

# Machine learning in Anti-microbial Resistance Prediction

## Abstract

Antimicrobial resistance (AMR) poses a significant and growing threat to global public health, rendering conventional treatment options ineffective against a widening spectrum of pathogens. In this context, machine learning, a subfield of artificial intelligence, has emerged as a promising tool for predicting and mitigating AMR. This abstract provides an overview of the application of machine learning techniques in the domain of AMR prediction. Antibacterial peptides (ABPs) are a family of peptides found in nature that play a crucial role in the innate immune systems of organisms. These ABPs offer several advantages over widely used antibiotics. As a result, they have recently received a lot of attention as potential replacements for currently available antibiotics. But it is expensive and time-consuming to identify ABPs from natural sources. Thus, wet lab researchers employ various tools to screen promising ABPs rapidly [10]

Machine learning leverages the power of computational algorithms and data-driven approaches to analyse and interpret vast datasets encompassing microbial genomics, clinical records, and epidemiological information. In the era of increasing antimicrobial resistance, the need for early identification and prompt treatment of multidrug-resistant infections is crucial for achieving favorable outcomes in critically ill patients. As traditional microbiological susceptibility testing requires at least 24 hours, automated machine learning (AutoML) techniques could be used as clinical decision support tools to predict antimicrobial resistance and select appropriate empirical antibiotic treatment. By discerning intricate patterns and relationships within this multidimensional data landscape, machine learning models can forecast the emergence and spread of antimicrobial resistance with unprecedented accuracy.

Supervised machine learning methods enable the development of predictive models by learning from labelled datasets, allowing the identification of genetic markers, mutations, and microbial traits associated with resistance. Unsupervised learning techniques, on the other hand, unveil hidden patterns and clusters within genomic and proteomic data, shedding light on novel resistance mechanisms. Reinforcement learning strategies facilitate the optimization of antimicrobial treatment regimens, aiming to minimize the risk of resistance development.

## **Topic Overview:**

Antimicrobial resistance (AMR) is a growing global health crisis that threatens our ability to effectively combat bacterial infections. The overuse and misuse of antibiotics have led to the emergence of antibiotic-resistant bacteria, rendering many conventional treatments ineffective. Addressing this crisis requires innovative approaches, and machine learning has emerged as a powerful tool in predicting and mitigating antimicrobial resistance.

## **Purpose of the Review:**

Machine learning models have the capacity to analyse vast and complex datasets, including genomic, clinical, and epidemiological information, to identify subtle patterns and associations that may not be apparent through traditional methods. Machine learning can aid in the early identification of high-risk patients or populations, allowing for targeted interventions to prevent the development and spread of resistant infections. ML-based models can help in tracking the spread of resistant strains and identifying emerging resistance hotspots. This information is valuable for public health agencies and can guide the allocation of resources for prevention and control.

## **Main Findings:**

- We found that Machine learning is being increasingly applied in various areas of medical domains, including Bacterial Resistance predictions, Classification of Micro-Organisms, and so forth.
- A variety of Ensemble Methods were utilized in these studies, with Random Forest. being the most used, followed by Support Vector Machines (SVM), and Naive Bayes.
- Other models, such as Logistic Regression, K-Nearest Neighbors (K-NN), and Clustering Algorithms, were also utilized.
- The data sources used in these studies included Scopus, WHO, and Surveys.

## **Implications and Conclusion:**

The results of the review showed that Artificial Intelligence and Machine Learning has the potential to improve the performance of Anti-Microbial Resistance. However, further research is needed to review provided a comprehensive overview of the current state of knowledge in the field of technological approach/tool for specific medical domain and offered new insights for researchers interested in this research area.

**Keywords:** AI, Infectious Diseases, Antimicrobial Resistance, Machine Learning, Drug Resistance, Neural Networks, Pathogens, Bacteria, Genome Analysis, Drug Sensitivity, Prediction Models, Antibiotics, Clinical Isolates, Genetic Markers, Epidemiology, Data Mining, Clustering, Algorithms, Diagnostics, Genomic Data.

## 1. Introduction

Bacterial antimicrobial resistance (AMR) is a global threat [1, 2], which resulted in an estimated 1.27 million deaths in 2019 [3]. One key strategy to tackle AMR is to optimise antimicrobial use and prolong current antimicrobials' therapeutic life. Clinical decision support systems (CDSSs) are software designed to provide information to healthcare professionals, patients, or other individuals in order to make informed clinical decisions. With the advent of artificial intelligence (AI) and the ever-increasing prevalence of electronic health records (EHRs), numerous CDSSs utilising machine learning (ML) trained on historical patient data have been developed to assist with managing infections (4). As highlighted by the World Health Organization (WHO), the emergence of drug-resistant pathogens poses a formidable challenge to healthcare systems worldwide (WHO, 2019). To address this critical issue, innovative approaches to AMR prediction and management are essential. In recent years, machine learning (ML) has emerged as a powerful tool in the fight against AMR, offering groundbreaking capabilities for the prediction of antimicrobial resistance (W.H.O.,2023).

Machine learning, a subset of artificial intelligence, has emerged as a formidable ally in our efforts to combat AMR. This discipline harnesses the power of algorithms and data-driven insights to unlock the secrets hidden within vast and complex datasets. By discerning patterns, relationships, and predictive indicators, machine learning empowers us to foresee and mitigate the emergence and spread of antimicrobial resistance.

This exploration delves into the intersection of machine learning and microbiology, elucidating how sophisticated computational methods can be employed to predict and understand AMR dynamics. As we navigate the intricate landscape of microbial genetics, clinical data, and epidemiological factors, machine learning stands as a beacon of hope, offering the potential to revolutionize our approach to antimicrobial stewardship.

This journey will delve into the diverse applications of machine learning in AMR prediction, from the identification of genetic markers associated with resistance to the optimization of antimicrobial treatment regimens. Furthermore, we will examine the role of advanced deep learning techniques in extracting hidden insights from multifaceted data sources, providing a comprehensive view of AMR's complexities.

Artificial intelligence (AI), through its ability to process data and information and turn it into insight and knowledge, facilitates data analysis that exceeds human mind capabilities and solves the problem of limited rational decision-making due to insufficient information and time constraints. AI, when properly designed, can also be free of behavioural constraints, including irrational deviations from guidelines, peer influence on hierarchical cultural norms, and fatigue. Algorithms can learn objectively and are often able to make more accurate predictions than those observed in everyday practice. In view of its potential promise in AMR, ML could greatly improve research efficiency, allowing scientists to focus on more complex scientific matters [7].

In this paper, we reviewed the existing literature on machine learning applications with respect to antimicrobial resistance prediction. This is a narrative review, where we discuss the applications of

ML methods in the field of AMR and their value as an adjunct tool in the antibiotic stewardship practice, mainly from the clinician's perspective. Recently, a systematic review of the literature with meta-analysis was published on the topic of ML-driven prediction of AMR. In the current review, we have updated the literature search up to December 2022, and we principally emphasize the clinical context of ML applications, addressing healthcare professionals not quite familiar with AI technologies [6].

## **Background on the Antimicrobial:**

### **What are Antimicrobials?**

Antimicrobials – including antibiotics, antivirals, antifungals and Antiparasitics – are medicines used to prevent and treat infections in humans, animals and plants[4].

### **What is antimicrobial resistance?**

Antimicrobial resistance (AMR) is the ability of a microorganism to resist the action of one or more antimicrobial agents. The consequences of AMR can be severe, and prompt treatment with effective antimicrobials is the most effective way of reducing the risk of poor outcome from serious infections. AMR is one of the biggest threats to public health today, both globally and in the WHO Globally leading to mounting health-care costs, treatment failure and death[4].

## **Shift in Healthcare Paradigm:**

Currently, AMR is principally diagnosed using two techniques in clinical microbiology. One is classical culture-based antimicrobial susceptibility testing (AST), and the other is whole-genome sequencing for antimicrobial susceptibility testing (WGS-AST) [Waddington]. Although the former approach is simpler and easier to use, it typically requires a day or more to produce the results, which significantly lengthens the empirical antibiotic regimen and raises the possibility of treatment failure due to ineffective therapy or the threat of antibiotic resistance caused by broad-spectrum antibiotics.

## **Advantages of the New Approach:**

The implementation of ML methods has substantially reduced the time of bacterial susceptibility profiling to less than three hours for the flow-cytometry AST method (FAST) [21] and only 30 min for the infrared (IR) spectrometry [22]. While these ML-assisted diagnostics can accelerate antimicrobial susceptibility testing, they require costly infrastructure and expert personnel to be carried out.

## **Role of Technology:**

Supervised machine learning algorithms are commonly used with linear and logistic regression, k-nearest neighbours (k-NN), support vector machine (SVM), decision tree (DT), random forest (RF), and gradient boosting machine (GBM), the most prevalent algorithms, followed by neural networks

and deep learning approaches. The area under receiver operating curve characteristic (AUROC) is the main performance metric used in ML-driven prediction models. Other performance metrics include accuracy, sensitivity, specificity, positive and negative predictive values, precision, recall, and F1 measure, all of which, however, can be derived from a confusion matrix, a simple tabular structure that essentially summarizes the hits and misses of a learning algorithm.

**Purpose of the Review:** The major drivers behind the occurrence and spread of AMR are the use of antimicrobial agents and transmission of antimicrobial-resistant microorganisms between humans, between animals, and between humans, animals and the environment. While antimicrobial use exerts ecological pressure on bacteria and contributes to the emergence and selection of AMR, poor infection prevention and control (IPC) practices favour the further spread of these bacteria. Prudent antimicrobial use and high standards of IPC in all health-care settings are therefore the cornerstones of an effective response to AMR. Our principal goal with this review was to dissect and amalgamate scholarly literature that employs Machine learning across varied applications of medical domain, with a quest to answer.

## 2. Methods

We carried out a literature exploration in well-known databases and embarked on a scoping review of the current literature concerning Machine learning in Anti-microbial Resistance Prediction.

### 2.1. Search Strategy

#### 2.1.1. Search Sources

For this review, Scopus were consulted. Predictive ML models typically learn the relationships between the peptide sequences and their biological functions, such as their antimicrobial activity, from medium to large datasets. The power of such models relies on several factors such as their input datasets, their independent features (descriptors) and the type of algorithms used (i.e., classical ML vs neural network architectures) as illustrated in Fig. 1. Choosing the right input dataset(s) is the first important step toward model development. Most input datasets are either from in-house experiments or publicly available databases such as APD3 [16] or DBAASP [17].

#### 2.1.2. Search Terms

After an exhaustive literature survey and interactions with domain experts, we finalized our search terminologies. The review was zeroed in on [specific criteria]; thus, the search string combined three keywords interlinked with AND: (Antimicrobial Resistance AND Machine Learning AND Genome Analysis AND Pathogens).

## **2.2. Search Eligibility Criteria**

Incorporated in this review are all studies that employ algorithms, including supervised and unsupervised learning. We did not confine our search to any methods, disease types, data sources, or clinical results. Furthermore, considerations such as age, gender, and ethnicity were unrestricted. Both peer-reviewed articles and scholarly contributions like book chapters and conference proceedings were considered, but only those in English.

## **2.3. Study Selection**

The cases fulfilling the following criteria were included in the final analysis: a). Cases with isolation of NTM from the culture of a sterile site (CSF, abscess, or tissue sample), compatible with the clinical–radiological syndrome (CNS disease) or b). Identification of NTM species by molecular methods (e.g., PCR or next-generation sequencing) with a compatible clinical–radiological syndrome. Disseminated infections with isolation of NTM from pulmonary samples were also included. Reports with culture positive NTM without evidence of clinical CNS disease were excluded. Both paediatric and adult cases were included. Data regarding clinical presentation, diagnostic modalities, treatment and outcomes were documented.

## **2.4. Data Extraction**

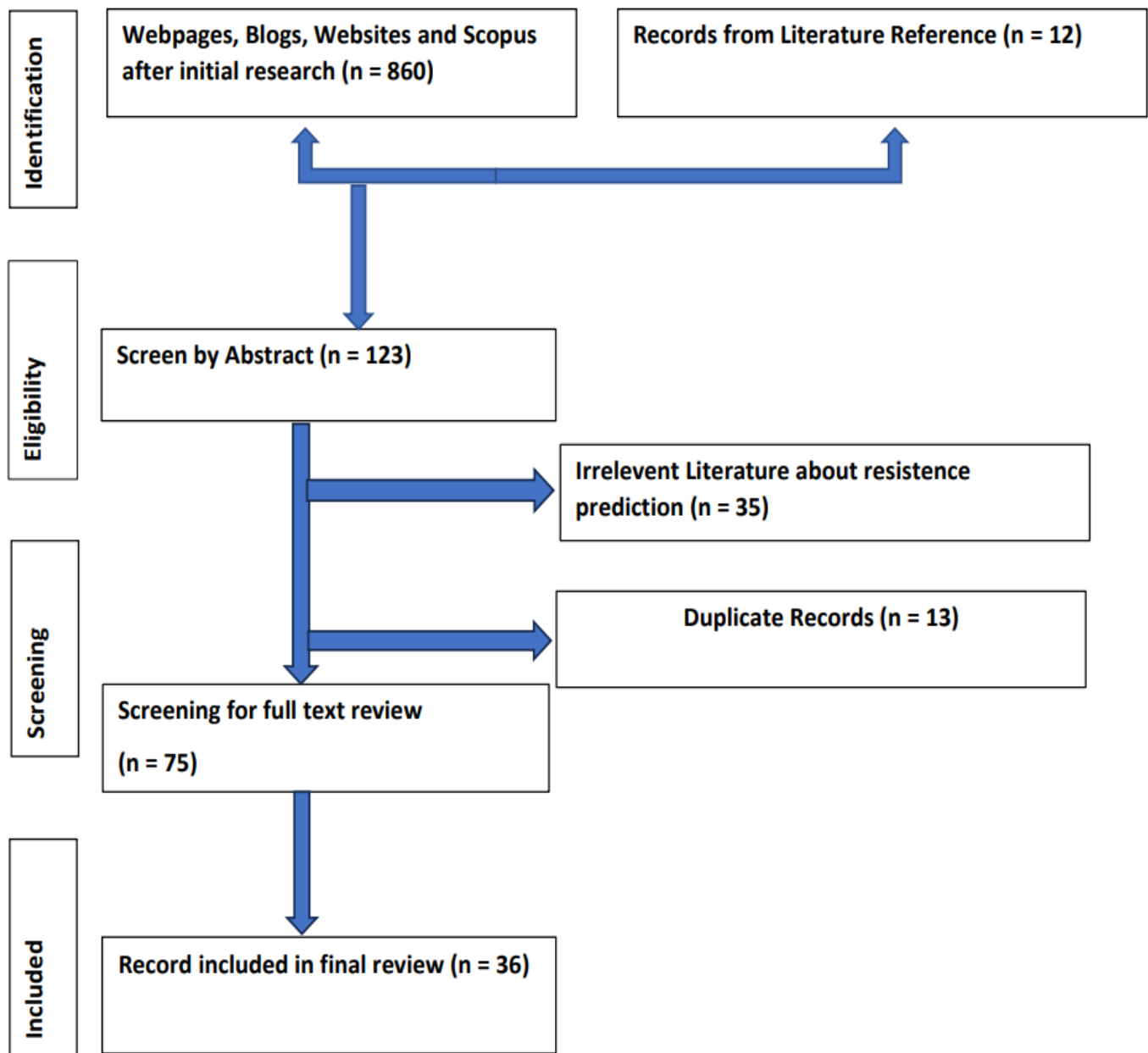
For streamlined and precise data mining, a standardized form was created, and pilot tested on a subset of Survey studies. Harvested data incorporated author details, publication year, institutional country, medical reports, and so on, with discrepancies addressed through dialogue and intervening when required.

The online software for systematic review (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia) was used to extract the data. The following data were extracted: clinical presentation, presence of risk factors, diagnostic procedures, antibiotic therapy and outcome with follow-up. The disparity between the authors was resolved by the discussion and consensus with the other reviewer authors (VM, VT, GKB). Case reports are associated with inherent bias; to decrease this, we adopted the standardized critical appraisal tool proposed by the Joanna Briggs Institute (JBI) [11].

## **3. Results**

### **3.1. Search Results :**

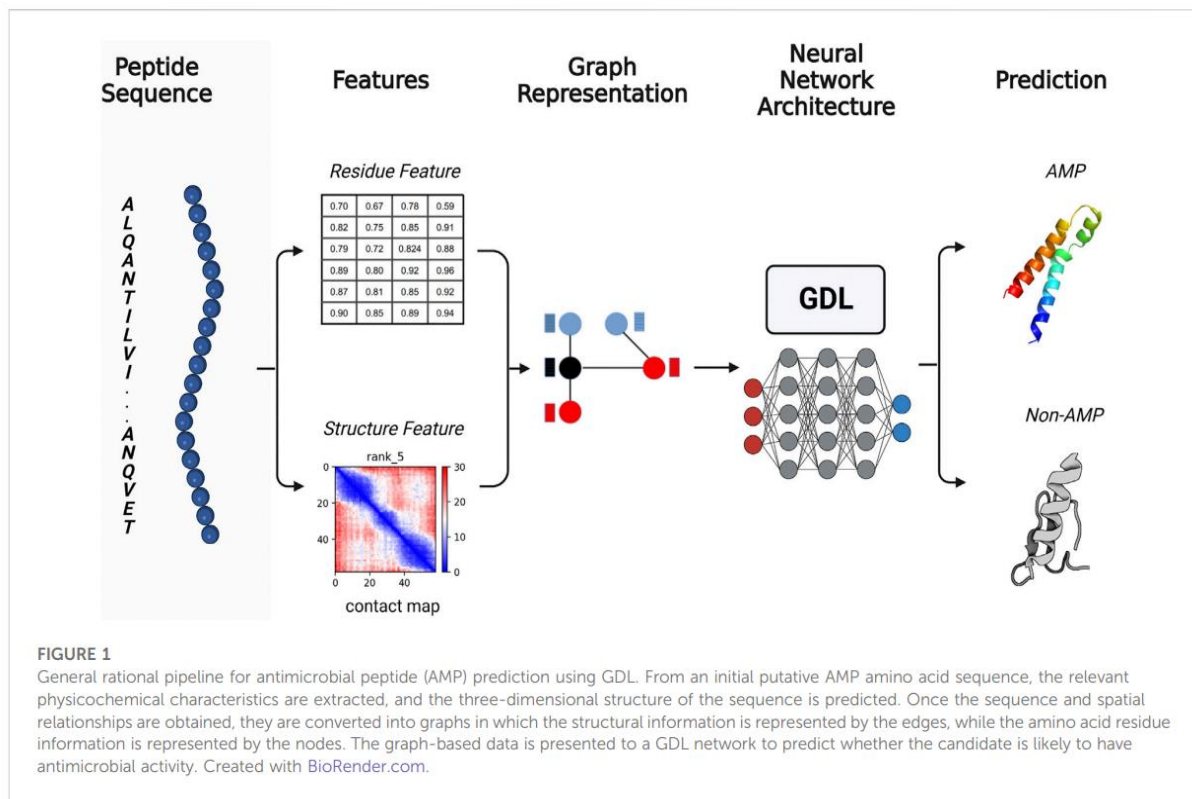
Our search initially identified studies, were retained after removing duplicates. Based on the inclusion criteria outlined in the Section, 198 studies were selected for full-text review. After further review, some studies were excluded, while some additional studies were identified by checking reference lists. In total, some studies met the inclusion criteria and were included in the data extraction and synthesis process. The study screening and selection process is summarized in Diagram.



### 3.2. Demographics of the Studies

Among deep learning methods leveraged for AMP predictions, the most widely used are convolutional neural networks (CNNs) [18]. The CNNs used in conventional deep learning assume that the data are related and organized as a regular grid, following the parameters of Euclidean geometry [18]. Nonetheless, the three-dimensional structure of peptides and proteins is better represented in a non-Euclidean space because its manifold data cannot be flattened without significant distortions. To implement deep learning for prediction in non-Euclidean systems, geometric deep learning emerges as a more efficient computational tool compared to several advanced and contemporary techniques.





### 3.4. Model development and software:

The model was applied on the MIMIC-IV EHR dataset, which was randomly split based on patients' "stay\_id" into training, validation, and testing sets (70%, 15%, and 15%, respectively). PyTorch [19] was used to create a bi-directional LSTM recurrent neural network (RNN) with a custom dataset class to extract labels and features. In order to address the mortality class imbalance, over-sampling was used during training. To be specific, those cases with positive mortality were replicated three times within the custom dataset class to achieve a more balanced mortality rate of 51.90% within the train dataset. The Adam optimiser [20] was used with binary cross-entropy loss for classification, mean squared error loss for regression, and Ray Tune for hyperparameter optimisation [21]. Training utilised 50 epochs, during which the model with the best performance on the validation dataset (RMSE or area under the receiver operating characteristic curve for LOS and mortality prediction, respectively) was selected as the final model. Two separate LSTM autoencoder models were trained on the whole training dataset to create embedding representations relevant to patients' LOS and mortality outcomes. Models were evaluated using functions and metrics from the Torch-Metrics, Scikit-learn, and SciPy libraries.



### **3.4.1. Synthetic Minority Oversampling Technique:**

SMOTE is a common oversampling method that was proposed to improve random oversampling, and its efficacy on high-dimensional data has been investigated in an earlier article [28]. After applying the a forementioned technique, the resulting dataset had 15,326 instances (an increase of about 33.3%), while the class attribute (antimicrobial susceptibility) contained 50.0% positive cases (sensitive to a specific antibiotic) and 50.0% negative cases (resistant to a specific antibiotic).

### **3.4.2. Understanding the drivers of AMR :**

The paradigm that the rapid expansion in AMR is a response to selective pressure from antimicrobial use drives AMR is well accepted [8]. WGS is being used to interrogate this selective evolution at the molecular level, across wide geographical areas, and between genetically diverse organisms. WGS was exploited to investigate AMR and the clinical epidemiology of atypical enteropathogenic *E. coli* isolated from symptomatic and asymptomatic children in South Asia and sub-Saharan Africa [22]. Despite broad geographical, symptomatic, and phylogenetic diversity, 65% of the bacterial isolates were resistant to three or more classes of antimicrobials. In this study, the best predictor of resistance profile was not the presence or absence of clinical symptoms or genetic lineage, but the geographical patterns of antimicrobial usage [22]

### **3.4.3. Random Forest Prediction of Phenotypes Resistance:**

Weka's datasets are used and stored in a unique file format known as attribute relation file format (ARFF). Due to the wide variety of file types used for biological data, it implements a format-conversion input layer that can transform common file types into the ARFF format. Weka filters any classes that can be applied to a dataset to alter it, and bio-Weka has filters for working with biological sequences. It enabled us to compare and match sequences with BLAST and other sequence alignment tools. In addition, alignment-based classification was performed using auto alignment score evaluation schemes.

A java-based machine learning algorithm called bio-Weka and RF was used to perform the predictive modelling. The DSK (k-mer counting software) [9] was used to generate K-mer profiles (abundance profiles of all unique words of length  $k$  in each genome) from the assembled contigs, with  $k = 31$ . This is a common length for analysing bacterial genomes [24]. In order to create the dataset, the 31-mer profiles of all strains were combined using the combine kmers tool in SEER [23].

### **3.4.4. Regression Statistics:**

Kappa statistics are reliable because they can be tested repeatedly [9, 25], ensuring that researchers have access to accurate, comprehensive data regarding research samples. It evaluates the predicted

classification accuracy against a random classification [25]. We used a kappa statistic that relies on binary values, where 0 is considered as a null value and 1 represents the predicted outcome of the evaluation as in equation [51]. It also serves as an indicator of the reliability of the evaluation. Not only that, but the LR variables help resolve the two-way binary classifications. When applied to the field of binary numbers, it makes predictions in the form of continuous values that allow for the preservation of sensitivity [26].

### 3.5. Machine Learning Datasets and Infection Science:

ML is a type of artificial intelligence (AI), described as a set of methods/algorithms that learn patterns from data, utilising these patterns to produce predictions (27). ML techniques have increasingly been applied in medicine and healthcare. Recently, research relating to the application of ML to IS has become more prominent, with ML technology having the potential to improve several aspects of IS including treatment, diagnosis, and management of patients, optimising and improving clinical laboratory workflows, and improving public health surveillance. Implications of applications in this area include improvement of patient outcomes, optimising for timely patient treatment, reduction in costs and unnecessary resource expenditure, and early identification and mitigation of emerging infectious disease to support readiness for potential public health emergencies [11]. The increase in ML applications for IS was made apparent during the height of the COVID-19 pandemic, where a significant body of research was dedicated to the application of ML for problems associated with COVID-19

Review title	Year	Focus area
Artificial intelligence and machine learning assisted drug delivery for effective treatment of infectious diseases (8)	2021	Drug delivery
Evolving Applications of Artificial Intelligence and Machine Learning in Infectious Diseases Testing (9)	2021	Infectious disease testing
Machine learning for clinical decision support in infectious diseases: a narrative review of current applications (10)	2020	Decision support systems
Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies (11)	2020	General infection management
Machine learning applications for COVID-19: a state-of-the-art review (12)	2021	COVID-19
Significant applications of machine learning for COVID-19 pandemic (13)	2020	COVID-19
Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review (14)	2020	COVID-19
Role of machine learning techniques to tackle the COVID-19 crisis: systematic review (15)	2021	COVID-19
Early prediction of sepsis in the ICU using machine learning: a systematic review (16)	2021	Sepsis in the ICU
Viral outbreaks detection and surveillance using wastewater-based epidemiology, viral air sampling, and machine learning techniques: a comprehensive review and outlook (17)	2022	Surveillance and epidemiology
Recent evolutions of machine learning applications in clinical laboratory medicine (18)	2021	Clinical laboratory
Machine learning in the clinical microbiology laboratory: has the time come for routine practice? (19)	2020	Clinical laboratory
Applications of machine learning in routine laboratory medicine: current state and future directions (20)	2022	Clinical laboratory
Use of artificial intelligence in infectious disease (21)	2020	Infectious disease: diagnosis, transmission, response to treatment, and resistance

**TABLE 2.** Recent literature reviews relating to the application of machine learning in infection science.

## 4. Discussion

Several studies have highlighted the increasing global prevalence of antimicrobial resistance [12–16]. This is related to the challenges of treating bacterial infections, the consequences of which can be severe. *P. aeruginosa* is one of the most common bacterial species, and its families are responsible for some of the most dangerous infections ever seen in humans. There is a correlation between the resistance of these bacteria to multiple antibiotic classes and the severity of the infection, which complicates treatment. Antibiotic resistance among these microorganisms has been rising steadily over the years, and it is now common to find clinical samples resistant to multiple drugs. The development of antibiotic resistance causes doctors to delay administering the most effective treatment methods and prescribe a larger dosage of antibiotics than is necessary.

Early detection of AMR remains challenging despite rapid diagnostic advances. A delay in diagnosis can prolong the period of ineffective antibiotic therapy. Statistical models for predicting drug resistance can play an important role, especially in settings where rapid diagnostic tests are unavailable or are difficult to perform due to a lack of resources.

## Principal Findings

Results across experiments were consistent, with stop control days often showing the greatest performance indicating our stop impact estimations, which occur on days where the true outcome upon stopping is unknown, are more reliable. The stop impact results from this retrospective study show that stopping antibiotics earlier can be associated with a statistically significant average LOS reduction of 2.71 days. Overall minimal impact on mortality was observed, which is to be expected given death can be caused by a large number of factors beyond those included as model features

Molecular AMR diagnostics significantly reduce the time to achieve results compared to classical phenotypic tests, but aside from the high setup costs, the need for technical infrastructure, and staff training in bioinformatics, their major drawback is that they can detect only known resistance genes or mutations [28]. ML techniques, such as whole-genome sequencing, seem to be useful in AMR surveillance projects; however, the existing evidence does not support their use in guiding clinical decision-making for most bacterial species, according to the European Committee on Antimicrobial Susceptibility Testing [28].

## **Practical and Research Implications**

The development of technically robust and safe ML systems is another area that has not been represented in the IS domain. Particularly, ensuring that they are resilient against adversarial examples. Understanding the role that the users and the various stakeholders have in the design and deployment of ML systems is an important area of future work to support the requirement for human agency and oversight. In the context of IS, this includes software engineers, clinical researchers, hospitals, clinical laboratories, the patients, regulatory agencies, and other domains of oversight. Furthermore, it is critical to consider additional user and human factor elements for developing trust in ML systems within the infection science context. For example, previous research has explored methods for user interface design to investigate user trust issues and the creation of trustworthy clinical decision support platforms (29).

## **Strengths and Limitations**

### **Technical limitations**

The concept of relatedness between bacterial isolates is based on a process of continuous evolution, but prospectively identifying genetic signals of AMR evolution from the background noise of genetic variation and sequencing error is difficult. Using sequence data to infer relatedness and transmission is highly variable, and mutation rates can vary because of different lifestyles/conditions such as biofilm formation, antimicrobial exposure, disease states, and environmental pressures such as starvation [30-31].

Additionally, the proposed methodology is consistent with the practices of patient cohering (placing patients who have been exposed to or infected with the same pathogen in the same inpatient room) or staff cohering (assigning specific healthcare providers to care only for patients/residents known to be colonized or infected), which constitute an effective surveillance measure for multidrug-resistant infections that may prevent inadvertent patient-to-patient dissemination.

Despite certain limitations of the study, our primary goal was to create an inexpensive ancillary tool to help clinicians rapidly identify patients carrying antibiotic-resistant strains and guide appropriate antibiotic treatment with greater confidence. In future work, dataset enhancement with clinical attributes will probably improve the AutoML algorithms' performance.

## **Conclusions**

In conclusion, the application of machine learning in predicting antimicrobial resistance represents a significant stride towards addressing one of the most pressing challenges in modern healthcare. Through the integration of advanced computational models and vast datasets, machine learning algorithms have demonstrated the capacity to discern patterns, identify risk factors, and forecast the emergence of antimicrobial resistance with remarkable accuracy. This chapter discusses the

importance of predictive and generative machine learning (ML) models in peptide drug design. Generative models produce artificial peptides with desired properties or functions, using algorithms such as RNNs, CNNs, VAEs, or GANs. Peptide drug design is a multi-objective problem, and the interplay between predictive and generative ML models provides a viable design strategy to address these challenges. This study evaluated the results of applying AutoML of the Microsoft Azure platform to two internal medicine departments' antimicrobial susceptibility datasets. In this article, we propose the use of AutoML as a decision tool for physicians since it can be more readily applied even by non experts (e.g., a data scientist may be needed for a full-blown investigation, but a physician can gain some insight with a relatively smooth learning curve) and, as we showed, the deduced models have good performance.

These applications have targeted a wide range of problems including the treatment and management of patients, the improvement of laboratory workflows, and predicting infection trends. However, there is a significant gap between the number of ML applications in research, and the effective implementation of ML systems. This paper has provided readers with an introduction to infection science, outlined the principles of trustworthy ML systems, presented an overview of the ML applications in IS that implement some of the principles of trustworthy AI, and identified the need for a change in focus towards implementing the practices of trustworthy ML systems to promote adoption of these systems in practice. As we move forward, the synergy between machine learning and antimicrobial resistance research holds the promise of revolutionizing our approach to infectious diseases. By harnessing the power of artificial intelligence, we can anticipate, adapt, and ultimately overcome the challenges posed by microbial evolution, ushering in a new era of precision medicine and global health security.

## References:

1. Nations U. Political declaration of the high level meeting of the general assembly on antimicrobial resistance: draft resolution/submitted by the president of the general assembly New York: UN (2016).
2. World Health Organization. Global action plan on antimicrobial resistance. World Health Organization (2015)
3. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Aguilar GR, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet*. (2022) 399:629–55. doi: 10.1016/S0140-6736(21)02724-0.
4. Levy, S.B. and Marshall, B. (2004) 'Antibacterial resistance worldwide: Causes, challenges and responses', *Nature Medicine*, 10(S12). doi:10.1038/nm1145.
5. Antimicrobial resistance [WWW Document], n.d. . World Health Organization. URL <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance/> (accessed 11.6.23).
6. Tang, R. *et al.* (2022) 'Machine learning in predicting antimicrobial resistance: A systematic review and meta-analysis', *International Journal of Antimicrobial Agents*, 60(5–6), p. 106684. doi:10.1016/j.ijantimicag.2022.106684.
7. Rajpurkar, P., Chen, E., Banerjee, O., Topol, E.J., 2022. Ai in Health and Medicine. *Nature Medicine* 28, 31–38.

8. Waddington, C.; Carey, M.E.; Boinett, C.J.; Higginson, E.; Veeraraghavan, B.; Baker, S. Exploiting genomics to mitigate the public health impact of antimicrobial resistance. *Genome Med.* 2022, 14, 15.
9. Saka Gianni, A. et al. (2023) Machine learning applications in antimicrobial resistance, *Encyclopaedia*. Available at: [https://encyclopedia.pub/entry/41695#ref\\_2](https://encyclopedia.pub/entry/41695#ref_2) (Accessed: 07 November 2023).
10. Sharma, R., Shrivastava, S., Singh, S.K., Kumar, A., Singh, A.K., Saxena, S., 2023. Artificial Intelligence-based model for predicting the minimum inhibitory concentration of antibacterial peptides against ESKAPEE pathogens. *IEEE Journal of Biomedical and Health Informatics* 1–10. (Sharma et al., 2023)
11. McFadden, B.R., Reynolds, M., Inglis, T.J., 2023. Developing machine learning systems worthy of trust for infection science: A requirement for future implementation into clinical practice. *Frontiers in Digital Health* 5.
12. Lluka, T., Stokes, J.M., 2022. Antibiotic discovery in the artificial intelligence era. *Annals of the New York Academy of Sciences* 1519, 74–93. (Lluka & Stokes, 2022)
13. Fernandes, F.C., Cardoso, M.H., Gil-Ley, A., Luchi, L.V., da Silva, M.G., Macedo, M.L., de la Fuente-Nunez, C., Franco, O.L., 2023. Geometric deep learning as a potential tool for antimicrobial peptide prediction. *Frontiers in Bioinformatics* 3.
14. Moola S, Munn Z, Tufan Aru C, Aromatases E, Sears K, Sfatcu R, et al. Chapter 7: Systematic reviews of aetiology and risk. In: Aromataris E, Munn Z (Editors). *JBIM Manual for Evidence Synthesis*. JBI, 2020. <https://doi.org/10.46658/JBIMES-20-08>
15. Wang G, Li X, Wang Z (2016) APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res* 44:D1087–D1093. <https://doi.org/10.1093/nar/gkv1278>.
16. Pirtskhalava M, Gabrielian A, Cruz P et al (2016) DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. *Nucleic Acids Res* 44:D1104–D1112. <https://doi.org/10.1093/nar/gkv1174>.
17. Luo, Y., and Song, Y. (2021). Mechanism of antimicrobial peptides: Antimicrobial, anti-inflammatory and antibiofilm activities. *Int. J. Mol. Sci.* 22, 11401. doi:10.3390/ijms222111401.
18. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: an imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems* 32 (NeurIPS 2019). Vol. 32. Vancouver Canada: Curran Associates, Inc. (2019). p. 8024–8035.
19. Kingma DP, Ba J. Adam: a method for stochastic optimization (2014). Available from: <https://arxiv.org/abs/1412.6980>.
20. Liaw R, Liang E, Nishihara R, Moritz P, Gonzalez JE, Stoica I. Tune: A research platform for distributed model selection and training [Preprint] (2018). Available at: <http://arxiv.org/1807.05118>.
21. Ingle DJ, Levine MM, Kotloff KL, Holt KE, Robins-Browne RM. Dynamics of antimicrobial resistance in intestinal *Escherichia coli* from children in community settings in South Asia and sub-Saharan Africa. *Nat Microbiol.* 2018;3(9):1063–73.
22. X. Su, G. Jing, Y. Zhang, and S. Wu, "Method development for cross-study microbiome data mining: challenges and opportunities," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 2075–2080, 2020.



23. A. L. Hicks, N. Wheeler, L. Sánchez-Busó, J. L. Rakeman, S. R. Harris, and Y. H. Grad, "Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data," *PLoS Computational Biology*, vol. 15, no. 9, Article ID e1007349, 21 pages, 2019.
24. M. N. Sohail, R. Jiadong, M. U. Muhammad, S. T. Chauhdary, J. Arshad, and A. J. Verghese, "An accurate clinical implication assessment for diabetes mellitus prevalence based on a study from Nigeria," *Processes*, vol. 7, 2019.
25. M. U. Muhammad, R. Jiadong, N. S. Muhammad, and B. Nawaz, "Stratified diabetes mellitus prevalence for the Northwestern Nigerian States, a data mining approach," *International Journal of Environmental Research and Public Health*, vol. 16, no. 21, p. 4089, 2019.
26. Huyen C. Designing machine learning systems: an iterative process for production-ready applications. O'Reilly Media, Incorporated (2022).
27. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin Microbial Infect* 2017;23(1).
28. Beltrão G, Paramonova I, Sousa S. User interface design for AI-based clinical decision-support system: preliminary study. 2022 17th Iberian Conference on Information Systems and Technologies (CISTI) (2022). p. 1–4. Available from: <https://doi.org/10.23919/CISTI54924.2022.9820378>.
29. Fitzgerald DM. Bacterial evolution: the road to resistance. *Elife*. 2019;8:e52092.
30. Croucher NJ, Didelot X. The application of genomics to tracing bacterial pathogen transmission. *Curr Opin Microbial*. 2015;23:62–7.