ELYAN, E., HUSSAIN, A., SHEIKH, A., ELMANAMA, A.A., VUTTPITTAYAMONGKOL, P. and HIJAZI, K. 2022. Antimicrobial resistance and machine learning: challenges and opportunities. *IEEE access* [online], 10, pages 31561-31577. Available from: https://doi.org/10.1109/ACCESS.2022.3160213

Antimicrobial resistance and machine learning: challenges and opportunities.

ELYAN, E., HUSSAIN, A., SHEIKH, A., ELMANAMA, A.A.,
VUTTPITTAYAMONGKOL, P. and HIJAZI, K.

2022

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 8, 2022, accepted March 14, 2022, date of publication March 16, 2022, date of current version March 25, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3160213

Antimicrobial Resistance and Machine Learning: Challenges and Opportunities

EYAD ELYAN[1], **AMIR HUSSAIN**[2], **AZIZ SHEIKH**[3], **ABDELRAOUF A. ELMANAMA**[4], **PATTARAMON VUTTIPITTAYAMONGKOL**[5], **AND KAROLIN HIJAZI**[6]
[1]School of Computing, Robert Gordon University, Aberdeen AB10 7AQ, U.K.
[2]School of Computing, Edinburgh Napier University, Edinburgh EH14 1DJ, U.K.
[3]Usher Institute, University of Edinburgh, Edinburgh EH8 9AG, U.K.
[4]Medical Technology Department, Faculty of Health Sciences, Islamic University of Gaza, Gaza, Palestine
[5]School of Information Technology, Mae Fah Luang University, Chiang Rai 57100, Thailand
[6]School of Medicine, Medical Sciences & Nutrition, University of Aberdeen, Aberdeen AB24 3FX, U.K.

Corresponding author: Eyad Elyan (e.elyan@rgu.ac.uk)

This work was supported in part by the Global Challenge Research Fund (GCRF).

ABSTRACT Antimicrobial Resistance (AMR) has been identified by the World Health Organisation (WHO) as one of the top ten global health threats. Inappropriate use of antibiotics around the world and in particular in Low-to-Middle-Income Countries (LMICs), where antibiotics use and prescription are poorly managed, is considered one of the main reasons for this problem. It is projected that the COVID-19 pandemic will accelerate the threat of AMR due to the increasing use of antibiotics across the world, and especially in countries with limited resources. In recent years, machine learning-based methods showed promising results and proved capable of providing the necessary tools to inform antimicrobial prescription and combat AMR. This timely paper provides a critical and technical review of existing machine learning-based methods for addressing AMR. First, an overview of the AMR problem as a global threat to public health, and its impact on countries with limited resources (LMICs) are presented. Then, a technical review and evaluation of existing literature that utilises machine learning to tackle AMR are provided with emphasis on methods that use readily available demographic and clinical data as well as microbial culture and sensitivity laboratory data of clinical isolates associated with multi-drug resistant infections. This is followed by a discussion of challenges and limitations that are considered barriers to scaling up the use of machine learning to address AMR. Finally, a framework for accelerating the use of AMR data-driven framework, and building a feasible solution that can be realistically implemented in LMICs is presented with a discussion of future directions and recommendations.

INDEX TERMS AMR, antimicrobial resistance, machine learning, LMICs.

I. INTRODUCTION

Inappropriate use of antibiotics around the world has contributed to the steady rise of antimicrobial resistance (AMR) [1]–[3] and in particular in Low-to-Middle-Income Countries (LMICs) [4]. AMR occurs as a result of genetic changes in microbes which encode for mechanisms through which microbes themselves can become less susceptible or resistant to one or multiple types of antimicrobials (the latter phenomenon known as multi-drug resistance). This, in turn, leads to the clinical failure of treatment against many types of infectious diseases [5]–[7]. It has been confirmed that mortality of patients with blood infections caused by multi-drug resistant bacteria is significantly higher than those where multi-drug resistant bacteria are not detected [8]. A surveillance review between the years 2002 and 2011 showed growing AMR trends in various countries across the world [9]. Recently, AMR was identified by the World Health Organisation (WHO) as one of the top ten global health threats (WHO, 2019). A review commissioned by the UK government estimated that AMR will result in 10 million yearly deaths by the year 2050, with a global financial impact estimated to be up to 100 trillion USD.[1]

AMR is dependent on an array of healthcare as well as socio-politico-economic factors. However, there is consensus that overuse and misuse of antimicrobials, particularly a

The associate editor coordinating the review of this manuscript and approving it for publication was Weiping Ding.

---

[1] https://amr-review.org/Publications.html

certain type of antibiotics known as broad-spectrum antibiotics, is at the heart of the rise of AMR globally [10]. On this basis, control of antibiotic use has been one of the pillars of action plans to combat AMR, along with the promotion of research of new antibiotics and non-antimicrobial alternatives together with improved infection-control and improved water, sanitation, and hygiene (WASH) conditions, both at national (UK five-year National Action Plan, 2019-2024) and global level (WHO, 2015). The framework of strategies seeking to measure and improve the use of antibiotics to reverse the rising rates of AMR is known as Antimicrobial Stewardship.

The strategies underpinning antimicrobial stewardship programmes around the world differ vastly amongst different countries, but the common aim is to reduce antibiotic prescription practices that have a detrimental effect on patient safety and outcomes [10]. These programmes are now very common in developed countries and are designed to rationalise the use of antibiotics by healthcare workers and to stop the spread of multi-drug infections [11].

Despite strategies and efforts to manage AMR. It must be pointed out that AMR rates can be significantly affected by factors, other than overuse and misuse of antimicrobials, such as pandemics. For example, in the context of the COVID-19 pandemic [12], it has been argued that "AMR represents a slow-moving disaster, and that an urgent response to mitigate its impact has become even more urgent". It is projected that the COVID-19 pandemic will compound AMR rates and compromise antimicrobial stewardship activities [13]. This is partly attributed to the increasing use of antibiotics across the world which is expected to accelerate the threat of AMR as suggested by recent literature [14]. A recent study found that more than 70% of COVID-19 cases in Asia received antimicrobial treatment [7]. Other studies also reported a similar increase in antibiotic usage and AMR rates in several countries and across income levels [15]–[17]. These studies suggest that there is an urgent need for collective efforts to handle and manage AMR, especially in LMICs where resources are very limited.

In addition to the traditional stewardship programs, in recent years, the role of machine learning and deep learning became increasingly important in combating the AMR problem [18], [19]. This is mainly due to the significant progress in machine learning and data-driven products development, and also due to the availability of large volumes of structured and unstructured data, especially clinical and experimental data [5]. Studies that use 'off-the-shelf' supervised machine learning algorithms to build predictive models to inform antibiotic prescription showed some promising results, and proved that machine learning-driven solutions can provide the necessary tools to help inform antimicrobial prescription and combat the AMR problem [3], [20], [21]. Despite these promising results in controlled environments ([3], [20]–[23]), existing literature suggest that the progress in the area of using predictive modeling to support clinical decisions for antibiotics prescription and antimicrobial management is still limited, and far from taking full advantage of the significant progress and development of the data and algorithms available [3].

This review will focus on studies that use clinical and demographic data as well as microbiology data to train and build supervised learning algorithms aimed at supporting clinicians in choosing the appropriate antimicrobial treatments. We expect machine learning-driven solutions to be particularly impactful in LMICs where adoption of costly point-of-care testing technologies is not realistic as discussed later in this review. The main contributions of the paper are outlined as follows:

- The paper provides an overview of the severe impact of AMR on countries with limited resources (LMICs) with attention to one case study flagging alarming AMR trends in a setting facing some of the direst public health conditions on the globe (Section II).
- In-depth technical review and evaluation of existing studies of machine learning methods to tackle the AMR problem with an emphasis on methods that use patient clinical and demographic data as well as bacterial genotypic information to train predictive models (Section III).
- Discussion of challenges and limitations that are considered barriers to exploiting the full potential of machine learning and deep learning methods and scaling up their use to impact AMR (Section IV).
- Recommendation of possible future directions toward accelerating research and development in machine learning models to inform antibiotic prescribing and build feasible solutions that can be realistically implemented in LMICs (Section V).

The rest of this paper is organised as follows: Section II presents an overview of AMR and its impact on public health, especially in LMICs where concerns have been reported in relation to the risk of epidemics of multidrug-resistant infections in countries like Palestine [24], where parts of it have been under blockade since 2007 and ranked amongst the most densely populated places in the world (Gaza) (Palestinian Central Bureau of Statistics).[2] In section III we present an overview and discussion of supervised machine learning algorithms, including an in-depth technical review of methods used to impact AMR, and a focus on learning methods exploiting clinical, demographic and genotype data. In Section IV detailed discussion of existing challenges is presented, and finally, in Section V possible future directions and conclusions are discussed.

II. AMR AND LMICs

Immediate antibiotic treatment wherever severe bacterial infections are suspected remains imperative for improvement of patient outcome [25]. However, antibiotic treatment should be specific to the microbe causing the infection. The standard laboratory methods to identify the bacterium causing

---

[2] http://www.pcbs.gov.ps/default.aspx

the infection, as well as the antibiotic required to treat the infection (known as phenotypic resistance testing), requires between 24 to 48 hours, with inevitably much longer timeframes in resource-limited settings. These long waits lead to the lengthening of treatments using all-encompassing antibiotics (broad-spectrum antibiotics) which are not specific to the microbe causing the infection [26].

While in resource-rich nations antimicrobial stewardship programmes are being widely implemented, very little has been accomplished in LMICs as a result of unfavourable political and socio-economic factors and fundamentally widespread poverty and inadequate resources. It is recognised that in resource-limited settings, the lack of control over sales of antibiotics and less than optimal antibiotic prescribing practices is a key factor to the steady rise of AMR [4].

The middle east region is considered a hotspot for AMR [27] likely as a result of suboptimal laboratory capacity, poor implementation of antimicrobial stewardship programmes and inadequate data and digital recording of patients information [11]. These are the countries on or near the Arabian Peninsula, including Kuwait, Saudi Arabia, Bahrain, Qatar, Oman, United Arab Emirates, Egypt, Sudan, Iraq, Iran, Yemen, Syria, Jordan, Palestine, Cyprus, Libya, Turkey, and Lebanon [28]. While the socio-economic landscape in this region differs greatly amongst different countries, gross inequalities in healthcare provision and widespread poverty remains a common feature across most countries, some of which have been devastated by decades of wars and political conflicts.

Specific concerns have been reported concerning the risk of epidemics of multidrug-resistant infections in Gaza [24], [29], under political blockade since 2007 and ranked amongst the most densely populated places in the world (Palestinian Central Bureau of Statistics).[3] Numerous outbreaks of antibiotic resistance that have been documented are likely the result of poor infection control, irrational prescription practices, and easy access to antimicrobials [30]–[32]. Naturally, management of serious infections in this setting is nowhere close to the standards called by most antimicrobial stewardship programmes. The exceptional pressure on the health care system and substandard laboratory infrastructure lead to over-use of antibiotics as antibiotic susceptibility testing is often not carried out at all.

A recent retrospective study of records of patients admitted to the European Gaza Hospital, which serves patients from all Palestinian Ministry of Health hospitals in Gaza, confirmed arbitrary prescription of a broad range of antibiotic regimes and little or no microbiological testing [33]. Over-use of antimicrobials in healthcare settings adds to poorly regulated antibiotic sales in Gaza and the Occupied Palestinian Territories, where antibiotics can be purchased without a need for a medical prescription [11], [27]. These practices will continue to exacerbate the current high AMR rates in Gaza [34] where the control of other critical factors contributing to AMR such as infection control and sanitation remains a struggle. Only recently, local health authorities have recognized the importance of regulation of antibiotic use with colistin being banned in the Gaza strip for animal use.

It should be noted that the development of rapid and cheap point-of-care diagnostic tools to guide prescribing practices in LMICs has been long advocated[4] as the benefit of these technologies is likely greater in under-resourced settings with reduced access to specialist expertise that can guide antibiotic prescribing [10]. However, little progress has been made in this area partly because of the general lack of interest of the pharmaceutical industry in the development of rapid diagnostics for infections.[5] In this context, the parallel development of machine learning and data-driven tools to inform antibiotic prescribing at the point of care could make a major contribution to overcoming irrational prescribing practices in areas under exceptional pressure such as Gaza. The urgent need for machine learning and intelligent methods to be 'put to good use in changing antibiotic practices' was acknowledged in the O'Neil report commissioned by the UK.[6] While little has been done to develop this area of work, the effectiveness of these methods to improve antimicrobial prescribing in hospitals has been proven [35]. For example, a relatively recent study in a children's hospital in Cambodia demonstrated the highly predictive value of simple machine learning algorithms applied to patient data readily available in a resource-limited setting and the superiority of these tools over traditional and common machine learning methods such as logistic regression models [3]. Existing literature suggests that with basic clinical information and some demographic patients data such as age, gender, and residential address, good results can be achieved in terms of informing drugs prescription and AMR management [36]–[38].

III. MACHINE LEARNING & AMR

Recent literature shows clearly that the use of datasets that capture patient's demographic and clinical data can provide decision support tools to determine whether a bacterium is resistant to a line particular antibiotic.

Existing data-driven methods based on machine learning approaches to tackle AMR can be broadly categorised into 1) methods that utilise clinical and demographic information of patients as well antimicrobial susceptibility data [3], [20]–[22] (Section III-B) and 2) methods utilising the whole genome- sequencing data [39]–[41] (Section III-C). Both types of methods have shown promising results and proved potential as valid tools to support stewardship programmes and inform antibiotic prescribing practices. However, whole-genome sequencing-based approaches will be unlikely adopted in LMICs due to the relatively high costs associated with this technology [42]. On this basis, the present review will focus mostly on approaches that utilise

---

[3] http://www.pcbs.gov.ps/default.aspx

[4] https://amr-review.org/Publications.html

[5] https://amr-review.org/Publications.html

[6] https://amr-review.org/Publications.html

demographic, clinical, and basic microbiological culture data. We precede this with a quick overview of supervised learning algorithms in the context of AMR.

A. MACHINE LEARNING

Recent development in Artificial Intelligence (AI) in areas such as machine learning [43] and in particular in the area of deep neural networks and deep learning [44] had significantly improved performance in data-driven applications across different domains. In recent years, the use of machine learning models and advanced deep learning-based methods is becoming increasingly common in health sciences. This can be attributed to three main reasons. Firstly, the significant advancement and development at the algorithmic level, where a wide range of learning algorithms are now readily available to use. These methods range from simple linear regression models up to advanced ensemble-based methods and deep neural network-based methods [45]. Secondly, the large amount of data available can provide a rich source for data analytics. The final reason for the widespread of AI-driven solutions in healthcare and across other domains is the availability of powerful computing machines that are capable of running large-scale experiments using complex models and a huge amount of data.

Health, in particular, has benefited greatly from the adoption of AI-driven solutions [46] in many areas of clinical decision making [47]–[49] and infectious disease [50]. Despite some recent and promising results in hospital settings [3], [35], antibiotic prescribing and management is the exception, it can be argued that more work is needed in this area to apply the recent advances in machine learning and deep learning to tackle the AMR problem. The relatively slow progress in this area can be largely attributed to the lack of public datasets that captures demographic and clinical information to drive research in the data-driven application for AMR management. This is unlike genome-based AMR-related datasets which can be easily accessed from the public domain [51], [52].

Machine learning is a very broad topic that can be categorised into different types including unsupervised learning, supervised learning, and reinforcement learning. In health science-related practice, including decision-making in antibiotic prescribing, supervised machine learning is used to learn from past observations. In a nutshell, supervised classification and regression methods use a set of features X (input variables) to predict a certain outcome y often referred to as a label. Formally, the aim is to find a function that maps X to y as $h(x) : X \rightarrow y$. As an example, consider a dataset A with m instances $x_1, x_2, \ldots, x_m$, where each instance x_i is a feature vector representing a patient's information and is defined by n features (e.g. age, gender, medical history, blood pressure, etc...) as $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$, and y_i denotes the final diagnosis of that patient (outcome or the class label). A typical AMR work in this context is reported in [53] where the authors used clinical, demographic, and living condition data of patients to predict bacteria's resistance/sensitivity to common antibiotics. Equation 1 depicts this scenario, and in a learning task, the aim is to build a model that can provide a support tool to clinicians or medical experts to support their assessment.

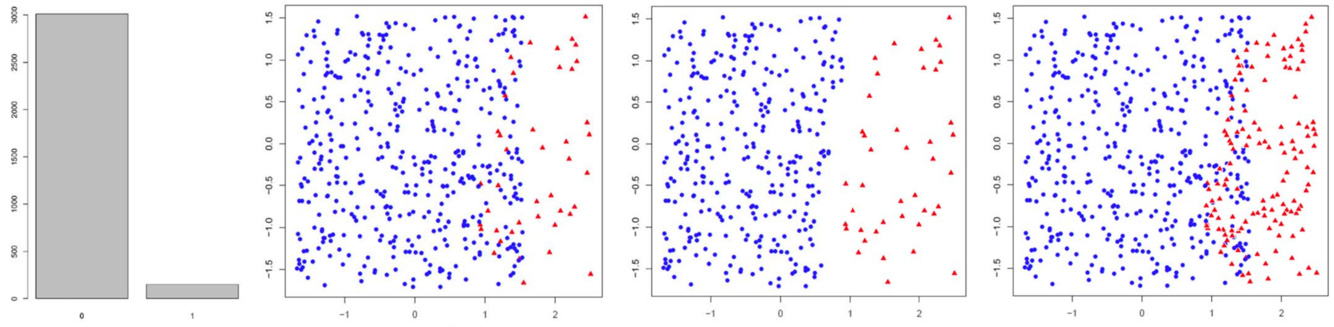
$$X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & \vdots & \ldots & x_{mn} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ .. \\ .. \\ .. \\ y_m \end{bmatrix} \tag{1}$$

It should be noted that the quality and accuracy of learning algorithms, and the prediction of a certain outcome given a set of input features (Equation 1), depend on several factors. The three key critical elements for good performance of machine learning solutions are 1) the dataset 2) the machine learning algorithm chosen to map the input variables to the output 3) and the evaluation method of the model's performance.

1) DATASET VARIATIONS

A dataset that captures more variations with appropriate class distribution often leads to better results [54]–[56]. For this reason, a major part of the work is commonly dedicated to preparing the dataset and hand-crafting a set of discriminating features to improve the performance of the learning algorithms [57], [58]. For example, in [58], the authors had to apply unsupervised machine learning algorithms to find within class similarities, before applying a predictive model to different health-related datasets for diagnostic purposes. This was achieved by applying a simple clustering technique to the data (k-means) before feeding it into the learning algorithm. The motivation was to increase the dataset diversity to improve the model's predictive power. Results showed, that by applying such a simple pre-processing method, accuracy can be significantly improved, especially, in health-related datasets.

One of the inherent and most challenging problems in healthcare data relating to a range of clinical areas, is class-imbalance [59]. This problem happens in supervised machine learning when the total number of samples in a class of data is far less than the number of samples in another class of data as depicted in Figure 1(a). This scenario is very common in various domains, and in particular in medical datasets [60]. The dominance of majority class instances often causes biased performance of the learning algorithms due to the objective to maximise the overall prediction accuracy. A high misclassification error on the minority class will occur as a result. The situation becomes more concerning when the minority class is the class of interest with a high error cost, wherein in this case, the minority class is realised as the positive class whereas the majority class is referred to as the negative class. This circumstance is often seen in the medical domain where the availability of the positive class data is limited. For example, consider a dataset that contains thousands of records where each represents patient clinical characteristics (input features) and the diagnoses (outcome/class), and only 1% of the record is labelled/diagnosed with the particular disease



(a) (b) (c) (d)

FIGURE 1. Data class-imbalance and application of data resampling, a) original dataset b) undersampling c) undersampling from the overlapping region and d) oversampling.

of interest. In such a scenario, the positive class will be under-represented in the dataset, and hence would be difficult to detect by a learning algorithm.

Antimicrobial resistance datasets as well often have the class-imbalance problem where commonly infective agents susceptible to antibiotics dominate the dataset when compared to resistant infections [22], [61], [62]. Recent literature confirms that the imbalanced class distribution in AMR datasets negatively affects the performance of machine learning algorithms [20], [22]. In [20], Martínez-Agüero *et al.* used machine learning to study six different bacterial species and their resistance/sensitivity to common antibiotics drugs. The study was carried out using more than 10,000 observations associated with antibiograms done for positive cultures. While they reported relatively good predictive power in relation to resistance of a certain bacterial species to a particular antibiotic drug, the authors noted that machine learning performs poorly in predicting antibiotic resistance profiles that are least represented in the dataset, for instance, resistance to polymixines.

Various methods to mitigate the impact of the class-imbalance problem on the performance of machine learning algorithms have been described. As mentioned above, one of them is data resampling, which is widely used in the field [54]. Data resampling consists of oversampling and undersampling. The former involves creating synthetic minority-class samples to improve their presence in the dataset whereas the latter is the practice of reducing majority-class samples. A typical example is illustrated in Figure 1, which represents a class-imbalanced dataset, where the number of negative instances is significantly in excess of the number of positive instances, and the dataset after being undersampled (c) and oversampled (d). Maguire *et al.* [62] identified the problem of class imbalance in predicting AMR phenotypic resistance from genomic data. Due to such an issue, only a subset of antibiotics with a reasonable balance of resistance was selected for the experiments, and SMOTE [63], a well-known, long-established oversampling method was applied to create a more balanced training data. Similarly, in [64], Chowdhury *et al.* adopted existing undersampling and oversampling methods to handle a class imbalance in predicting AMR protein sequences for Gram-negative bacteria using machine learning and achieved high accuracy.

Other approaches to deal with class-imbalanced datasets include algorithm-level methods and ensemble-based methods. Algorithm-level methods are related to modifying existing learning algorithms to emphasise the learning more of the minority class. A typical example of these is cost-sensitive methods where a larger weight (cost) is associated with misclassification of positive instances [65]. Ensemble-based methods are combinations of data-level and algorithm-level techniques. Hence, they will carry the benefits of both approaches. In an AMR-related problem, Kavvas *et al.* [66] proposed an ensemble SVM with adjusted class weight to account for the imbalance of resistant and susceptible strains in identifying signatures of AMR evolution to 13 antibiotics. In [61], Hyun *et al.* followed a similar approach by using ensemble SVM with random oversampling and undersampling to improve the performance in predicting AMR determinants for microbial pathogens. Similarly, Van *et al.* [67] presented the use of random oversampling with XGBoost, which is gradient boosted decision trees, to deliver fast estimations of the antibiogram for Gram-negative bacteria with highly imbalanced data.

It can be argued that most of the methods presented in the past decade for handling the class-imbalance problem are largely based on data sampling methods or its variants, e.g. random oversampling, random oversampling, cluster-based data sampling, and others [54]. Most of these data-sampling methods mainly rely on reducing the number of majority class instances, increasing the number of minority class instances, or a combination of both. However, in medical-related data, undersampling methods may cause information loss, and so reduce the overall model's accuracy [56].

New oversampling/ data augmentation methods appeared recently in the literature that uses Generative Adversarial Neural Networks (GANs) to synthesize data and improve the learning outcomes. GANs were initially introduced by J. Goodfellow in 2014 [68] and ever since has been successfully applied to generate data, in particular images and videos.

A typical example is presented in [69] where the authors used GANs-based method to generate minority class instances and handle the class imbalance. The key advantage of using GANs-based methods is that more diverse data instances can be obtained, unlike traditional data synthesis methods.

For additional literature on potential strategies for handling class imbalance in any future work on imbalanced AMR-related datasets, the reader is referred to previous studies [70]–[72] and to [54] for a comprehensive and recent review on class-imbalance and methods to handle it.

2) LEARNING METHODS

The fundamental element required for building a machine learning solution that follows the pre-processing and preparation of the dataset, is the choice of an appropriate model to fit the data. There is a wide range of machine learning algorithms [73] and no one recipe fits all types of data. The choice of a particular algorithm is often informed by several factors, such as the type, size, and complexity of the data. For example, linearly separable data can be easily handled by a simple regression model. For more complex and non-linear datasets Support Vector Machines (SVM) [74], Ensemble-based methods such as Random Forests (RF) [75], Artificial Neural Networks (ANN) [76], Deep Learning (DL) [[44], provide better alternatives. For a systematic review of machine learning methods used in AMR datasets captured using mass spectrometry, the reader is referred to [77] which also provides a gentle introduction to these models. Despite the superior performance of complex models such as ANN, SVM, RF, simple methods such as logistic regression (LR), and decision trees (DT) are still common in AMR and other health applications [3], [20], [22], [78], [79] because of their easier interpretation (explainable models). Other common methods include Naïve Bayes (NB), which is based on Bayes' theorem [73].

Ensemble methods such as RF and Gradient Boosting showed superior performance over other traditional machine learning methods [80]. These methods are constructed based on training and combining more than one machine learning model (e.g. SVM, Logistic Regression, Decision Tree, ...). Results of these different models are then aggregated to obtain the final prediction [75].

The performance of the ensemble-based method largely depends on the model's diversity. The models within an ensemble need to be diverse enough and independent from one another. One simple approach is to train more than one learning algorithm (e.g. SVM, ANN, LR, etc...) and aggregate the predictions of each model. However, the most commonly used approach to diversify the ensemble is to train the same algorithm (e.g. a Decision Tree) using different subsets of data (diverse subsets sampled from the training set) [81]. This is often achieved by a simple data sampling method with replacement called Bootstrap aggregating (Bagging). As can be seen in Figure 2, using Bagging, we can obtain diverse subsets, where certain instances can be sampled more than one time for the same model.

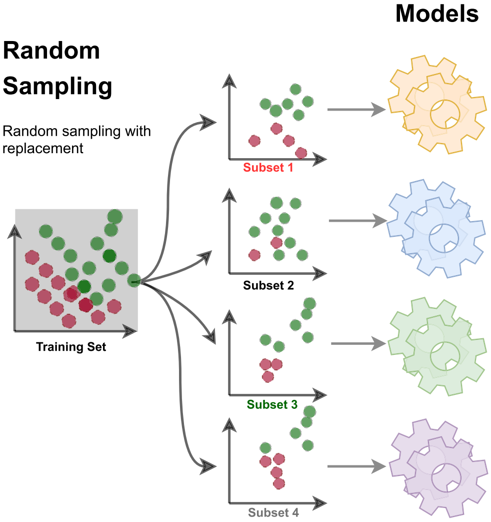


FIGURE 2. Sampling with replacement and ensemble learning.

Random Forest, developed by Leo Breiman in 2001 [75], is the most common ensemble and has gained significant interest by the machine learning research community due to its performance across different complex datasets from various domains. In one of the largest experimental studies and when comparing almost 180 different machine learning algorithms using 121 different datasets from the public domain, RF ranked top, followed by SVM with Gaussian Kernel, and Gradient Boosting [45]. RF generates a diversified and good performing ensemble using two different strategies:

- First, Bagging that applies sampling with replacement to the original training set is carried out to construct each tree in the ensemble as can be seen in Figure 2. Using this sampling method, it can be proved statistically that only 63% unique instances/ observations appear in the training set and the rest are repeated [75].
- Then, at each node split, only one subset of features is drawn randomly to assess the goodness of each feature/attribute.

A key advantage of using RF over other methods is that it requires minimum parameters tunning. In fact, only two parameters need to be set, namely, the number of trees and number of features assessed for the goodness of split at each node in the tree. The number of trees is often set between 100 to 500, while the number of features can be either set experimentally [57] or to ($\sqrt{F}$ or $\log_2 F$) where F is the total number of features in the dataset.

Because of its superior performance and ease of implementation, RF has been widely used across a whole range of applications in life sciences [58], [82], medical diagnosis [83], oil and gas [84], and others including its use for antimicrobial resistance prediction [3], [20] as shown in Table 1 and in the following section.

3) EVALUATION

Various methods are used to evaluate the performance of machine learning models. These methods are designed to

TABLE 1. Common ML methods used in AMR.

Work	Method	Data size	Data types	Year
Feretzakis et al. [36]	Ensembles (Stacking, Boosting, ...)	11496 instances	susceptibility and demographic	2021
Rawson et al. [85]	SVM, NB, ANN	1186 patients	demographic, microbiology, blood cultures	2021
Feretzakis et al. [21]	RF, DT, LR	2307 instances	susceptibility and demographic	2020
Moran et al. [86]	Extreme Gradient Boosting	9352 Patients	urine, blood, demographic data	2020
Chowdhury et al. [87]	Support Vector Regression	-	movement, demographic & billing data	2020
Martínez-Agüero et al. [20]	LR, KNN, RF	2630 patients	clinical and demographic data	2019
Cánovas-Segura et al. [22]	LR	1393 positive *Enterococci* cultures	cultures and demographic data	2019
Yelin et al. [78]	Gradient Boosting, LR	315,047	urine cultures and patients data	2019
Oonsivilai et al. [3]	RF	243 patients	blood stream infections	2018
Kulshrestha et al. [88]	Association Rules	-	patients data	2018
Vazquez-Guillamet et al. [79]	LR, DT	1618 patients	patients data with sepsis	2017
Wang et al. [89]	Regression Models	9000 Instances	clinical and demographic data	2017
Chow et al. [90]	Decision Support System	1886 patients	antimicrobial prescriptions	2015
Papageorgiou and Elpiniki [91]	Heuristics / Rule-based	-	urine tract infection data	2012
Puuronen et al. [23]	Ensembles and Features Selection	4000 patients	sensitivity tests and patients data	2008

assess model performances on unseen data observations. Typical methods include holdout and cross validation. Additionally, different evaluation metrics are commonly used in this context to measure the performance of the model (e.g. accuracy).

a: HOLDOUT

One of the most common ways to gain insight on how well a supervised machine learning model can generalise to unseen examples/ data observations is by splitting the whole dataset into two sets: training and testing sets. The training set is used to train the model to best fit the data, while the testing set is used to assess the performance and quality of the trained model. Having a testing set that has not been seen or used during the training process provides a good estimate of the performance of the model on new data instances/ unseen examples model [73]. This practice continues to be the most common approach to evaluate a model's performance.

The training process can include experimenting with different parameter settings, comparing different models, and applying different pre-processing steps such as feature selection, dimension reductions, and others to improve performance. Such a process is known as model selection. One of the common errors that occur during this process is the reuse of the testing set, which can lead to model overfitting. This is a very common problem in machine learning, and overfitting happens when the model performs well on the training data, and its performance drops significantly on the testing/ unseen data. Wide range of methods available in the literature to handle this problem, including regularisation, data augmentation, early stopping of the training process, and others [92].

A safer approach is to split the data into three different datasets, namely training, validation, and testing sets. The advantage of such a split is that the training and model selection process is entirely separated from observations in the testing set. However, this approach is often dependent on the availability of sufficiently large datasets.

b: K-FOLD CROSS-VALIDATION

One of the main disadvantages of the hold-out method is its sensitivity to the way the dataset is partitioned. In other words, the model will be biased toward the way the training and validation sets are partitioned, and its performance will vary accordingly. When the dataset is limited, then a small validation set won't provide a good estimate for the model's performance. Therefore, one solution to such scenario is to use K-fold cross-validation. The K-fold cross validation, also known as repeated K-fold cross validation, is another common approach to split the data. Here, the dataset is randomly split into k folds where k-1 folds are used for training the model, and the other one is used for testing. The procedure is then repeated k times, resulting in k different models, and an average performance of all the models is then calculated. This approach provides better estimates of model's performance on unseen data observations. A major drawback of this approach is the need for more computing power, as the number of training runs that must be carried out increases by a factor of K [43].

c: STRATIFIED K-FOLD CROSS-VALIDATION

A widely-used variant of K-fold cross-validation is stratified K-fold cross-validation. The splitting procedure is similar to the repeated K-fold cross-validation except that the class distribution of the dataset is preserved in the splits. That is, each fold will have the same proportions of samples in each class as in the original dataset. Hence, the stratified K-fold cross validation will be more a suitable method for classification tasks with imbalanced class distributions [93].

d: LEAVE-ONE-OUT CROSS-VALIDATION

Another common cross-validation technique is leave-one-out cross-validation (LOOCV). LOOCV is basically a configuration of K-fold cross-validation where K is equal to the number of instances in the dataset. In each fold, one instance is left out as the testing instance while all other instances are used for training the model. The performance of the model is then calculated from the average performance off all folds. LOOCV produces a reliable and unbiased estimate of model performance; however, it is computationally expensive. Hence, it will be a good choice when the dataset is not too large or a strict model estimation is required.

e: JACKKNIFE CROSS-VALIDATION

The process of Jackknife cross-validation is similar to LOOCV. The only difference is that in Jackknife cross-validation, in each fold, the model is evaluated based on the kept samples, not the left-out sample. The method is simple, but computationally expensive and may not be suitable for large datasets.

It should be pointed out that all these validation methods aim at providing an accurate and non-biased estimate of the model's performance on unseen data, and the choice of which validation method to use is largely depending on the problem and the size of data. A technical discussion on different model's validation methods can be found at [94].

f: EVALUATION METRICS

Various evaluation metrics are used for evaluating ML model's performance. The most obvious and common one is accuracy, which can be simply defined as the number of correctly classified examples divided by the total number of examples in the testing set and is defined as in Equation 2

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{2}$$

where TP denotes a True Positive instance, for example, a data instance representing a bacterial species that is resistant to particular antibiotic, while TN denotes a True Negative instance (bacterial species sensitive to a particular antibiotic). However, for modelling of datasets with imbalanced class distribution, such as AMR-related data, accuracy can be a misleading metric, and other metrics are used to provide a better estimation of the model performance [95]–[98].

In AMR-related datasets, where the data are often hugely imbalanced [20], [22], it is crucial to assess the model performance in terms of detecting the class of interest (prediction of resistance of a certain bacterial species to an antibiotic). Therefore, in addition to accuracy, metrics that are often reported in AMR-related datasets are sensitivity, often referred to as the True Positive Rate (TPR) or recall as defined in Equation 3, and specificity which captures the True Negative Rate as defined in Equation 4

$$sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$specificity = \frac{TN}{TN + FP} \tag{4}$$

Balanced accuracy (Equation. 5) and G-mean (Equation 6) are also common metrics to evaluate model's performance, where balanced accuracy is the arithmetic mean of the accuracy over each class in the dataset [99]–[102], while the G-mean represents the geometric mean of sensitivity and specificity.

$$balanced\ accuracy = \frac{sensitivity + specificity}{2} \tag{5}$$

$$G - mean = \sqrt{specificity * sensitivity} \tag{6}$$

Area Under the Curve (AUC), harmonic mean (F1-score) (Equation 7) and other metrics that aim to capture the overall performance of the model across all classes are often used in AMR-related studies as explained in Section III-B. F1-score or F measure/ score is a common metric for assessment of the model performance in medical datasets by combining both precision and recall of the model into one measurement:

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{7}$$

It should be noted that the formulas above (e.g. Equations 2, 3, 4, ...) are often multiplied by 100 to obtain results in percentage format. For additional discussion of these various metrics the reader is referred to [54].

B. DEMOGRAPHIC AND CLINICAL DATA

Standard practices to detect AMR are based on a microbiological culture of clinical samples performed in dedicated diagnostic laboratories which require up to 72 hours [103]. Using standard laboratory methods this time for identifying the bacterium causing an infection as well as the antibiotic required to treat can be reduced 24 hours. A reduction of this time frame is critical for the detection of multidrug-resistant infections requiring patient isolation and for clinicians to minimize empirical prescribing and initiate targeted antimicrobial therapy as quickly as possible [20]. Machine learning [43] provides a promising solution for an accurate prediction tool to inform the physician's decision-making process. However, a recent systematic review [104] confirmed the clear gap between the availability of large volumes of structured and unstructured data to train predictive models, and the use of predictive modelling to support clinical decisions for antibiotics prescription and antimicrobial management [3]. It also confirmed the urgent need for the development of an evidence-based framework for informing and reporting clinical decision support systems for antimicrobial management.

Recent literature shows several attempts to utilise data for building machine learning algorithms to improve antibiotic prescribing practices. A recent study [3] shows that using 'off-the-shelf' machine learning algorithms trained on a relatively small patient dataset (243 patients with blood stream infections) can provide an informed prediction on antibiotic susceptibilities to guide antibiotic prescribing practices. It has been widely reported that most machine learning algorithms outperform the commonly used logistic regression models. Random Forest [75] ranked top in comparative studies. Interestingly, the results showed that input features such as time (from admission to blood culture), patient age, and how the infection was acquired (hospital vs community) were the most important covariates for predicting susceptibility. It should be noted that such information can easily be recorded if not already available at most points of care. Despite the relatively small dataset used in the experiments, this study [3] shows clearly that good results can be obtained with a fairly straightforward application of a suite of machine learning algorithms. In other words, no extensive data pre-processing features engineering, or algorithmic modification/ optimisation were carried out, which shows clearly the potential use of machine learning in this context.

Feretzakis *et al.* developed a clinical decision support tool to predict antimicrobial resistance and inform the appropriate choice of antibiotic treatment [36]. A dataset of 11,496 instances representing 499 patients was used in the development of different Ensemble-based learning methods. The dataset contains the previous susceptibility testing such as Gram stain, susceptibility results, and with some patients, demographic information including age and sex, and accuracy ranging between 75.5% to 76.9% was reported using stacked ensemble-based method. Unlike the majority-voting approach in bagging (e.g. RF), stacked ensemble constructs the ensemble by training the model to perform the aggregation of the predictions of the different models [105]. It should be pointed out that these results were achieved without considering the patient's clinical information, such as the source of infection, presence of active infection, and other factors. Furthermore, the authors had to apply oversampling methods using variants of SMOTE to handle the class-imbalance in the dataset.

Martínez-Agüero *et al.* [20] used clinical and demographic data as well as data from cultures and antibiograms to develop machine learning models to identify antibiotic resistance profiles of bacteria isolated from patients in intensive care units. The authors explored several machine learning algorithms including Logistic Regression, *k*-Nearest Neighbour (KNN), and RF. A dataset representing information of 2630 patients was used in the study to determine the resistance of *Pseudomonas* to six families of antimicrobials: Aminoglycosides (AMG), Carbapenemics (CAR), Fourth-generation Cephalosporins (CF4), Broad-spectrum antibiotics (PAP), Polymixines (POL) and Quinolones (QUI). Promising results were shown in terms of accuracy reaching 90% in some cases (QUI), and *F1*-score ranging between 58% and 90%. The authors reported various pre-processing steps before applying machine learning algorithms. These include features selections and engineering to reduce the number of features in the data, and the handling of class imbalance using undersampling-based methods, aiming at reducing the number of majority class instances in the dataset. It is worth pointing out that, undersampling methods, although easy to implement and can lead to better results in some domains, often cause information loss, and may result in reducing the overall accuracy of the model [54], [56], especially in medical-related data.

Similarly, Feretzakis *et al.* [21] used demographic information and Gram-stain data, and other clinical patients characterising features to inform the choice of antibiotics in intensive care units. A dataset representing patients admitted to the intensive care units in a Greek hospital was created over a period of two years. Features such as age, gender, sample type, and Gram stain were used as predictors of antimicrobial susceptibility. Using a dataset of 2307 instances and an open-source java-based framework [106], the authors built an experimental framework utilising different methods including RF, Decision Trees, SVM, KNN and Multilayer Perceptron (MLP). MLP and Decision Trees achieved top performance in terms of $F1-$Score, while RF ranked top with respect to area under the curve (ROC). It should be pointed out that the authors in this work used data from the microbiology laboratory that already exists in the hospital, and were able to achieve results that allow anticipation of culture sensitivity results from the labs. These results again suggest that readily available clinical and antimicrobial susceptibility data can be used to predict resistance and provide a decision support tool to clinicians. This approach is clearly much less resource-consuming than machine learning and deep learning methods that employ genome-based data [39]–[41], [107] which proved to be very accurate, yet expensive and more difficult to implement.

Canovas-Segura *et al.* [22] used a range of machine learning algorithms to build predictive models for antimicrobial resistance using different learning algorithms including Logistic Regression, Decision Trees and Rule-based models, implemented using the caret-package in R [108]. The motivation behind these choices, is the use of models that can be easily interpreted, unlike RF, MLP, and other more complicated models. In implementing the Logistic Regression, the authors used a regularisation approach (Least Absolute Shrinkage and Selection Operator) to account for overfitting and improve the model's generalisation. A dataset representing 1393 positive Enterococci cultures was used for evaluation purposes. Similar to other health-related datasets, the dataset used in this experiment was imbalanced, and the authors applied random oversampling to obtain a balanced dataset. The authors also used a sliding window technique to account for possible concept drift (i.e. considering more recent examples to be more relevant). Using Area Under the Curve (AUC) to evaluate the models, Logistic Regression and oversampling with sliding window achieved the best results of 0.79.

In [88], the authors built a model to predict the antimicrobial susceptibility of clinical bacterial isolates. Due to the lack of historical data, they used the results of the antibiotic susceptibility test for evaluation and utilised association rules to perform the predictions. Relatively high accuracy was reported (between 85% to 90%) but no details of the dataset (for example the size of the data, class distribution) and no evaluation metrics other than accuracy were provided. Furthermore, association rules, and as the name suggests are rule-based mining methods, that may not generalise on unseen examples in this context, unlike other complex models, for example, ensemble-based methods.

Concept drift is an inherently challenging problem, in particular when handling and analysing streaming or non-stationary data [109], where data may evolve or change over time. Puuronen *et al.* [23] used Ensemble Learning methods with local features selection to predict the sensitivity of a pathogen to an antibiotic based on data about the antibiotic, the isolated pathogen, and the demographic and clinical features of the patient. To evaluate the proposed methods, data collected over a period of two years (2002 to 2004) representing more than 4000 patients were used. Each record of the data represents one antimicrobial susceptibility test which includes the isolated pathogen, the antibiotic panel used in the test, and the result of the test (sensitive, resistant). Demographic features (sex, age) and clinical information such as department, days spent in the ICU, days at the hospital, were used to build the predictive model. The proposed methods were implemented using an open-source Java platform [106] and the results show that features selection can improve the performance of ensemble learning in this context. It should be noted, however, that a period of two years (period of the study), may not pose a challenge in terms of concept drift, and therefore studies over longer periods are needed.

To overcome the problem of 'empirical' antimicrobial treatment, in the absence of susceptibility testing, Yelling *et al.* [78] used demographic data (age, gender, home residence), date of sampling, past urine culture results, and patients history of drug purchases to build a predictive model using Gradient Boosting and Logistic Regression. A dataset of more than 700,000 records representing 315, 047 patients were used for evaluation purposes. Findings showed strong associations of AMR with demographic information of patients, past urine culture, and their history of drug purchase.

Elpiniki [91] proposed a fuzzy-based method for recommending appropriate antibiotic treatment for uncomplicated urinary tract infections. A set of heuristics (hard-coded rules) based on clinical data was used to build the system. Typical rules used in this work include conditions of patient's age and their allergy to a particular drug. The overall accuracy of 86% was reported. However, the hard-coded rules used and the size of the dataset (38 patients) make it difficult to assess the model performance across unseen and more diverse patient datasets.

Moran *et al.* [86] used a dataset of 9352 patients with 15695 admissions collected over six years in Birmingham city-UK to predict the presence of antibiotic resistance to co-amoxiclav and piperacillin/tazobactam. The dataset used contains information about urine and blood cultures, demographic and prescribing data. The authors used Extreme Gradient Boosting (XGBoost), another ensemble learning-based method [110], to build the predictive model. Comparable results with medical staff selection of antibiotics for patients were reported. Interestingly, the authors reported that the model was performing better on urine cultures data compared to blood cultures. AUC for piperacillin/tazobactam and co-amoxiclav were 0.70 and 0.71, respectively, for urine culture, while for blood cultures AUC values were 0.66 and 0.67, respectively.

Vazquez-Guillamet *et al.* [79] used stepwise Logistic Regression with backward elimination and Decision Trees to predict antimicrobial resistance in Gram-negative bacteria. Three types of antibiotics were considered in this study piperacillin-tazobactam (PT), cefepime (CE), and meropenem (ME) using a dataset representing 1618 patients with sepsis caused by Gram-negative bacteria. The model achieved AUC values of 0.68, 0.63, and 0.83 for PT, CE, and ME, respectively. The authors reported that with simple features, decision trees provide a potential solution to distinguish patients at low, intermediate, or high risk of resistance to PT, CE, and ME. Clearly, the results reported are not high but show the potential use of machine learning to inform antibiotic prescribing and practices, and suggest that with larger and more diverse datasets, performance comparable to clinicians can be achieved.

Chow *et al.* [90] presented a study to assess the impact of using 'Antibiotic computerised decision support systems' (CDSSs) on antimicrobial management and prescriptions. Methods used are based on rule-based algorithms to provide guidance on antibiotic selection and dosing, and were developed based on hospitals antimicrobial stewardship committee. Using a dataset of 1886 patients, 25% of which received antibiotic treatment based on CDSSs recommendation, they showed that CDSSs reduced the mortality risk of patients under the age of 65, and without increasing the risk for older patients.It should be pointed out that the authors used only first prescription for 'empirical therapy' per patient during the study period (between October 2011 and September 2021), in other words before of the infection-causing microorganism.

To inform antimicrobial prescribing practices, the authors of [87] proposed a machine-learning-based approach to assess Antimicrobial Utilisation (AU). A dataset collected from patient health records was used, where antimicrobial information, movement data, demographic, and billing data over a period of two years were used to build the machine learning models. Extensive features engineering were applied to the data. This includes removing duplicates, removing highly correlated features, using statistical methods (e.g. chi-squared test) to remove redundant categorical features, and so on. Several experiments using various learning algorithms were carried out, and Support Vector Regression and Cubic Regression Models proved to be superior over other machine learning methods in this context.

Wange *et al.* [89] used multivariate regression to understand the relation between antibiotic usage and antibiotic resistance in hospitals using a dataset of more than 9000 isolates collected from 37 hospitals. Data used in the regression model included antibiotic resistance indices, hospital type, patient-days, and antibiotic types, and the primary outcome (label) is antibiotic susceptibility. Results showed an unexpected increase in antibiotic susceptibility for hospital-specific antibiotic usage.

In an attempt to mitigate the risk of antimicrobial resistance for COVID-19 patients, Rawson *et al.* [85] presented a method to predict bacterial infection in hospitalised COVID-19 patients. The authors used demographic and microbiology data of 1186 patients (60% COVID-19 positive). A 10-fold stratified cross-validation used to train Gaussian Naive Bayes, SVM, and ANN models. SVM and using 21 commonly available blood tests variables showed superior performance over other models.

Despite the relatively small size of the datasets used in the literature and the limited number of patients (between 2000 to 10,000), it is evident that machine learning algorithms based on demographic and clinical data provide a promising tool to enhance antibiotic prescribing practices. In addition, it is clear that certain features proved to be extremely useful in building predictive models. In particular, age, gender, how the infection were acquired, time of admission, and others. It should be also noted, most of the clinical and demographic data used in the work above, is readily available at hospitals, or can be easily collected and recorded, which provides unique opportunity to help address the AMR problem.

C. BACTERIAL GENOMIC DATA

In recent years, promising results were reported using bacterial genome-based machine learning methods, mainly due to the increasing availability of whole-genome sequence datasets [6]. Several methods showed promising potential as a decision support tool for clinicians [6], [39]–[41], [111].

Zhichang *et al.* [6] used SVM and Set Covering Machine (SCM) models to predict the resistance of the five drugs (Tetracycline, Ampicillin, Sulfisoxazole, Trimethoprim, and Enrofloxacin). Input data included genomes of the isolates and the reference genes, and an accuracy greater than 90% was reported on the testing sets.

Similarly, Nguyen, *et al.* [39] used a dataset of almost 5000 non-typhoidal *Salmonella* genomes collected over a period of 15 years in the US to train XGBoost model for predicting the Minimum Inhibitory Concentration (MIC) for 15 antibiotics and reported an average accuracy of 95%.

Moradigaravand *et al.* [40] used different machine learning models to predict resistance to four different lines of antibiotic drugs using genome sequences of Escherichia coli strains. Results showed that decision trees and gradient boosting (ensemble-based method) consistently outperformed other methods with the highest accuracy on the testing set reaching 91%.

Current evidence would suggest that machine learning methods based on genomic data provide superior results over approaches based on clinical data. This can be partly attributed to the several datasets that are available in the public domain. For example Antibiotic Resistance Genes Database (ARDB) [51], and the Comprehensive Antibiotic Research Database (CARD) integrates disparate molecular and sequence data [52]. A mini-review that appeared recently [18] showed that a combination of machine learning algorithms and lab testing not only accelerates the process of discovering new antimicrobials but also produces accurate results that outperformed traditional methods. However, whole-genome sequencing is relatively costly and unlikely to be available in LMICs for diagnostic purposes, unlike demographic, clinical, and antimicrobial susceptibility data often routinely available through hospital and laboratory registries.

IV. CHALLENGES & OPPORTUNITIES

The latest developments in the field of machine learning and the availability of large volumes of structured and unstructured data to train predictive models are expected to advance existing practices across a range of healthcare domains including control of antimicrobial prescribing [107]. Recent literature shows that the training of machine learning models to use readily-available information at medical facilities such as patient demographic, clinical, and microbiological data provide a promising and inexpensive solution to inform antibiotic prescription practices and in turn contain the AMR problem.

A. CHALLENGES

Despite the recent attempts to use machine learning to predict antimicrobial resistance, the application of machine learning in this context is still limited. Existing research shows some key barriers to taking full advantage of the significant progress in the area of machine learning and applying it to tackle AMR at the local and global levels. Key challenges based on the reviewed literature be outlined as follows:

- Machine learning and deep learning methods largely depend on data availability and data quality. Unfortunately, most attempts either use limited and proprietary datasets (not available in the public domain), or datasets available to a particular geographic area. This makes it difficult to benchmark and compare different methods. To accelerate research and development in this area, datasets need to be made available to the public domain after meeting all ethical and legal requirements. In particular, data related to cultures, isolates, and corresponding demographic and clinical data. Existing work shows promising results in informing antibiotics prescription and management. However, the relatively small datasets used (e.g. 38 patients [91], 243 patients [3], 499 patients [36], . . .), makes it difficult to assess how these solutions can be generalised to unseen data. In other words, larger, and more diverse datasets need to be created and made available in the public domain, similar to bacterial genomes with AMR-related datasets, were several of these already available on the public domain [51], [52], [107].
- The inherent problematic characteristic of the data itself, and in particular the class imbalance or the skewed representation of data. In AMR datasets, this is a common problem [20], [22], and therefore such imbalance must be handled before modelling the data to avoid biased results. Various methods exist in the literature that rely on data sampling methods or algorithmic modifications can be used to improve the quality and data distribution [54]. However, it should be noted also that most data sampling-based methods, can lead to information loss and reduce the overall accuracy of the model. In particular, if undersampling based methods were applied to reduce the data imbalance [56]. Data oversampling

methods are common approaches that are also extensively used to handle the imbalance by synthesising more data instances. Methods such as SMOTE [68] or its variants [112]. However, the flawed application of these methods can lead to overfitting [113]. Ongoing recording of data at the points of care can help build a rich dataset that is more diverse and more representative of patients' demographic and clinical information.

- Concept drift is another challenging problem [114], in particular when handling and analysing streaming or non-stationary data [109]. Concept-drift means that data evolve or change over time, a likely feature of AMR datasets, given that bacteria develop resistance to antimicrobials over time.Existing work showed promising results in terms of accounting of this challenge [22], [23] using various machine learning methods. However, data collected over longer periods of time would provide a unique opportunity to study bacteria resistance over time, as well as push the research boundaries in a very challenging machine learning topic (detecting concept drift)..
- The significant progress that took place in the area of machine learning, can be largely attributed to the use of more complex models. Some of these complex models are even considered black boxes, especially DL-based methods. [115]. In the medical domain, and related applications such as AMR, explaining and understanding the model's outcome is important. Simple models such as Logistic Regression and Decision Trees are easier to understand, and results can be easily interpreted. Literature also shows that such methods can achieve reasonable results with limited clinical and demographic data [22], [79]. However, and as can be seen in Table 1, kernel-based models such as SVM, or ensemble-based methods such as RF, proved to be more accurate in AMR settings, although at the expense of results understanding and interpretation. In other words, there is a trade-off between performance, results understanding, and the model's explainability. For an extensive review of explainability methods, the reader is referred to [116].

It should be noted that almost all of the aforementioned challenges are related to data quality and data availability as well as the lack of coordinated efforts between experts from different domains (e.g. machine learning community, medical staff, and government and regulatory bodies) make the clinical application of data-driven solutions that utilise machine learning a distant prospect. Such coordination between the different stakeholders can also help streamline the process of ethical approvals, patient's privacy, and other ethical considerations [117].

B. OPPORTUNITIES

The role of machine learning may prove critical in the fight against AMR. Machine learning-based methods proved capable of significant predictive power to inform rational and targeted antimicrobial prescribing choices that will minimize misuses of antibiotics. To scale up and accelerate research and development in this area and make a contribution to addressing this global threat to public health, machine learning and data-driven solutions must become incorporated in the strategies underpinning Antimicrobial Stewardship programmes across different countries.

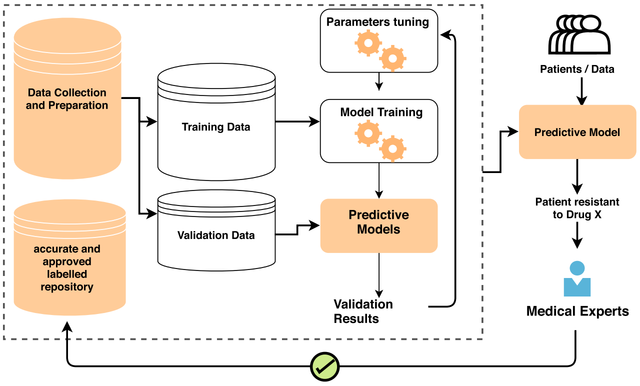


FIGURE 3. Machine learning-driven framework.

Figure 3 shows a schematic diagram for a possible pipeline of work that could be part of any antibiotics stewardship programme. This includes strategies for collecting, recording and digitising demographic, clinical, and microbiological data. Such registries can provide a rich source for the creation of large, diverse, and rich datasets and unique opportunities to accelerate research and development of data-driven solutions for AMR-informed prescription and management.

Continuous collection and recording of relevant data will help train and deploy accurate predictive models based on supervised machine learning algorithms. The literature so far shows clearly that even off-the-shelf machine learning algorithms provide reasonable results. We expect an improved performance by predictive models based on larger and more diverse datasets. The wide range of open-access tools that implement different advanced supervised learning algorithms, for example, R (e.g. caret package) [108], and Python (scikit-learn library [118]) facilitates the adoption of such strategy in resource-limited countries by reducing infrastructure costs associated with data collection, storage, and machine learning solution development and deployment.

A key element to the successful implementation of such an ambitious framework is a collaborative approach between medical staff and machine learning specialists. As shown in Figure 3, data collection and model training and evaluation are followed by validation by medical staff before models are refined and improved accordingly. Such an iterative and collaborative approach will provide a large and accurately labelled dataset that can be also a very rich source for pushing the research boundaries for the application of state-of-the-art machine learning algorithms in AMR and beyond.

V. CONCLUSION

AMR is considered one of the top ten global health threats according to WHO. In response to this growing threat, resource-rich countries adopted and implemented antimicrobial stewardship programs to manage the use and prescription of antibiotics. However, in LMICs where AMR rates are the highest, very little has been done to combat AMR as a result of unfavourable political and socio-economic factors, widespread poverty, and inadequate resources.

Recent development in the field of machine learning and deep learning provides a unique opportunity to improve existing practices in relation to managing and controlling the use of antibiotics and combating AMR. In this paper, we presented a critical and in-depth technical review of recent studies that used machine learning-driven solutions to combat AMR. The review shows clearly that using off-the-shelf learning algorithms can provide a better and more informed approach to antibiotic prescribing. These promising results were achieved despite the relatively small datasets used for training the models, and the simple learning methods that were used such as Logistic Regression and Decision Trees.

Machine learning solutions can provide a data-driven framework to improve antibiotic prescribing practices in LMICs where management of AMR remains most challenging. Relevant datasets required to train machine learning models to inform antibiotic prescribing are readily available at most points of care (e.g. demographic and clinical information of patients), and were not available, collection and digitisation of such datasets is neither technically challenging nor expensive. In addition, open-source platforms can then be used to build and deploy such solutions. Therefore, the application of such solutions in LMICs is relatively inexpensive and provides a unique opportunity to LMICs countries to handle such a threat.

The paper also presented and critically discussed technical and practical barriers to the scale up the use of machine learning-driven solutions to combat antimicrobial resistance. These include data-related challenges such as the size and the diversity of datasets and concept drift. We also highlighted potential factors hampering the clinical application of machine learning-driven solutions and suggested a framework that brings together policy/ decision-makers, medical experts and machine learning engineers to work together toward the integration of intelligent solutions with antibiotic stewardship programmes.

Finally, a future direction towards more coordinated efforts between the different stakeholders and ongoing recording and monitoring of AMR-related data is presented in this paper. Such data can be the driving force to manage and minimise the threat of AMR, taking into consideration, that existing literature shows that basic demographic information of the patients (e.g. age, gender, residential information, and access to sanitation facilities), along with clinical information such as times of admission to hospitals, lengths of stay, source of infection, antibiotic received and others, proved to be powerful features, and even with 'off-the-shelf' machine learning algorithms, reasonable results were achieved.

REFERENCES

[1] S. Mayor, "First WHO antimicrobial surveillance data reveal high levels of resistance globally," *Brit. Med. J.*, vol. 360, p. k462, Jan. 2018. [Online]. Available: https://www.bmj.com/content/360/bmj.k462

[2] B. T. Tadesse, E. A. Ashley, S. Ongarello, J. Havumaki, M. Wijegoonewardena, I. J. González, and S. Dittrich, "Antimicrobial resistance in Africa: A systematic review," *BMC Infectious Diseases*, vol. 17, no. 1, p. 616, Dec. 2017, doi: 10.1186/s12879-017-2713-1.

[3] M. Oonsivilai, Y. Mo, N. Luangasanatip, Y. Lubell, T. Miliya, P. Tan, L. Loeuk, P. Turner, and B. S. Cooper, "Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia," *Wellcome Open Res.*, vol. 3, p. 131, Oct. 2018.

[4] B. P. Bougnom and L. J. V. Piddock, "Wastewater for urban agriculture: A significant factor in dissemination of antibiotic resistance," *Environ. Sci. Technol.*, vol. 51, no. 11, pp. 5863–5864, Jun. 2017, doi: 10.1021/acs.est.7b01852.

[5] M. N. Anahtar, J. H. Yang, and S. Kanjilal, "Applications of machine learning to the problem of antimicrobial resistance: An emerging model for translational research," *J. Clin. Microbiol.*, vol. 59, no. 7, Jun. 2021, Art. no. e01260, doi: 10.1128/jcm.01260-20.

[6] Z. Liu, D. Deng, H. Lu, J. Sun, L. Lv, S. Li, G. Peng, X. Ma, J. Li, Z. Li, T. Rong, and G. Wang, "Evaluation of machine learning models for predicting antimicrobial resistance of actinobacillus pleuropneumoniae from whole genome sequences," *Frontiers Microbiol.*, vol. 11, p. 48, Feb. 2020.

[7] T. M. Rawson, L. S. P. Moore, N. Zhu, N. Ranganathan, K. Skolimowska, M. Gilchrist, G. Satta, G. Cooke, and A. Holmes, "Bacterial and fungal coinfection in individuals with coronavirus: A rapid review to support COVID-19 antimicrobial prescribing," *Clin. Infectious Diseases*, vol. 71, pp. 2459–2468, May 2020, doi: 10.1093/cid/ciaa530.

[8] R. Sirijatuphat, K. Sripanidkulchai, A. Boonyasiri, P. Rattanaumpawan, O. Supapueng, P. Kiratisin, and V. Thamlikitkul, "Implementation of global antimicrobial resistance surveillance system (glass) in patients with bacteremia," *PLoS ONE*, vol. 13, no. 1, pp. 1–15, 2018, doi: 10.1371/journal.pone.0190132.

[9] I. Morrissey, M. Hackel, R. Badal, S. Bouchillon, S. Hawser, and D. Biedenbach, "A review of ten years of the study for monitoring antimicrobial resistance trends (SMART) from 2002 to 2011," *Pharmaceuticals*, vol. 6, no. 11, pp. 1335–1346, Nov. 2013, doi: 10.3390/ph6111335.

[10] K. Hijazi, C. Joshi, and I. M. Gould, "Challenges and opportunities for antimicrobial stewardship in resource-rich and resource-limited countries," *Expert Rev. Anti-Infective Therapy*, vol. 17, no. 8, pp. 621–634, Aug. 2019, doi: 10.1080/14787210.2019.1640602.

[11] S. Devi, "AMR in the middle east: 'A perfect storm,'" *Lancet*, vol. 394, no. 10206, pp. 1311–1312, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140673619323062

[12] L. A. Wilson, S. R. Van Katwyk, P. Fafard, A. M. Viens, and S. J. Hoffman, "Lessons learned from COVID-19 for the post-antibiotic future," *Globalization Health*, vol. 16, no. 1, pp. 1–3, Oct. 2020, doi: 10.1186/s12992-020-00623-x.

[13] *Antimicrobial Stewardship Programmes in Health-Care Facilities in Low- and Middle-Income Countries: A WHO Practical Toolkit*, World Health Organization, Geneva, Switzerland, 2019.

[14] J. Hsu, "How COVID-19 is accelerating the threat of antimicrobial resistance," *Brit. Med. J.*, vol. 369, p. m1983, May 2020, doi: 10.1136/bmj.m1983.

[15] S. Tomczyk, A. Taylor, A. Brown, M. E. A. de Kraker, A. El-Saed, M. Alshamrani, R. S. Hendriksen, M. Jacob, S. Löfmark, O. Perovic, N. Shetty, D. Sievert, R. Smith, J. Stelling, S. Thakur, A. C. Vietor, and T. Eckmanns, "Impact of the COVID-19 pandemic on the surveillance, prevention and control of antimicrobial resistance: A global survey," *J. Antimicrobial Chemotherapy*, vol. 76, no. 11, pp. 3045–3058, Oct. 2021, doi: 10.1093/jac/dkab300.

[16] G. M. Knight, R. E. Glover, C. F. McQuaid, I. D. Olaru, K. Gallandat, Q. J. Leclerc, N. M. Fuller, S. J. Willcocks, R. Hasan, E. van Kleef, and C. I. Chandler, "Antimicrobial resistance and COVID-19: Intersections and implications," *eLife*, vol. 10, p. e64139, Feb. 2021, doi: 10.7554/eLife.64139.

[17] B. Egyir, N. Obeng-Nkrumah, and G. B. Kyei, "COVID-19 pandemic and antimicrobial resistance: Another call to strengthen laboratory diagnostic capacity in Africa," *Afr. J. Lab. Med.*, vol. 9, no. 1, pp. 1–4, Sep. 2020, doi: 10.4102/ajlm.v9i1.1302.

[18] H. J. Lau, C. H. Lim, S. C. Foo, and H. S. Tan, "The role of artificial intelligence in the battle against antimicrobial-resistant bacteria," *Current Genet.*, vol. 67, no. 3, pp. 421–429, Feb. 2021, doi: 10.1007/s00294-021-01156-5.

[19] N. Macesic, F. Polubriaginof, and N. P. Tatonetti, "Machine learning," *Current Opinion Infectious Diseases*, vol. 30, no. 6, pp. 511–517, Dec. 2017, doi: 10.1097/qco.0000000000000406.

[20] S. Martínez-Agüero, I. Mora-Jiménez, J. Lérida-García, J. Álvarez-Rodríguez, and C. Soguero-Ruiz, "Machine learning techniques to identify antimicrobial resistance in the intensive care unit," *Entropy*, vol. 21, no. 6, p. 603, Jun. 2019. [Online]. Available: https://www.mdpi.com/1099-4300/21/6/603

[21] G. Feretzakis, E. Loupelis, A. Sakagianni, D. Kalles, M. Martsoukou, M. Lada, N. Skarmoutsou, C. Christopoulos, K. Valakis, A. Velentza, S. Petropoulou, S. Michelidou, and K. Alexiou, "Using machine learning techniques to aid empirical antibiotic therapy decisions in the intensive care unit of a general hospital in Greece," *Antibiotics*, vol. 9, no. 2, p. 50, Jan. 2020.

[22] B. Cánovas-Segura, A. Morales, A. Lopez Martinez-Carrasco, M. Campos, J. M. Juarez, L. López Rodríguez, and F. Palacios, "Improving interpretable prediction models for antimicrobial resistance," in *Proc. IEEE 32nd Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2019, pp. 543–546.

[23] S. Puuronen, M. Pechenizkiy, and A. Tsymbal, "Effectiveness of local feature selection in ensemble learning for prediction of antimicrobial resistance," in *Proc. 21st IEEE Int. Symp. Computer-Based Med. Syst.*, Jun. 2008, pp. 632–637.

[24] R. Kanapathipillai, N. Malou, K. Baldwin, P. Marty, C. Rodaix, C. Mills, P. Herard, and M. Saim, "Antibiotic resistance in palestine: An emerging part of a larger crisis," *Brit. Med. J.*, vol. 363, p. k4273, Oct. 2018. [Online]. Available: https://www.bmj.com/content/363/bmj.k4273

[25] V. X. Liu, V. Fielding-Singh, J. D. Greene, J. M. Baker, T. J. Iwashyna, J. Bhattacharya, and G. J. Escobar, "The timing of early antibiotics and hospital mortality in sepsis," *Amer. J. Respiratory Crit. Care Med.*, vol. 196, no. 7, pp. 856–863, Oct. 2017.

[26] M. R. Pulido, M. Garcia-Quintanilla, R. Martin-Pena, J. M. Cisneros, and M. J. McConnell, "Progress on the development of rapid methods for antimicrobial susceptibility testing," *J. Antimicrobial Chemotherapy*, vol. 68, no. 12, pp. 2710–2717, Dec. 2013, doi: 10.1093/jac/dkt253.

[27] N. Abu-Shaban, "Gaza shootings: An orthopaedic crisis and mass disability," *Brit. Med. J.*, vol. 362, p. k3295, Aug. 2018, doi: 10.1136/bmj.k3295.

[28] Z. Alaali and A. S. Bin Thani, "Patterns of antimicrobial resistance observed in the Middle East: Environmental and health care retrospectives," *Sci. Total Environ.*, vol. 740, Oct. 2020, Art. no. 140089. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969720336093

[29] M. Zeitoun, G. Abu-Sittah, R. A. Shomar, and N. El Achi, "AMR and COVID-19 on the frontline: A call to rethink war, WASH, and public health," *Ann. Global Health*, vol. 87, no. 1, p. 21, Feb. 2021.

[30] A. A. Elmanama, N. A. A. Laham, and G. A. Tayh, "Antimicrobial susceptibility of bacterial isolates from burn units in Gaza," *Burns*, vol. 39, no. 8, pp. 1612–1618, Dec. 2013, doi: 10.1016/j.burns.2013.04.011.

[31] Z. Astal, A. El-Manama, and F. A. Sharif, "Antibiotic resistance of bacteria associated with community-acquired urinary tract infections in the southern area of the gaza strip," *J. Chemotherapy*, vol. 14, no. 3, pp. 259–264, Jan. 2002, doi: 10.1179/joc.2002.14.3.259.

[32] Z. Astal, "Susceptibility patterns inPseudomonas aeruginosaCausing nosocomial infections," *J. Chemotherapy*, vol. 16, no. 3, pp. 264–268, Jun. 2004, doi: 10.1179/joc.2004.16.3.264.

[33] S. Alyacoubi, Y. Abuowda, L. Albarqouni, B. Böttcher, and K. Elessi, "Inpatient management of community-acquired pneumonia at the European gaza hospital: A clinical audit," *Lancet*, vol. 391, p. S40, Feb. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140673618304069

[34] R. H. Rida, N. A. Al Laham, and A. A. Elmanama, "Carbapenem resistance among clinical and environmental gram-negative isolates recovered from hospitals in gaza strip, Palestine," *Germs*, vol. 8, no. 3, pp. 147–154, Sep. 2018.

[35] M. T. Baysari, E. C. Lehnbom, L. Li, A. Hargreaves, R. O. Day, and J. I. Westbrook, "The effectiveness of information technology to improve antimicrobial prescribing in hospitals: A systematic review and meta-analysis," *Int. J. Med. Informat.*, vol. 92, pp. 15–34, Aug. 2016.

[36] G. Feretzakis, A. Sakagianni, E. Loupelis, D. Kalles, N. Skarmoutsou, M. Martsoukou, C. Christopoulos, M. Lada, S. Petropoulou, A. Velentza, S. Michelidou, R. Chatzikyriakou, and E. Dimitrellos, "Machine learning for antibiotic resistance prediction: A prototype using off-the-shelf techniques and entry-level data to guide empiric antimicrobial therapy," *Healthcare Informat. Res.*, vol. 27, no. 3, pp. 214–221, Jul. 2021, doi: 10.4258/hir.2021.27.3.214.

[37] L. R. Kolozsvári, J. Kónya, J. Paget, F. G. Schellevis, J. Sándor, G. J. Szőllősi, S. Harsányi, Z. Jancsó, and I. Rurik, "Patient-related factors, antibiotic prescribing and antimicrobial resistance of the commensal Staphylococcus aureus and streptococcus pneumoniae in a healthy population–Hungarian results of the APRES study," *BMC Infectious Diseases*, vol. 19, no. 1, Mar. 2019, doi: 10.1186/s12879-019-3889-3.

[38] R. Ben-Ami, J. Rodríguez-Baño, H. Arslan, J. D. D. Pitout, C. Quentin, E. S. Calbo, Ö. K. Azap, C. Arpin, A. Pascual, D. M. Livermore, J. Garau, and Y. Carmeli, "A multinational survey of risk factors for infection with extended-spectrum β-lactamase–producing enterobacteriaceae in nonhospitalized patients," *Clin. Infectious Diseases*, vol. 49, no. 5, pp. 682–690, Sep. 2009, doi: 10.1086/604713.

[39] M. Nguyen, S. W. Long, P. F. McDermott, R. J. Olsen, R. Olson, R. L. Stevens, G. H. Tyson, S. Zhao, and J. J. Davis, "Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella," *J. Clin. Microbiol.*, vol. 57, no. 2, Feb. 2019. [Online]. Available: https://jcm.asm.org/content/57/2/e01260-18

[40] D. Moradigaravand, M. Palm, A. Farewell, V. Mustonen, J. Warringer, and L. Parts, "Prediction of antibiotic resistance in *Escherichia Coli* from large-scale pan-genome data," *PLOS Comput. Biol.*, vol. 14, no. 12, pp. 1–17, 2018. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1006258

[41] R. Colbaugh and K. Glass, "Predicting antimicrobial resistance via lightly-supervised learning," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 2428–2433.

[42] M. Helmy, M. Awad, and K. A. Mosa, "Limited resources of genome sequencing in developing countries: Challenges and solutions," *Appl. Transl. Genomics*, vol. 9, pp. 15–19, Jun. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212066116300205

[43] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[44] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org.

[45] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" *J. Mach. Learn. Res.*, vol. 15, pp. 3133–3181, Jan. 2014. [Online]. Available: http://jmlr.org/papers/v15/delgado14a.html

[46] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Med.*, vol. 25, no. 1, pp. 24–29, Jan. 2019, doi: 10.1038/s41591-018-0316-z.

[47] R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: Current issues and guidelines," *Int. J. Med. Inform.*, vol. 77, no. 2, pp. 81–97, Feb. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1386505606002747

[48] P. J. Lisboa and A. F. G. Taktak, "The use of artificial neural networks in decision support in cancer: A systematic review," *Neural Netw.*, vol. 19, no. 4, pp. 408–415, May 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608005002844

[49] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications—Clinical benefits and future directions," *Int. J. Med. Informat.*, vol. 73, no. 1, pp. 1–23, Feb. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1386505603002119

[50] N. K. Tran, S. Albahra, L. May, S. Waldman, S. Crabtree, S. Bainbridge, and H. Rashidi, "Evolving applications of artificial intelligence and machine learning in infectious diseases testing," *Clin. Chem.*, vol. 68, no. 1, pp. 125–133, Dec. 2021, doi: 10.1093/clinchem/hvab239.

[51] B. Liu and M. Pop, "ARDB-antibiotic resistance genes database," *Nucleic Acids Res.*, vol. 37, pp. D443–D447, Jan. 2009, doi: 10.1093/nar/gkn656.

[52] A. G. McArthur, N. Waglechner, F. Nizam, A. Yan, M. A. Azad, A. J. Baylay, K. Bhullar, M. J. Canova, G. De Pascale, L. Ejim, and L. Kalan, "The comprehensive antibiotic resistance database," *Antimicrobial Agents Chemotherapy*, vol. 57, no. 7, pp. 3348–3357, Jul. 2013, doi: 10.1128/AAC.00419-13.

[53] M. Oonsivilai, Y. Mo, N. Luangasanatip, Y. Lubell, T. Miliya, P. Tan, L. Loeuk, P. Turner, and B. S. Cooper, "Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia," *Wellcome Open Res.*, vol. 3, p. 131, Oct. 2018, doi: 10.12688/wellcomeopenres.14847.1.

[54] P. Vuttipittayamongkol, E. Elyan, and A. Petrovski, "On the class overlap problem in imbalanced data classification," *Knowl.-Based Syst.*, vol. 212, Jan. 2021, Art. no. 106631. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705120307607

[55] P. Vuttipittayamongkol and E. Elyan, "Neighbourhood-based undersampling approach for handling imbalanced and overlapped data," *Inf. Sci.*, vol. 509, pp. 47–70, Jan. 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025519308114

[56] E. Elyan, C. F. Moreno-Garcia, and C. Jayne, "CDSMOTE: Class decomposition and synthetic minority class oversampling technique for imbalanced-data classification," *Neural Comput. Appl.*, vol. 33, no. 7, pp. 2839–2851, Apr. 2021, doi: 10.1007/s00521-020-05130-z.

[57] E. Elyan and M. M. Gaber, "A genetic algorithm approach to optimising random forests applied to class engineered data," *Inf. Sci.*, vol. 384, pp. 220–234, Apr. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025516305783

[58] E. Elyan and M. M. Gaber, "A fine-grained random forests using class decomposition: An application to medical diagnosis," *Neural Comput. Appl.*, vol. 27, no. 8, pp. 2279–2288, Nov. 2016, doi: 10.1007/s00521-015-2064-z.

[59] P. Vuttipittayamongkol and E. Elyan, "Improved overlap-based undersampling for imbalanced dataset classification with application to epilepsy and Parkinson's disease," *Int. J. Neural Syst.*, vol. 30, no. 8, Jul. 2020, Art. no. 2050043, doi: 10.1142/s0129065720500434.

[60] P. Vuttipittayamongkol and E. Elyan, "Overlap-based undersampling method for classification of imbalanced medical datasets," in *Artificial Intelligence Applications and Innovations*, I. Maglogiannis, L. Iliadis, and E. Pimenidis, Eds. Cham, Switzerland: Springer, 2020, pp. 358–369.

[61] J. C. Hyun, E. S. Kavvas, J. M. Monk, and B. O. Palsson, "Machine learning with random subspace ensembles identifies antimicrobial resistance determinants from pan-genomes of three pathogens," *PLOS Comput. Biol.*, vol. 16, no. 3, Mar. 2020, Art. no. e1007608.

[62] F. Maguire, M. A. Rehman, C. Carrillo, M. S. Diarra, and R. G. Beiko, "Identification of primary antimicrobial resistance drivers in agricultural nontyphoidal Salmonella enterica serovars by using machine learning," *mSystems*, vol. 4, no. 4, p. e00211, Aug. 2019.

[63] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, Jun. 2002. [Online]. Available: http://dl.acm.org/citation.cfm?id=1622407.1622416

[64] A. S. Chowdhury, D. R. Call, and S. L. Broschat, "Antimicrobial resistance prediction for gram-negative bacteria via game theory-based feature evaluation," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Dec. 2019.

[65] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data," *J. Big Data*, vol. 5, no. 1, pp. 1–30, Nov. 2018, doi: 10.1186/s40537-018-0151-6.

[66] E. S. Kavvas, E. Catoiu, N. Mih, J. T. Yurkovich, Y. Seif, N. Dillon, D. Heckmann, A. Anand, L. Yang, V. Nizet, J. M. Monk, and B. O. Palsson, "Machine learning and structural analysis of mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance," *Nature Commun.*, vol. 9, no. 1, pp. 1–9, Dec. 2018.

[67] P.-J. Van Camp, D. B. Haslam, and A. Porollo, "Prediction of antimicrobial resistance in gram-negative bacteria from whole-genome sequencing data," *Frontiers Microbiol.*, vol. 11, p. 1013, May 2020.

[68] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[69] A. Ali-Gombe and E. Elyan, "MFC-GAN: Class-imbalanced dataset classification using multiple fake class generative adversarial network," *Neurocomputing*, vol. 361, pp. 212–221, Oct. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231219309257

[70] S. Das, S. Datta, and B. B. Chaudhuri, "Handling data irregularities in classification: Foundations, trends, and future challenges," *Pattern Recognit.*, vol. 81, pp. 674–693, Sep. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320318300931

[71] P. Branco, L. Torgo, and R. P. Ribeiro, "A survey of predictive modeling on imbalanced domains," *ACM Comput. Surv.*, vol. 49, no. 2, p. 31, Aug. 2016, doi: 10.1145/2907070.

[72] J. Stefanowski, *Dealing with Data Difficulty Factors While Learning From Imbalanced Data*. Cham, Switzerland: Springer, 2016, pp. 333–363, doi: 10.1007/978-3-319-18781-5_17.

[73] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[74] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 1998.

[75] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[76] F. Amato, A. López, E. M. Peña-Méndez, P. Vanhara, A. Hampl, and J. Havel, "Artificial neural networks in medical diagnosis," *J. Appl. Biomed.*, vol. 11, no. 2, pp. 47–58, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1214021X14600570

[77] C. V. Weis, C. R. Jutzeler, and K. Borgwardt, "Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: A systematic review," *Clin. Microbiol. Infection*, vol. 26, no. 10, pp. 1310–1317, Oct. 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1198743X20301580

[78] I. Yelin, O. Snitser, G. Novich, R. Katz, O. Tal, M. Parizade, G. Chodick, G. Koren, V. Shalev, and R. Kishony, "Personal clinical history predicts antibiotic resistance of urinary tract infections," *Nature Med.*, vol. 25, no. 7, pp. 1143–1152, Jul. 2019. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31273328

[79] M. C. Vazquez-Guillamet, R. Vazquez, S. T. Micek, and M. H. Kollef, "Predicting resistance to piperacillin-tazobactam, cefepime and meropenem in septic patients with bloodstream infection due to gram-negative bacteria," *Clin. Infectious Diseases*, vol. 65, no. 10, pp. 1607–1614, Oct. 2017.

[80] L. Rokach, *Pattern Classification Using Ensemble Methods*. Singapore: World Scientific, Nov. 2009, doi: 10.1142/7238.

[81] K. Fawagreh, M. M. Gaber, and E. Elyan, "Random forests: From early developments to recent advancements," *Syst. Sci. Control Eng.*, vol. 2, no. 1, pp. 602–609, 2014, doi: 10.1080/21642583.2014.956265.

[82] W. G. Touw, J. R. Bayjanov, L. Overmars, L. Backus, J. Boekhorst, M. Wels, and S. A. F. T. van Hijum, "Data mining in the life sciences with random forest: A walk in the park or lost in the jungle?" *Briefings Bioinf.*, vol. 14, no. 3, pp. 315–326, Jul. 2012, doi: 10.1093/bib/bbs034.

[83] B. Dai, R.-C. Chen, S.-Z. Zhu, and W.-W. Zhang, "Using random forest algorithm for breast cancer diagnosis," in *Proc. Int. Symp. Comput., Consum. Control (ISC)*, Dec. 2018, pp. 449–452.

[84] E. Elyan, C. M. Garcia, and C. Jayne, "Symbols classification in engineering drawings," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[85] T. M. Rawson, B. Hernandez, R. C. Wilson, D. Ming, P. Herrero, N. Ranganathan, K. Skolimowska, M. Gilchrist, G. Satta, P. Georgiou, and A. H. Holmes, "Supervised machine learning to support the diagnosis of bacterial infection in the context of COVID-19," *JAC-Antimicrobial Resistance*, vol. 3, no. 1, Jan. 2021, doi: 10.1093/jacamr/dlab002.

[86] E. Moran, E. Robinson, C. Green, M. Keeling, and B. Collyer, "Towards personalized guidelines: Using machine-learning algorithms to guide antimicrobial selection," *J. Antimicrobial Chemotherapy*, vol. 75, no. 9, pp. 2677–2680, Sep. 2020.

[87] A. S. Chowdhury, E. T. Lofgren, R. W. Moehring, and S. L. Broschat, "Identifying predictors of antimicrobial exposure in hospitalized patients using a machine learning approach," *J. Appl. Microbiol.*, vol. 128, no. 3, pp. 688–696, Mar. 2020.

[88] S. Kulshrestha, S. Panda, D. Nayar, V. Dohe, and A. Jarali, "Prediction of antimicrobial resistance for disease-causing agents using machine learning," in *Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Jun. 2018, pp. 972–975.

[89] A. Wang, N. Daneman, C. Tan, J. S. Brownstein, and D. R. MacFadden, "Evaluating the relationship between hospital antibiotic use and antibiotic resistance in common nosocomial pathogens," *Infection Control Hospital Epidemiol.*, vol. 38, no. 12, pp. 1457–1463, Dec. 2017.

[90] A. L. P. Chow, D. C. Lye, and O. A. Arah, "Mortality benefits of antibiotic computerised decision support system: Modifying effects of age," *Sci. Rep.*, vol. 5, no. 1, pp. 1–11, Dec. 2015.

[91] E. I. Papageorgiou, "Fuzzy cognitive map software tool for treatment management of uncomplicated urinary tract infection," *Comput. Methods Programs Biomed.*, vol. 105, no. 3, pp. 233–245, Mar. 2012.

[92] X. Ying, "An overview of overfitting and its solutions," *J. Phys., Conf.*, vol. 1168, Feb. 2019, Art. no. 022022, doi: 10.1088/1742-6596/1168/2/022022.

[93] S. Arlot and A. Celisse "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, pp. 40–79, Jul. 2010. [Online]. Available: https://doi.org/10.1214/09-SS054

[94] S. Saeb, L. Lonini, A. Jayaraman, D. C. Mohr, and K. P. Kording, "The need to approximate the use-case in clinical machine learning," *GigaScience*, vol. 6, no. 5, p. gix019, Mar. 2017, doi: 10.1093/gigascience/gix019.

[95] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced data sets," *J. Inf. Eng. Appl.*, vol. 3, no. 10, pp. 1–13, 2013.

[96] L. A. Jeni, J. F. Cohn, and F. De La Torre, "Facing imbalanced data–recommendations for the use of performance metrics," in *Proc. Humaine Assoc. Conf. Affective Comput. Intell. Interact.*, Sep. 2013, pp. 245–251.

[97] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process*, vol. 5, no. 2, pp. 1–11, Mar. 2015.

[98] Q. Gu, L. Zhu, and Z. Cai, "Evaluation measures of the classification performance of imbalanced data sets," in *Proc. Int. Symp. Intell. Comput. Appl.* Springer, 2009, pp. 461–471.

[99] P. Vuttipittayamongkol, E. Elyan, A. Petrovksi, and C. Jayne, "Overlap-based undersampling for improving imbalanced data classification," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.* Springer, 2018, pp. 1–9.

[100] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3121–3124.

[101] A. Amin, S. Anwar, A. Adnan, M. Nawaz, N. Howard, J. Qadir, A. Hawalah, and A. Hussain, "Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study," *IEEE Access*, vol. 4, pp. 7940–7957, 2016.

[102] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017.

[103] C. Weis, A. Cuénod, B. Rieck, F. Llinares-López, O. Dubuis, S. Graf, C. Lang, M. Oberle, M. Brackmann, K. K. Søgaard, M. Osthoff, K. Borgwardt, and A. Egli, "Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning," *Biorxiv*, 2021. [Online]. Available: https://www.biorxiv.org/content/early/2021/07/12/2020.07.30.228411

[104] T. M. Rawson, L. S. P. Moore, B. Hernandez, E. Charani, E. Castro-Sanchez, P. Herrero, B. Hayhoe, W. Hope, P. Georgiou, and A. H. Holmes, "A systematic review of clinical decision support systems for antimicrobial management: Are we failing to investigate these interventions appropriately?" *Clin. Microbiol. Infection*, vol. 23, no. 8, pp. 524–532, Aug. 2017.

[105] B. Zenko, L. Todorovski, and S. Dzeroski, "A comparison of stacking with meta decision trees to bagging, boosting, and stacking with other methods," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2001, pp. 669–670.

[106] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explor. Newslett.*, vol. 11, no. 1, pp. 10–18, Nov. 2009, doi: 10.1145/1656274.1656278.

[107] J. Lv, S. Deng, and L. Zhang, "A review of artificial intelligence applications for antimicrobial resistance," *Biosafety Health*, vol. 3, no. 1, pp. 22–31, Feb. 2021. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2590053620300872

[108] M. Kuhn, "Building predictive models in R using the caret package," *J. Statist. Softw.*, vol. 28, no. 5, pp. 1–26, 2008. [Online]. Available: https://www.jstatsoft.org/v028/i05

[109] S. Wares, J. Isaacs, and E. Elyan, "Data stream mining: Methods and challenges for handling concept drift," *Social Netw. Appl. Sci.*, vol. 1, no. 11, Oct. 2019, doi: 10.1007/s42452-019-1433-0.

[110] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[111] L. Collineau, P. Boerlin, C. A. Carson, B. Chapman, A. Fazil, B. Hetman, S. A. McEwen, E. J. Parmley, R. J. Reid-Smith, E. N. Taboada, and B. A. Smith, "Integrating whole-genome sequencing data into quantitative risk assessment of foodborne antimicrobial resistance: A review of opportunities and challenges," *Frontiers Microbiology*, vol. 10, p. 1107, May 2019, doi: 10.3389/fmicb.2019.01107.

[112] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "DBSMOTE: Density-based synthetic minority over-sampling technique," *Int. J. Speech Technol.*, vol. 36, no. 3, pp. 664–684, Apr. 2012.

[113] M. S. Santos, J. P. Soares, P. H. Abreu, H. Araujo, and J. Santos, "Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]," *IEEE Comput. Intell. Mag.*, vol. 13, no. 4, pp. 59–76, Nov. 2018.

[114] S. Wares, J. Isaacs, and E. Elyan, "Burst detection-based selective classifier resetting," *J. Inf. Knowl. Manage.*, vol. 20, no. 2, Apr. 2021, Art. no. 2150027, doi: 10.1142/s0219649221500271.

[115] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020. [Online]. Available: https://www.mdpi.com/1099-4300/23/1/18

[116] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–42, Sep. 2019, doi: 10.1145/3236009.

[117] D. S. Char, M. D. Abràmoff, and C. Feudtner, "Identifying ethical considerations for machine learning healthcare applications," *Amer. J. Bioethics*, vol. 20, no. 11, pp. 7–17, Nov. 2020, doi: 10.1080/15265161.2020.1819469.

[118] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2011.

EYAD ELYAN received the degree in computer science from Al-Quds University, Palestine, in 1999, the M.Sc. degree in software engineering from Bradford University, U.K., in 2004, and the Ph.D. degree from Bradford University, for his work on 3D facial modeling and recognition, in 2008. He is currently a Professor in machine learning and computer vision with the School of Computing, Robert Gordon University. He is leading the machine learning and vision applications research theme with the School of Computing. His research interests include machine learning, deep learning, applied computer vision, and ensemble learning. He is a Fellow Member of the British Higher Education Academy. He works with the Scotland Data Laboratory, Innovation Centre Ambassador.

AMIR HUSSAIN received the B.Eng. (Hons.) and Ph.D. degrees from the University of Strathclyde, Glasgow, U.K., in 1992 and 1997, respectively. Following his postdoctoral and academic positions with the University of the West of Scotland (an EPSRC Postdoctoral Fellow, from 1996 to 1998), Dundee (a Research Lecturer, from 1998 to 2000), and Stirling (a Lecturer, from 2000 to 2004, a Senior Lecturer, from 2004 to 2008, a Reader, from 2008 to 2012, and a Professor, from 2012 to 2018). He joined as a Professor with Edinburgh Napier University, U.K., in 2018, and the Founding Head of the Data Science and Cyber Analytics (DSCA) Research Group (managing over 20 academics and research staff). He is also the Founding Head of the Cognitive Big Data Analytics (CogBiD) Research Laboratory.

AZIZ SHEIKH is currently the Director of the Usher Institute, University of Edinburgh. He is also the Director of the Asthma U.K. Centre for Applied Research (AUKCAR), the Director of the Scottish Allergy and Respiratory Academy (SARA), the Director of the NIHR Global Respiratory Health Unit (RESPIRE), and the Co-Director of the NHS Digital Academy. He is currently holding a number of visiting chairs, such as the University of Birmingham, U.K., Queen Mary's University of London, U.K., and Maastricht University, The Netherlands, respectively. He was previously a Visiting Professor in medicine with the Harvard Medical School.

ABDELRAOUF A. ELMANAMA received the B.Sc. degree in medical technology from Southwestern University, Philippines, in 1990, the M.Sc. degree in microbiology from the University of Santo Tomas, Philippines, in 1993, and the Ph.D. degree in microbiology from Ain Shams University through the Alaqsa Joint Program, in 2004. He is currently working as a Lecturer in general, diagnostic and environmental microbiology with the Medical Technology Department, Islamic University of Gaza. He was appointed by the British Council as a Credited Trainer in food microbiology and HACCP system. His research interests include seawater pollution, antibiotic resistance, ethnobotany, and drinking and wastewater. He is also a Researcher with more than 20 years of experience and teaching courses in both undergraduate and postgraduate levels. He supervised the B.A. thesis, the M.A. thesis, and the Ph.D. dissertations. He has edited and coauthored monographs on contemporary trends in political thought. He has published over 100 articles in peer-reviewed journals.

PATTARAMON VUTTIPITTAYAMONGKOL was born in Lampang, Thailand. She received the B.Sc. degree in electrical engineering from the University of Illinois at Urbana–Champaign, Illinois, USA, in 2011, the M.Sc. degree in electrical engineering from the University of Southern California, California, USA, in 2013, and the Ph.D. degree from the School of Computing, Robert Gordon University, Aberdeen, U.K., in 2020. Since 2013, she has been a Faculty Member of the School of Information Technology, Mae Fah Luang University, Thailand. Her research interests include data analytics and machine learning with focus on supervised learning on class-imbalanced datasets.

KAROLIN HIJAZI received the Graduate degree *(cum laude)* from the University of Siena, in 2004, and the Ph.D. degree in pathology from the King's College London, in 2008. She is currently a Clinical Reader with the University of Aberdeen. During her dental undergraduate program, she trained in molecular microbiology and developed a special interest for host-microbe interactions. She worked as a Postdoctoral Research Associate and subsequently continued her involvement in three European commission-funded projects aimed at the development of anti-HIV-1 microbicides. She is a member of the Infection & Immunity Program with the Institute of Medical Sciences, and leads the Microbial Diseases Research Group, Institute of Dentistry, University of Aberdeen. She is a Section Editor of the *Journal of Global Antimicrobial Resistance*. She serves on European commission expert evaluator panels for funding calls relating to infectious diseases and AMR.

• • •