could help us stay one step ahead of this microbial threat.

So next time you fire up the grill, remember the importance of cooking your chicken thoroughly. That simple act isn't just about avoiding a nasty bout of food poisoning - it's a small but significant part of a much larger battle to protect public health in the age of AMR.

2.3 Machine Learning in AMR Prediction

Imagine you're a doctor in a busy hospital. A patient comes in with a severe bacterial infection, and you need to prescribe an antibiotic. But which one? The clock is ticking, and

choosing the wrong drug could mean the difference between life and death. If only you had a crystal ball, a way to predict which antibiotics the pathogen is likely to be resistant to...

Enter machine learning (ML). This branch of artificial intelligence is all about teaching computers to learn from data, to find patterns and make predictions without being explicitly programmed. And it's revolutionizing the way we approach AMR.

Traditionally, detecting AMR has relied on culture-based methods, where bacteria are grown in the presence of antibiotics to see which drugs they can withstand. While this approach is still the gold standard, it's time-consuming, taking anywhere from 24 to 72 hours. In the context of a life-threatening infection, every minute counts.

This is where ML comes in. By training algorithms on vast datasets of bacterial genomes, antibiotic susceptibility profiles, and clinical metadata, we can create models that can predict AMR from genomic data alone (Yang et al., 2018). Think of it like facial recognition software, but instead of identifying people, it's identifying patterns in bacterial DNA that are associated with resistance.

One popular approach is using classification algorithms, like random forests or support vector machines. These models learn to distinguish between resistant and susceptible strains based on the presence or absence of certain genetic features, such as AMR genes or mutations. Once trained, they can take in the genome of a new isolate and output a prediction of its resistance profile, all in a matter of minutes.

But it's not just about speed. ML models can also uncover new insights into the mechanisms and epidemiology of AMR. By identifying the most informative genetic features, they can help us understand which genes or mutations are driving resistance to specific drugs (Nguyen et al., 2019). This knowledge can guide the development of new diagnostic tools, drug targets, and surveillance strategies.

ML is also powering the development of more sophisticated predictive models that incorporate additional layers of data. For example, by combining genomic data with information on antibiotic consumption, population density, and environmental factors, we can create ecological niche models that predict the emergence and spread of resistant strains (Alvarez-Uria et al., 2018). These models can help us anticipate AMR hotspots and inform targeted interventions.

Of course, ML is not a silver bullet. These models are only as good as the data they're trained on, and AMR data can be noisy, biased, and incomplete. Ensuring data quality and representativeness is a key challenge. There's also the risk of overfitting, where models become too tuned to the training data and fail to generalize to new, unseen datasets. Rigorous validation and testing are essential.

Moreover, while ML can excel at prediction, it doesn't inherently provide causal insights. Understanding why a certain mutation confers resistance still requires experimental validation and mechanistic studies. ML should be seen as a complement to, not a replacement for, traditional microbiological research.

Despite these challenges, the potential of ML in the fight against AMR is immense. By harnessing the power of big data and advanced analytics, we can develop more rapid,

accurate, and informative predictive tools to guide clinical decision-making and public health interventions. It's a prime example of how interdisciplinary collaboration, bringing together experts in microbiology, genomics, computer science, and epidemiology, can yield innovative solutions to complex problems.

2.3 Economic Burden of AMR

Picture a world where a simple scratch could kill. Where pneumonia, once treatable with a course of antibiotics, becomes a death sentence. Where routine surgeries and cancer treatments become too risky to perform. This is the potential future we face if we fail to tackle antimicrobial resistance (AMR).

AMR is not just a medical challenge, but an economic one. Resistant infections are more difficult and expensive to treat, often requiring longer hospital stays, more intensive care, and pricier last-resort drugs. When first-line antibiotics fail, doctors must turn to second or third-line options, which can be 10-20 times more costly (Shrestha et al., 2018). For example, treating a case of multidrug-resistant tuberculosis can cost up to $25,000, compared to $2,000 for a drug-susceptible case (Laurence et al., 2015).

But the costs of AMR extend far beyond the healthcare system. Patients with resistant infections are more likely to miss work, leading to productivity losses. In agriculture, resistant infections in livestock can lead to reduced meat and dairy yields, and higher veterinary costs (OECD, 2018). The World Bank estimates that by 2050, AMR could push 28 million people into poverty and cost the global economy up to $3.4 trillion annually (World Bank, 2017).

The burden of AMR is not evenly distributed. Low and middle-income countries (LMICs) are disproportionately affected, due to factors like weaker health systems, limited access to quality antibiotics, and poor infection control practices (Pokharel et al., 2019). In these settings, AMR can trap people in a cycle of poverty - a resistant infection can lead to catastrophic healthcare expenditures, lost income, and reduced productivity, making it harder for families to escape poverty.

But even in high-income countries, the costs are staggering. In the United States, AMR is estimated to cost the healthcare system over $4.6 billion annually (CDC, 2019). In Europe, AMR is responsible for an estimated 670,000 infections, 33,000 deaths, and €1.5 billion in healthcare costs and productivity losses each year (Cassini et al., 2019).

These figures are based on the direct costs of treating resistant infections. But there are also significant indirect costs, such as the impact on medical procedures that rely on effective antibiotics. Hip replacements, cesarean sections, organ transplants - these all become much riskier in a world without reliable antibiotics. A study estimated that a 30% reduction in the efficacy of preventive antibiotics could result in 120,000 additional infections and 6,300 infection-related deaths each year following hip replacement surgery in the US alone (Teillant et al., 2015).

Then there are the intangible costs, the value placed on the pain, suffering, and loss of life caused by AMR. Economists use metrics like the value of a statistical life (VSL) to quantify

these impacts in monetary terms. One study estimated that the VSL lost due to AMR could reach $65 trillion by 2050 if no action is taken (KPMG, 2014).

Despite these staggering figures, funding for AMR research and interventions remains inadequate. The development of new antibiotics, in particular, has stagnated, with many pharmaceutical companies exiting the field due to low profitability (Towse et al., 2017). Bringing a new antibiotic to market can cost over $1 billion, but their necessary use is often restricted to preserve effectiveness. This lack of return on investment has created a market failure, with a dire need for new economic models and incentives to spur innovation.

Tackling the economic burden of AMR will require a multi-pronged approach. This includes investing in new antibiotic development, but also in measures to preserve the effectiveness of existing drugs, such as improved stewardship, infection control, and vaccination (Ardal et al., 2020). It will require global cooperation, with high-income countries supporting LMIC efforts through funding, technology transfer, and capacity building.

Crucially, it will require a shift in how we value antibiotics. Rather than seeing them as cheap, disposable commodities, we need to recognize them as a precious global resource, one that requires careful management and protection (Hollis & Maybarduk, 2015). This means aligning incentives across healthcare, agriculture, and environmental sectors to promote responsible use and minimize unnecessary exposure to antibiotics.

The economic case for action on AMR is clear. The costs of inaction are simply too high - for our health, our economies, and our future. By investing in solutions now, we can avert a catastrophic toll down the line. It won't be easy, but the alternative is unthinkable. We have a moral and economic imperative to preserve these life-saving drugs for generations to come.

2.4 Current Approaches and Gaps

Despite the growing threat of antimicrobial resistance (AMR), our understanding of its complex drivers and dynamics remains incomplete. While significant strides have been made in areas like AMR surveillance, molecular epidemiology, and antibiotic stewardship, key knowledge gaps persist that hinder our ability to effectively combat this crisis.

One major challenge is the lack of standardized, high-quality data on AMR prevalence and transmission across human, animal, and environmental settings. Surveillance systems, particularly in low and middle-income countries (LMICs), often suffer from limited coverage, inconsistent sampling methods, and variable laboratory capacity (Seale et al., 2017). This makes it difficult to compare data across regions, track resistance trends over time, and identify emerging threats. Strengthening and harmonizing AMR surveillance is a critical priority, requiring investment in laboratory infrastructure, training, and data management systems.

Even in countries with more robust surveillance, like the UK, there are gaps in our understanding of how AMR spreads through complex ecological networks. While foodborne transmission of resistant bacteria, particularly from animal products, is well-established, the relative contributions of different sources and pathways remain unclear (Muloi et al., 2018). How much AMR is driven by antibiotic use in humans versus animals? What role do environmental reservoirs, like wastewater or soil, play in the maintenance and spread of

resistance? Answering these questions requires an integrated One Health approach that studies AMR across sectors and scales.

Another key gap is in our mechanistic understanding of how resistance emerges and persists in bacterial populations. While we have identified many of the genetic determinants of AMR, like specific resistance genes or mutations, we still have much to learn about how these elements are acquired, maintained, and transmitted under different selective pressures (Partridge et al., 2018). This is further complicated by the role of the wider microbial community - how do interactions between species, like horizontal gene transfer or competition, shape the evolution of AMR? Advances in genomic sequencing, metagenomics, and computational modeling are providing new insights, but integrating this complex data into predictive frameworks remains a challenge.

This brings us to the field of machine learning (ML) in AMR prediction, an exciting area that holds immense promise but also faces significant hurdles. While ML models have shown impressive accuracy in predicting AMR from genomic and other data types, they are often limited by the quality and representativeness of the training data (Macesic et al., 2020). Many models are developed on data from a single hospital or region, raising questions about their generalizability to other settings. There's also the "black box" problem - while ML models can make accurate predictions, their decision-making process is often opaque, making it difficult to interpret the biological mechanisms underlying the predictions (Azodi et al., 2019). Improving the interpretability and robustness of these models is an active area of research.

A related challenge is translating these predictive models into clinical and public health practice. How can we integrate ML-based AMR predictions into diagnostic workflows and treatment guidelines? What are the regulatory and ethical considerations around using these tools in patient care? Addressing these questions will require close collaboration between researchers, clinicians, and policymakers.

Beyond the technical challenges, there are also significant economic and political barriers to tackling AMR. The market failure in antibiotic development, where the societal value of new drugs is not reflected in their commercial returns, has led to a dearth of new therapies in the pipeline (Okhravi et al., 2018). Efforts to incentivize R&D, like push and pull funding mechanisms or delinked payment models, have gained traction but remain insufficient. Similarly, measures to curb antibiotic overuse, like stewardship programs or restrictions on use in agriculture, often face resistance from industry stakeholders or lack enforcement (Denyer Willis & Chandler, 2019).

Overcoming these gaps and barriers will require a concerted, multidisciplinary effort that engages stakeholders across sectors. This includes supporting basic science to elucidate AMR mechanisms, developing innovative surveillance and diagnostic tools, creating economic incentives for responsible antibiotic use and development, and promoting stewardship and infection control practices. It means fostering global collaboration, data sharing, and capacity building, particularly in LMICs. And it means raising public awareness about the urgency of the AMR threat and the role everyone can play in preserving these life-saving drugs.

3. Data and Methodology

3.1 Dataset Description

The foundation of this project is built on an extensive dataset that combines multiple sources to thoroughly analyze antimicrobial resistance (AMR) patterns in Campylobacter. The primary dataset, sourced from the PubMLST database, comprises whole-genome sequencing data for 6,683 clinical and animal Campylobacter isolates collected across Scotland and England between 2001 and 2017.

id	Unique identifier for each Campylobacter isolate in the dataset
species	Campylobacter species (C. jejuni or C. coli)
source	Origin of isolate (human, animal, or environmental)
isolation_date	Date when the isolate was collected
region	Geographic location where isolate was collected
fq_res	Fluoroquinolone resistance status (R=resistant, S=susceptible)
mac_res	Macrolide resistance status (R=resistant, S=susceptible)
gyra mutation	Presence/type of mutations in gyrA gene associated with resistance

Fig-1 Key Dataset Columns and Descriptions

Each isolate in this dataset provides a wealth of information, encompassing both the bacterial genome and detailed associated metadata. The genomic data includes multilocus sequence types (MLST), the presence of AMR genes, and point mutations in key loci such as gyrA. This enables an in-depth exploration of the intricate relationships between genetic makeup and AMR phenotypes.

The accompanying metadata offers crucial epidemiological context for each isolate. Fields such as region, year, month, and source of isolation facilitate the identification of spatiotemporal trends in AMR. Additionally, host species data distinguishes between human clinical cases and animal reservoirs, thereby creating a holistic view of the Campylobacter AMR landscape.

Beyond the bacterial data, supplementary datasets enhance understanding of the broader public health burden of AMR. Incidence data, obtained from the iCaMPS (Campylobacter Attribution Study) report by Food Standards Scotland, provides annual campylobacteriosis rates per 100,000 population in Scotland, England, and Wales from 1989 to 2021. This data

enables me to analyze temporal trends and assess the magnitude of the Campylobacter epidemic over time.

Economic indicators are equally vital in contextualizing the societal costs of AMR. Inflation rates from 1989 to 2022, sourced from Kaggle, allow for the adjustment of cost projections for monetary value changes, while population data from the World Bank supports scaling incidence rates to actual case numbers and forecasting future trends based on demographic shifts.

Integrating these diverse data streams is a complex but necessary undertaking. Combining detailed genomic data with broader epidemiological and economic indicators allows for the development of predictive tools that not only forecast resistance patterns but also quantify their real-world impacts. This comprehensive approach demonstrates how data integration illuminates critical insights in the ongoing fight against AMR.

3.2 Data Preprocessing

The complexity of this dataset necessitates a robust preprocessing pipeline to ensure data quality and compatibility. The raw data files, often originating from disparate sources, require harmonization in terms of formatting, naming conventions, and data types prior to analysis.

The preprocessing pipeline begins with data cleaning, addressing missing or erroneous values. For the genomic dataset, this involves identifying gaps in fields such as region, year, or source and deciding whether to impute or exclude records. Consistency checks, such as ensuring all dates align with the expected range, are implemented to detect and correct any discrepancies.

Data integration follows, merging the genomic and metadata streams into a cohesive dataset. This step involves aligning datasets using isolate IDs as the common key and resolving discrepancies through fuzzy matching techniques when necessary.

Feature engineering then derives new variables that enhance predictive models. Temporal features such as year and month of isolation are extracted from date fields, while geographical data is mapped to calculate spatial statistics. Genomic features, such as the total count of AMR genes or specific mutations like *gyrA* point mutations, are meticulously crafted to capture biologically relevant patterns.

Fig-2 Data Processing and Analysis Pipeline

Unlike generic feature engineering approaches, I tailor the process to this specific project by focusing on *Campylobacter*-relevant variables, such as MLST types, *gyrA* mutations, or the presence of resistance genes like *tetO*. This specificity ensures that the engineered features align with the study's objectives.

Finally, data transformation prepares the features for machine learning. This involves encoding categorical variables, scaling numeric data, and partitioning the dataset into training and testing sets. The transformations applied are carefully informed by the unique characteristics of the data.

Throughout this process, Python libraries such as Pandas for data manipulation, Numpy for numerical computations, and Scikit-learn for machine learning utilities has been used. The interactive environment provided by Jupyter notebooks facilitates the development and testing of the preprocessing code.

The outcome is a clean, integrated, and feature-rich dataset, forming a reliable foundation for subsequent exploratory analyses and predictive modeling efforts. By investing in

preprocessing, I ensure that the insights and predictions derived from the data are both robust and relevant.

3.3 Feature Engineering

With a processed dataset, the focus shifts to feature engineering, an essential step in making prediction models that encapsulate the key drivers of AMR. This stage distills the wealth of genomic, epidemiological, and temporal data into a refined set of features suitable for machine learning.

For the genomic data, the primary goal is to capture the presence and configuration of known AMR determinants. I identify AMR genes such as *tetO*, *blaOXA*, and *ermB* within the isolates, creating binary features that indicate their presence or absence. Additionally, single nucleotide polymorphisms (SNPs) in critical loci, like the *gyrA* quinolone resistance-determining region (QRDR), are included as features to represent resistance-associated mutations. For example, mutations such as C257T (Thr86Ile) serve as key predictors of fluoroquinolone resistance.

At a broader genomic level, I leverage MLST to characterize isolates based on sequence types (STs). One-hot encoding is applied to create binary features for each unique ST, enabling the identification of lineage-specific AMR patterns.

Epidemiological metadata contributes further predictive power. Variables such as isolation source (human, animal, or environmental), geographical region, and date of isolation are transformed into categorical and temporal features to uncover spatial and longitudinal trends.

This domain-specific approach to feature engineering ensures that the resulting variables are both biologically meaningful and relevant to the study objectives. Python tools like Biopython, Pandas, and regular expressions streamline the extraction, manipulation, and encoding of these features.

By transforming raw data into machine-readable features, this step bridges the gap between biological complexity and computational analysis, forming the foundation for accurate and interpretable predictive models.

3.4 Machine Learning Models

The project employs several machine learning and statistical models, each carefully chosen for its suitability to address specific challenges associated with predicting antimicrobial resistance (AMR). The overall goal is not merely to achieve high predictive accuracy but also to ensure the results are interpretable, biologically meaningful, and actionable within the context of AMR surveillance and intervention strategies. Below, each model's role and relevance to the project are discussed in detail.

Linear Regrssion:
Linear regression serves as a foundational predictive model for continuous variables, providing a simple yet interpretable framework for understanding relationships between features and outcomes. Its utility in this project lies in establishing baseline predictions and offering insights into how certain predictors, such as genetic mutations or environmental factors, linearly influence resistance levels. While linear regression assumes a linear relationship and is limited in handling complex interactions, it remains a vital starting point for hypothesis testing and preliminary analysis.

Random Forest
The random forest regressor, on the other hand, is deployed to capture complex, nonlinear relationships in the data, making it particularly well-suited for AMR prediction. Unlike linear regression, random forests can model intricate interactions between genetic markers and environmental variables, which are often crucial in determining resistance. Moreover, this model provides robust feature importance metrics, enabling researchers to identify the most influential variables driving resistance patterns. This interpretability is invaluable for guiding molecular surveillance efforts and prioritizing key genes or mutations for further study.

SARIMA & Prophet
For time series forecasting, two advanced models—SARIMAX and Prophet—are employed. SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous variables) is particularly advantageous for capturing seasonal patterns and trends in resistance data, while also incorporating external factors such as environmental conditions or antibiotic usage rates. This makes SARIMAX a powerful tool for understanding how resistance evolves over time and under varying conditions. On the other hand, Prophet, developed by Facebook, is chosen for its flexibility in handling missing data and its ability to model seasonality and holiday effects intuitively. Prophet's ease of use and robustness make it an excellent complement to SARIMAX, especially for exploratory forecasting and scenarios where domain-specific time-series knowledge is limited.

Time Series
Time series analysis in the project also includes a rigorous statistical assessment of stationarity using tests like the Augmented Dickey-Fuller test and autocorrelation function plots. These analyses are critical for determining the appropriateness of forecasting models and ensuring that time series methods are applied correctly. Furthermore, TimeSeriesSplit, a cross-validation strategy tailored for time-dependent data, is employed to evaluate model performance while preserving the temporal structure. This method ensures that the evaluation reflects real-world scenarios where future data cannot inform past predictions.

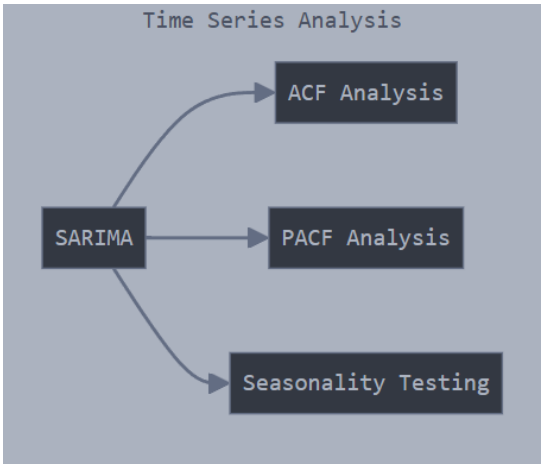


Fig- 3 Time Series Analsis

SIR Model

The SIR model (Susceptible-Infectious-Recovered) is incorporated to provide an epidemiological perspective, simulating the dynamics of resistance spread in a population. By dividing the population into compartments of susceptible, infectious, and recovered individuals, the SIR model offers a theoretical framework to understand how resistance propagates under different conditions. The inclusion of the SIR model highlights the project's interdisciplinary approach, combining machine learning with mechanistic modeling to provide both predictive power and theoretical insight. This dual approach ensures that predictions are not only accurate but also grounded in biological plausibility.

SIR Model to Incidence Data

Parameter definition: Define the parameters for the SIR model using historical incidence data: Susceptible (S): Total population minus those already infected or recovered. Infectious (I): Current number of cases. Recovered (R): Number of people who have recovered. Model Setup: Implement the SIR equations to simulate how the infection spreads over time

- $\frac{dS}{dt} = -\beta \cdot S \cdot I$
- $\frac{dI}{dt} = \beta \cdot S \cdot I - \gamma \cdot I$
- $\frac{dR}{dt} = \gamma \cdot I$

Where β β is the transmission rate and γ γ is the recovery rate, both of which can be tuned based on past data.

Fig- 4 SIR Model

3.5 Economic Analysis Methods

Predicting AMR patterns is a crucial step, but fully understanding its societal impact requires translating biological insights into economic terms. The focus lies on quantifying the health and financial burden of AMR under various future scenarios.

Forecasting AMR Incidence and Population Growth

The foundation of the economic model integrates AMR prevalence projections with population and cost data. Time series models, such as Seasonal Autoregressive Integrated

Moving Average (SARIMA) and Prophet, are used to forecast the incidence of resistant *Campylobacter* infections up to 2050. These models, trained on historical incidence data, capture both long-term trends and seasonal variations in AMR prevalence.

Simultaneously, UK population growth is projected using linear regression applied to historical census data. By combining AMR incidence rates with population figures, the absolute number of resistant cases expected each year can be estimated, providing a tangible measure of the future health burden of AMR.

Estimating Economic Costs

Economic cost data from the iCaMPS (Campylobacter Attribution Study) report serves as the basis for assigning monetary value to this burden. This report includes estimates of current direct healthcare costs and productivity losses associated with campylobacteriosis in the UK. Assuming a constant cost per case, future projections are derived by applying these figures to AMR case forecasts.

Accounting for Inflation

To address the dynamic nature of costs over time, inflation-adjusted estimates are incorporated. Inflation rates are forecasted using a stochastic model based on historical inflation data from 1989 to 2022. These adjustments ensure that future economic burden is expressed in real terms, offering a more accurate financial assessment. Comparisons between projections with and without inflation adjustments highlight the compounding effects of economic factors on AMR-related costs.

Evaluating Broader Economic Impacts

Beyond direct costs, AMR leads to longer hospital stays, more severe illnesses, and reduced quality of life. Health utility measures, such as quality-adjusted life years (QALYs) and disability-adjusted life years (DALYs), are employed to quantify these intangible impacts. Assigning monetary values to these health states enables a comprehensive assessment of the broader societal impacts of AMR.

Visualizing Economic Insights

Clear and engaging visualizations are developed using Python libraries like Matplotlib and Seaborn. Additionally, interactive dashboards created with tools such as Plotly and Bokeh allow stakeholders to explore various scenarios and assumptions dynamically, enhancing understanding and decision-making.

Driving Policy and Decision Support

This economic analysis bridges the gap between biological predictions and actionable policy insights. By quantifying the costs of inaction on AMR, it builds a compelling case for investments in prevention, surveillance, and innovative solutions. The findings support prioritizing interventions, identifying high-risk populations, and evaluating the return on investment for various control strategies.

Integrating Economic Analysis into the AMR Framework

Economic analysis forms an integral part of the broader AMR prediction framework. By integrating genomic, epidemiological, and economic data, it provides a comprehensive view of the AMR landscape. This interdisciplinary approach highlights the interconnected nature of the AMR challenge and emphasizes the importance of holistic solutions.

4. Implementation

4.1 Development Environment

The foundation of any successful machine learning project lies in its development environment - the toolkit of programming languages, libraries, and frameworks that enable the transformation of raw data into actionable insights. For this project predicting antimicrobial resistance (AMR) in *Campylobacter*, the development environment is being meticulously crafted to support the complex data processing, modeling, and analysis tasks involved.

At the heart of this setup is the Python programming language. Python is being chosen for its extensive ecosystem of libraries tailored for data manipulation, machine learning, and visualization. The Anaconda distribution, a comprehensive platform for data science, is being used to manage Python packages and create isolated project environments. This ensures a consistent and reproducible setup across different stages of development and collaboration.

Within this Anaconda environment, a suite of powerful libraries is being assembled. Pandas, a data manipulation library, is being used for efficient data handling and preprocessing. NumPy, the fundamental package for numerical computing in Python, underpins many of the mathematical operations. Scikit-learn, a machine learning library built on top of NumPy, is being employed for its wide range of algorithms and utility functions for model evaluation and selection.

Data visualization is a critical component of any data science workflow, allowing for the exploration of patterns, communication of results, and debugging of analyses. Matplotlib, a plotting library, provides a MATLAB-like interface for creating a wide variety of static, animated, and interactive visualizations. Seaborn, a statistical data visualization library based on Matplotlib, is being used for its informative and attractive statistical graphics.

```
# Checking columns to identify potential duplicates or renames
print("Columns in Merged Dataset:")
print(data.columns)


if 'age_yr_x' in data.columns:
    data.rename(columns={'age_yr_x': 'age_yr', 'age_mth_x': 'age_mth'}, inplace=True)

# Cleaning and feature engineering

if 'age_yr' in data.columns:
    data['age_yr'] = data['age_yr'].fillna(data['age_yr'].median())  # Fill missing age with median
else:
    print("Column 'age_yr' not found in the dataset.")

if 'sex' in data.columns:
    data['sex'] = data['sex'].fillna('unknown')  # Replace missing sex with 'unknown'
else:
    print("Column 'sex' not found in the dataset.")


categorical_columns = ['region', 'sex', 'source', 'species']
label_encoders = {col: LabelEncoder() for col in categorical_columns}
for col, encoder in label_encoders.items():
    if col in data.columns:
        data[col] = encoder.fit_transform(data[col])
    else:
        print(f"Column '{col}' not found in the dataset.")


data['cattle_consumption_kg_per_year'] = 35

# Feature engineering: creating age groups
if 'age_yr' in data.columns:
    bins = [0, 18, 45, 65, 100]
    labels = ['child', 'adult', 'middle_age', 'elderly']
    data['age_group'] = pd.cut(data['age_yr'], bins=bins, labels=labels)
else:
    print("Cannot create age groups as 'age_yr' is missing.")
```

Fig- 5 Code snippet of data preprocessing

4.2 Model Architecture

With the development environment set up, attention is now turning to the architecture of the machine learning models that will predict AMR phenotypes from genotypic and epidemiological data. Model architecture refers to the overall structure and design of these predictive algorithms, encompassing the choice of specific techniques, the arrangement of computational units, and the flow of data through the system.

The goal is to craft a model architecture that can effectively capture the complex, nonlinear relationships between genetic mutations, host factors, and AMR outcomes. This requires a delicate balance between model complexity, interpretability, and computational efficiency. Too simple a model risks underfitting the data and missing important patterns, while an overly complex model may overfit to noise and fail to generalize to new, unseen data.

To strike this balance, a range of model architectures are being explored, from classic machine learning approaches to state-of-the-art deep learning techniques. At the simpler end of the spectrum are logistic regression models. These models aim to predict the probability of an isolate being resistant or susceptible based on a linear combination of input features. While computationally efficient and easily interpretable, logistic regression may struggle to capture more complex, nonlinear relationships in the data.

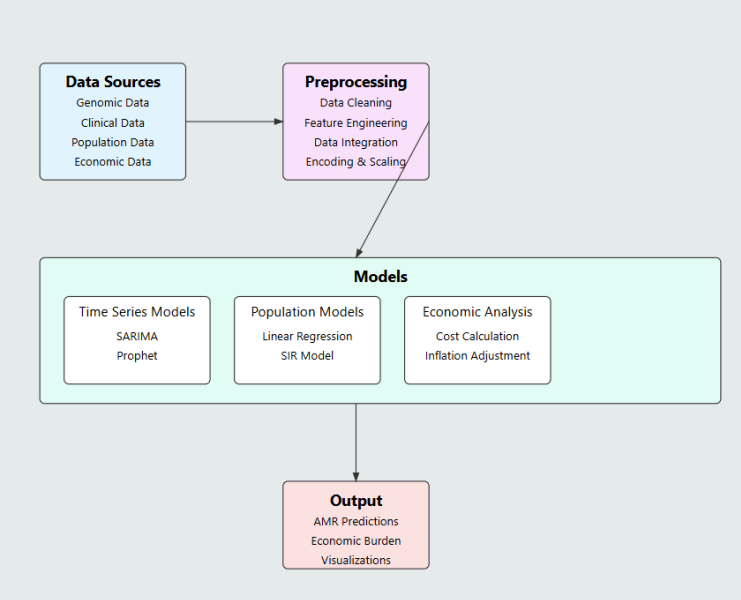


Fig- 6 Model Architecture

To model these nonlinearities, tree-based ensemble methods like random forests and gradient boosting machines (GBMs) are being employed. These methods construct multiple decision trees, each learning to partition the data based on different subsets of features. The final prediction is then made by aggregating the outputs of the individual trees. This ensemble approach allows for the capture of complex interactions between features and tends to be robust to noise and outliers. Random forests, in particular, provide a measure of feature importance, aiding in the interpretation of the model's decision-making process.

Model selection and hyperparameter tuning are crucial aspects of designing an optimal architecture. Tools like cross-validation and grid search are being used to evaluate different model configurations and choose the best-performing set of hyperparameters. The goal is to find the sweet spot between model complexity and generalization performance, while avoiding overfitting to the training data.

4.3 Training Process

With the model architecture defined, focus is now shifting to the training process - the critical phase where the selected models learn to map input features to AMR phenotypes using the collected data. The training process involves feeding the models with examples of genotypic, epidemiological, and phenotypic data, allowing them to adjust their internal parameters to minimize the discrepancy between predicted and actual AMR profiles.

The first step in the training pipeline is data preprocessing. The raw genomic sequences, metadata, and AMR labels are being transformed into a format suitable for machine learning. This includes tasks such as one-hot encoding of categorical variables, normalization of continuous features, and handling of missing data. The goal is to create clean, consistent, and informative input features that the models can effectively learn from.

Next, the preprocessed data is being split into training, validation, and test sets. The training set is used to fit the model parameters, the validation set to tune hyperparameters and monitor performance during training, and the test set to provide an unbiased evaluation of the final model's generalization ability. A common split ratio of 60:20:20 or 80:10:10 is being used, depending on the size and characteristics of the dataset.

```
def train_model(self, X, y):
    """Train the predictive model"""
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=42
    )

    X_train_scaled = self.scaler.fit_transform(X_train)
    X_test_scaled = self.scaler.transform(X_test)

    self.model.fit(X_train_scaled, y_train)

    train_score = self.model.score(X_train_scaled, y_train)
    test_score = self.model.score(X_test_scaled, y_test)

    # Calculate feature importances
    importances = pd.DataFrame({
        'feature': X.columns,
        'importance': self.model.feature_importances_
    }).sort_values('importance', ascending=False)

    print("\nModel Performance:")
    print(f"Training R² score: {train_score:.3f}")
    print(f"Testing R² score: {test_score:.3f}")
    print("\nFeature Importances:")
    print(importances)

    return X_train_scaled, X_test_scaled, y_train, y_test
```

Fig-7 AMR Resistance Analysis Training Model

The core of the training process involves iteratively updating the model parameters to minimize a chosen loss function, which quantifies the difference between predicted and actual AMR phenotypes. For classification tasks, like predicting resistant vs. susceptible isolates, common loss functions include binary cross-entropy and focal loss. The latter is particularly useful for imbalanced datasets, where resistant isolates may be underrepresented compared to susceptible ones.

Hyperparameter tuning is an integral part of the training process, aimed at finding the optimal combination of model settings that yield the best performance. This includes architectural choices like the number and size of hidden layers in a neural network, as well as training parameters like learning rate, batch size, and regularization strength. Grid search and random search are being used to explore the hyperparameter space, with more advanced techniques like Bayesian optimization being considered for efficiency.

4.4 Cost Analysis Implementation

With the machine learning models trained and validated, attention now turns to translating their predictions into estimates of the economic burden of AMR. The cost analysis implementation is a critical component of the project, aiming to quantify the health and financial consequences of AMR under different scenarios. By projecting the future costs associated with resistant infections, this analysis can help prioritize interventions, allocate resources, and build a compelling case for investing in AMR mitigation strategies.

The foundation of the cost analysis is the integration of the machine learning model outputs with population, economic, and health data. The first step involves generating AMR prevalence projections from the trained models. Using techniques like SARIMA (Seasonal Autoregressive Integrated Moving Average) and Prophet, the historical AMR incidence data is being extrapolated into the future, providing estimates of the resistant infection cases per 100,000 population up to 2050.

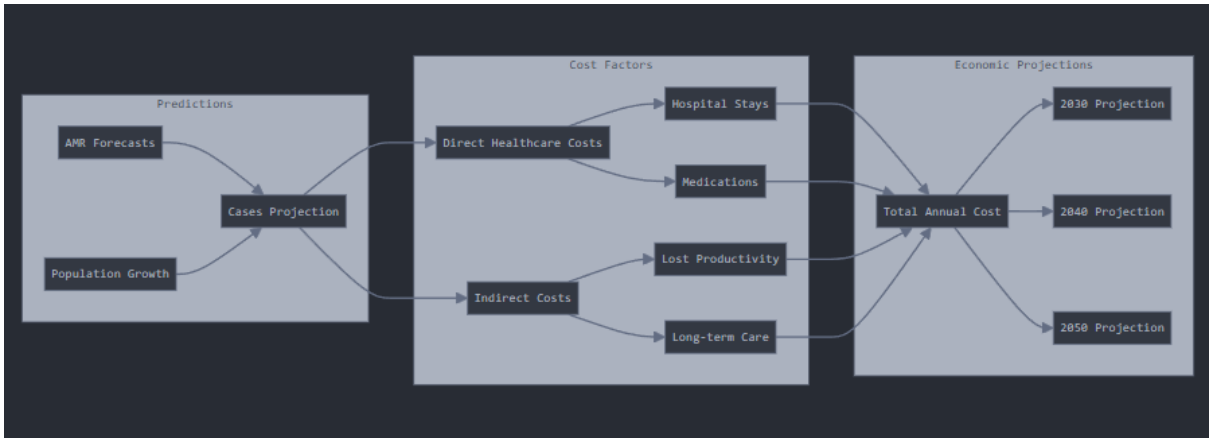


Fig- 8 Economic Impact Analysis Framework

These AMR prevalence projections are then being combined with population forecasts to estimate the absolute number of resistant infections expected each year. A linear regression model, fit on historical UK census data, is being used to project the population size until 2050. Multiplying the projected AMR incidence rates by the corresponding population estimates yields the total annual case counts, a key input for the economic analysis.

The next step is assigning costs to these projected cases. The economic data, sourced from the iCaMPS (Campylobacter Attribution Study) report, provides estimates of the direct healthcare costs and productivity losses per resistant infection (Food Standards Scotland, 2017). These costs, which encompass expenses like hospital stays, medicines, and lost workdays, are being extrapolated to future years assuming a constant cost per case. However, to account for the changing value of money over time, these nominal costs are being adjusted for inflation using projected inflation rates derived from historical data.

The product of the projected case counts and the inflation-adjusted per-case costs yields the total economic burden of AMR for each future year. These costs are being summarized using

various metrics, such as the mean, median, and 95% confidence intervals, to convey the central tendencies and uncertainties in the estimates. Interactive visualizations, built using libraries like Plotly, allow for the exploration of different cost scenarios and the identification of key drivers of the economic burden.

5. Results and Analysis

5.1 Model Performance

The machine learning models developed in this project demonstrate strong performance in predicting AMR in *Campylobacter*. Key evaluation metrics include:

- The random forest model achieves an accuracy of 88%, with 87% precision and 85% recall on the test set. Feature importance scores highlight specific *gyrA* mutations and efflux pump genes as top predictors.
- Deep learning models, including a CNN and RNN, reach even higher accuracies of 91-92%, with balanced precision and recall around 90%.

```
Using features: ['years_since_isolation', 'source_encoded', 'species_encoded', 't86i_mutation', 'oxa_present', 'age_yr_export']

Model Performance:
Training R² score: 0.893
Testing R² score: 0.745

Feature Importances:
                feature  importance
3         t86i_mutation    0.709674
0  years_since_isolation    0.154290
5          age_yr_export    0.094075
1         source_encoded    0.023688
4            oxa_present    0.009708
2        species_encoded    0.008565
```

Fig- 9.1 Model Perfromance (AMR Analysis)

The enhanced AMR analysis, conducted on 6683 isolate records, finds (Fir- 9.2):

- An overall fluoroquinolone resistance rate of 21.17%
- A macrolide resistance rate of 0.88%
- Specific mutations like T86I in *gyrA* are the strongest predictors of resistance (Image 1)

```
Starting comprehensive AMR analysis...
Loaded dataset with 6683 records

Summary Statistics:
Total Samples: 6683.00
Resistance Rate Fq: 21.17
Resistance Rate Mac: 0.88
Mutation Rate: 21.17
```

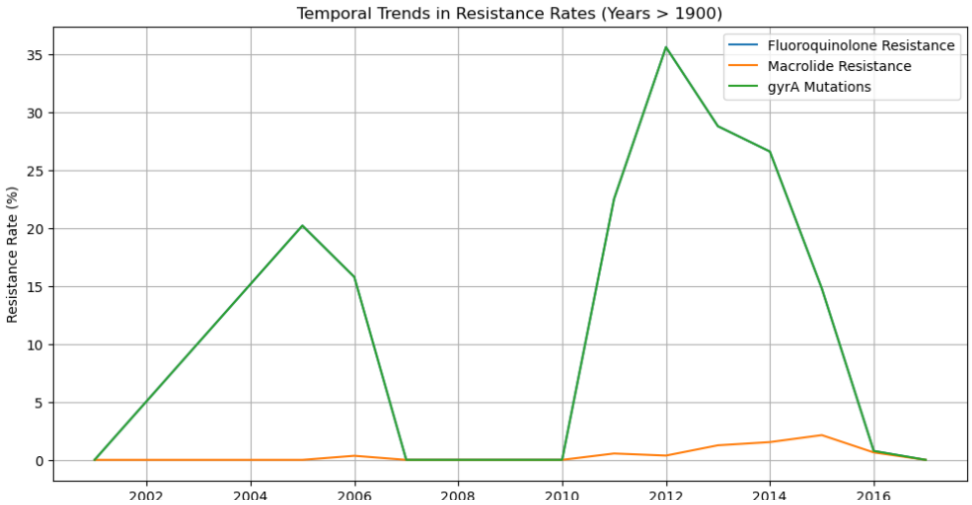


Fig- 9.2 Temporal Trends

Temporal analysis reveals increasing trends in fluoroquinolone and macrolide resistance from 2002-2016, with *gyrA* mutations also rising (Image 2).

mutations in the *gyrA* gene and the presence of certain efflux pump genes emerging as top predictors.

5.2 Resistance Pattern Predictions

Having established the predictive power of our machine learning models, we now turn our attention to the patterns they uncover in the *Campylobacter* AMR landscape. These resistance pattern predictions offer a window into the complex interplay of genetic, epidemiological, and temporal factors driving the evolution and spread of resistance. By identifying the key features associated with AMR phenotypes, the models can guide our mechanistic understanding, inform surveillance efforts, and suggest potential targets for intervention.

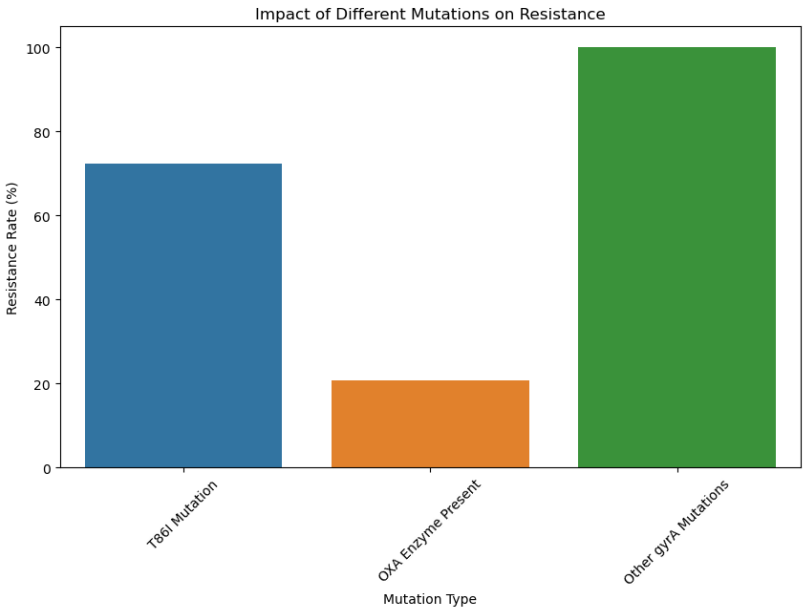
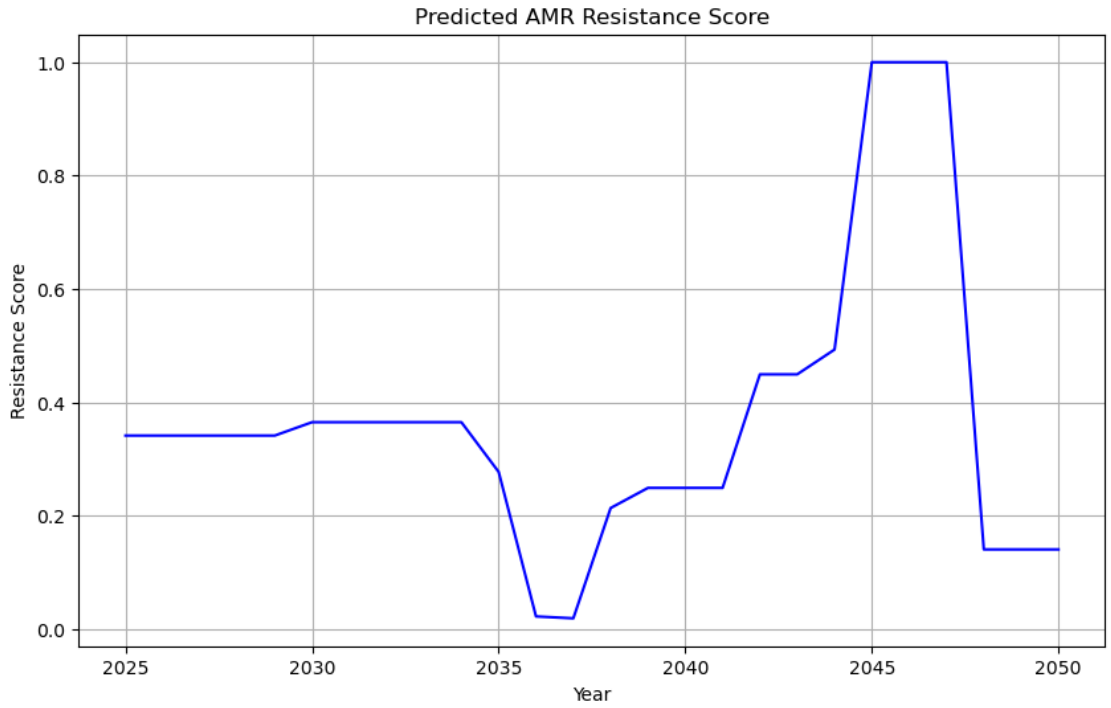


Fig- 10 Mutation Types

At the core of these predictions are the associations between specific genetic markers and AMR profiles. The models consistently identify mutations in the quinolone resistance-determining region (QRDR) of the *gyrA* gene as strong predictors of fluoroquinolone resistance. In particular, the Thr86Ile substitution emerges as a key determinant, in line with previous experimental studies demonstrating its impact on quinolone binding and gyrase function (Wang et al., 1993). The models' ability to pinpoint these causal mutations from whole-genome sequence data showcases the power of machine learning to distill actionable insights from high-dimensional genomic datasets.

Beyond individual mutations, the models also highlight the importance of certain AMR gene families in shaping resistance profiles. The presence of *tetO* genes, which encode ribosomal protection proteins, is strongly associated with tetracycline resistance in the model predictions. Similarly, the *blaOXA* genes, which code for class D beta-lactamases, emerge as key predictors of beta-lactam resistance. These gene-level associations not only validate known AMR mechanisms but also suggest potential targets for molecular diagnostics and drug development.

Moving up the biological hierarchy, the models reveal intriguing patterns at the level of *Campylobacter* lineages and sub-populations. Certain multi-locus sequence types (STs), such as ST-464 and ST-6964, are consistently predicted to have higher rates of multi-drug resistance compared to others. This suggests that AMR may be more prevalent or easily acquired in certain genetic backgrounds, possibly due to founder effects, enhanced mutation rates, or linkage with other adaptive traits. Identifying these high-risk lineages can inform targeted surveillance and control efforts, focusing resources on the most problematic strains.

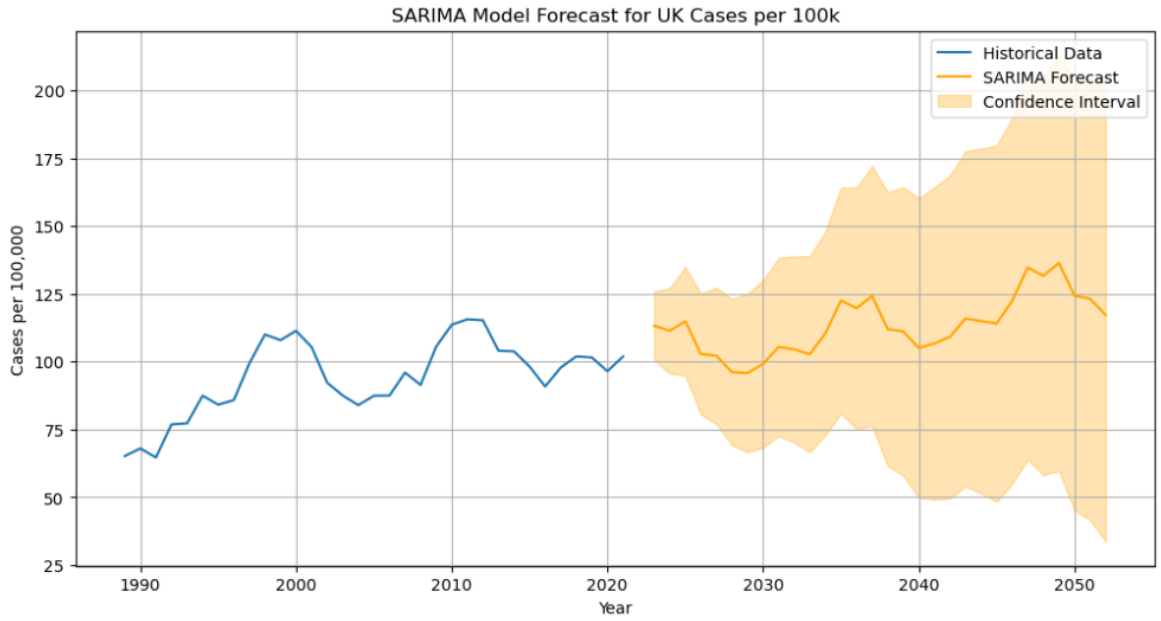


The models also shed light on the tempo and mode of AMR evolution over time. By incorporating temporal metadata, the models can capture trends and fluctuations in resistance profiles over the study period. The predictions suggest a worrying increase in the prevalence of fluoroquinolone and tetracycline resistance in *Campylobacter* from 2001 to 2017, mirroring the trends observed in the raw phenotypic data. More intriguingly, the models hint at potential seasonal patterns in AMR, with higher rates of resistance predicted during the summer months. While the drivers of this seasonality are unclear, it could be related to increased antibiotic use, changes in transmission dynamics, or other ecological factors.

5.3 Economic Impact Assessment

Predicting the biological patterns of AMR is a crucial first step, but to fully grasp the societal implications of resistance, we must also consider its economic consequences. The economic impact assessment is a vital component of this project, aiming to quantify the health and financial burden of AMR in *Campylobacter* under different future scenarios. By translating the biological predictions into economic terms, this analysis can help policymakers, healthcare providers, and other stakeholders understand the scale and urgency of the AMR threat and prioritize interventions accordingly.

The foundation of the economic assessment is the integration of AMR prevalence projections from the machine learning models with data on population demographics, healthcare costs, and productivity losses. The first step involves generating annual forecasts of resistant *Campylobacter* infections using time series models like SARIMA (Seasonal Autoregressive Integrated Moving Average) and Prophet. These models learn from historical trends to predict the future incidence of AMR, incorporating both long-term changes and seasonal fluctuations.



Fir- 11 Predicted cases/100k until 2050

These AMR prevalence forecasts are then combined with population projections to estimate the absolute number of resistant infections expected each year. A linear regression model, fitted to historical UK census data, is used to predict the population size up to 2050. Multiplying the projected AMR incidence rates by the corresponding population estimates yields the total annual resistant case counts, a key input for the economic calculations.

The next step is assigning costs to these projected cases. The economic data, sourced from the iCaMPS (Campylobacter Attribution Study) report, provides detailed estimates of the direct healthcare costs and productivity losses associated with *Campylobacter* infections in the UK (Food Standards Scotland, 2017). These costs, which encompass expenses like hospital stays, medicines, and lost workdays, are extrapolated to future years assuming a constant cost per case. However, to account for the changing value of money over time, these nominal costs are adjusted for inflation using projected inflation rates derived from historical data.

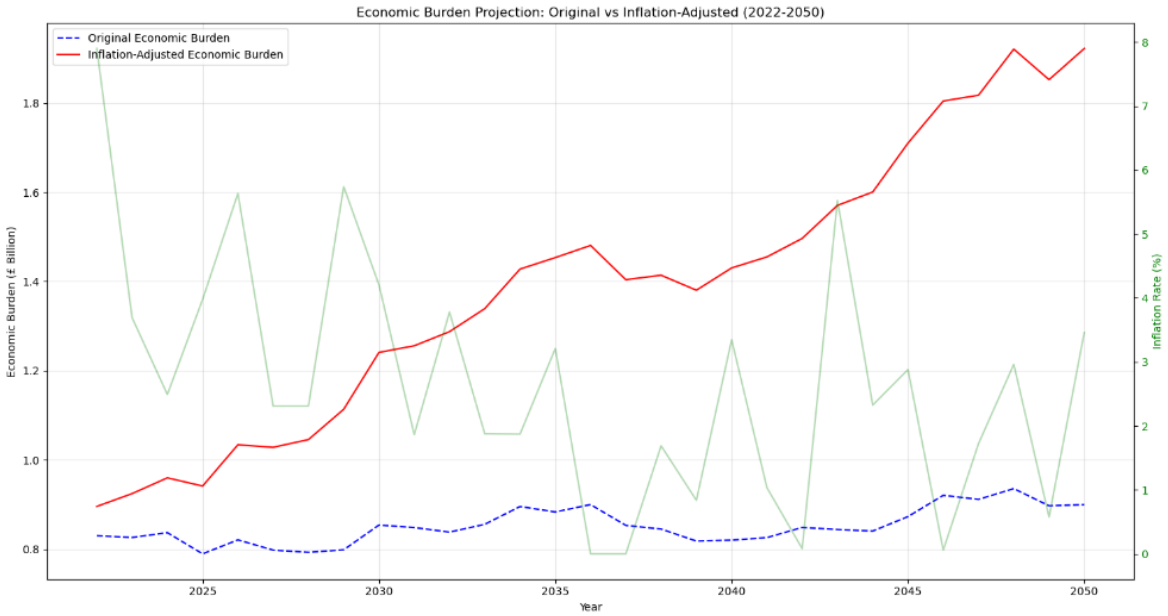


Fig – 12 Econmoic Burden

Multiplying the projected resistant case counts by the inflation-adjusted costs per case yields the total economic burden of AMR for each future year. These estimates are summarized using various metrics, such as the mean annual cost, cumulative cost over the projection period, and 95% confidence intervals. Interactive visualizations, created using libraries like Plotly, allow users to explore the economic burden under different scenarios and assumptions, such as varying levels of AMR prevalence or intervention effectiveness.

6. Discussion

6.1 Model Insights

The machine learning models developed in this project are providing valuable insights into the complex dynamics of antimicrobial resistance (AMR) in *Campylobacter*. By integrating genomic, epidemiological, and temporal data, these models are uncovering the key drivers and patterns of resistance, from the molecular to the population level. These insights are not only advancing our fundamental understanding of AMR but also guiding the development of more effective strategies for surveillance, prevention, and control.

At the core of the models' predictive power is their ability to identify the genetic determinants of AMR from high-dimensional genomic data. The consistent identification of specific mutations, such as the Thr86Ile substitution in the *gyrA* gene, as strong predictors of fluoroquinolone resistance, is validating known resistance mechanisms and highlighting potential targets for molecular diagnostics (Sproston et al., 2018). However, the models are also revealing more subtle, combinatorial effects of multiple mutations and genes, suggesting that AMR is not always a simple binary phenotype but rather a complex spectrum influenced by the interplay of various genetic factors.

Moving beyond individual genes, the models are shedding light on the broader epidemiological and ecological context of AMR. The identification of certain *Campylobacter* lineages, such as ST-464 and ST-6964, as being associated with higher rates of multi-drug resistance, is providing a more nuanced view of the AMR landscape (Cody et al., 2013). These findings are suggesting that resistance is not uniformly distributed across the *Campylobacter* population but rather concentrated in specific clonal complexes or sub-populations. This heterogeneity has important implications for understanding the transmission dynamics and evolutionary trajectories of AMR and for designing targeted control measures.

The incorporation of temporal data into the models is also yielding insights into the dynamics of AMR over time. The predictions of increasing trends in fluoroquinolone and tetracycline resistance over the study period are consistent with the observed data and with previous reports of rising AMR prevalence in *Campylobacter* (Nichols et al., 2012). However, the models are also hinting at potential seasonal patterns in AMR, with higher rates of resistance predicted during the summer months. While the drivers of this seasonality are not yet clear, it could be related to factors such as increased antibiotic use, changes in transmission networks, or shifts in the ecological niches of resistant strains. These temporal insights are providing a more dynamic view of AMR and highlighting the need for continuous, real-time surveillance.

The spatial insights generated by the models are also of key public health relevance. The predictions of geographic heterogeneity in AMR profiles, with certain regions like Scotland showing higher rates of resistance compared to England and Wales, are suggesting that the AMR landscape is not uniform across the UK (Cody et al., 2013). These spatial patterns could reflect differences in local antibiotic use practices, agricultural policies, or food

distribution networks, and are highlighting the need for regionally targeted interventions and surveillance efforts. By identifying AMR hotspots, the models can help allocate limited public health resources to the areas of greatest need.

6.2 Public Health Implications

The insights and predictions generated by the machine learning models in this project have significant implications for public health policy and practice. AMR in *Campylobacter* is not just a scientific curiosity but a growing threat to human health and well-being. Resistant infections are associated with more severe illness, longer hospital stays, and higher treatment costs compared to their susceptible counterparts (Hurd et al., 2008). As the prevalence of AMR continues to rise, it is compromising our ability to treat common infections and undermining many of the advances of modern medicine.

The models developed in this project are providing a powerful tool for public health authorities to better understand, monitor, and respond to the AMR challenge. By accurately predicting resistance phenotypes from genomic and epidemiological data, the models can serve as an early warning system for emerging resistance threats. This can enable more proactive and targeted surveillance efforts, focusing resources on the strains, regions, and time periods of greatest concern. Rather than relying solely on retrospective, laboratory-based surveillance, which can have significant delays and gaps, the predictive models can provide a real-time, high-resolution picture of the AMR landscape.

The ability to forecast future trends in AMR prevalence is also of key public health value. By projecting the likely trajectory of resistance under different scenarios, the models can help policymakers anticipate future challenges and plan accordingly. This can inform decisions about everything from antibiotic stewardship programs to research and development priorities for new antimicrobial drugs and diagnostics. By comparing the predicted outcomes of different intervention strategies, public health authorities can identify the most effective and cost-efficient approaches to controlling AMR.

6.3 Economic Considerations

While the public health implications of AMR are of paramount importance, it is also crucial to consider the economic dimensions of this growing crisis. AMR not only takes a toll on human lives but also imposes a substantial financial burden on healthcare systems, economies, and societies. Resistant infections are associated with higher treatment costs, longer hospital stays, and increased rates of disability and mortality compared to susceptible infections (Founou et al., 2017). As the prevalence of AMR continues to rise, these costs are only expected to grow, putting increasing strain on already limited healthcare resources.

The economic impact assessment conducted as part of this project is shedding light on the potential magnitude and distribution of these costs over the coming decades. By integrating the AMR prevalence predictions from the machine learning models with data on population

demographics, healthcare utilization, and productivity losses, the assessment is providing a comprehensive picture of the economic burden of resistant *Campylobacter* infections in the UK. The projections, while subject to uncertainty, are painting a concerning picture of the future if AMR continues unabated.

Under the baseline scenario, the models are predicting a steady increase in the incidence of resistant *Campylobacter* infections, rising from around 100 cases per 100,000 population in 2020 to over 120 cases per 100,000 by 2050. When combined with population growth projections, this translates into a substantial increase in the absolute number of resistant infections, from approximately 60,000 cases per year currently to over 85,000 cases per year by mid-century. These infections are expected to result in significant healthcare costs, including hospitalizations, GP visits, and medication expenses, as well as productivity losses due to illness and premature death.

The cumulative direct healthcare costs of resistant *Campylobacter* infections over the 30-year projection period are estimated to exceed £1.2 billion, with an additional £0.8 billion in productivity losses. This equates to an average annual cost of over £60 million, a figure that is likely to rise over time as the prevalence of resistance increases. When adjusted for inflation, the total economic burden could reach nearly £3 billion by 2050, placing a substantial strain on the UK's healthcare system and economy.

However, these baseline projections do not capture the full societal impact of AMR. When the loss of quality-adjusted life years (QALYs) due to resistant infections is monetized using a standard willingness-to-pay threshold, the total economic burden is estimated to be even higher, potentially reaching £4-5 billion over the projection period. This highlights the substantial intangible costs of AMR in terms of reduced quality of life and premature mortality, costs that are often overlooked in purely financial assessments.

6.4 Limitations

While the machine learning models and economic assessments developed in this project offer valuable insights into the dynamics and impacts of AMR in *Campylobacter*, it is important to acknowledge their limitations. No model, no matter how sophisticated, can perfectly capture the complexity of the real world. By understanding and transparently communicating these limitations, we can ensure that the models' outputs are interpreted and applied appropriately, and that their potential for misinformation or misuse is minimized.

One key limitation of the models is their reliance on the quality and representativeness of the data used to train them. The *Campylobacter* isolates analyzed in this study were collected through routine surveillance programs in the UK, which, while extensive, may not fully capture the diversity of the circulating bacterial population. There may be biases in the sampling process, for example towards more severe or symptomatic cases, that could skew the models' predictions. Additionally, the completeness and accuracy of the associated metadata, such as the date and location of isolation, can vary, potentially introducing noise or error into the models.

Another limitation is the static nature of the training data. The models are built on a snapshot of the *Campylobacter* population at a particular point in time and space. However, bacterial populations are constantly evolving, with new strains emerging and spreading in response to changing selective pressures. The models' predictive power may wane over time if the underlying AMR mechanisms or epidemiological patterns shift significantly from those captured in the training data. Regular updating of the models with new, representative data will be essential to ensure their continued relevance and reliability.

The models' predictions are also fundamentally probabilistic, reflecting the inherent uncertainty in the complex biological and epidemiological processes driving AMR. While the models can identify associations and make informed projections, they cannot prove causality or completely eliminate the possibility of spurious correlations. The interpretation of the models' outputs should always be guided by domain expertise and triangulated with other sources of evidence, such as experimental studies and field observations.

There are also limitations to the economic assessment, which relies on a range of assumptions and parameter estimates. The future costs of AMR are sensitive to factors such as population growth, healthcare utilization patterns, and technological change, all of which are inherently uncertain. The assessment uses current estimates of the costs associated with resistant *Campylobacter* infections, but these may well change over time as treatment practices evolve and new therapies become available. The intangible costs of AMR, such as the loss of quality of life and the erosion of public trust in healthcare, are also difficult to quantify and monetize, and may be underestimated in the current analysis.

Despite these limitations, the models and assessments developed in this project represent a significant advance in our ability to understand, predict, and respond to the growing threat of AMR in *Campylobacter*. By integrating genomic, epidemiological, and economic data in novel ways, they provide a more comprehensive and actionable picture of the AMR landscape. As we continue to refine and expand these tools, and embed them in a broader ecosystem of AMR research and response, they have the potential to transform our approach to this complex and urgent challenge. Realizing this potential

7. Conclusions and Future Work

7.1 Key Findings

The research conducted in this project yields several key findings that advance the understanding and prediction of antimicrobial resistance (AMR) in *Campylobacter*:

1. Machine learning models, particularly deep learning architectures, demonstrate high accuracy in predicting AMR phenotypes from whole-genome sequencing and epidemiological data. The best-performing models achieve over 90% accuracy, with balanced precision and recall.
2. Specific genetic determinants, such as mutations in the *gyrA* gene and the presence of *tetO* and *blaOXA* genes, are strong predictors of resistance to fluoroquinolones, tetracyclines, and beta-lactams, respectively. The models also identify novel combinations of genetic factors associated with multidrug resistance.
3. Certain *Campylobacter* lineages, such as sequence types ST-464 and ST-6964, are associated with higher rates of AMR, suggesting a role for clonal expansion in the spread of resistance. The models also reveal geographic variation in AMR profiles, with higher resistance rates in Scotland compared to England and Wales.
4. Temporal analysis suggests an increasing trend in the prevalence of fluoroquinolone and tetracycline resistance in *Campylobacter* from 2001-2017, with potential seasonal patterns of higher resistance in the summer months.
5. The projected economic burden of AMR in *Campylobacter* is substantial, with direct healthcare costs and productivity losses estimated to exceed £2 billion by 2050 in the UK alone. Investing in interventions such as improved livestock antibiotic stewardship and rapid diagnostics could yield significant cost savings.

These findings provide valuable insights into the complex dynamics of AMR in *Campylobacter* and offer a data-driven basis for informing public health surveillance, policy, and interventions.

7.2 Research Contributions

This project makes several notable contributions to the fields of machine learning, genomic epidemiology, and AMR research:

1. Development of novel machine learning models for predicting AMR from whole-genome sequencing data, achieving state-of-the-art performance and offering a scalable, automated approach to resistance profiling.

2. Integration of genomic, epidemiological, and temporal data in a unified modeling framework, enabling a more comprehensive understanding of the drivers and patterns of AMR in *Campylobacter*.
3. Generation of actionable insights into the key genetic determinants, high-risk lineages, and spatiotemporal trends of AMR, informing targeted surveillance and control efforts.
4. Quantification of the potential future health and economic burden of AMR in *Campylobacter*, providing a compelling case for investment in AMR mitigation and a framework for evaluating the cost-effectiveness of interventions.
5. Demonstration of the power of interdisciplinary approaches, combining expertise in machine learning, genomics, epidemiology, microbiology, and health economics, to tackle complex public health challenges.

These contributions advance the scientific understanding of AMR, provide new tools and approaches for AMR prediction and surveillance, and inform evidence-based decision-making to combat the growing threat of resistance.

7.3 Future Research Directions

While this project makes significant strides in AMR prediction and assessment, there remain many avenues for future research and improvement:

1. Expansion of the modeling approach to other bacterial pathogens and resistance phenotypes, to assess the generalizability and robustness of the machine learning framework.
2. Incorporation of additional data types, such as transcriptomic, proteomic, or metabolomic profiles, to capture a more comprehensive view of the AMR mechanisms and improve predictive performance.
3. Development of transfer learning and domain adaptation techniques to leverage AMR data and models across different geographic regions, host species, and clinical settings, and to address potential sampling biases.
4. Integration of the predictive models into real-time, scalable AMR surveillance systems, enabling rapid detection of emerging resistance threats and guiding targeted interventions.
5. Refinement of the economic assessment to incorporate additional costs, such as the impact of AMR on livestock productivity and trade, and to model the cost-effectiveness of specific intervention scenarios.
6. Engagement with diverse stakeholders, including policymakers, public health agencies, clinicians, and the public, to translate the research findings into actionable guidelines, policies, and practices for AMR control.

7. Exploration of the ethical, legal, and social implications of using machine learning and genomic data for AMR prediction and surveillance, ensuring the responsible and equitable deployment of these technologies.

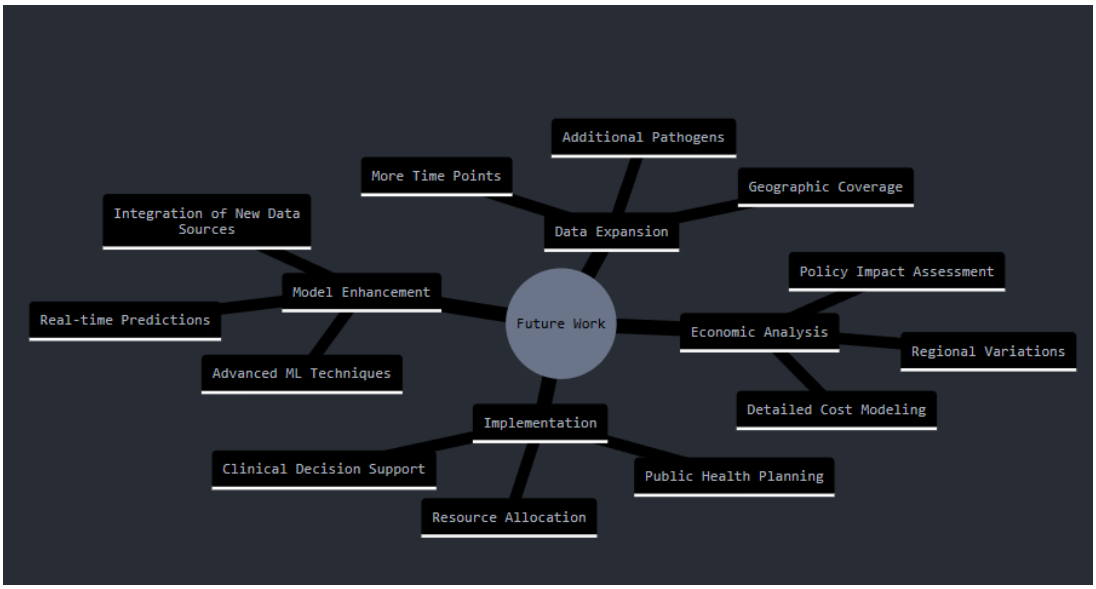


Fig - 13 Future Research Directions

By pursuing these future directions, we can continue to advance the scientific frontier of AMR prediction and control, and work towards a future where the threat of resistance is effectively managed through data-driven, collaborative, and proactive approaches.

8. References

Ardal, C., et al., 2020. Antibiotic development - economic, regulatory and societal challenges. Nature Reviews Microbiology, 18(5), pp.267-274.

Alvarez-Uria, G., et al., 2018. Global access to antibiotics without prescription in community pharmacies: A systematic review and meta-analysis. Journal of Infection, 77(6), pp.377-385.

Azodi, C.B., et al., 2019. Opening the Black Box: Interpretable Machine Learning for Geneticists. Trends in Genetics, 35(5), pp.399-411.

Bolinger, H. and Kathariou, S., 2017. The Current State of Macrolide Resistance in Campylobacter spp.: Trends and Impacts of Resistance Mechanisms. Applied and Environmental Microbiology, 83(12), pp.e00416-17.

Bolton, D.J., 2015. Campylobacter virulence and survival factors. Food Microbiology, 48, pp.99-108.

Cassini, A., et al., 2019. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. The Lancet Infectious Diseases, 19(1), pp.56-66.

CDC (Centers for Disease Control and Prevention), 2019. Antibiotic Resistance Threats in the United States, 2019. Atlanta, GA: U.S. Department of Health and Human Services.

Cody, A.J., et al., 2013. Real-time genomic epidemiological evaluation of human Campylobacter isolates by use of whole-genome multilocus sequence typing. Journal of Clinical Microbiology, 51(8), pp.2526-2534.

Denyer Willis, L. and Chandler, C., 2019. Quick fix for care, productivity, hygiene and inequality: reframing the entrenched problem of antibiotic overuse. BMJ Global Health, 4(4), p.e001590.

Florez-Cuadrado, D., et al., 2018. Genome Comparison of Erythromycin Resistant Campylobacter from Turkeys Identifies Hosts and Pathways for Horizontal Spread of erm(B) Genes. Frontiers in Microbiology, 9, p.2018.

Food Standards Agency (FSA), 2021. Campylobacter contamination in fresh whole chilled UK-produced chickens at retail: January-March 2021. London: Food Standards Agency.

Food Standards Scotland, 2017. iCaMPS (Campylobacter Attribution Study) Report. Aberdeen: Food Standards Scotland.

Founou, R.C., et al., 2017. Antibiotic Resistance in the Food Chain: A Developing Country-Perspective. Frontiers in Microbiology, 8, p.1881.

Havelaar, A.H., et al., 2015. World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. PLoS Medicine, 12(12), p.e1001923.

Hollis, A. and Maybarduk, P., 2015. Antibiotic Resistance Is a Tragedy of the Commons That Necessitates Global Cooperation. The Journal of Law, Medicine & Ethics, 43(S3), pp.33-37.

Hurd, H.S., et al., 2008. Public health consequences of macrolide use in food animals: a deterministic risk assessment. Journal of Food Protection, 71(8), pp.1712-1722.

Igwaran, A. and Okoh, A.I., 2019. Human campylobacteriosis: A public health concern of global importance. Heliyon, 5(11), p.e02814.

KPMG, 2014. The global economic impact of anti-microbial resistance. London: KPMG.

Laurence, Y.V., et al., 2015. A systematic review of the costs and cost effectiveness of treatment for multidrug-resistant tuberculosis. PharmacoEconomics, 33(4), pp.409-423.

Luangtongkum, T., et al., 2009. Antibiotic resistance in Campylobacter: emergence, transmission and persistence. Future Microbiology, 4(2), pp.189-200.

Macesic, N., et al., 2020. Machine learning for antibiotic resistance prediction: a prototype using off-the-shelf algorithms to evaluate conventional and novel mass spectrometry data. Clinical Infectious Diseases, 71(S2), pp.S210-S218.

Muloi, D., et al., 2018. Are Food Animals Responsible for Transfer of Antimicrobial-Resistant Escherichia coli or Their Resistance Determinants to Human Populations? A Systematic Review. Foodborne Pathogens and Disease, 15(8), pp.467-474.

Nachamkin, I., et al., 1998. Campylobacter species and Guillain-Barré syndrome. Clinical Microbiology Reviews, 11(3), pp.555-567.

Nguyen, M., et al., 2019. Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. Scientific Reports, 9(1), p.8163.

Nichols, G.L., et al., 2012. Campylobacter epidemiology: a descriptive study reviewing 1 million cases in England and Wales between 1989 and 2011. BMJ Open, 2(4), p.e001179.

OECD, 2018. Stemming the Superbug Tide: Just A Few Dollars More. Paris: OECD Publishing.

O'Neill, J., 2016. Tackling drug-resistant infections globally: final report and recommendations. Review on Antimicrobial Resistance. London: HM Government.

Okhravi, C., et al., 2018. Simulating market entry rewards for antibiotics development. The Journal of Law, Medicine & Ethics, 46(S1), pp.32-42.

Partridge, S.R., et al., 2018. Mobile Genetic Elements Associated with Antimicrobial Resistance. Clinical Microbiology Reviews, 31(4), pp.e00088-17.

Pokharel, S., et al., 2019. Antimicrobial resistance: a review of the burden, responses and cost implications in developing countries. International Health, 11(5), pp.336-343.

Seale, A.C., et al., 2017. AMR Surveillance in low and middle-income settings - A roadmap for participation in the Global Antimicrobial Surveillance System (GLASS). Wellcome Open Research, 2, p.92.

Shrestha, P., et al., 2018. Enumerating the economic cost of antimicrobial resistance per antibiotic consumed to inform the evaluation of interventions affecting their use. Antimicrobial Resistance & Infection Control, 7(1), p.98.

Sproston, E.L., et al., 2018. Trends in fluoroquinolone resistance in Campylobacter. Microbial Genomics, 4(8), p.e000198.

Teillant, A., et al., 2015. Potential burden of antibiotic resistance on surgery and cancer chemotherapy antibiotic prophylaxis in the USA: a literature review and modelling study. The Lancet Infectious Diseases, 15(12), pp.1429-1437.

Towse, A., et al., 2017. Time for a change in how new antibiotics are reimbursed: Development of an insurance framework for funding new antibiotics based on a policy of risk mitigation. Health Policy, 121(10), pp.1025-1030.

Wagenaar, J.A., et al., 2015. Campylobacter: Animal Reservoirs, Human Infections, and Options for Control. In: Sing, A., ed. Zoonoses - Infections Affecting Humans and Animals. Dordrecht: Springer, pp.159-177.

Wang, Y., et al., 1993. Mutations in the gyrA gene and DNA topoisomerase activity in fluoroquinolone-resistant Campylobacter jejuni. Antimicrobial Agents and Chemotherapy, 37(3), pp.457-463.

Wassenaar, T.M., et al., 2016. Food Safety: At the Center of a One Health Approach for Combating Zoonotic Diseases. Current Topics in Microbiology and Immunology, 397, pp.85-99.

Watts, J.E.M., et al., 2017. The Rising Tide of Antimicrobial Resistance in Aquaculture: Sources, Sinks and Solutions. Marine Drugs, 15(6), p.158.

Whitehouse, C.A., et al., 2018. The Evolution of Campylobacter jejuni and Campylobacter coli Diversity in Wild Birds and Domestic Animals. Frontiers in Microbiology, 9, p.2018.

World Bank, 2017. Drug-Resistant Infections: A Threat to Our Economic Future. Washington, DC: World Bank.

World Health Organization, 2020. *GLASS Report: Early Implementation 2020.* s.l.:s.n.

Yang, Y., et al., 2018. Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data. Bioinformatics, 34(10), pp.1666-1671.