






Review

# The Synergy of Machine Learning and Epidemiology in Addressing Carbapenem Resistance: A Comprehensive Review

Aikaterini Sakagianni <sup>1</sup>, Christina Koufopoulou <sup>2</sup>, Petros Koufopoulos <sup>3</sup>, Georgios Feretzakis <sup>4</sup>,  
Dimitris Kalles <sup>4</sup>, Evgenia Paxinou <sup>4</sup>, Pavlos Myrianthefts <sup>5</sup> and Vassilios S. Verykios <sup>4,\*</sup>

<sup>1</sup> Intensive Care Unit, Sismanogleio General Hospital, 15126 Marousi, Greece; sakagianni@sismanoglio.gr

<sup>2</sup> Anesthesiology Department, Aretaieio Hospital, National and Kapodistrian University of Athens, 11528 Athens, Greece; ckoufopoulou@uoa.gr

<sup>3</sup> Internal Medicine Department, Sismanogleio General Hospital, 15126 Marousi, Greece; smd1600078@uoa.gr

<sup>4</sup> School of Science and Technology, Hellenic Open University, 26335 Patras, Greece; georgios.feretzakis@ac.eap.gr (G.F.); kalles@eap.gr (D.K.); paxinou.evgenia@ac.eap.gr (E.P.)

<sup>5</sup> Faculty of Nursing, School of Health Sciences, National and Kapodistrian University of Athens, 11527 Athens, Greece; pmiriant@nurs.uoa.gr

\* Correspondence: verykios@eap.gr

**Abstract:** Background/Objectives: Carbapenem resistance poses a significant threat to public health by undermining the efficacy of one of the last lines of antibiotic defense. Addressing this challenge requires innovative approaches that can enhance our understanding and ability to combat resistant pathogens. This review aims to explore the integration of machine learning (ML) and epidemiological approaches to understand, predict, and combat carbapenem-resistant pathogens. It examines how leveraging large datasets and advanced computational techniques can identify patterns, predict outbreaks, and inform targeted intervention strategies. Methods: The review synthesizes current knowledge on the mechanisms of carbapenem resistance, highlights the strengths and limitations of traditional epidemiological methods, and evaluates the transformative potential of ML. Real-world applications and case studies are used to demonstrate the practical benefits of combining ML and epidemiology. Technical and ethical challenges, such as data quality, model interpretability, and biases, are also addressed, with recommendations provided for overcoming these obstacles. Results: By integrating ML with epidemiological analysis, significant improvements can be made in predictive accuracy, identifying novel patterns in disease transmission, and designing effective public health interventions. Case studies illustrate the benefits of interdisciplinary collaboration in tackling carbapenem resistance, though challenges such as model interpretability and data biases must be managed. Conclusions: The combination of ML and epidemiology holds great promise for enhancing our capacity to predict and prevent carbapenem-resistant infections. Future research should focus on overcoming technical and ethical challenges to fully realize the potential of these approaches. Interdisciplinary collaboration is key to developing sustainable strategies to combat antimicrobial resistance (AMR), ultimately improving patient outcomes and safeguarding public health.

**Keywords:** carbapenem resistance; machine learning; epidemiology; antimicrobial resistance; predictive modeling; public health



**Citation:** Sakagianni, A.; Koufopoulou, C.; Koufopoulos, P.; Feretzakis, G.; Kalles, D.; Paxinou, E.; Myrianthefts, P.; Verykios, V.S. The Synergy of Machine Learning and Epidemiology in Addressing Carbapenem Resistance: A Comprehensive Review. *Antibiotics* **2024**, *13*, 996. <https://doi.org/10.3390/antibiotics13100996>

Academic Editors: Delia Muntean, Adrian Man and Florin George Horhat

Received: 19 September 2024

Revised: 16 October 2024

Accepted: 19 October 2024

Published: 21 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Antimicrobial resistance (AMR) represents a critical and escalating threat to global health, with significant implications for the treatment and prevention of infectious diseases. The rise of AMR is primarily driven by the misuse and overuse of antimicrobials in human medicine, agriculture, and veterinary practices, leading to the emergence of drug-resistant pathogens that are increasingly difficult to treat with existing medications. According to the World Health Organization (WHO), AMR is among the top-ten global public health

threats facing humanity. In 2019, bacterial AMR directly caused 1.27 million deaths and contributed to 4.95 million deaths worldwide [1].

The impact of AMR extends beyond health, affecting economic stability and development. Projections indicate that if no effective measures are taken, AMR could lead to 10 million deaths annually by 2050 and could cost the global economy up to USD 100 trillion due to increased healthcare costs, loss of productivity, and other factors [2]. High-income countries are experiencing significant issues with healthcare-associated infections caused by resistant bacteria, such as methicillin-resistant *Staphylococcus aureus* (MRSA), while low- and middle-income countries bear a disproportionate burden of resistant infections due to weaker healthcare infrastructure and surveillance systems [3]. In response to this global crisis, coordinated international efforts are crucial. The WHO emphasizes a One Health approach, integrating actions across human health, animal health, and environmental sectors to combat AMR comprehensively [4]. Strengthening antimicrobial stewardship programs, enhancing infection prevention and control measures, and investing in the development of new antibiotics, vaccines, and diagnostics are pivotal strategies recommended by global health authorities to address AMR effectively [5,6].

While AMR spans a wide range of antimicrobial drugs and pathogens, carbapenem resistance (CR) is particularly alarming. Carbapenems are considered one of the last lines of defense against multidrug-resistant bacterial infections, making CR a significant clinical challenge. The emergence of carbapenem-resistant organisms threatens the effectiveness of this vital class of antibiotics, especially in treating severe hospital-acquired infections [7]. The motivation behind this review is to explore the intersection of machine learning (ML) and epidemiology in addressing CR. The growing complexity and volume of AMR data, combined with the increasing need for rapid predictions, present an opportunity to leverage ML techniques in enhancing the surveillance, prediction, and control of carbapenem-resistant infections (CRIs).

This review aims to synthesize the current state of research on the use of ML in AMR prediction, particularly focusing on CR, and to propose future research directions that combine these approaches. The review's relevance lies in its potential to inform public health policies and enhance clinical decision-making in managing CRIs.

The key contributions of this review include the following:

1. Comprehensive overview of synergy. A detailed analysis of how ML and epidemiological methods can complement each other in addressing carbapenem resistance is provided.
2. Identification of gaps in traditional approaches. The review outlines the limitations of traditional epidemiological methods in capturing the complexity of resistance mechanisms and transmission patterns and discusses how ML can fill these gaps.
3. Evaluation of ML applications. It examines the current state of ML applications in antimicrobial resistance, particularly in predicting CR, and the potential effectiveness of these models in clinical and public health settings.
4. Proposals for future research. The review identifies key areas for future research, including the need for more robust data integration, model validation, and the development of real-time surveillance systems.
5. Clinical and public health implications. The review emphasizes the clinical and public health benefits of integrating ML and epidemiology to improve predictions, patient outcomes, and intervention strategies.

The remainder of this paper is organized as follows: Section 2 discusses the epidemiology and mechanisms of CR, including the clinical and public health implications. Section 3 delves into the traditional epidemiological approaches used in studying AMR, highlighting their strengths and limitations. Section 4 focuses on the potential of ML to enhance CR predictions, discussing existing models and applications. Section 5 explores how ML and epidemiological methods can be integrated to provide a comprehensive approach to tackling carbapenem resistance. Section 6 outlines future research directions and recom-

recommendations for improving AMR surveillance and prediction systems. Finally, Section 7 concludes the review by summarizing key insights and contributions.

## 2. Specific Focus on Carbapenem Resistance

AMR presents a wide array of challenges, but resistance to carbapenem antibiotics is particularly concerning due to their role as the last effective treatment option for many bacterial infections. The emergence and rapid spread of CR in organisms, such as *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*, exemplifies the severity of the AMR crisis. This specific form of resistance underscores the broader challenge posed by AMR and highlights the urgent need for improved surveillance, prevention, and treatment strategies.

Carbapenem-resistant organisms often exhibit complex resistance mechanisms that make infections difficult to treat and control. Therefore, understanding these mechanisms is essential for developing targeted strategies to combat CR on a global scale [7].

### 2.1. Mechanisms of Resistance

Carbapenem resistance arises through several mechanisms, primarily involving enzymatic degradation, efflux pumps, and porin mutations [8,9]. One of the most critical mechanisms is the production of carbapenemases, a group of enzymes that can hydrolyze carbapenems and other  $\beta$ -lactams, rendering these potent antibiotics ineffective. Notable carbapenemases include KPC (*Klebsiella pneumoniae* carbapenemase), NDM (New Delhi metallo- $\beta$ -lactamase), VIM (Verona integron-encoded metallo- $\beta$ -lactamase), and OXA-48 (Oxacillinase-48) [7,8]. These enzymes differ in their genetic origins and geographical distribution, but they all confer high levels of resistance, making infections difficult to treat [9]. The widespread dissemination of these enzymes is largely driven by plasmids, which facilitate the transfer of resistance genes across different bacterial species, thus exacerbating the problem of multidrug resistance [8,9].

Furthermore, efflux pumps are able to actively extrude a wide range of different antibiotics out of the bacterial cells, being one of the most important causes of resistance to carbapenems. In general, this decreases the intracellular drug content to subtherapeutic levels. This not only leads to reduced efficacy of carbapenems but also ensures that bacteria can survive even in environments of high antibiotic pressure. Overexpression of efflux pumps can be driven by genetic changes, or an alteration in regulation, often through the influence of the presence of antibiotics. This underscores the dynamic nature of bacterial adaptation [10].

Besides enzymatic degradation and active efflux mechanisms, the third important mechanism of resistance is changes in the permeability of bacterial cell membranes. It most frequently occurs in the form of mutations in the porin proteins, leading to decreased expression or a complete loss of specific porin channels, acting as portals for antibiotics to cross the bacterial cell membrane. These mutations result in a reduced uptake of carbapenems, preventing the antibiotics from reaching their intracellular targets, such as penicillin-binding proteins [11]. The combined effect of porin loss and carbapenemase production can result in extremely high levels of resistance, posing a significant challenge to treatment [12].

Resistance mechanisms are further complicated by the ability of bacteria to share resistance genes through horizontal gene transfer (HGT). This process allows for the rapid spread of resistance determinants between different bacteria, even across species and genera. HGT occurs through various means, including transformation, transduction, and conjugation, with conjugative plasmids being particularly important in the spread of carbapenem resistance [13]. The global spread of carbapenemase-producing Enterobacteriaceae (CPE) is a testament to the role of HGT in the dissemination of resistance, making it a critical factor in the ongoing battle against antibiotic resistance [7].

The combination of the above resistance mechanisms creates a complex challenge in the control of carbapenem-resistant infections (CRIs) [14]. A multifaceted approach involving new antibiotic development, use of combination therapy, and strict infection

control practices is required to limit the dissemination of the resistant bacterial strains. In addition, ongoing surveillance and research on the molecular mechanisms of resistance maintain an edge in this continuing battle [12].

## 2.2. Epidemiology and Incidence of Carbapenem-Resistant Organisms

Resistance to carbapenem has escalated into a health problem worldwide, especially during the last decade. The most threatening ones are the carbapenem-resistant Enterobacteriaceae (CRE), carrying the most concerning ability to cause severe and frequently untreatable infections. The presence of these bacteria has been documented in multiple geographical areas, with strikingly high rates in countries that face specific problems related to poor antibiotic stewardship practices, as well as where overuse of antibiotics occurs both in healthcare and agriculture [15,16].

However, the effects of the emergence and spread of CRE have varied among different European countries. For example, in countries such as Greece and Italy, challenges related to CRE have been especially prominent [16]. The combination of factors, such as heavy antibiotic consumption, high rates of hospital-acquired infections, and poor infection control measures, has contributed to the wide dissemination of this type of resistance, with the majority of outbreaks occurring in healthcare environments [17]. Recent studies showed that hospital-acquired infections in Greece account for approximately 316,000 hospital bed days and EUR 73 million in costs due to resistant pathogens, particularly *Klebsiella pneumoniae* and *Escherichia coli*, with over 50% of *K. pneumoniae* isolates being carbapenem-resistant [18].

It is no different in Asia, as high levels of carbapenem resistance have been reported from a number of countries, such as India and China, where antibiotics are freely available without prescription, healthcare facilities are possibly limited or under-resourced, and there is massive use of antibiotics in agriculture and aquaculture. The carbapenemase-producing organisms (CPOs) have presented tremendous challenges to public health. In India, there have been reports of a high prevalence of CPOs, particularly those carrying the New Delhi metallo- $\beta$ -lactamase (NDM) gene, in both community-acquired and nosocomial infections [15]. For instance, a study reported that 90.3% of carbapenem-resistant bacteria were carbapenemase producers, with NDM-1 being the most dominant at 47% [19]. Equally burdened by a huge population and significant infectious disease load, China has likewise reported increasing incidences of CRE. Between 2012 and 2016, 85.7% of CRE strains in China carried carbapenemase genes, with KPC being predominant in *K. pneumoniae*, and NDM in *E. coli* and *Enterobacter cloacae*. Moreover, the resistance rates of *K. pneumoniae* to meropenem in China have reached 24.2%, underscoring the severity of the problem [19].

The United States remains no exception to the impending threat of CR pathogens. Reports of CRE infections are increasing all over the country, posing significant threats to public health. The CDC has defined CRE as an urgent threat, and there has been much concern expressed over the potential to spread in healthcare settings. More than 2.8 million antibiotic-resistant infections occur in the U.S. each year, resulting in over 35,000 deaths. CRE, in particular, has been responsible for 13,100 cases and 1100 deaths annually [20]. The outbreaks of CRE in the U.S. healthcare facilities have been associated with high morbidity and mortality, pointing to the need for an improved infection control practice and antibiotic stewardship [20].

Besides CRE, other CR pathogens include *Pseudomonas aeruginosa* and *Acinetobacter baumannii*. These are part of the problem and considerably affect the outcomes of treatment, the prevalence of which has increased with time. Community-acquired pneumonia due to CR *A. baumannii* (CRAB) is also rampant in healthcare settings, especially in the ICUs, with resistance rates described at over 30–90% in regions such as Asia, Eastern Europe, and Latin America [21].

Even though it varies across different regions, CR *P. aeruginosa* (CR-PA) shows similar alarming resistance rates worldwide [22]. In the United States, CR-PA is a significant healthcare-associated pathogen, responsible for 10–30% of *P. aeruginosa* isolates. A study

highlighted that carbapenemase-producing *P. aeruginosa* is frequently found in ventilator-associated pneumonia (VAP) and catheter-related urinary tract infections, contributing to longer hospital stays and higher mortality rates [22]. Both these non-fermenting CR pathogens, including *A. baumannii*, lead to increased healthcare burdens, accounting for over 80% of CR cases in a five-year U.S.-based study [23].

A systematic review of carbapenem resistance in animals, foods, and the environment on the African continent further emphasized the widespread nature of this issue. The review found a pooled prevalence of 19.1% across animal, environmental, and food ecosystems, highlighting *Escherichia* spp. (53.5%), *Klebsiella* spp. (35.4%), and *Pseudomonas* spp. (15.7%) as the predominant CRB species. The most common carbapenemase genes reported were from the blaOXA (52.4%) and blaNDM (40.5%) families. These findings suggest that animal–environment–food ecosystems play a significant role as reservoirs for CRE and other carbapenem-resistant bacteria, further driving their dissemination into human populations [24].

This global and multisectoral spread of carbapenem resistance highlights the urgent need for enhanced surveillance, infection control, and new antibiotics targeting both fermenting and non-fermenting CR pathogens, as well as stricter antibiotic stewardship practices across all sectors.

### 2.3. Clinical Implications

Carbapenem-resistant infections, especially those caused by CPOs, such as *Klebsiella pneumoniae*, have significant clinical relevance due to their association with higher mortality rates. A study conducted in Italian hospitals between 2010 and 2013 reported a 14-day mortality rate of 34.1% among patients with KPC-producing *K. pneumoniae* (KPC-Kp) infections. Risk factors associated with this high mortality include bloodstream infections (BSIs), septic shock, chronic renal failure, and the use of colistin-resistant isolates. Inadequate empirical therapy further elevates these risks [25]. A meta-analysis of worldwide studies on CRE infections reported that 26–44% of deaths in patients with CRIs were directly attributable to resistance [23]. This highlights the profound clinical consequences of CRE, particularly in the context of BSIs, which are associated with even higher mortality rates [26].

Furthermore, CPO outbreaks are increasingly being observed in healthcare settings, leading to endemic outbreaks in some regions [26,27]. For example, a prolonged outbreak of NDM-producing *Klebsiella pneumoniae* occurred in Tuscany, Italy, from 2018 to 2021. Genomic sequencing of 117 isolates from 76 patients revealed the spread of a high-risk clone (ST-147), resistant to nearly all antibiotics, highlighting the regional transmission of this multidrug-resistant organism [28]. These infections disproportionately affect critically ill and immunocompromised patients, further deteriorating their clinical outcomes and stressing the urgent need for robust infection control measures, surveillance, and the development of novel therapeutic interventions [25,27].

The substantial mortality rates associated with CRE, particularly in bloodstream infections, underscore the need for effective therapeutic interventions. In response, newer combination therapies, such as  $\beta$ -lactam/ $\beta$ -lactamase inhibitors, have emerged as promising treatment options, demonstrating improved survival rates in clinical studies [25]. However, the growing prevalence of CRE and the phenomenon of heteroresistance, i.e., the presence of antibiotic-resistant subpopulations within a seemingly sensitive population, which may contribute to treatment failure, together form a critical challenge in clinical settings, reinforcing the need for an aligned international response to control the spread of these resistant pathogens [29].

## 3. Epidemiological Methods

### 3.1. Introduction to Epidemiology

Epidemiology is the study of the distribution of diseases and health events within a population, focusing on the factors that influence their prevalence and distribution. This field is essential for informing public health interventions, policy development, and clinical

practice aimed at reducing the burden of disease and improving health outcomes. Over time, knowledge gained from epidemiological studies has contributed greatly to fighting infectious diseases, treating chronic diseases, and extending the human lifespan [30].

The integration of advanced data analytics and computational tools in epidemiology has greatly expanded our ability to analyze health data and model disease dynamics. The use of advanced statistical methods, ML algorithms, and big data analytics has improved the identification of risk factors, outbreak monitoring, and evaluation of interventions [31,32]. These innovations will not only provide more accurate predictions, but also open up new research opportunities, potentially revolutionizing public health strategies and improving health outcomes around the world.

### 3.2. Traditional Epidemiological Approaches to Studying AMR

Traditional epidemiological methods have played a key role in understanding the dissemination and impact of AMR. Epidemiological approaches often involve observational studies, such as cohort, case-control, and cross-sectional studies, which help identify risk factors associated with the emergence and distribution of resistance. Surveillance systems, such as the WHO Global Antimicrobial Resistance Surveillance System (GLASS), collect and analyze data on AMR trends in different regions, relying on laboratory data to monitor resistance models [16]. These systems provide critical information to guide public health interventions and antibiotic stewardship programs.

Statistical modeling is another cornerstone of traditional epidemiological methods in AMR research. Regression models, for instance, are commonly used to examine the relationship between antibiotic exposure and the emergence of resistance, adjusting for potential confounding factors. While traditional approaches have provided valuable insights into the epidemiology of AMR, they often rely on certain assumptions and may struggle to capture the complexity and non-linearity of resistance dynamics [31]. Nonetheless, these methods remain fundamental for generating evidence that guides effective public health strategies to combat AMR.

### 3.3. Strengths and Limitations of Epidemiological Approaches in the Context of Rapidly Evolving Resistance Patterns

As in most branches of science, the traditional approach to epidemiology involves the use of frequentist-based statistical approaches that rely on hypothesis testing and computations of probability related to an association or treatment effect. In this area, classic regression models either predict the outcome variable based on a number of other variables or model the relation of individual variables to the outcome [30]. These models are, however, based on certain assumptions; for instance, linearity and the absence of multicollinearity, which present many challenges when the research questions become sophisticated and the amounts of data increase, the so-called “curse of dimensionality” [31].

Despite these challenges, traditional epidemiology has produced successful studies and standardized surveillance systems that have significantly advanced our understanding of AMR. For example, the European Antimicrobial Resistance Surveillance Network (EARS-Net) has been instrumental in tracking resistance trends across Europe, leading to informed public health interventions. The burden of antibiotic-resistant infections in the EU and European Economic Area (EEA) was assessed by another study focusing on cases, deaths, and disability-adjusted life years (DALYs) [33]. Using data from the EARS-Net, the researchers estimated 671,689 infections, with 63.5% linked to healthcare. These infections led to approximately 33,110 deaths and 874,541 DALYs [34]. Similarly, the Global Antimicrobial Resistance Surveillance System (GLASS), launched in 2015, has provided valuable data on AMR patterns worldwide, facilitating global comparisons and targeted responses [16].

However, traditional methods also have limitations, particularly in the context of rapidly evolving resistance patterns. These limitations include the following:

- Data lag. The time required to collect, process, and analyze data can result in delays, making it challenging to respond promptly to emerging resistance threats [35].
- Data completeness. Incomplete data collection and reporting can lead to gaps in understanding the full scope of AMR. Variability in laboratory capacities and surveillance systems across regions further complicates this issue [36].
- Complexity of AMR. AMR is influenced by a multitude of factors, including antibiotic usage, infection control practices, and genetic mechanisms. Traditional methods may struggle to account for these complex, multifactorial influences without integrating more advanced analytical techniques [37].
- Predictive limitations. Traditional epidemiological methods often focus on descriptive and retrospective analyses, which may not be sufficient for predicting future resistance trends or for real-time surveillance [38].

To overcome these challenges, the integration of machine learning and other advanced computational techniques with traditional epidemiological methods is increasingly being advocated. Machine learning can enhance the ability to analyze large and complex datasets, identify hidden patterns, and make real-time predictions, thereby complementing and extending the capabilities of traditional epidemiological approaches [39,40].

## 4. Machine Learning in Healthcare

### 4.1. Introduction to Machine Learning

Machine learning (ML) is a branch of artificial intelligence that involves training algorithms to recognize patterns in data and make predictions or decisions without explicit programming for each task. There are three main types of ML:

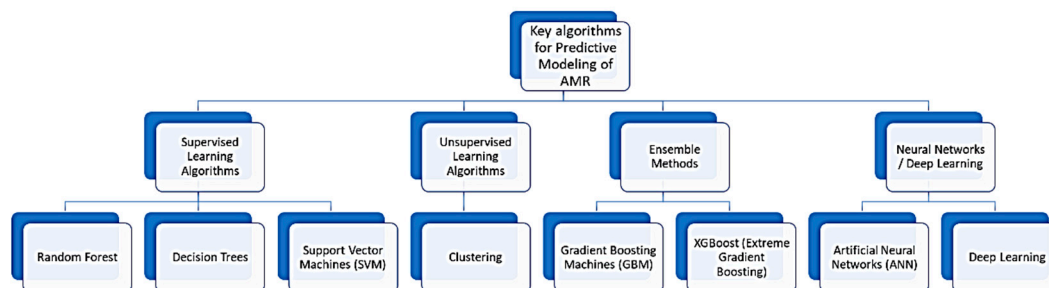
- Supervised learning, which involves training an algorithm on a labeled dataset, where the input–output pairs are known. The algorithm learns to map inputs to the correct output. Common algorithms include linear regression, decision trees, and support vector machines [41].
- Unsupervised learning, where the algorithm is trained on data without labeled responses and aims to find hidden patterns or intrinsic structures in the input data. Key techniques include clustering (e.g., k-means and hierarchical clustering) and association (e.g., Apriori algorithm) [42].
- Reinforcement learning, where the algorithm learns by interacting with an environment, receiving rewards or penalties based on the actions it takes. It aims to maximize cumulative rewards over time. Examples include Q-learning and deep reinforcement learning [43].

### 4.2. Key Algorithms and Applications

Machine learning algorithms vary in complexity and application. Some commonly used algorithms for building predictive models to forecast AMR trends, evaluate resistance risk, and provide decision support for treatment planning include the following:

- Linear regression, which is used for predicting a continuous target variable based on one or more predictor variables [44].
- Decision trees, with a flowchart-like structure, where each internal node represents a decision based on an attribute, and each leaf node represents an outcome [45].
- Support vector machine (SVM), which is a classification method that finds the hyperplane that best separates the data into classes [46].
- Neural networks and deep learning models are inspired by the human brain's structure, capable of learning complex patterns from large datasets, used extensively in image and speech recognition. In the context of deep learning, an artificial neural network with more than one hidden layer is referred to as deep learning, distinguishing it from simpler models with fewer layers [47].

A diagram that outlines some commonly used algorithms in predictive modeling of AMR is shown in Figure 1.



**Figure 1.** A diagram that outlines some commonly used algorithms in predictive modeling of AMR.

#### 4.3. Bridging Terminology: Aligning Epidemiology and Machine Learning Concepts

In the fields of epidemiology and ML, certain terminologies and concepts align closely, despite their usage in different contexts. In epidemiology and biostatistics, terms such as dependent variable, outcome variable, and response variable are used to refer to the variable that is being measured or predicted. In ML and statistical modeling, this concept corresponds to the label or class that the model aims to predict. Conversely, independent variables, predictor variables, and explanatory variables in epidemiology are equivalent to features in ML, which are the inputs or attributes used to predict the label [31].

A common tool in epidemiological studies is the contingency table or  $2 \times 2$  table, which displays the relationship between two categorical variables. In ML, this is referred to as a confusion matrix, which is used to evaluate the performance of classification models by showing the actual versus predicted classifications. Sensitivity in epidemiology, also known as recall in ML, measures the true-positive rate of a test. The positive predictive value in epidemiology, which assesses how many of the positive test results are true positives, is analogous to precision in ML [31].

When discussing outcome groups, the majority class in ML represents the outcome group with the highest frequency, while the minority class refers to the outcome group with the lowest frequency. In epidemiology, this concept is reflected in the proportion of cases in each category of the outcome variable, particularly when the outcome is categorical. This is similar to the concept of class balance in machine learning, which denotes the distribution of cases among different classes [31]. Understanding these terms and their equivalents across both fields improves clarity and facilitates effective communication when merging epidemiological insights in ML methods.

#### 4.4. Benefits of Machine Learning in Analyzing Complex Biological Data and Predicting Trends

Machine learning offers numerous benefits in healthcare, especially in analyzing complex biological data and predicting trends. One significant advantage is its ability to handle big data. ML algorithms can process and analyze vast amounts of information from various sources, such as electronic health records (EHRs), genomic data, and medical imaging [48]. Additionally, ML models excel in predictive analytics by identifying patterns and correlations in historical data, enabling the prediction of disease outbreaks, patient outcomes, and treatment responses [49].

Another key benefit is in personalized medicine, where ML helps tailor medical treatment to individual patients based on their genetic profile, lifestyle, and other factors, thereby improving treatment efficacy and reducing adverse effects [40]. Furthermore, advanced ML models significantly enhance diagnostics, offering high accuracy in diagnosing diseases from medical images, pathology slides, and other diagnostic tests [50,51].

### 5. Integration of Machine Learning and Epidemiology

Integrating ML with epidemiological data enhances the ability to predict and respond to resistance trends. This integration involves handling diverse data sources, developing predictive models, and conducting real-time surveillance.



### 5.1. Data Sources and Preprocessing Techniques

Data sources: Both machine learning and epidemiology rely on several data sources to offer good models and meaningful outputs. Some of the main categories these data sources fall under include the following:

- Genomic data. Genomic sequences, which comprise DNA or RNA of both pathogens and hosts, are helpful in the identification of genetic markers responsible for specific traits, such as drug resistance and virulence, which become indispensable for full comprehension of infectious disease mechanisms and epidemiology [52].
- Clinical data. Patients' electronic health records (EHRs) are a rich source of vital information, such as demographics, diagnosis, treatment, and results/outcomes. This dataset captures detailed patient histories that can be used to track disease progression and treatment responses [53].
- Environmental data. Environmental factors, such as air quality, water quality, and climatic variables, may affect dissemination of infectious diseases. Such information may even indicate changing environmental conditions and, thus, the impact on disease transmission [54].
- Sociodemographic data. Information on aspects such as the population's economic status, density, and education level is critical for understanding disease transmission within populations. More so, such elements can bring to light health-related disparities and susceptibilities [55].

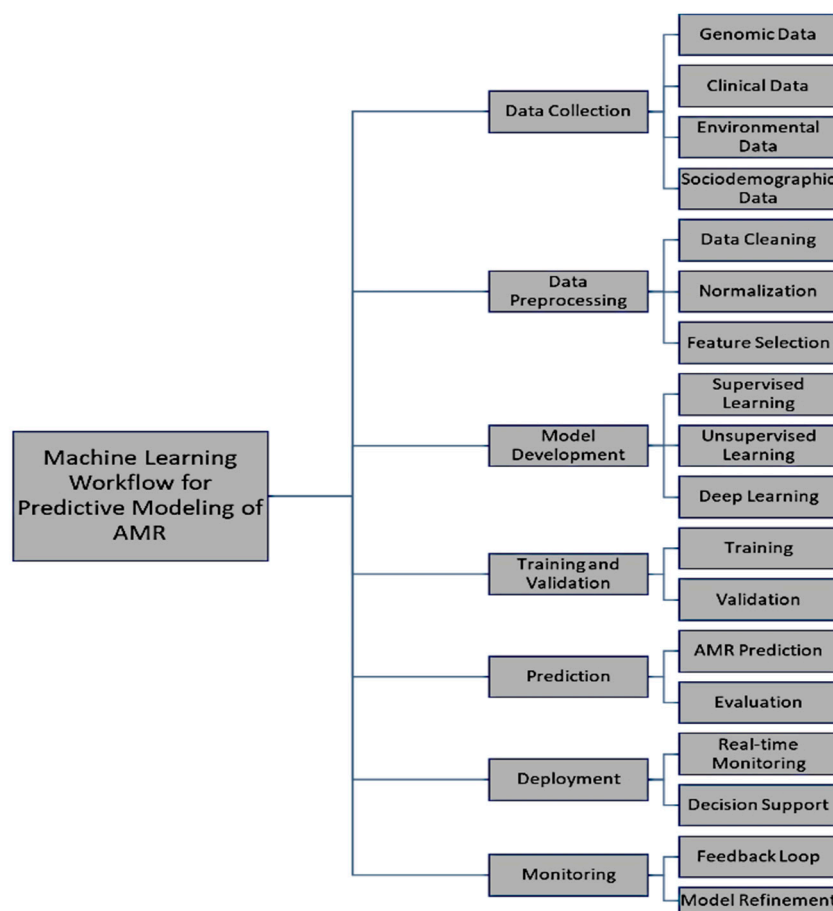
Preprocessing techniques for data preparation: To ensure the accuracy and usability of these diverse datasets, several preprocessing steps are critical, as follows:

- Data cleaning. Identification and correction of errors, inconsistencies, or incompleteness. Cleaning the data assures dependability and quality within the data and, hence, validation of ML models [56].
- Normalization. When several datasets are combined, normalization is necessary to standardize their scale. Certain algorithms are sensitive to the range of the data; thus, normalization ensures that no feature dominates the model because of a difference in its scale [57].
- Feature selection, which determines the most relevant variables. This, in turn, aids model performance by reducing dimensionality and weeding out insignificant or redundant data. It is totally focused on the most important part of the data and yields a higher performance with low computational complexity [58]. Figure 2 outlines the steps for a machine learning workflow for predictive modeling of AMR. This workflow illustrates how machine learning models handle diverse datasets for predicting AMR trends and providing clinical decision support.

Challenges and solutions in data integration: Integrating ML with epidemiological data allows for more precise modeling of disease dynamics and resistance patterns. For example, linking genomic data with clinical outcomes can reveal genetic determinants of drug resistance, while environmental and sociodemographic data provide a broader context within public health frameworks [59–61]. Genomic data, in particular, have become an important component of epidemiological analysis, especially in understanding the molecular mechanisms behind resistance and disease transmission. The integration with traditional epidemiological approaches has transformed how infectious diseases, such as carbapenem resistance, are studied [62].

However, combining these diverse data sources poses several challenges. A key issue is data format inconsistencies, as genomic, clinical, and environmental datasets often use different formats, complicating integration. Standardizing formats can help streamline the integration process and improve the efficiency of analysis.

Another challenge is data privacy and security, particularly when dealing with sensitive clinical and sociodemographic information. Additionally, missing data present a hurdle for comprehensive analysis. All these challenges and possible solutions are discussed in Section 7.



**Figure 2.** Machine learning workflow for predictive modeling of AMR.

### 5.2. Predictive Modeling

Development of predictive models: ML has become a powerful tool in the fight against carbapenem resistance by enabling the development of sophisticated predictive models. These models utilize various ML techniques to analyze large and complex datasets, aiming to forecast resistance patterns and guide effective interventions [63–65]. Supervised learning algorithms are commonly employed, where models are trained on historical data that include information on infection cases, antibiotic usage, and resistance outcomes [66]. This training helps the models identify patterns and correlations that indicate the likelihood of resistance [67,68].

Key features incorporated into these predictive models include patient demographics, clinical history, hospital environment, and the genetic characteristics of pathogens [69]. By considering these variables, ML models provide valuable insights into the emergence and spread of carbapenem-resistant infections. For example, models can predict how changes in antibiotic prescribing practices or hospital infection control measures might impact resistance trends [70]. This proactive approach allows healthcare providers to implement timely and targeted interventions to manage and mitigate the spread of resistance [71–73].

Evaluation metrics for model performance: The performance of predictive models is evaluated using various metrics to ensure their accuracy and reliability. Common evaluation metrics include the following:

- Accuracy. The proportion of true results (both true positives and true negatives) among the total number of cases examined. It indicates the overall correctness of the model [74].
- Precision. The proportion of true-positive results among all positive results predicted by the model. It measures the model's ability to correctly identify true resistance cases without including false positives [75].

- Recall (sensitivity). The proportion of true-positive results among all actual positive cases. It assesses the model's ability to detect true resistance cases [76].
- F1 score. The harmonic mean of precision and recall, providing a single metric that balances both. It is particularly useful when the data are imbalanced, meaning the number of positive cases is much smaller than the number of negative cases [77].
- Area under the receiver operating characteristic (ROC) curve (AUROC). A plot of the true-positive rate against the false-positive rate at various threshold settings. The AUROC provides a single measure of the model's ability to discriminate between positive and negative cases [78]. In the context of predicting antibiotic resistance, it measures how effectively the model can differentiate between cases where bacteria are resistant to an antibiotic and cases where they are not.

Case studies demonstrating successful predictions and early detection of resistance trends: Several case studies exemplify the successful application of ML in predicting and managing carbapenem resistance. Machine learning has proven highly effective in predicting BSIs and detecting antimicrobial resistance trends, particularly for carbapenem-resistant Gram-negative bacteria (CRGNB) in intensive care unit (ICU) patients. In a multicenter study from China, a random forest algorithm achieved an AUROC of 0.88 for CRGNB prediction and 0.86 for overall BSI prediction, enabling early intervention and targeted antibiotic therapies [63]. Similarly, a 2022 study demonstrated that ML algorithms, particularly random forest, could predict CRGNB carriage with 85.92% accuracy [79]. These models integrate variables, such as prior antibiotic use, mechanical ventilation, and invasive procedures, to provide real-time monitoring of resistance patterns. The early detection of CRGNB infections allows hospitals to optimize antimicrobial stewardship, reduce unnecessary use of broad-spectrum antibiotics, and better target high-risk patients, significantly improving clinical outcomes and infection control efforts.

Another study focused on developing and validating a ML-based algorithm to predict CR bacterial infections at the time of culture collection, achieving a sensitivity of 30%, a positive predictive value (PPV) of 30%, and a negative predictive value (NPV) of 99%, with *Pseudomonas* species accounting for 58% of the resistant infections. Integration of the model into the EHR system could enable real-time predictions, improving antibiotic stewardship by allowing early intervention and reducing unnecessary use of last-resort antibiotics [80].

Despite limitations, including reliance on single-center datasets, in most of the studies, these models showed promise for broader application, particularly in high-risk healthcare settings, as they can be easily retrained with additional data to reflect changing microbiological trends. This early detection is crucial for improving antimicrobial stewardship, reducing the unnecessary use of broad-spectrum antibiotics, and focusing treatments on high-risk patient groups. As these tools evolve, they promise to further optimize infection control strategies and enhance patient outcomes in hospital settings [81].

### 5.3. Epidemiological Insights

Enhancing traditional epidemiological analysis with machine learning: Machine learning (ML) significantly enhances traditional epidemiological analysis by providing advanced tools for data processing, pattern recognition, and predictive modeling [31,82]. Traditional epidemiology often relies on statistical methods that may not fully capture complex interactions within large datasets [31]. In contrast, ML algorithms can handle high-dimensional data, identify non-linear relationships, and uncover hidden patterns that are not apparent through conventional methods. This capability allows for more accurate risk assessments and targeted interventions.

Identifying risk factors and transmission patterns: ML models can integrate diverse data sources, such as genomic, clinical, environmental, and sociodemographic data, to identify risk factors and transmission patterns of AMR. For example, ML algorithms can analyze patient records to determine the factors associated with higher risks of infection with antibiotic-resistant pathogens. By mapping these factors, ML helps in understanding how resistance spreads within communities and healthcare settings [71,83–85].

**Real-time surveillance and outbreak prediction:** Real-time surveillance and outbreak prediction have become increasingly effective through the application of ML algorithms [86,87]. These algorithms excel at continuously analyzing incoming data to detect early signs of an outbreak, providing timely alerts to public health authorities and enabling them to take immediate action [88,89]. Predictive models can forecast the spread of antimicrobial resistance based on current trends, allowing for proactive measures to mitigate the impact of potential outbreaks. ML-driven surveillance systems, in particular, offer significant advantages in monitoring hospital data for unusual antibiotic resistance patterns, facilitating rapid responses to emerging threats [90,91]. For instance, a study by Caglayan developed a predictive framework using ML to identify ICU patients at risk of colonization with multi-drug-resistant organisms, including CRE [92]. The analysis of 4670 ICU admissions showed that the best-performing model achieved 82% sensitivity and 83% specificity. Among the key risk factors identified were prior stays in long-term care facilities and recent isolation procedures. This tool can be instrumental for clinicians in implementing timely infection control measures for high-risk patients, ultimately improving patient outcomes and preventing the spread of resistant infections.

#### 5.4. Real-World Applications

**Examples of machine learning in hospitals and public health:** Machine learning has made notable strides in improving health outcomes by analyzing large-scale data to predict AMR trends. For instance, ML models have been applied to monitor and predict outbreaks of multidrug-resistant tuberculosis (MDR-TB). By integrating diverse data sources, such as patient records, radiomic features (such as cavitation), and sociodemographic information, ML enhances the ability to track and control the spread of MDR-TB more effectively than traditional methods [93,94]. This approach allows public health officials to allocate resources more efficiently, focusing on high-risk areas and implementing targeted interventions to curb resistance. Such predictive modeling empowers public health systems to act preemptively, potentially reducing the overall burden of drug-resistant infections.

**Success stories and lessons learned from integrating machine learning with epidemiology:** There are several success stories where ML integration with traditional epidemiological tools has led to significant improvements in patient care and infection control. For instance, in one large healthcare system, ML algorithms outperformed traditional scoring systems, such as the Modified Early Warning Score (MEWS), Sequential Organ Failure Assessment (SOFA), and Systemic Inflammatory Response Syndrome (SIRS), in predicting severe sepsis. The ML model used only patient age and six vital signs from electronic health records (EHRs) to enhance early sepsis detection, while also reducing alarm fatigue, a prevalent concern in patient safety [95]. This demonstrates the ability of ML to provide more precise alerts, improving both detection and clinical workflow.

Similarly, the InSight algorithm, developed at the University of California, San Francisco Medical Center, achieved impressive results in detecting sepsis and septic shock. It reached an area under the receiver operating characteristic curve (AUROC) of 0.92 for sepsis detection and 0.96 for predicting septic shock four hours before onset [96]. This ML model not only outperformed existing sepsis scoring systems but also proved to be robust in the face of missing data, adaptable across institutions through transfer learning, and generalizable to various clinical settings. Its strong performance underscores the potential of ML to drive improvements in early diagnosis and intervention, which are critical in conditions such as sepsis, where early treatment significantly improves outcomes.

Machine learning has also contributed to better antibiotic stewardship programs. One study demonstrated the effectiveness of the XGBoost algorithm (<https://xgboost.readthedocs.io/en/latest/index.html>, accessed on 19 August 2024) in predicting antibiotic resistance for three Gram-negative bacteria: *Escherichia coli*, *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa*. Using data from 15,695 hospital admissions in the UK, the ML model slightly outperformed clinicians in selecting appropriate antibiotics, achieving an AUROC of 0.70 [97]. Importantly, this approach could reduce the use of broad-spectrum

antibiotics by up to 40%, a key step in combating the development of further antibiotic resistance. Despite these promising results, the study called for further validation through prospective trials to ensure its effectiveness and acceptance in clinical practice.

Another notable example of ML-driven antibiotic stewardship is the successful reduction of extended-spectrum beta-lactamase (ESBL)-targeted therapies in hospital settings. An ML program identified patients at low risk for ESBL-producing pathogens, allowing for more targeted use of antibiotics instead of relying on broad empirical carbapenem use [72]. This precision-guided treatment approach not only helps preserve the efficacy of carbapenems but also reduces the risk of fostering additional resistance.

The lessons learned from these implementations emphasize several key points for integrating ML into healthcare. First, high-quality data are essential for accurate and reliable predictions. Second, interdisciplinary collaboration between data scientists, healthcare professionals, and public health officials is critical to ensure that ML models are not only computationally robust but also clinically relevant. Third, continuous evaluation and refinement of ML models are necessary to maintain their effectiveness, especially as healthcare environments and microbial landscapes evolve. Effective communication between all stakeholders ensures that the full potential of ML can be realized in improving patient outcomes and public health [98].

### 5.5. Case Studies

Specific instances where machine learning and epidemiology have been used to address carbapenem resistance: Several case studies highlight the successful integration of ML and epidemiological methods to combat carbapenem resistance. One notable example is a study conducted in a large urban hospital, where ML algorithms were used to predict the occurrence of carbapenem-resistant *Klebsiella pneumoniae* (CRKP) infections. The predictive model analyzed patient data, including demographics, medical history, and previous antibiotic use, to identify individuals at high risk of developing CRKP infections [99]. This approach allowed for early intervention and targeted infection control measures, significantly reducing the incidence of CRKP infections. Another case study involved the use of ML to analyze national surveillance data on CRIs in the United States [80]. A machine learning model was developed to predict CRIs using data from 68,472 patients. Built with extreme gradient boosting, the model achieved an AUC of 0.846, with a 99% negative predictive value. Despite moderate sensitivity, it effectively ruled out CR infections, aiding in early detection and intervention in healthcare settings.

Impact on patient outcomes and public health interventions: The integration of ML and epidemiology offers significant theoretical advantages, such as enhancing patient outcomes and public health interventions. In the hospital setting, predictive models could enable healthcare providers to implement timely and appropriate infection control measures, potentially reducing the spread of carbapenem-resistant pathogens and improving patient outcomes [92]. On a broader scale, ML-driven epidemiological studies have the potential to significantly influence public health policies and resource allocation. By accurately predicting areas at high risk for carbapenem resistance, it is postulated that public health authorities can more effectively prioritize interventions, such as enhanced surveillance, targeted education campaigns, and stricter antibiotic stewardship programs [100,101]. This targeted approach could lead to a notable reduction in the incidence of carbapenem-resistant infections and a subsequent improvement in overall public health outcomes [102]. A multinational cohort study by Giannella et al. developed a risk prediction model for CRE infections following liver transplantation [90]. The model identifies several risk factors, including prior antibiotic use, specific comorbidities, and healthcare exposure, providing clinicians with a valuable tool to predict CRE infections and implement preventive measures in high-risk patients. Freire et al. extended this work by proposing a predictive risk score for CRE colonization prior to liver transplantation, using clinical and epidemiological data [91]. This risk score could guide preventive measures, such as targeted antibiotic prophylaxis, to address the challenge of identifying CRE colonization in patients on the

waiting list. Public health authorities can use these findings to prioritize resources, focusing efforts on hospitals with higher incidences of CRE carriage among vulnerable populations, thereby reducing overall healthcare costs and improving patient safety.

The following table (Table 1) summarizes prediction models for carbapenem resistance, highlighting data sources, accuracy, and ML algorithms used. These studies illustrate how combining ML techniques with epidemiological data enhances early detection and intervention for CRIs.

**Table 1.** Summary of prediction models for carbapenem resistance: integration of machine learning and epidemiological data across studies.

No.	Author	Geographical Setting	Publication Year	Medical Setting	Data Source	ML Algorithms	Performance Evaluation	Bacterial Species
1	Timothy Sullivan [99]	United States (Single Center)	2018	Hospital setting	EHR data, <i>Klebsiella pneumoniae</i> bacteremia cases	Multiple logistic regression	AUROC: 0.731, Sensitivity: 73%, Specificity: 59%, PPV: 16%, NPV: 95%	<i>Klebsiella pneumoniae</i> (Carbapenem-resistant)
2	Ariane Khaledi [71]	Germany, Spain	2020	Clinical settings, multicenter	Whole genome sequencing (WGS), transcriptomic data, gene presence/absence, expression profiles	Machine Learning (unspecified classifiers)	Sensitivity: 0.8–0.9, Predictive values: >0.9	<i>Pseudomonas aeruginosa</i> (Carbapenem-resistant)
3	Ed Moran [97]	United Kingdom (Single Center)	2020	Hospital setting	Blood and urine cultures, demographics, microbiology and prescribing data	XGBoost	AUROC: 0.70, Point-scoring tools: AUROC 0.61 to 0.67, estimated reduction in broad-spectrum antibiotic use by 40%	<i>Escherichia coli</i> , <i>Klebsiella pneumoniae</i> , <i>Pseudomonas aeruginosa</i>
4	Ryan J. McGuire [80]	United States (Single Center)	2021	Tertiary-care academic medical center	Demographics, medications, vital signs, procedures, lab results, cultures	Extreme gradient boosting (XGBoost)	AUROC: 0.846, Sensitivity: 30%, PPV: 30%, NPV: 99%	Carbapenem-resistant bacteria
5	Maddalena Giannela [90]	Multinational	2021	Liver transplantation units (multicenter)	Demographics, clinical data, mechanical ventilation, acute renal injury, surgical reintervention	Multivariable logistic regression, Fine-Gray subdistribution hazard model	AUROC: 74.6 (derivation), AUROC: 73.9 (bootstrapped validation), Brier Index: 16.6	Carbapenem-resistant Enterobacteriaceae (CRE)
6	Qiqiang Liang [79]	China (Single Center)	2022	Intensive care unit (ICU)	Demographics, screening records, clinical data, vitals	Random forest, XGBoost, decision tree, logistic regression	AUROC: 0.91 (random forest), 0.89 (XGBoost, decision tree), 0.78 (logistic regression)	Carbapenem-resistant Gram-negative bacteria (CRGNB)
7	Maristela Pinheiro Freire [91]	Brazil, Italy	2022	Liver transplantation units (multicenter)	Antibiotic use, hepato-renal syndrome, CLIF-SOFA scores, cirrhosis complications	Machine learning (unspecified)	Sensitivity: 66%, Specificity: 83%, NPV: 97%	Carbapenem-resistant Enterobacterales (CRE)

Table 1. Cont.

No.	Author	Geographical Setting	Publication Year	Medical Setting	Data Source	ML Algorithms	Performance Evaluation	Bacterial Species
8	Çağlar Çağlayan [92]	United States (Single Center)	2022	Intensive care unit (ICU)	EHR, MDRO screening program, sociodemographic and clinical factors	Logistic regression (LR), random forest (RF), XGBoost	Sensitivity: VRE 80%, CRE 73%, MRSA 76%, MDRO 82%; Specificity: VRE 66%, CRE 77%, MRSA 59%, MDRO 83%	MRSA, VRE, Carbapenem-resistant Enterobacteriaceae (CRE)
9	Qiqiang Liang [63]	China (Single Center)	2024	Intensive care unit (ICU)	Demographics, mechanical ventilation, invasive catheterization, carbapenem use history	Random forest, XGBoost, SVM	AUROC: random forest 0.86, XGBoost (infection): 0.86, SVM: 0.88, RF (CRGNB): 0.87	Carbapenem-resistant Gram-negative bacteria (CRGNB)
10	Yun Li [65]	China/USA	2024	Intensive care unit (ICU)	Electronic health record data (PLAGH-ICU, MIMIC-IV)	Machine learning models	AUROC: 0.786 (PLAGH-ICU), 0.744 (MIMIC-IV)	Multidrug-resistant organisms (MDRO), including carbapenem-resistant species
11	Bing Liu [64]	China (Single Center)	2024	Multiple hospital settings	Whole-genome sequencing (WGS) data, metagenomic sequencing (MGS), genomic features	Machine learning (unspecified algorithms)	AUROC: 0.906 (IPM), 0.925 (MEM), PPV: 0.897 (IPM), 0.889 (MEM)	<i>Pseudomonas aeruginosa</i> (Carbapenem-resistant)

AUROC: area under the receiver operating characteristic curve; CLIF-SOFA: Chronic Liver Failure–Sequential Organ Failure Assessment; EHR: electronic health record; SVM: support vector machine; LR: logistic regression; RF: random forest; XGBoost: extreme gradient boosting; CRE: carbapenem-resistant Enterobacteriaceae; MRSA: methicillin-resistant *Staphylococcus aureus*; VRE: vancomycin-resistant Enterococci; MDRO: multidrug-resistant organisms; NPV: negative predictive value; PPV: positive predictive value; CRGNB: carbapenem-resistant Gram-negative bacteria; PLAGH-ICU: a Chinese hospital ICU dataset; MIMIC-IV: Medical Information Mart for Intensive Care IV; IPM: imipenem; MEM: meropenem; WGS: whole-genome sequencing; MGS: metagenomic sequencing.

## 6. Challenges and Future Directions

### 6.1. Technical and Ethical Challenges

**Data quality and completeness:** One of the primary technical challenges in integrating ML with epidemiological studies is ensuring the quality and completeness of the data. Addressing data quality is crucial, as accurate, comprehensive datasets are fundamental to reliable modeling and outcomes. Inconsistent data collection methods, missing values, and errors in data entry can lead to biased models and inaccurate conclusions. This challenge is amplified when integrating data from diverse sources, such as genomic, clinical, and environmental datasets, which often suffer from data format inconsistencies. Standardizing data collection protocols and formats—using systems such as HL7 for clinical data—can streamline integration and improve data quality [103]. Robust data cleaning, validation processes, and advanced imputation techniques, such as k-nearest neighbors (KNN) or multiple imputation by chained equations (MICE), are crucial for addressing these issues and ensuring the reliability of the datasets. [104].

**Interpretability of machine learning models:** ML models, especially deep learning algorithms, often deliver high accuracy but lack interpretability, which limits their adoption in clinical and epidemiological settings. Healthcare professionals need to trust and understand how decisions are made. This “black box” nature becomes a barrier, particularly when integrated with complex epidemiological data, such as genomic or environmental information. Developing interpretable models and techniques to explain predictions can help build trust among healthcare providers and epidemiologists, ensuring that decisions are transparent and justifiable [105].

Ethical considerations in data use and patient privacy: Ethical concerns about patient privacy and data security are frequently raised in relation to large datasets in ML. The integration of clinical and sociodemographic data further complicates these matters, as sensitive patient information must be securely stored and anonymized. Strict compliance with data protection regulations, such as the General Data Protection Regulation (GDPR) and Health Insurance Portability and Accountability Act (HIPAA), is essential, as are robust encryption techniques to safeguard data.

The GDPR sets strict rules for the processing of sensitive patient data, including medical records, and requires that personal data be anonymized or pseudonymized before transfer. This is essential to ensure patient confidentiality and prevent reidentification. However, these requirements can limit the availability of detailed data, which are essential for effective ML models in AMR monitoring. For example, while anonymization can protect individual privacy, it reduces the granularity of the data, which can impact the ability of models to detect specific AMR trends or patterns.

Additionally, the GDPR places limitations on the transfer of personal data outside the European Economic Area (EEA). In the case of AMR monitoring, which often requires international cooperation, these regulations may prevent data sharing with countries that do not have equivalent data protection laws. Ensuring compliance requires setting up complex legal frameworks, such as standard contractual clauses, which can slow down data exchange and collaboration efforts.

Furthermore, the integration of clinical, genomic, and sociodemographic data for AMR surveillance raises ethical issues regarding consent and transparency [106]. Under the GDPR, patients must explicitly consent to the use of their data and have the right to withdraw this consent. Balancing these ethical considerations with the need for large, diverse datasets to improve predictive modeling for AMR poses a significant challenge.

Addressing biases in data and algorithms: Bias in data and algorithms poses a significant challenge, leading to unfair or discriminatory outcomes. Biases may originate at various stages, from data collection to model training and deployment. For example, if a training dataset is not representative of the broader population, models may perform poorly for underrepresented groups, resulting in inequitable healthcare delivery. When integrating diverse datasets, such as genomic, clinical, environmental, and sociodemographic data, it is crucial to monitor biases continuously. Algorithms should be developed with fairness in mind, ensuring equitable outcomes across all population groups [107].

Model generalizability in diverse healthcare contexts: The generalizability of machine learning models remains a significant challenge, particularly in healthcare settings with varying resources, patient populations, and infrastructure [108]. Factors contributing to CR may differ widely across regions and environments, further complicating the application of predictive models across diverse contexts. The variability in EHRs and laboratory data from different institutions or countries adds to the complexity, as these data may not be easily reproducible or standardized. To address this, it is crucial to tailor models to specific local contexts and continuously refine them with locally sourced data to ensure that predictions remain accurate and effective. Even when algorithms are developed, their prospective validation in diverse clinical environments is essential to ensure generalizability and reliability across various healthcare systems [105].

Addressing the limitations and operational challenges of machine learning in clinical implementation: While the integration of ML into healthcare holds great promise for enhancing patient outcomes and streamlining public health initiatives, it is crucial to acknowledge the limitations and operational challenges that accompany its implementation. One significant barrier is the requirement for large, high-quality datasets, which can be difficult to obtain, especially in settings with limited data-sharing practices [106]. Furthermore, integrating ML into existing clinical workflows may encounter resistance from staff due to concerns about reliability and accountability, as well as the need for comprehensive training on new technologies. Addressing these challenges through targeted strategies and



ongoing evaluation will be essential for realizing the full potential of ML in improving healthcare delivery [107].

### 6.2. Future Directions

Potential advancements in machine learning algorithms and computational power: Advancements in ML algorithms and computational power are poised to significantly enhance the capabilities of epidemiological studies. The development of more sophisticated algorithms, such as deep learning models with improved architectures, can lead to better performance in detecting and predicting patterns of AMR [109]. Additionally, the increasing availability of high-performance computing resources, including cloud-based platforms and specialized hardware, such as GPUs and TPUs, allows for the processing of larger datasets and the training of more complex models [110].

Opportunities for integrating other emerging technologies (e.g., AI and big data analytics) with epidemiology: The integration of other emerging technologies with epidemiology presents numerous opportunities for advancing the field. Artificial intelligence (AI), big data analytics, and the Internet of Things (IoT) can provide valuable insights into the spread and control of AMR. For example, AI can enhance the analysis of genomic data to identify resistance genes, while big data analytics can uncover trends and correlations in vast datasets from various sources, including electronic health records and environmental sensors [111]. The IoT can facilitate real-time monitoring of environmental conditions and the spread of infectious diseases, enabling more timely and effective public health interventions [112].

Recommendations for policy and practice to maximize the impact of these interdisciplinary approaches: To maximize the impact of interdisciplinary approaches combining ML and epidemiology, several key recommendations should be implemented, as follows:

- Standardization of data collection. The establishment of standardized protocols for data collection and reporting is critical. This will enhance the quality and comparability of datasets, which are vital for the effectiveness of ML models in epidemiological studies [113].
- Investment in infrastructure. Governments and organizations should invest in the necessary infrastructure, including high-performance computing resources and secure data storage solutions, to support the integration of ML and epidemiological methods [114].
- Interdisciplinary collaboration. It is essential to foster collaboration among data scientists, epidemiologists, healthcare professionals, and policymakers. Such interdisciplinary partnerships can drive the creation of effective and practical solutions for tackling antimicrobial resistance (AMR) and other public health challenges [40].
- Ethical and regulatory frameworks. Developing comprehensive ethical and regulatory frameworks is crucial. These frameworks should address privacy concerns, data security, and the responsible use of AI and ML technologies, thereby ensuring public trust and the successful deployment of these approaches in real-world settings [115].

### 6.3. The Role of Interdisciplinary Collaboration in Advancing This Field

Interdisciplinary collaboration plays a crucial role in advancing the integration of ML and epidemiology. By bringing together experts from various fields, including computer science, public health, medicine, and social sciences, collaborative efforts can leverage diverse perspectives and expertise to tackle complex challenges associated with AMR. Such collaborations can lead to the development of innovative models, the identification of novel intervention strategies, and the creation of comprehensive public health policies that are informed by data-driven insights [116].

In essence, the integration of ML and epidemiology is not just about combining tools from different disciplines, but about creating a synergistic framework where each discipline informs and enhances the other. This collaborative approach can lead to breakthroughs in understanding and combating AMR, ultimately leading to more effective interventions,

better patient outcomes, and stronger public health systems. By leveraging the strengths of various fields, interdisciplinary collaboration can drive the innovation needed to address one of the most pressing global health challenges of our time.

## 7. Conclusions

The integration of ML and epidemiology has the potential to transform public health interventions by offering precise predictions and tailored approaches to infection control. ML-driven models help healthcare systems and public health authorities respond more effectively to the growing challenge of AMR, reducing the spread of resistant infections, improving patient outcomes, and optimizing the use of resources. This data-driven approach promises to significantly enhance both patient care and public health efforts.

Looking ahead, the future of combining ML and epidemiology holds great promise for combating AMR. As ML algorithms and computational power continue to advance, their application in healthcare will become increasingly sophisticated and impactful. The integration of emerging technologies, such as AI and big data analytics, with epidemiological methods will further enhance the ability to predict, monitor, and respond to resistance trends. This synergy will enable more efficient resource allocation, early detection of resistance patterns, and timely interventions, ultimately helping to curb the growing threat of antimicrobial resistance on a global scale.

Furthermore, the broader impact of this integration will foster a more proactive and informed approach to public health, enabling healthcare systems to implement targeted interventions and policies that not only address current challenges but also anticipate future outbreaks of AMR. By leveraging the power of ML in epidemiology, we can build a more resilient and responsive healthcare infrastructure that significantly improves global health outcomes.

**Author Contributions:** Conceptualization, A.S., G.F. and V.S.V.; methodology, A.S., C.K., D.K. and G.F.; software, P.K. and E.P.; validation, D.K., V.S.V., P.M. and G.F.; formal analysis, A.S.; investigation, P.K.; resources, C.K. and E.P.; data curation, A.S. and E.P.; writing—original draft preparation, A.S.; writing—review and editing, A.S., C.K., P.K. and E.P.; visualization, P.K.; supervision, E.P. and V.S.V.; project administration, G.F., P.M. and V.S.V.; funding acquisition, Not applicable. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We acknowledge the use of the AI tool Grammarly for language refinement and coherence in the preparation of this manuscript. The intellectual content, writing, analysis, and interpretations remain entirely the work of the authors.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *Lancet* **2022**, *399*, 629–655. [CrossRef] [PubMed]
2. O'Neill, J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*; Government of the United Kingdom and Wellcome Trust: London, UK, 2016. Available online: <https://wellcomecollection.org/works/thvwsuba> (accessed on 8 April 2021).
3. Walsh, T.R.; Gales, A.C.; Laxminarayan, R.; Dodd, P.C. Antimicrobial resistance: Addressing a global threat to humanity. *PLoS Med.* **2023**, *20*, e1004264. [CrossRef] [PubMed]
4. World Health Organization; Food and Agriculture Organization of the United Nations; United Nations Environment Programme; World Organisation for Animal Health. *A One Health Priority Research Agenda for Antimicrobial Resistance*; WHO: Geneva, Switzerland, 2023. Available online: <https://www.who.int/publications/i/item/9789240075924> (accessed on 2 August 2024).

5. Anderson, M.; Panteli, D.; Mossialos, E. *Strengthening the EU Response to Prevention and Control of Antimicrobial Resistance (AMR): Policy Priorities for Effective Implementation*; WHO Regional Office for Europe: Geneva, Switzerland, 2024.
6. Laxminarayan, R.; Impalli, I.; Rangarajan, R.; Cohn, J.; Ramjeet, K.; Trainor, B.W.; Strathdee, S.; Sumpradit, N.; Berman, D.; Wertheim, H.; et al. Expanding antibiotic, vaccine, and diagnostics development and access to tackle antimicrobial resistance. *Lancet* **2024**, *403*, 2534–2550. [[CrossRef](#)] [[PubMed](#)]
7. Nordmann, P.; Naas, T.; Poirel, L. Global spread of carbapenemase-producing Enterobacteriaceae. *Emerging Infectious Diseases* **2011**, *17*, 1791–1798. [[CrossRef](#)] [[PubMed](#)]
8. Queenan, A.M.; Bush, K. Carbapenemases: The versatile beta-lactamases. *Clin. Microbiol. Rev.* **2007**, *20*, 440–458. [[CrossRef](#)]
9. Poirel, L.; Pitout, J.D.; Nordmann, P. Carbapenemases: Molecular diversity and clinical consequences. *Future Microbiol.* **2007**, *2*, 501–512. [[CrossRef](#)]
10. Piddock, L.J.V. Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria. *Clin. Microbiol. Rev.* **2006**, *19*, 382–402. [[CrossRef](#)]
11. Fernández, L.; Hancock, R.E. Adaptive and mutational resistance: Role of porins and efflux pumps in drug resistance. *Clin. Microbiol. Rev.* **2012**, *25*, 661–681. [[CrossRef](#)]
12. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 593–656. [[CrossRef](#)]
13. Michaelis, C.; Grohmann, E. Horizontal Gene Transfer of Antibiotic Resistance Genes in Biofilms. *Antibiotics* **2023**, *12*, 328. [[CrossRef](#)]
14. Mó, I.; da Silva, G.J. Tackling Carbapenem Resistance and the Imperative for One Health Strategies—Insights from the Portuguese Perspective. *Antibiotics* **2024**, *13*, 557. [[CrossRef](#)] [[PubMed](#)]
15. Logan, L.K.; Weinstein, R.A. The epidemiology of carbapenem-resistant Enterobacteriaceae: The impact and evolution of a global menace. *J. Infect. Dis.* **2017**, *215* (Suppl. S1), S28–S36. [[CrossRef](#)] [[PubMed](#)]
16. World Health Organization. *Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report: 2022*; World Health Organization: Geneva, Switzerland, 2022; Available online: <https://www.who.int/publications/i/item/9789240062702> (accessed on 2 August 2024).
17. van Duin, D.; Doi, Y. The global epidemiology of carbapenemase-producing Enterobacteriaceae. *Virulence* **2017**, *8*, 460–469. [[CrossRef](#)] [[PubMed](#)]
18. Barmpouni, M.; Gordon, J.P.; Miller, R.L.; Dennis, J.W.; Grammelis, V.; Rousakis, A.; Souliotis, K.; Poulakou, G.; Daikos, G.L.; Al-Taie, A. Clinical and Economic Value of Reducing Antimicrobial Resistance in the Management of Hospital-Acquired Infections with Limited Treatment Options in Greece. *Infect. Dis. Ther.* **2023**, *12*, 1891–1905. [[CrossRef](#)] [[PubMed](#)]
19. Ma, J.; Song, X.; Li, M.; Yu, Z.; Cheng, W.; Yu, Z.; Zhang, W.; Zhang, Y.; Shen, A.; Sun, H.; et al. Global spread of carbapenem-resistant Enterobacteriaceae: Epidemiological features, resistance mechanisms, detection and therapy. *Microbiol. Res.* **2023**, *266*, 127249. [[CrossRef](#)]
20. Centers for Disease Control and Prevention (CDC). *Antibiotic Resistance Threats in the United States, 2019*; U.S. Department of Health and Human Services: Washington, DC, USA, 2019. Available online: <https://www.cdc.gov/antimicrobial-resistance/data-research/threats/index.html> (accessed on 19 August 2024).
21. Zhang, S.; Di, L.; Qi, Y.; Qian, X.; Wang, S. Treatment of infections caused by carbapenem-resistant *Acinetobacter baumannii*. *Front. Cell. Infect. Microbiol.* **2024**, *14*, 1395260. [[CrossRef](#)]
22. Tenover, F.C.; Nicolau, D.P.; Gill, C.M. Carbapenemase-producing *Pseudomonas aeruginosa*—An emerging challenge. *Emerg. Microbes Infect.* **2022**, *11*, 811–814. [[CrossRef](#)]
23. Cai, B.; Echols, R.; Magee, G.; Arjona Ferreira, J.C.; Morgan, G.; Ariyasu, M.; Sawada, T.; Nagata, T.D. Prevalence of carbapenem-resistant gram-negative infections in the United States predominated by *Acinetobacter baumannii* and *Pseudomonas aeruginosa*. *Open Forum Infect. Dis.* **2017**, *4*, ofx176. [[CrossRef](#)]
24. Dossouvi, K.M.; Ametepe, A.S. Carbapenem Resistance in Animal-Environment-Food from Africa: A Systematic Review, Recommendations and Perspectives. *Infect. Drug Resist.* **2024**, *17*, 1699–1728. [[CrossRef](#)]
25. Tumbarello, M.; Trecarichi, E.M.; De Rosa, F.G.; Giannella, M.; Giacobbe, D.R.; Bassetti, M.; Losito, A.R.; Bartoletti, M.; Del Bono, V.; Corcione, S.; et al. Infections caused by KPC-producing *Klebsiella pneumoniae*: Differences in therapy and mortality in a multicentre study. *J. Antimicrob. Chemother.* **2015**, *70*, 2133–2143. [[CrossRef](#)]
26. Falagas, M.E.; Tansarli, G.S.; Karageorgopoulos, D.E.; Vardakas, K.Z. Deaths attributable to carbapenem-resistant Enterobacteriaceae infections. *Emerg. Infect. Dis.* **2014**, *20*, 1170–1175. [[CrossRef](#)] [[PubMed](#)]
27. Bonomo, R.A.; Burd, E.M.; Conly, J.; Limbago, B.M.; Poirel, L.; A Segre, J.; Westblade, L.F. Carbapenemase-producing organisms: A global scourge. *Clin. Infect. Dis.* **2018**, *66*, 1290–1297. [[CrossRef](#)] [[PubMed](#)]
28. Martin, M.J.; Corey, B.W.; Sannio, F.; Hall, L.R.; MacDonald, U.; Jones, B.T.; Mills, E.G.; Harless, C.; Stam, J.; Maybank, R.; et al. Anatomy of an extensively drug-resistant *Klebsiella pneumoniae* outbreak in Tuscany, Italy. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2110227118. [[CrossRef](#)] [[PubMed](#)]
29. Lin, C.K.; Page, A.; Lohsen, S.; Haider, A.A.; Waggoner, J.; Smith, G.; Babiker, A.; Jacob, J.T.; Howard-Anderson, J.; Satola, S.W. Rates of resistance and heteroresistance to newer  $\beta$ -lactam/ $\beta$ -lactamase inhibitors for carbapenem-resistant Enterobacterales. *JAC-Antimicrob. Resist.* **2024**, *6*, dlac048. [[CrossRef](#)]
30. Epidemiology is a science of high importance. *Nat. Commun.* **2018**, *9*, 1703. [[CrossRef](#)]

31. Wiemken, T.L.; Kelley, R.R. Machine learning in epidemiology and health outcomes research. *Annu. Rev. Public Health* **2020**, *41*, 21–36. [CrossRef]
32. Saqib, M.; Iftikhar, M.; Neha, F.; Karishma, F.; Mumtaz, H. Artificial intelligence in critical illness and its impact on patient care: A comprehensive review. *Front. Med.* **2023**, *10*, 1176192. [CrossRef]
33. European Centre for Disease Prevention and Control. *Antimicrobial Resistance in the EU/EEA (EARS-Net)—Annual Epidemiological Report 2022*; ECDC: Stockholm, Switzerland, 2023. Available online: <https://www.ecdc.europa.eu/en/publications-data/surveillance-antimicrobial-resistance-europe-2022> (accessed on 16 August 2024).
34. Cassini, A.; Högberg, L.D.; Plachouras, D.; Quattrocchi, A.; Hoxha, A.; Simonsen, G.S.; Colomb-Cotinat, M.; Kretzschmar, M.E.; Devleeschauwer, B.; Cecchini, M.; et al. 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 modeling analysis. *Lancet Infect. Dis.* **2019**, *19*, 56–66. [CrossRef]
35. Ayobami, O.; Willrich, N.; Suwono, B.; Eckmanns, T.; Markwart, R. The epidemiology of carbapenem-non-susceptible *Acinetobacter* species in Europe: Analysis of EARS-Net data from 2013 to 2017. *Antimicrob. Resist. Infect. Control* **2020**, *9*, 89. [CrossRef]
36. Musa, K.; Okoliegbe, I.; Abdalaziz, T.; Aboushady, A.T.; Stelling, J.; Gould, I.M. Laboratory surveillance, quality management, and its role in addressing antimicrobial resistance in Africa: A narrative review. *Antibiotics* **2023**, *12*, 1313. [CrossRef]
37. Holmes, A.H.; Moore, L.S.; Sundsfjord, A.; Steinbakk, M.; Regmi, S.; Karkey, A.; Guerin, P.J.; Piddock, L.J.V. Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* **2016**, *387*, 176–187. [CrossRef] [PubMed]
38. Wall, S. Prevention of antibiotic resistance—An epidemiological scoping review to identify research categories and knowledge gaps. *Glob. Health Action* **2019**, *12*, 1756191. [CrossRef] [PubMed]
39. Leek, J.T.; Peng, R.D. Statistics: P values are just the tip of the iceberg. *Nature* **2015**, *520*, 612. [CrossRef] [PubMed]
40. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [CrossRef] [PubMed]
41. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.
42. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2009.
43. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
44. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*; Springer: Berlin/Heidelberg, Germany, 2013.
45. Breiman, L.; Friedman, J.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*, 1st ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 1984. [CrossRef]
46. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
47. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
48. Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358. [CrossRef]
49. Obermeyer, Z.; Emanuel, E.J. Predicting the future: Big data, machine learning, and clinical medicine. *N. Engl. J. Med.* **2016**, *375*, 1216–1219. [CrossRef]
50. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [CrossRef]
51. Sakagianni, A.; Feretzakis, G.; Kalles, D.; Koufopoulou, C.; Kaldis, V. Setting up an Easy-to-Use Machine Learning Pipeline for Medical Decision Support: A Case Study for COVID-19 Diagnosis Based on Deep Learning with CT Scans. *Stud. Health Technol. Inform.* **2020**, *272*, 13–16. [CrossRef]
52. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [CrossRef] [PubMed]
53. Murphy, S.N.; Weber, G.; Mendis, M.; Gainer, V.; Chueh, H.C.; Churchill, S.; Kohane, I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J. Am. Med. Assoc.* **2010**, *303*, 124–130. [CrossRef] [PubMed]
54. Brown, M.A.; Southworth, F.; Sarzynski, A. The geography of metropolitan carbon footprints. *Policy Soc.* **2009**, *27*, 285–304. [CrossRef]
55. Marmot, M. Social determinants of health inequalities. *Lancet* **2005**, *365*, 1099–1104. [CrossRef] [PubMed]
56. Rahm, E.; Do, H.H. Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.* **2000**, *23*, 3–13.
57. Jain, A.K.; Duin, R.P.W.; Mao, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 4–37. [CrossRef]
58. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182. [CrossRef]
59. Ali, T.; Ahmed, S.; Aslam, M. Artificial Intelligence for Antimicrobial Resistance Prediction: Challenges and Opportunities towards Practical Implementation. *Antibiotics* **2023**, *12*, 523. [CrossRef]
60. Bianconi, I.; Aschbacher, R.; Pagani, E. Current Uses and Future Perspectives of Genomic Technologies in Clinical Microbiology. *Antibiotics* **2023**, *12*, 1580. [CrossRef]
61. Robertson, A.J.; Mallett, A.J.; Stark, Z.; Sullivan, C. It is in our DNA: Bringing electronic health records and genomic data together for precision medicine. *JMIR Bioinform. Biotechnol.* **2024**, *5*, e55632. [CrossRef]

62. Armstrong, G.L.; MacCannell, D.R.; Taylor, J.; Carleton, H.A.; Neuhaus, E.B.; Bradbury, R.S.; Posey, J.E.; Gwinn, M. Pathogen Genomics in Public Health. *N. Engl. J. Med.* **2019**, *381*, 2569–2580. [[CrossRef](#)] [[PubMed](#)]
63. Liang, Q.; Ding, S.; Chen, J.; Chen, X.; Xu, Y.; Xu, Z.; Huang, M. Prediction of carbapenem-resistant gram-negative bacterial bloodstream infection in intensive care unit based on machine learning. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 123. [[CrossRef](#)] [[PubMed](#)]
64. Liu, B.; Gao, J.; Liu, X.F.; Rao, G.; Luo, J.; Han, P.; Hu, W.; Zhang, Z.; Zhao, Q.; Han, L.; et al. Direct prediction of carbapenem resistance in *Pseudomonas aeruginosa* by whole genome sequencing and metagenomic sequencing. *J. Clin. Microbiol.* **2023**, *61*, e0061723. [[CrossRef](#)] [[PubMed](#)]
65. Li, Y.; Cao, Y.; Wang, M.; Wang, L.; Wu, Y.; Fang, Y.; Zhao, Y.; Fan, Y.; Liu, X.; Liang, H.; et al. Development and validation of machine learning models to predict MDRO colonization or infection on ICU admission by using electronic health record data. *Antimicrob. Resist. Infect. Control* **2024**, *13*, 74. [[CrossRef](#)] [[PubMed](#)]
66. Feretzakis, G.; Loupelis, E.; Sakagianni, A.; Kalles, D.; Martsoukou, M.; Lada, M.; Skarmoutsou, N.; Christopoulos, C.; Valakis, K.; Velentza, A.; et al. Using Machine Learning Techniques to Aid Empirical Antibiotic Therapy Decisions in the Intensive Care Unit of a General Hospital in Greece. *Antibiotics* **2020**, *9*, 50. [[CrossRef](#)] [[PubMed](#)]
67. Feretzakis, G.; Sakagianni, A.; Loupelis, E.; Kalles, D.; Skarmoutsou, N.; Martsoukou, M.; Christopoulos, C.; Lada, M.; Petropoulou, S.; Velentza, A.; et al. Machine Learning for Antibiotic Resistance Prediction: A Prototype Using Off-the-Shelf Techniques and Entry-Level Data to Guide Empiric Antimicrobial Therapy. *Healthc. Inform. Res.* **2021**, *27*, 214–221. [[CrossRef](#)]
68. Tang, R.; Luo, R.; Tang, S.; Song, H.; Chen, X. Machine learning in predicting antimicrobial resistance: A systematic review and meta-analysis. *Int. J. Antimicrob. Agents* **2022**, *60*, 106684. [[CrossRef](#)]
69. Kim, J.I.; Maguire, F.; Tsang, K.K.; Gouliouris, T.; Peacock, S.J.; McAllister, T.A.; Beiko, R.G. Machine learning for antimicrobial resistance prediction: Current practice, limitations, and clinical perspective. *Clin. Microbiol. Rev.* **2022**, *35*, e0017921. [[CrossRef](#)]
70. Amin, D.; Garzón-Orjuela, N.; Garcia Pereira, A.; Parveen, S.; Vornhagen, H.; Vellinga, A. Artificial Intelligence to Improve Antibiotic Prescribing: A Systematic Review. *Antibiotics* **2023**, *12*, 1293. [[CrossRef](#)]
71. Khaledi, A.; Weimann, A.; Schniederjans, M.; Asgari, E.; Kuo, T.; Oliver, A.; Cabot, G.; Kola, A.; Gastmeier, P.; Hogardt, M.; et al. Predicting antimicrobial resistance in *Pseudomonas aeruginosa* with machine learning-enabled molecular diagnostics. *EMBO Mol. Med.* **2020**, *12*, e10264. [[CrossRef](#)]
72. Ravkin, H.D.; Ravkin, R.M.; Rubin, E.; Neshler, L. Machine-learning-based risk assessment tool to rule out empirical use of ESBL-targeted therapy in endemic areas. *J. Hosp. Infect.* **2024**, *149*, 90–97. [[CrossRef](#)] [[PubMed](#)]
73. Sophonsri, A.; Lou, M.; Ny, P.; Minejima, E.; Nieberg, P.; Wong-Beringer, A. Machine learning to identify risk factors associated with the development of ventilated hospital-acquired pneumonia and mortality: Implications for antibiotic therapy selection. *Front. Med.* **2023**, *10*, 1268488. [[CrossRef](#)] [[PubMed](#)]
74. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63. [[CrossRef](#)]
75. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [[CrossRef](#)]
76. Davis, J.; Goadrich, M. The relationship between precision-recall and ROC curves. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 233–240. [[CrossRef](#)]
77. Goutte, C.; Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *Advances in Information Retrieval*; Losada, D.E., Fernández-Luna, J.M., Eds.; ECIR 2005; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2005; Volume 3408, pp. 345–359. [[CrossRef](#)]
78. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [[CrossRef](#)]
79. Liang, Q.; Zhao, Q.; Xu, X.; Zhou, Y.; Huang, M. Early prediction of carbapenem-resistant gram-negative bacterial carriage in intensive care units using machine learning. *J. Glob. Antimicrob. Resist.* **2022**, *29*, 225–231. [[CrossRef](#)]
80. McGuire, R.J.; Yu, S.C.; Payne, P.R.O.; Lai, A.M.; Vazquez-Guillamet, M.C.; Kollef, M.H.; Michelson, A.P. A Pragmatic Machine Learning Model To Predict Carbapenem Resistance. *Antimicrob. Agents Chemother.* **2021**, *65*, e0006321. [[CrossRef](#)]
81. Olawade, D.B.; Wada, O.J.; David-Olawade, A.C.; Kunonga, E.; Abaire, O.; Ling, J. Using artificial intelligence to improve public health: A narrative review. *Front. Public Health* **2023**, *11*, 1196397. [[CrossRef](#)]
82. Branda, F.; Scarpa, F. Implications of artificial intelligence in addressing antimicrobial resistance: Innovations, global challenges, and healthcare's future. *Antibiotics* **2024**, *13*, 502. [[CrossRef](#)]
83. Raghupathi, W.; Raghupathi, V. Big data analytics in healthcare: Promise and potential. *Health Inf. Sci. Syst.* **2014**, *2*, 3. [[CrossRef](#)]
84. Lepper, H.C.; Woolhouse, M.E.J.; van Bunnik, B.A.D. The role of the environment in dynamics of antibiotic resistance in humans and animals: A modelling study. *Antibiotics* **2022**, *11*, 1361. [[CrossRef](#)]
85. Li, L.G.; Yin, X.; Zhang, T. Tracking antibiotic resistance gene pollution from different sources using machine-learning classification. *Microbiome* **2018**, *6*, 93. [[CrossRef](#)]
86. Zhang, T.; Rabhi, F.; Chen, X.; Paik, H.Y.; MacIntyre, C.R. A machine learning-based universal outbreak risk prediction tool. *Comput. Biol. Med.* **2024**, *169*, 107876. [[CrossRef](#)]
87. Cho, G.; Park, J.R.; Choi, Y.; Ahn, H.; Lee, H. Detection of COVID-19 epidemic outbreak using machine learning. *Front. Public Health* **2023**, *11*, 1252357. [[CrossRef](#)]

88. Zeng, D.; Cao, Z.; Neill, D.B. Artificial intelligence-enabled public health surveillance—From local detection to global epidemic monitoring and control. *Artif. Intell. Med.* **2021**, *437–453*. [[CrossRef](#)]
89. MacIntyre, C.R.; Chen, X.; Kunasekaran, M.; Quigley, A.; Lim, S.; Stone, H.; Paik, H.Y.; Yao, L.; Heslop, D.; Wei, W.; et al. Artificial intelligence in public health: The potential of epidemic early warning systems. *J. Int. Med. Res.* **2023**, *51*, 3000605231159335. [[CrossRef](#)]
90. Giannella, M.; Freire, M.; Rinaldi, M.; Abdala, E.; Rubin, A.; Mularoni, A.; Gruttadauria, S.; Grossi, P.; Shbaklo, N.; Tandoi, F.; et al. Development of a risk prediction model for carbapenem-resistant Enterobacteriaceae infection after liver transplantation: A multinational cohort study. *Clin. Infect. Dis.* **2021**, *73*, e955–e966. [[CrossRef](#)]
91. Freire, M.P.; Rinaldi, M.; Terrabuio, D.R.B.; Furtado, M.; Pasquini, Z.; Bartoletti, M.; de Oliveira, T.A.; Nunes, N.N.; Lemos, G.T.; Maccaro, A.; et al. Prediction models for carbapenem-resistant Enterobacteriales carriage at liver transplantation: A multicenter retrospective study. *Transpl. Infect. Dis.* **2022**, *24*, e13920. [[CrossRef](#)]
92. Çağlayan, Ç.; Barnes, S.L.; Pineles, L.L.; Harris, A.D.; Klein, E.Y. A data-driven framework for identifying intensive care unit admissions colonized with multidrug-resistant organisms. *Front. Public Health* **2022**, *10*, 853757. [[CrossRef](#)]
93. Li, Y.; Wang, B.; Wen, L.; Li, H.; He, F.; Wu, J.; Gao, S.; Hou, D. Machine learning and radiomics for the prediction of multidrug resistance in cavitary pulmonary tuberculosis: A multicentre study. *Eur. Radiol.* **2023**, *33*, 391–400. [[CrossRef](#)]
94. Zhang, F.; Zhang, F.; Li, L.; Pang, Y. Clinical utilization of artificial intelligence in predicting therapeutic efficacy in pulmonary tuberculosis. *J. Infect. Public Health* **2024**, *17*, 632–641. [[CrossRef](#)]
95. Burdick, H.; Pino, E.; Gabel-Comeau, D.; Gu, C.; Roberts, J.; Le, S.; Slote, J.; Saber, N.; Pellegrini, E.; Green-Saxena, A.; et al. Validation of a machine learning algorithm for early severe sepsis prediction: A retrospective study predicting severe sepsis up to 48 h in advance using a diverse dataset from 461 US hospitals. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 276. [[CrossRef](#)]
96. Mao, Q.; Jay, M.; Hoffman, J.L.; Calvert, J.; Barton, C.; Shimabukuro, D.; Shieh, L.; Chettipally, U.; Fletcher, G.; Kerem, Y.; et al. Multicenter validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward, and ICU. *BMJ Open* **2018**, *8*, e017833. [[CrossRef](#)]
97. Moran, E.; Robinson, E.; Green, C.; Keeling, M.; Collyer, B. Towards personalized guidelines: Using machine-learning algorithms to guide antimicrobial selection. *J. Antimicrob. Chemother.* **2020**, *75*, 2677–2680. [[CrossRef](#)]
98. Beam, A.L.; Kohane, I.S. Big data and machine learning in health care. *JAMA* **2018**, *319*, 1317–1318. [[CrossRef](#)]
99. Sullivan, T.; Ichikawa, O.; Dudley, J.; Li, L.; Aberg, J. The rapid prediction of carbapenem resistance in patients with Klebsiella pneumoniae bacteremia using electronic medical record data. *Open Forum Infect. Dis.* **2018**, *5*, ofy091. [[CrossRef](#)]
100. Arzilli, G.; De Vita, E.; Pasquale, M.; Carloni, L.M.; Pellegrini, M.; Di Giacomo, M.; Esposito, E.; Porretta, A.D.; Rizzo, C. Innovative techniques for infection control and surveillance in hospital settings and long-term care facilities: A scoping review. *Antibiotics* **2024**, *13*, 77. [[CrossRef](#)]
101. Elbehiry, A.; Marzouk, E.; Abalkhail, A.; El-Garawany, Y.; Anagreyah, S.; Alnafea, Y.; Almuzaini, A.M.; Alwarhi, W.; Rawway, M.; Draz, A. The development of technology to prevent, diagnose, and manage antimicrobial resistance in healthcare-associated infections. *Vaccines* **2022**, *10*, 2100. [[CrossRef](#)]
102. OECD. *Embracing a One Health Framework to Fight Antimicrobial Resistance*; OECD Health Policy Studies: Paris, France, 2023. [[CrossRef](#)]
103. AlQudah, A.A.; Al-Emran, M.; Shaalan, K. Medical data integration using HL7 standards for patient’s early identification. *PLoS ONE* **2021**, *16*, e0262067. [[CrossRef](#)]
104. Kahn, M.G.; Brown, J.S.; Chun, A.T.; Davidson, B.N.; Meeker, D.; Ryan, P.B.; Schilling, L.M.; Weiskopf, N.G.; Williams, A.E.; Zozus, M.N. Transparent reporting of data quality in distributed data networks. *Egems (Gener. Evid. Methods Improv. Patient Outcomes)* **2015**, *3*, 1052. [[CrossRef](#)]
105. Ribeiro, M.T.; Singh, S.; Guestrin, C. Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’16), San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144. [[CrossRef](#)]
106. Mittelstadt, B.D.; Allo, P.; Taddeo, M.; Wachter, S.; Floridi, L. The ethics of algorithms: Mapping the debate. *Big Data Soc.* **2016**, *3*. [[CrossRef](#)]
107. Mehrabi, N.; Morstatter, F.; Saxena, N.A.; Lerman, K.; Galstyan, A.G. A survey on bias and fairness in machine learning. *ACM Comput. Surv. (CSUR)* **2019**, *54*, 1–35. [[CrossRef](#)]
108. Sakagianni, A.; Koufopoulou, C.; Feretzakis, G.; Kalles, D.; Verykios, V.S.; Myrianthefs, P.; Fildisis, G. Using machine learning to predict antimicrobial resistance—A literature review. *Antibiotics* **2023**, *12*, 452. [[CrossRef](#)]
109. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
110. Dean, J.; Ghemawat, S. MapReduce: Simplified data processing on large clusters. *Commun. ACM* **2019**, *51*, 107–113. [[CrossRef](#)]
111. Schadt, E.E.; Linderman, M.D.; Sorenson, J.; Lee, L.; Nolan, G.P. Computational solutions to large-scale data management and analysis. *Nat. Rev. Genet.* **2010**, *11*, 647–657. [[CrossRef](#)]
112. Osama, M.; Ateya, A.A.; Sayed, M.S.; Hammad, M.; Pławiak, P.; El-Latif, A.A.A.; Elsayed, R.A. Internet of medical things and healthcare 4.0: Trends, requirements, challenges, and research directions. *Sensors* **2023**, *23*, 7435. [[CrossRef](#)]
113. Gliklich, R.E.; Leavy, M.B.; Dreyer, N.A. (Eds.) *Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User’s Guide*, 3rd ed.; Addendum 2; Agency for Healthcare Research and Quality (US): Rockville, MD, USA, 2019. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK551879/> (accessed on 19 August 2024).

114. OECD. *Health at a Glance 2019: OECD Indicators*; OECD Publishing: Paris, France, 2019. [[CrossRef](#)]
115. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* **2018**, *28*, 689–707. [[CrossRef](#)]
116. Krause-Jüttler, G.; Weitz, J.; Bork, U. Interdisciplinary Collaborations in Digital Health Research: Mixed Methods Case Study. *JMIR Hum. Factors* **2022**, *9*, e36579. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.