We are IntechOpen,
the world's leading publisher of Open Access books
Built by scientists, for scientists

5,800
Open access books available

142,000
International authors and editors

180M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Machine Learning for Antimicrobial Resistance Research and Drug Development

Shamanth A. Shankarnarayan, Joshua D. Guthrie and Daniel A. Charlebois

Abstract

Machine learning is a subfield of artificial intelligence which combines sophisticated algorithms and data to develop predictive models with minimal human interference. This chapter focuses on research that trains machine learning models to study antimicrobial resistance and to discover antimicrobial drugs. An emphasis is placed on applying machine learning models to detect drug resistance among bacterial and fungal pathogens. The role of machine learning in antibacterial and antifungal drug discovery and design is explored. Finally, the challenges and prospects of applying machine learning to advance basic research on and treatment of antimicrobial resistance are discussed. Overall, machine learning promises to advance antimicrobial resistance research and to facilitate the development of antibacterial and antifungal drugs.

Keywords: machine learning, antimicrobial resistance, fungi, bacteria, infection, drug discovery and design

1. Introduction

Antimicrobials are the agents used to prevent and treat the infection caused by bacteria, fungi, viruses, and parasites in plants, animals, and humans. Sir Alexander Fleming in his Nobel Prize lecture emphasized the importance of avoiding resistance to antibiotics [1]. Antimicrobial resistance (AMR) is a phenomenon that occurs when infectious microorganisms do not respond to antimicrobial agents, leading to treatment failure, the spread of the infectious disease, and severe illness and death [2]. Among microorganisms, bacteria and fungi are the most encountered pathogens with resistance in clinical settings. Patients infected with resistant bacteria or fungi have worse clinical outcomes compared to patients with infections caused by the same bacteria or fungi without resistance [3]. It is estimated that by the end of year 2050, if unmitigated, AMR will result in 10 million lives lost per year and cumulative cost of 100 trillion USD [4]. The global burden associated with bacterial AMR alone, considering 204 countries and territories, 23 bacterial pathogens, and 88 drug-pathogen combinations, was 4.95 million deaths during the year 2019 [5]. The majority of these patients succumbed to lower respiratory tract and blood stream infections associated with drug-resistant

bacteria, with highest mortality rate of 27.3 per 100,000 patients [5]. Among elderly patients in the USA, the treatment of methicillin resistant *Staphylococcus aureus* (MRSA) infection costs $22,293 more per patient compared to patients infected with non-resistant *Staphylococcus aureus*. Similarly, treating patients infected with resistant carbapenem-resistant *Acinetobacter* species costs $57,390 more per patient compared to patients infected with non-resistant *Acinetobacter* species. These extra costs are attributed to the increased length of hospital stays and health complications, which lead to more medical interventions and higher mortality rates [6].

The most common bacterial pathogens associated with hospital acquired infections and AMR are the ESKAPE pathogens. ESKAPE is an acronym for *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter species* [7]. The priority pathogens recognized by the World Health Organization are extended spectrum beta lactamases (ESBL) producing *Escherichia coli*, MRSA, ESBL-producing *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, carbapenem-resistant *Acinetobacter baumannii*, and multidrug-resistant (MDR; organism resistant to at least one agent in three or more antimicrobial classes) *P. aeruginosa* and vancomycin-resistant *Enterococcus fecalis* [5, 8, 9]. Antimicrobial resistance among fungi is a serious issue because of the limited number of classes of antifungal agents available for treating invasive fungal infections, as compared to antibacterial agents (**Table 1**). Moreover, due to variety of socio-economic reasons it has been over a decade that no new class of antifungal drug has been developed [10]. Global warming and climate change is also predicted to increase the prevalence of fungal infections (as fungi adapt to higher temperatures, humans and animals may lose their

Mechanism of action	Antibacterial class
Inhibitor of cell wall synthesis	β-Lactams, Carbapenems, Cephalosporins, Monobactams, Penicillin, Glycopeptide
Cell membrane depolarizer	Lipopeptides
Inhibitor of protein synthesis	Aminoglycosides, Tetracyclines, Chloramphenicol, Lincosamides, Macrolides, Oxazolidinones, Streptogramins
Inhibitor of nucleic acid synthesis	Quinolones
Inhibitor of metabolic pathways	Sulfonamides, Trimethoprim
Mechanism of action	**Antifungal class**
Inhibitors of ergosterol synthesis	Azoles
Aqueous pores in cell membrane	Polyenes
Inhibitor of glucan synthase	Echinocandins
Inhibitor of squalene epoxidase	Allylamines
Inhibitor of nucleic acid	5-Flurocytosine

Table 1.
Different classes of antibacterial and antifungal drugs and their mechanism of action.
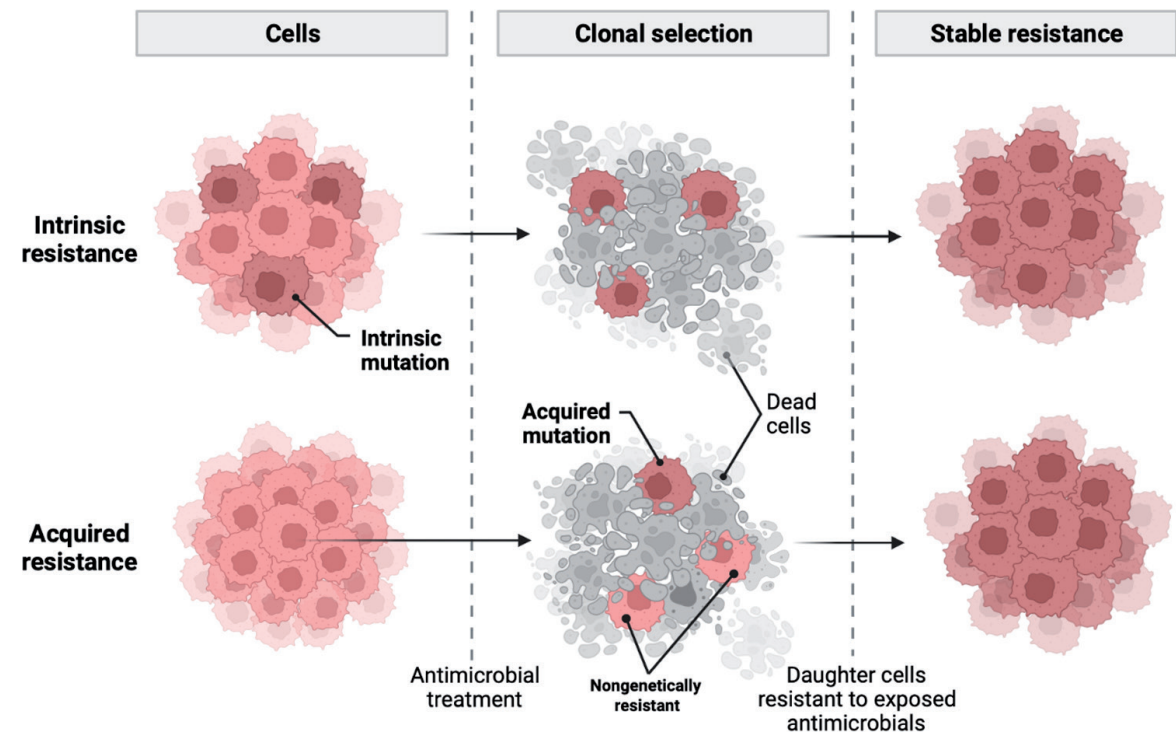
Figure 1.
Depicting the difference between intrinsic and acquired resistance. Microorganisms that are intrinsically resistant can propagate from the moment that they are exposed to the antimicrobial agent. Microorganisms can also acquire resistance during exposure to an antimicrobial agent through genetic and nongenetic mechanisms. Adapted from 'Intrinsic and acquired drug resistance', by BioRender.com (2022). Retrieved from https://app.biorender.com/ biorender-templates.

thermal protection provided by their elevated body temperatures) [11]. The majority of the invasive fungal infections are caused by yeasts, especially *Candida albicans*, which can cause mild symptomatic infection to acute sepsis with a mortality rate over 70% in immunocompromised patients [12]. Over the last decade, *Candia auris* has been reported on all continents and in more than 44 countries [13, 14]. The first known appearance of *Candida auris* dates back to 1996 in South Korea, when it was originally misidentified as *Candida hemulonii* (and then later correctly identified as *Candida auris*) [15]. This fungus displays intrinsic resistance and acquired resistant (**Figure 1**) to the major classes of antifungals and hospital disinfectants and has caused several outbreaks [16–19]. The main reason that *Candida auris* attention across globe is due to high mortality rate (45%) among patients with bloodstream infections [20]. Interestingly, *Candida auris* has different resistance profiles based on the genomic sequences identified in different countries; presently, *Candida auris* is classified into four discrete clades, as well as a potential fifth clade [21, 22]. *Candida auris* is less virulent than *C. albicans* because of the 'fitness cost' associated with its MDR nature; as a consequence, *Candida auris* has not been observed to revert back to its susceptible form in the absence of antimicrobial pressure [23]. Recently in the United States the identification of pandrug-resistant (resistant to all agents in all classes of antimicrobial agents) [24] *Candida auris* among skin colonizers has raised alarm [25]. Mycelial fungi, which consisting of network of fine filaments known as hyphae, such as *Aspergillus* species are ubiquitous in nature and commonly cause respiratory disorders. *Aspergillus* species resistant to the azole class of antifungals are a serious threat, as azoles are first line of therapy against *Aspergillus* infection [26]. Another mycelial fungi, *Trichophyton indotinea,* which causes skin infection is spreading across the globe [27, 28].

The emergence of AMR in high-income countries is mainly associated with use, misuse, and overuse of antibiotics in hospitals, agriculture, and communities [29]. Whereas in low- and middle-income countries unhygienic practices, contaminated water supplies, civil conflicts, and an increased number of immunocompromised patients (especially among HIV infections) are the main contributors to AMR [30]. Increased infections, and in turn increased use of antimicrobial agents, has imposed selection pressures that result in the retention of resistant strains. Identifying infectious agents early helps clinicians to promptly choose the appropriate antimicrobial agent to treat the infection based on the intrinsic resistance profiles and local epidemiology data on resistance [31]. Resistance profiling methods, such as culture-based and molecular biology-based methods, currently take up to 72 h from the time of sample collection. During this time, patients often receive broad-spectrum antibiotics, which may lead to acquired resistance (**Figure 1**). Several novel strategies have been developed for rapid detection of AMR. However, most of these methods are based on molecular biology, immunology, biochemistry, and rapid culture techniques [32]. Importantly, the cost and the expertise involved in establishing and maintaining these techniques and related devices is often too high for many hospitals and institutions, especially those in remote and impoverished communities.

Machine learning (ML) has been around for decades, as optical character recognition gained popularity during 1990s with its application as spam filters. A seminal paper by Geoffery Hinton in 2006 on recognizing handwritten digits using 'deep learning' (a ML technique implemented in artificial neural networks) rekindled interest in ML. Recently, during the 14th Critical Assessment of Protein Structure Prediction (CASP14) competition [33], a neural network based model called AlphaFold predicted protein structures with high accuracy (i.e., comparable to the experimental structures), outperforming other protein structural deduction methods [34]. Furthermore, deep learning is increasingly being applied to solve complex multidimensional problems, such as speech recognition [35] and image classification [36].

Machine learning is the application of advanced algorithms that enable a computer to 'learn' and generate predictive mathematical models from data. Arthur Samuel in 1959 described ML as 'the field of study that gives computers the ability to learn without being explicitly programmed' [37]. Tom Mitchell in 1997 provided a more engineer-oriented definition, when he stated that a 'computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E' [38]. Machine learning can be divided into supervised, unsupervised, and reinforcement learning. In supervised learning, the ML model is trained using labeled datasets, with the resulting model being a function that can take new data and predict an output. To determine the reliability of the trained model, a test set of complete input/output data which was not used during training is employed to determine an unbiased estimate of model performance. Whereas, in unsupervised learning, the training data are supplied without labels. Unsupervised learning algorithms find the similarity among data points and cluster them together. Reinforcement learning (RL) uses algorithms that learn from the accumulation of 'rewards' that a computational agent receives through interactions with its environment. Reinforcement learning, which is often combined with other ML methods such as deep neural networks, has led to some of the most successful artificial intelligence systems ever developed. These range from systems that beat human professionals in the game of Go [39] to systems that help control nuclear fusion reactions [40].

Recent advances in digitizing medical records and data generated in experiments have paved the way for ML applications in the fields of biology and medicine. Many

clinical trials are leveraging ML processes to improve the efficiency and quality of clinical research and pre-clinical drug development [41]. Machine learning is also being applied to assess the risk of developing sepsis based on patients' clinical records [42]. Machine learning has also found applications at the cellular level. For instance, convolutional neural networks (CNNs) can predict the interactions of transcription factors and histones within chromosome structures, which in turn aids in analyzing genome architecture as well as gene regulation [43]. Other examples include using neural networks to identify the role of non-coding DNA in humans in regulating gene expression [44] and applying recurrent neural networks (RNNs) to characterize chromatin folding in *Drosophila melanogaster* [45]. Furthermore, the availability of large-scale high-throughput genomic and epigenomic data has led to several studies that have highlighted the potential applications of ML in the field of genomics [46] as well as non-coding RNAs [47]. Machine learning has also been used to assist clinicians treating infectious diseases [48]. However, the use of ML in studying drug-resistant pathogens is less developed.

In this chapter, we first discuss the mechanisms of underlying bacterial and fungal AMR, followed by an overview of ML methods used to detect drug-resistant pathogens. We then highlight the application of ML in the discovery and design of antimicrobial drugs. Finally, we present the challenges and prospects of applying ML to AMR research and drug development.

2. Mechanisms of antibiotic resistance

The major burden of AMR in hospital settings is due to bacteria and fungi. Antimicrobial resistance can be classified into different types, including 'intrinsic resistance' and 'acquired resistance' (**Figure 1**) [49]. Intrinsic resistance occurs when bacteria or fungi are naturally resistant to an AMR drug or to a class of AMR drugs [50]. Bacteria and fungi which were previously susceptible to an antimicrobial drug can acquire resistance, for instance, by modifying the target site of the drug or by gaining a resistance mutation (**Figure 1**). In these scenarios, the microorganism develops resistance post-exposure to the drug. Whereas, if the microorganism does not have a target site for the drug or has a preexisting resistance mutation, then it is classified as intrinsically resistant. Other forms of AMR exist, such as 'clinical resistance', whereby a microorganism is susceptible to a drug *in-vitro,* but the drug is ineffective against the same microorganism in *in-vivo*. Clinical resistance can occur in a patient due to pharmacokinetic and pharmacodynamic factors.

Another aspect of AMR is 'persistence' and 'tolerance', which are phenomena that allow non-growing or slow growing bacterial and yeast pathogens to survive antimicrobial treatment [51, 52]. In the case of genetic resistance to a drug, all the progeny of the resistant microorganism stably inherit resistance to the drug (**Figure 1**). Whereas persistence occurs when a small fraction of a clonal bacterial population is resistant to an antibiotic, but the persistent cells do not harbor resistance mutations or genes. Rather, these persister cells are in a stationary or dormant phase, which reduces the effectiveness of antibiotics that target growth processes [53–55]. Antibiotic persistence is a heterogenous response of a bacterial population to an antibiotic and causes a delay in the clearance of the infection [56]. In contrast, tolerant cells require more time to be affected by an antimicrobial drug compared to susceptible cells [56]. Systemic infections due to persistent and tolerant organisms lead to higher mortality rates compared to infections caused by susceptible microorganisms [57]. Nongenetic

drug resistance is another form of AMR. Nongenetically drug-resistant phenotypes can be found in clonal cell populations [58] and results from genetically identical cells differentially expressing genes that confer resistance, along with various epigenetic mechanisms [59, 60].

Bacteria and fungi belong to different kingdoms, have differences in cellular components, and antibacterial and antifungal agents target different sites. Despite this, there are similarities between the AMR agents that are used to treat antifungal and antibacterial infections. For instance, cell wall inhibitors of bacteria target peptidoglycan, an important component of the bacterial cell wall, whereas some antifungal agents inhibit ergosterol, an important component of fungal cell membrane. Antibacterial agents have diverse mechanisms of action, including inhibiting cell wall synthesis, depolarizing cell membranes, as well as inhibiting of protein synthesis, nucleic acid synthesis, and metabolic pathways (**Table 1**) [61]. However, in contrast to many antibacterial agents, antifungal analogues for protein inhibitors, topoisomerase inhibitors, and metabolic pathways inhibitors are not available. Only a limited number of antifungal agents are available that target ergosterol synthesis, cell membrane integrity, glucan synthase, nucleic acid synthesis, and the squalene epoxidase enzyme.

2.1 Antibacterial resistance mechanisms

The main mechanisms of antibiotic resistance among bacteria are (i) limiting uptake of a drug; (ii) modifying a drug target; (iii) inactivating a drug; and (iv) active drug efflux (**Figure 2a**). Limiting uptake due to natural permeability barriers imposed by the cell membrane, drug inactivation by antibiotic inactivating enzymes, and drug efflux resulting non-specific protein efflux pumps are mechanisms of intrinsic resistance. Whereas the transfer of genes between bacteria that encode drug efflux pumps or enzymes that inactivate antibiotics, as well as drug target modifications, are acquired resistance mechanisms. Antibiotic resistance mechanisms differ between gram-negative and gram-positive bacteria due to differences in their cell wall composition. Gram-negative bacteria employ all the drug resistance mechanisms, whereas gram-positive bacteria mainly limit the uptake of a drug [62]. Due to the hydrophobic nature of the cell wall, many of the hydrophilic antibiotic cannot bind to the cell wall and the high lipid content among mycobacteria restricts the entry of hydrophilic antibiotics [63]. However, porin channels found within the cell membrane allow certain hydrophilic antibiotics to enter the cell. Modifications to these porin channels limits drug uptake [64]. Mutations in the gene responsible for porin proteins alter the selectivity of hydrophilic drugs [65]. Drug intake is also restricted by the thickening of cell wall [63]. Another widely observed phenomenon that restricts drug uptake is the formation of bacterial and fungal biofilms. The thick outer layer of a biofilm is composed of extracellular polymeric substances and is impenetrable to many antimicrobial drugs [66].

Antibiotics target multiple cellular components and bacteria can modify these targets leading to AMR. One of the major targets is the cell wall, which is commonly targeted by ß-lactam drugs, specifically among gram positive bacteria. Resistance to ß-lactam antibiotics results from modifications in the cell wall structures as well as a number of penicillin-binding-proteins [67]. Bacteria can alter the precursor of the target by mutating the gene responsible for these precursors, eventually leading to an altered target site. This results in the antibiotic failing to bind to the target site [68]. Ribosomes are also commonly targeted by antibiotics to inhibit protein synthesis.
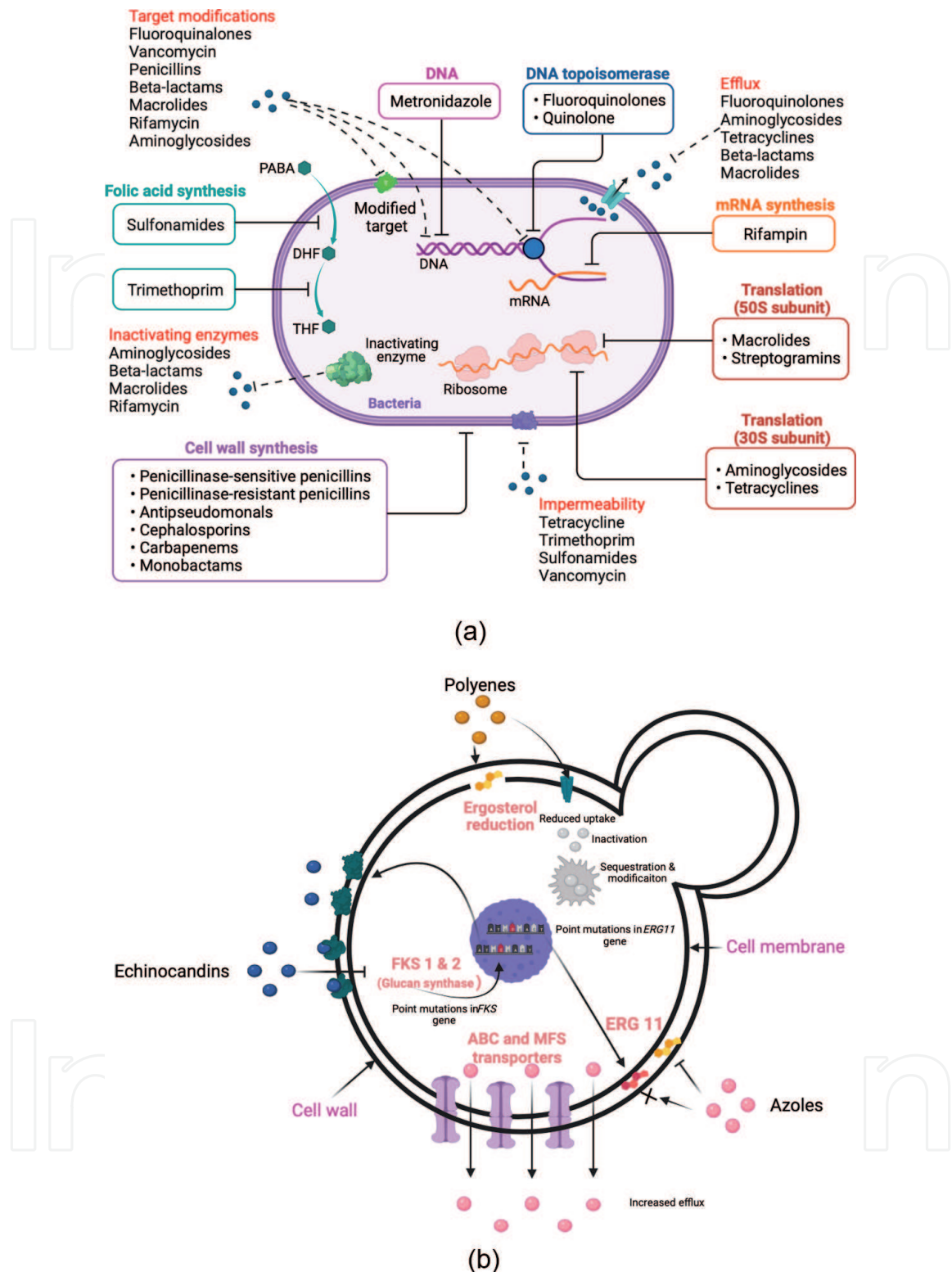
Figure 2.
Mechanisms of action of antimicrobial drugs in bacteria and fungi. (a) Effect of antibacterial drugs on bacterial cellular components and the corresponding resistance mechanism developed by bacteria. Created with Bio-Render. com. (b) Effect of antifungal drugs on fungal cellular components and the resistant mechanisms developed by the fungi. Adapted from "Antimicrobial Therapy Strategies", by BioRender.com (2022). Retrieved from https://app. biorender.com/biorender-templates.

Mutations in the ribosomal gene leading to the protection of the ribosomes and methylation of the ribosomal subunits lower the binding affinity of antibiotics, leading to resistance [69]. Similarly, modifications in the DNA gyrase or topoisomerase enzyme,

nucleic acid synthesis inhibitors fail to bind to these enzymes [70]. Drugs that inhibit metabolic pathways inhibit important metabolic byproducts that are essential for bacterial survival. These antibiotics competitively bind to the active sites of enzymes responsible for the synthesis essential metabolites. Mutations in the gene responsible for these enzymes restricts antibiotics from binding [71]. Another mechanism of AMR is the inactivation of the drug by the pathogens. Degrading or transferring a chemical group to the antibiotics modifies its structure and affinity towards the target [72]. Efflux pumps remove toxic substances from the bacterial cell; some efflux pumps are constitutively expressed and others are induced or overexpressed in the presence of antibiotics. There are majorly five families of efflux pumps depending on the energy source they utilize and their structure [64]. Namely, the ATP-binding cassette (ABC) family, the multidrug and toxic compound extrusion family, the small multidrug resistance family, the major facilitator superfamily (MFC), and the resistance-nodulation-cell division family. The majority of the bacteria resistant to antibiotics overexpress efflux pumps from one of these families during antibiotics exposure [73].

2.2 Antifungal resistance mechanisms

Antifungal resistance mechanisms are not as extensively studied as antibacterial resistance mechanisms. Several factors including immunosuppressive treatments, indiscriminate use of broad-spectrum antibiotics, and immune suppressive diseases like HIV led to a surge in fungal infections during 1970s and 1980s [74]. Antifungal drugs including imidazoles and azoles were subsequently approved during late 1980s and 1990. Extensive use, misuse, and overuse of these antifungal drugs since then have led to the emergence of AMR in fungal pathogens. Determining if a fungal isolate is resistant is based on the minimum inhibitory concentration (MIC) of the antifungal drug. The MIC of a fungus isolated from a clinical sample informs the decision on the appropriate course of antifungal therapy.

Currently three major classes of anti-fungal drugs used for treating systemic fungal infections. Namely, azoles (itraconazole, voriconazole, posaconazole, and isavuconazole), polyenes (amphotericin B) and echinocandins (caspofungin, micafungin, and anidulafungin) (**Table 1**). The limited number of classes of antifungal drugs and AMR in fungi restricts treatment options. The emergence of MDR fungal species further hinders treatment options. Azoles target ergosterol biosynthetic pathway, as ergosterol is necessary in the cell membrane to maintain the stability, permeability and the activity of membrane bound enzymes (**Figure 2b**) [75]. The substitution of an amino acid in the binding site of the enzyme is a common mechanism of azole resistance among *Candida* species. Overexpression of *ERG11* gene is also common among azole-resistant strains [76]. Furthermore, the overexpression of drug targets decreases the effectiveness of a drug, as more drug is required for inhibition [77]. Like bacterial efflux pumps, fungi have two main membrane associated efflux pumps superfamilies, the ABC superfamily and the MFC superfamily. Overexpression of *Candida* drug resistance (CDR) genes such as *CDR1* and *CDR2* of the ABC superfamily lead to the efflux of azoles and decreased drug accumulation [78, 79]. Gain-of-function mutation in the gene responsible for a transcription factor UPC2 leads to upregulation of many ergosterol biosynthesis genes, conferring azole resistance [80]. Another transcription factor TAC1 regulates the activity of efflux pumps in *Candida* species. TAC1 is responsible for upregulation of *CDR1* and *CDR2* in the presence of azoles [81]. Chromosomal abnormalities and mitochondrial defects also contribute to azole resistance [82, 83]. Stress response pathways related to the heat shock protein Hsp90

provide critical strategies for the survival in the presence azoles leading to resistance [84]. Echinocandin resistance is mainly due to mutations in the *FKS* gene. *FKS* gene is responsible for the synthesis of glucan synthase enzyme involved in the synthesis of ß-glucan in the fungal cell wall [85, 86]. In certain cases, echinocandin induces chitin synthesis via protein kinase-C, high osmolarity glycerol, and calcineurin pathways [87] by activating two chitin synthases (Chs2 and Chs8) [88], leading to masked target sites. Polyene resistance in fungal pathogens is less understood because of its various mechanisms of action on the fungal cell. Polyenes act on the fungal cell membrane by interacting with ergosterol and impairs the membrane barrier function [89]. Polyene resistance is mainly attributed to the alterations in the sterol content of the cell membrane, a defense mechanism developed against oxidative stress created by the drug and reorientation of ergosterol structures within the cell membrane [90]. Furthermore, *Candida* species harboring mutations in the *ERG3* and *ERG6* genes exhibited polyene resistance [91]. However, increased catalase activity by the fungal cell also reduces the oxidative stress imparted by the amphotericin leading to resistance [92]. Polyene and azole resistance in combination has been reported among *Candida* species as well as *Cryptococcus neoformans*, and has mostly been attributed to the reduction of ergosterol in the cell membrane and accumulation of its intermediates [93].

Current methods for detecting AMR among the infecting pathogens take up to 72 h from the time of sample collection. All the isolated bacterial and fungal pathogens must undergo standard antimicrobial susceptibility testing (AST) as recommended by the European Committee on Antimicrobial Susceptibility Testing and the Clinical Laboratory Standards Institute [94, 95]. Early detection of the infecting pathogen along with its drug resistance profile are critical for initiating prompt antimicrobial therapy. However, several challenges are faced during this process, such identifying the pathogen, differentiating between commensal and pathogenic microorganisms in a clinical sample [96]. After successful isolation of the pathogen, a round of subculture must be performed so that contamination can be excluded before commencing AST. Microbroth dilution and disk diffusion AST methods can get delayed due to contamination, leading to delays in initiating the appropriate antimicrobial therapy. Several new technologies and methods are being used for early and rapid detection of AMR. For example, technologies based on nucleic acid amplification, hybridization, microscopy, electrochemical, mass spectroscopy, and nanotechnology [97, 98]. However, these methods require sophisticated instruments, expertise, and expensive consumables restricts their deployment in low-income countries. Point-of-care tests (POCTs) used at patient bedsides are now being used to determine AMR; POCTs can be also used among outpatients. Some types of POCTs like microscopy stations, single molecule biosensors, and microfluidic platforms are being tested [99, 100]. The drawbacks of POCTs, including small sample size, lack of internal standards, and their inability to detect nongenetic forms of AMR resistance still need to be resolved. More advanced methods such as ML approaches to detect AMR could further reduce turn-around times and could be deployed across diagnostic laboratories. Machine learning methods can be also applied to detect certain features that are present in resistant bacteria and fungi, but absent in sensitive isolates, which the human eye or other diagnostic technologies may fail to recognize [101]. For instance, real-time high-throughput screening of modified proteins within the resistant isolates [102] has been less explored and is an ideal application for ML methods. The application of ML methods (Section 3) may lead to a deeper understanding of AMR mechanisms, which in turn could lead to rapidly detecting AMR pathogens in patients (Section 4) and to developing new drugs (Section 5).
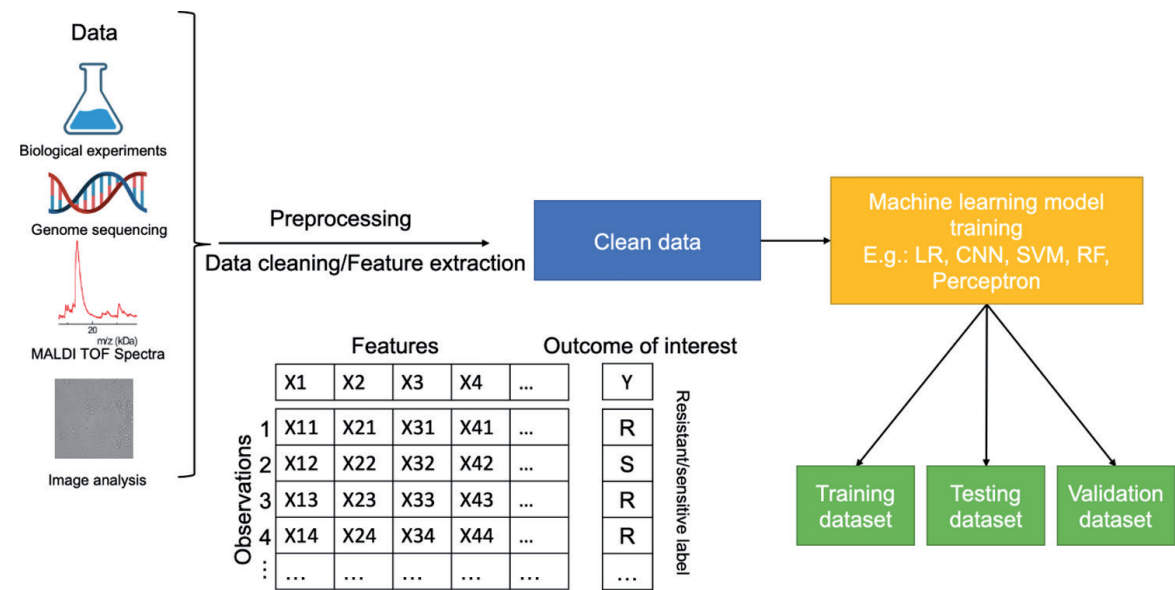
3. Machine learning basics

Machine learning enables us to investigate and draw conclusions from information contained in data that would otherwise be inaccessible to humans. Problems that benefit from the application of ML are endless, but they have a few defining features [103]. First, the problem may have a known solution, but converting it into a computer program is not feasible or requires extensive resources. For example, humans can easily identify a dog within a group of other four-legged animals but writing a computer program to explicitly describe all possible aspects of a dog and its differences to other similar animals would be error prone and practically infeasible. On the other hand, training a ML algorithm to identify a dog may only take a few lines of code, given modern ML software tools. Second, complex problems where traditional methods have failed to identify a solution may benefit from the use of ML algorithms (**Figures 3** and **4**), such as the use of deep learning systems to master the game of Go [104] or to make highly accurate predictions of protein structure [34]. Not only does this enable the use of the resulting ML model in practical applications, but it can also guide researchers towards a deeper understanding of the system they are studying. For instance, ML can guide mathematicians by finding patterns and relations between mathematical objects that can lead to the formation of new conjectures and theorems [105].

Although the defining feature of all ML approaches is to learn from a given dataset, ML techniques can be separated into three broad categories based on the amount of human input: Supervised learning, unsupervised learning, and reinforcement learning [103, 106–108]. Each of these approaches have their own concepts, techniques, and areas of applicability, with the differences between them not always clear. Nonetheless, these categories are useful to provide a means to determine the best approach for a particular problem at hand. Understanding the available tools is
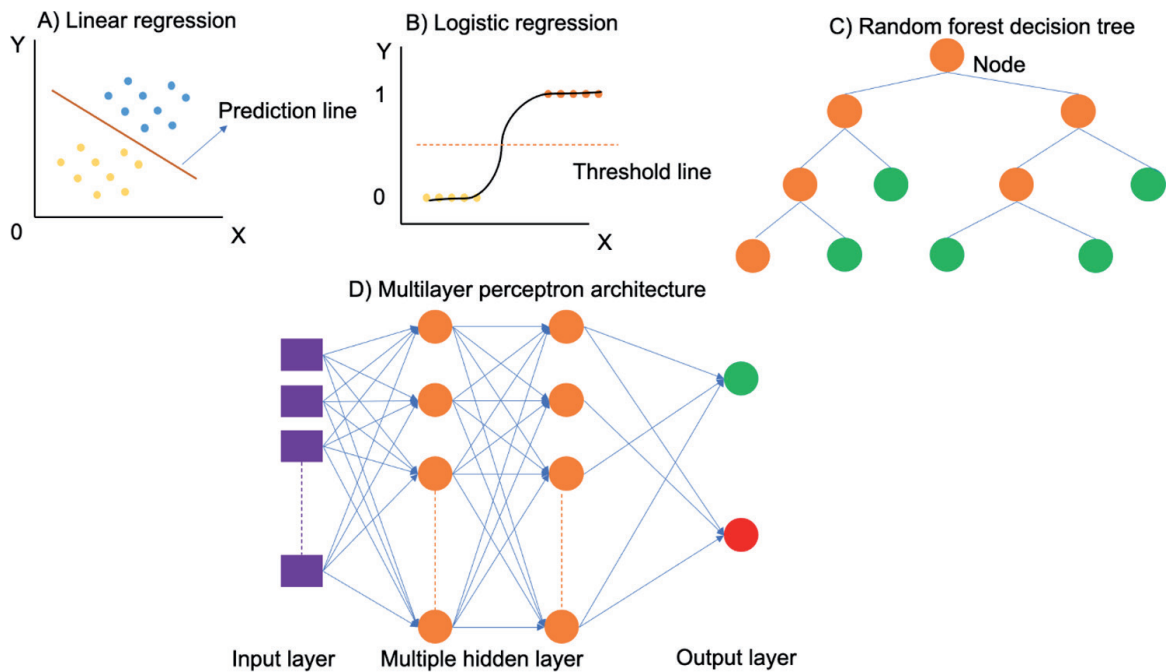
Figure 3.
A selection of common machine learning methods. (A) Linear regression model using a prediction line to distinguish the test dataset. (B) Logistic regression model using a threshold to distinguish the test dataset into two groups. (C) Random forest model using a visually generated decision tree for datapoints to estimate each samples outcome by voting. (D) Multilayer perceptron architecture consisting of an input layer, multiple hidden layers, and an output layer.

Figure 4.
The machine learning pipeline. This pipeline consists of data originating from different biological experiments, preprocessing steps for cleaning the data, along with the feature extraction process. Machine learning methods are then applied to the clean data by dividing this data into training, testing, and validation sets. 'MALDI TOF' stands for 'matrix assisted laser desorption ionization time of flight', 'LR' for 'logistic regression, 'CNN' for 'convoluted neural network, 'SVM' for 'support vector machine', and 'RF' for 'random forest'.

crucial for choosing the best ML technique to solve a particular problem. Although an extensive overview of each ML category is outside the scope of this chapter, we provide an overview of some of the common ML methods below.

3.1 Supervised learning

Supervised learning consists of algorithms that learn using a training set consisting of labeled data [106]. The goal of supervised learning is to find a model for the relationship between the inputs (called 'features') and known outputs, which can then be used to predict outputs for future inputs, where the actual outputs are unknown. Supervised learning techniques can be separated into two categories, 'classification' and 'regression' [109, 110].

Classification problems generally aim to classify future inputs into predefined categories through training on examples, where the inputs are labeled with their corresponding category [107]. Given enough quality training data, models created with classification techniques can provide accurate classification of future data, without requiring the details of the input data to be explicitly programed [103, 106–108]. For instance, a researcher may desire to have a computer take a microscopy image of a cell and return the name of the species, without requiring a human to identify the species. Using a training set of microscopy images for a variety of different species labeled with the name of the species, a classification model can be trained to learn the relationships between the visual aspects of the species and their labels. The model produced can then be used on unlabeled microscopy images to determine the species, saving researchers time and effort, along with producing a model that can be shared in the scientific community. Classification learning algorithms are not restricted to images; any form of data that can be separated into predefined categories can be fed into a classification learning algorithm for training to produce a classifier model [107, 108].

While classification methods aim to predict discrete class labels for inputs, regression methods aim to predict continuous numerical values for given numerical inputs [107, 108]. Regression techniques also learn from training data containing inputs and outputs, but in this case the data consists of numerical inputs and their corresponding numerical outputs, with the resulting model being a continuous mathematical relationship between inputs (independent variables) and outputs (dependent variables) [107]. The resulting model can then be provided with future inputs to make numerical predictions. For example, a researcher may be interested in finding a mathematical relationship between the inputs of an experiment (e.g., preset voltages) and the corresponding outputs they detect (e.g., electrical currents), for systems where theory is unable to make accurate predictions. By training a regression model on a large amount of set inputs and detected outputs, the researcher may be able to find a model that accurately predicts numerical outputs when given future inputs. Not only is this useful in a practical sense, but the resulting model can also be used to guide fundamental research by providing an accurate mathematical and physical relationships that can be further analyzed and understood in terms of theoretical ideas [105, 111].

Through extensive research on supervised learning, many different learning algorithms for classification and regression have been developed and programmed into readily available software packages. Linear regression, logistic regression [107, 108], support vector machines (SVMs) [112], decision trees and random forests [113] and most artificial neural networks [114] are some examples of supervised learning systems, each having their own advantages and disadvantages.

3.2 Unsupervised learning

Unsupervised learning methods, unlike supervised learning, attempt to learn from unlabeled data [115]. This often takes the form of data clustering, but other methods such as anomaly detection and dimensionality reduction also fall under this category [107, 108]. Clustering algorithms attempt to separate unlabeled data into groups with similar components, which can be useful for extracting information from high-dimensional data, which is often infeasible for a human to do. Anomaly detection involves finding anomalous outliers in large datasets by comparing data points to learned patterns, which can be helpful when working with noisy experimental data [116, 117]. Dimensionality reduction methods attempt to simplify high-dimensional data without losing important information, making the analysis and use of such data easier [118, 119]. Unsupervised learning methods can also be combined with supervised learning, referred to as 'semi-supervised' learning, to learn from data that is partially labeled [120, 121]. This is useful when working with large amounts of data, where labeling every data point is infeasible. Some examples of unsupervised learning methods include k-means clustering [122, 123], hierarchical clustering [124, 125], DBSCAN [126], isolation forests [127], principal component analysis [128], autoencoders [107, 108], locally linear embedding [129], and expectation-maximization algorithms [130].

3.3 Reinforcement learning

Reinforcement learning approaches rely on the idea of learning from 'rewards' obtained through interactions with an environment [131]. Reinforcement learning problems are formulated as a discrete-time stochastic control processes known as

'Markov decision processes', with the goal of training a computational system (or 'agent') to determine the best strategy (or 'policy') for reaching a defined goal [132]. The environment is defined by 'states' that the agent can be in, while the agent is able to perform certain 'actions' to interact with the environment. As the agent interacts with its environment, numerical values called rewards that model performance are collected for performing certain actions [132]. The goal of the agent is then to maximize these rewards (using sophisticated statistical methods) by learning the best policy for making decisions in particular situations through repeated interactions with its environment [132]. For example, a reinforcement learning system may be programmed into a cleaning robot to maximize the amount of cleaning it can do while still being able to return to its charging station. In this case, a positive reward would be given for picking up trash, while a negative reward would be given for letting its battery die without reaching the charging station. Using reinforcement learning methods, the robot can learn to optimize its own behavior through repeated experience with its environment.

3.4 Validating machine learning models

To ensure the model created using ML is accurate it must be validated on data independent of the training set [103, 106–108, 133]. Applying the trained model directly to a certain problem is one method of testing, but this is often impractical for real-world applications where model performance matters. The usual method of validation is to split the initial dataset into training and testing sets, where the model is trained on the training set and its accuracy is determined by comparing its predictions using the testing set inputs to the true outputs from the test set [107, 108]. This analysis provides the 'generalization error' estimate of the model, which is used to determine whether the model is accurate, and the errors associated with using the model on new data [107]. Many different metrics are used to determine the generalization error, such as the root mean square error or false-positive/false-negative rates [103, 107, 108], and the choice of method depends on the problem and the learning algorithm. Through iterative training and testing cycles, model performance is improved until a satisfactory accuracy is achieved.

A major issue when using ML is overfitting the model to the training set [103, 106–108, 133]. This corresponds to the case where the 'training error' (i.e., how well the model matches the training data) is low, but the generalization error (i.e., how well the model can predict outcome values for previously unseen data) is high [107, 108]. This is a common occurrence, especially when using models that are more complex than the actual relationships contained in the data. For example, if the actual relationship between inputs and outputs is linear but we attempt to fit a third-degree polynomial to the data, we may produce a model that passes through each of the training set data points exactly (low training error) but cannot generalize to data outside of the training set (high generalization error). Avoiding overfitting (as well as underfitting) requires the use of appropriate training and validation methods to determine model performance before deploying a trained ML model. The quantity of training data is also important. A lack of training data can lead to inaccurate or biased predictions. The amount of data required to create accurate models ultimately depends on the problem and ML method being used [103, 106–108, 133].

During the testing stage, it is important to tune the 'hyperparameters' of the model to improve training accuracy [103, 106–108, 133–135]. Hyperparameters refer to the parameters that are not being learned, such as gradient time steps or data batch

size. Many cross-validation techniques for hyperparameter tuning are available, such as k-fold cross validation [135], and can be implemented directly in ML software packages. It is also often necessary for datasets to be pre-processed before applying ML techniques [136]. Pre-processing is application/software dependent and involves converting the collected data into data structures that can be read by the ML algorithm/software package being used.

3.5 Machine learning software

The extensive and increasing use of ML in industry and scientific research has led to the development of many tools for applying ML techniques quickly and accurately. With almost every well-established ML algorithm being implemented in free dedicated software packages, deploying a ML solution has in some cases become as simple as writing a few lines of code. Although the researcher must determine whether their problem may benefit from the application of ML, the availability of extensively tested and optimized tools to apply ML has made doing so much easier once the relevant data has been collected and organized.

Python is currently the most used programming language for ML, as it contains well-developed and optimized ML libraries. However, other languages such as Julia are also becoming popular with ML researchers. Below is a list of some of the free software packages used for ML applications, along with the programming languages they can be used with.

- TensorFlow (https://www.tensorflow.org/) [137]. Developed by Google, TensorFlow can be used with a variety of programming languages, including Python, C++, Julia, and Java.

- Keras (https://keras.io/) [138]. Keras is a widely used, user-friendly Python interface for the TensorFlow library.

- Scikit-learn (https://scikit-learn.org/) [139]. Scikit-learn is a Python library that contains many ML algorithms, optimized for Python data structures. Wrappers to use Scikit-learn with other programming languages, such as Julia, are also available.

- PyTorch (https://pytorch.org/) [140]. Developed by Facebook, PyTorch is a ML framework primarily for Python, but it also has a C++ version.

4. Machine learning for detecting drug resistance

Over the last decade, an increase in AMR has occurred across the world. At the same time, ML methods have been successfully applied in numerous scientific fields. The availability of large datasets from whole genome sequencing (WGS), matrix assisted laser desorption ionization time of flight mass spectroscopy (MALDI TOF MS), transcriptional response to antibiotics and proteome profiles have facilitated the application of ML algorithms to detect AMR. Specifically, ML methods have been used to detect AMR in bacterial and fungal pathogens based on the data obtained from WGS and MALDI TOF MS (**Figure 4**) [102, 141–143]. Reduced genomic sequencing cost and high-throughput data from WGS has enabled application of

ML methods to sequence data. A few studies have utilized genome sequencing data to predict resistance phenotypes among bacterial pathogens using ML methods [144–149]. A ML method called 'adaptive boosting' was employed to detect carbapenem resistance in *A. baumannii*, MRSA, and beta-lactam and co-trimoxazole resistance in *S. pneumoniae* with accuracies ranging from 88 to 99% [145]. Similarly, another ML method called 'gradient-boosting' was able to detect MIC in *K. pneumoniae* against 20 antibiotics [146]. A software package called 'Mykrobe predictor' detected resistance in *S. aureus* and *Mycobacterium tuberculosis* against 12 antibiotics [147]. These models were able to classify the pathogens as either resistant or sensitive, however, the features used by the algorithm to classify them are not known. In this regard, classification and regression trees (CART) and set covering machines (SCM) models were employed to detect resistance among 12 bacterial species against 56 antibiotic combinations. Both CART and SET are rule-based learning algorithms, which helped to interpret the resistance mechanisms by identifying the presence or absence of 'k-mers' (all of a gene sequence's subsequences of length *k*). These type of methods help to interpret the model's results based on the features it has used, thus overcoming the 'interpretability problem' (i.e., non-availability of data or features used to reach the conclusion by the ML method) [150]. MALDI TOF MS is being extensively used for identifying bacteria and fungi in diagnostic laboratory across the world. The fluconazole resistance in *C. albicans* was detected using three ML methods (Random Forest, Logistic regression and Linear discriminant analysis (LDA)) using spectral data. Of these three models, authors found that LDA was most robust method in detecting AMR with the accuracy, sensitivity, and specificity of 85.7%, 88.9%, and 83.3% respectively. Furthermore, another study employed the MALDI TOF spectral data from *S. aureus*, *E. coli*, and *K. pneumoniae* to predict the resistance phenotype. They used multilayer perceptron and gradient boost methods to get an area under receiver operator curve (AUROC) of 0.80, 0.74, and 0.74 [102]. AUROC is the metric used to measure the accuracy of the ML model in predicting the label (in this case, sensitive or resistant). A few studies have utilized patient data to predict if patients could develop resistant infections along with suitable therapies based on the local epidemiology of the pathogens. Microsoft's Azure ML algorithm determined the appropriate therapy based on patient demographic data and the resistance profiles of previously isolated microorganisms [151]. Another study applied ML methods to patients' medical records to predict antibiotic resistance against five antibiotics [152]. Patient demographic data and previous clinical and antibiotic history was used to predict AMR in pathogens isolated from urinary tract infection, such that the appropriate antibiotic could be prescribed [153].

5. Machine learning in drug design and drug discovery

The success rate of a potential therapeutic drug is extremely very low. Between 2000 and 2015, the success rate of drug development in oncology alone was as low as 3.4% [154]. Drug discovery involves various steps from target identification, optimization, validation, and hit discovery [155]. Machine learning is being implemented in the drug discovery process, from identifying the potential molecules or compounds against a particular disease to clinical trials [156]. A new drug, from its discovery through to clinical trials, involves huge cost (approximately 2.5 billion USD) and may take up to 10–15 years to come to market [157, 158]. The advent of high-throughput screening methods and the associated 'omics' data, along with the computer-assisted

drug design (CADD) technologies, encouraged pharmaceutical companies to focus on leveraging ML methods to identify potential drug targets as well as new drugs. These *in-silico* methods not only provide the molecular properties of the potential drug molecules, but they also have an impact on the attrition rate in the drug discovery pipeline, especially in pre-clinical experiments.

The first step in the drug discovery is to associate the target with the disease of interest. Here, it is hypothesized that inhibiting or modifying the target results in the alleviation of the disease. Machine learning has been applied to find the target using protein-protein, transcriptional, and metabolic interactions within cells and tissues. In this regard, semi-supervised learning models based on drug-protein interaction network information, chemical structures and genomic sequence data were able to predicted drug-protein interactions on enzyme, ion channel, GPCR (G protein coupled receptor), and nuclear receptor datasets [159]. A decision tree-based meta-classifier was employed to predict genes based on the aforementioned interactions that are associated with morbidity and that can be used as targets [160]. Similarly, a SVM model was able to classify proteins as drug targets and non-drug targets, for breast, pancreatic, and ovarian cancers [156]. In this study, after predicting multiple targets, two of the predicted targets were validated using peptide inhibitors, which had antiproliferative activity on cell culture models. Other studies have utilized ML methods for identifying drug targets, including for Huntington's disease [161]. The drug-protein interaction (DPI) databases consist of drugs that interact with therapeutic protein targets. However, these drugs might interact with the non-target proteins *in-vivo*, leading to side-effects or toxicity. Furthermore, knowledge on the drug and non-target interaction is limited. To address this knowledge gap, a study used a pool of 35 ML methods to predict DPIs based on the similarities between drugs and protein targets [162].

Support vector machines have been extensively used in drug development. The SVM method has been applied to raw data to predict the radiation protection function and toxicity for radioprotectors targeting p53 [163]. A regression-SVM model was used to assess target-ligand interactions [164]. Support vector machines were also able to predict the 'druggability' based on the structure of target [165] and have been used for other applications such as identifying drug-target interaction [109], cancer cell properties, drug resistance [110], selection of therapeutic compounds from public database [166], predicting properties of organic compound [167], designing new ligands [168], and virtual screening [169]. Random forest algorithms have been used to improve scoring function performance in ligand-protein binding affinity [169]. Random forest approaches have also been used to select molecular descriptors to achieve better accuracy for the compounds designed for drugs used in immune network technology [170]. Multilayer perceptron (MLP) algorithm is another ML approach that has been mainly used to generate compounds automatically for *de novo* drug design [171]. Yavuz et al. used MLP approach to predict the secondary structure of the proteins, which are used in drug design [133]. Deep learning approaches such as deep neural networks (DNNs), CNNs, RNNs, and autoencoders have been exploited in the drug discovery process. Deep learning algorithms increase the prediction performance on quantitative structure-activity relationship by retrieving feature extractions and capabilities in chemical characters automatically. 'DeepChem' is a multi-task neural network platform that helps in performing drug development process [172]. Convolutional neural networks have been utilized to predict affinities in protein-ligand binding [114, 173, 174]. Additionally, RNNs have been employed to virtually screen of molecular libraries to find anti-cancer agents via molecular

fingerprints [175]. Finally, autoencoders have been used to generate molecules in *de novo* drug design [176, 177].

Machine learning approaches have been used to discover antibiotics. Stokes et al. discovered an antibiotic from the 'Drug Repurposing Hub' called halicin. This drug is effective against *E. coli*, *Clostridioides difficile*, and pan-resistant *Acinetobacter bahumanii* [178]. Machine learning methods can mine large databases of genes and metabolites to identify molecule types that may include novel antibiotics [179, 180]. Machine learning methods are also being applied to the databases such as 'ChEMBL', which contains 1.9 million compounds with biological activity against 12,500 targets [181], 'BindingDB', which consists of 805,000 compounds with their binding affinities and 7500 protein targets [182], and 'AnitbioticDB', which consists of 1100 compounds that are in different stage of development for therapeutic use [183]. Antimicrobial peptides (AMPs) are found in all classes of life and are an important component of the innate immune response. Xiao et al. used fuzzy k-nearest neighbor algorithm to identify and define the functions of AMPs [184]. Another study used a semi-supervised density-based clustering algorithm model on linear AMPs that are active against gram-negative strains. Wang et al., applied four ML methods to discover new agents against MRSA. In this study, the authors derived *in-silico* models from 5451 cell-based anti-MRSA assay data using Bayesian, SVM, recursive partitioning, and k-nearest neighbor methods. By applying a ML approach to the 'Guangdong Small molecule Tangible Library' (which contains over 7500 small molecules), 56 hits were found, of which 12 novel anti-MRSA compounds were reported [185]. Targeting components in bacteria that are absent in humans can lead to new treatments against infections. DNA gyrase present in bacteria was targeted by Li et al. to discover anti-DNA gyrase compound using a ML approach [186]. In the same study, the authors also used *in-vitro* models to verify the virtual hits to check the hit activities against *E. coli*, MRSA, and other bacteria. Machine learning approaches have also been applied to discover antifungal drugs. For instance, a ML approach was employed to generate genome-wide gene essentiality predictions for *C. albicans* using a functional genomics resource named 'Gene Replacement and Conditional Expression' to identify three primary targets out of 866 genes. These three genes were involved in kinetochore function, mitochondrial integrity, and translation; glutaminyl-tRNA synthetase Gln4 was then identified as the target of N-pyrimidinyl-β-thiophenylacrylamide, which is an antifungal compound [187]. Temporal convolutional networks (TCNs) have been developed and deployed for antifungal peptide (AFP) prediction using deep learning models [188]. Similarly, Mousavizadegan et al. used pseudo amino acid composition to predict AFPs using a SVM algorithm [138]. Three peptides with highest prediction score were subsequently used in *in-vitro* assays. Sharma et al. proposed 'Deep-AFPpred', a deep learning classifier that predicts AFPs from protein sequence data [189].

6. Challenges and prospects

Antimicrobial resistance is an emerging global health crisis. As infectious microorganisms are evolving resistance through genetic and nongenetic mechanisms, new methods are required to rapidly diagnose and treat drug-resistant infections. The recent discovery of novel forms of AMR, including tolerance, persistence, and nongenetic resistance highlights the ingenuity of pathogenic microorganisms as well as the multifaceted nature of this problem. Digitization of clinical records presents opportunities for leveraging ML methods for fast and accurate identification of resistant

microorganisms. However, applying ML methods to detect AMR is still in the nascent stage. Importantly, the quantity and quality of the data required to detect resistance among bacteria and fungi are still limited. Furthermore, ML models currently used elsewhere require optimization to successfully detect AMR. Advancement in the areas of laboratory diagnosis of infectious agents and sharing of data across different centers could pave the way forward for using ML methods identify and detecting drug-resistant microorganisms.

Machine learning has played an important role in the discovery of drugs by identifying novel drug targets and drug molecules. Several new drugs discovered using ML methods have been successful in clinical trials after spending comparatively less time in the drug discovery pipeline. Though ML methods are proving to useful in drug design and drug discovery, several challenges still exist. For instance, the absence of sufficient training data as well as biased, faulty, or noisy training data results in poor ML model predictions. To address this, methods to remove outliers, and filter out unwanted features are being developed to increase the predictive power of ML models.

Another issue is that ML algorithms employ a 'black box' approach to train ML models. Specifically, how the features are being interpreted during each stage of the training to come to an accurate prediction is largely still not understood. An area of research called explainable artificial intelligence (XAI) has emerged to address this issue. XAI consists of processes and methods that help the human users to comprehend the results generated by ML algorithms. Also, XAI helps to characterize the model accuracy, transparency, and outcomes [190]. Applying XAI in the field of AMR research may lead to the discovery of novel resistance mechanisms. Finally, the heterogeneity of many databases restricts the incorporation of ML algorithms to these databases. However, the data on disease, drug compounds, and AMR mechanisms are growing day-by-day, leading to the continuous curation of ML models. Other challenges for deploying ML algorithms include cross-platform normalization, statistical issues, and the division of testing datasets. Many of these issues may be resolved through sophisticated data preprocessing methods. Importantly, these data and interpretability issues will need to be resolved before ML methods are more widely adopted in scientific research and trusted in clinical settings.

Acknowledgements

Author details

Shamanth A. Shankarnarayan[1], Joshua D. Guthrie[1] and Daniel A. Charlebois[1,2]*

1 Department of Physics, University of Alberta, Edmonton, Canada

2 Department of Biological Sciences, University of Alberta, Edmonton, Canada

*Address all correspondence to: dcharleb@ualberta.ca

IntechOpen

19

References

[1] Fleming A. Sir Alexander Fleming—Nobel Lecture: Penicillin. Nobel Lect; 1945

[2] WHO. Antimicrobial Resistance [Internet]. 2021. Available from: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance

[3] MacGowan AP. Clinical implications of antimicrobial resistance for therapy. The Journal of Antimicrobial Chemotherapy. 2008;**62**(SUPPL. 2):105-114

[4] O'Neill J. Review on Antimicrobial Resistance: Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. London: Wellcome Trust; 2016. p. 80

[5] Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. Lancet. 2022;**6736**(21):629-655

[6] Nelson RE, Hatfield KM, Wolford H, Samore MH, Scott RD, Reddy SC, et al. National estimates of healthcare costs associated with multidrug-resistant bacterial infections among hospitalized patients in the United States. Clinical Infectious Diseases. 2007;**2021**(72):S17-S26

[7] Mulani MS, Kamble EE, Kumkar SN, Tawre MS, Pardesi KR. Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: A review. Frontiers in Microbiology. 2019;**10**(APR):539

[8] Jernigan JA, Hatfield KM, Wolford H, Nelson RE, Olubajo B, Reddy SC, et al. Multidrug-resistant bacterial infections in U.S. hospitalized patients, 2012-2017.

The New England Journal of Medicine. 2020;**382**(14):1309-1319

[9] Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, et al. Discovery, research, and development of new antibiotics: The WHO priority list of antibiotic-resistant bacteria and tuberculosis. The Lancet Infectious Diseases. 2018;**18**(3):318-327

[10] Wall G, Lopez-Ribot JL. Current antimycotics, new prospects, and future approaches to antifungal therapy. Antibiotics. 2020;**9**(8):1-10

[11] Nnadi NE, Carter DA. Climate change and the emergence of fungal pathogens. PLoS Pathogens. 2021;**17**(4):1-6

[12] Pappas PG, Lionakis MS, Arendrup MC, Ostrosky-Zeichner L, Kullberg BJ. Invasive candidiasis. Nature Reviews Disease Primers. 2018;**4**:18026

[13] Tracking Candida auris | Candida auris | Fungal Diseases | CDC [Internet]. 2022. Available from: https://www.cdc.gov/fungal/candida-auris/tracking-c-auris.html#historical

[14] Centers for Disease Control and Prevention. Tracking Candida auris: Candida auris Fungal Diseases CDC [Internet]. Centers for Disease Control and Prevention. 2019. Available from: https://www.cdc.gov/fungal/candida-auris/tracking-c-auris.html

[15] Oh BJ, Shin JH, Kim MN, Sung H, Lee K, Joo MY, et al. Biofilm formation and genotyping of *Candida haemulonii*, *Candida pseudohaemulonii*, and a proposed new species (*Candida auris*) isolates from Korea. Medical Mycology. 2010;**49**(1):98-102

[16] Rhodes J, Fisher MC. Global epidemiology of emerging *Candida auris*. Current Opinion in Microbiology. 2019;**52**:84-89

[17] Biswal M, Rudramurthy SM, Jain N, Shamanth AS, Sharma D, Jain K, et al. Controlling a possible outbreak of *Candida auris* infection: Lessons learnt from multiple interventions. The Journal of Hospital Infection. 2017;**97**(4):363-370

[18] European Centre for Disease Prevention and Control. Candida Auris Outbreak in Healthcare Facilities in Northern Italy, 2019-2021. ECDC: Stockholm; 2022

[19] Schelenz S, Hagen F, Rhodes JL, Abdolrasouli A, Chowdhary A, Hall A, et al. First hospital outbreak of the globally emerging *Candida auris* in a European hospital. Antimicrobial Resistance and Infection Control. 2016;**5**(1):35

[20] Chen J, Tian S, Han X, Chu Y, Wang Q, Zhou B, et al. Is the superbug fungus really so scary? A systematic review and meta-analysis of global epidemiology and mortality of *Candida auris*. BMC Infectious Diseases. 2020;**20**(1):1-10

[21] Du H, Bing J, Hu T, Ennis CL, Nobile CJ, Huang G. Candida auris: Epidemiology, biology, antifungal resistance, and virulence. PLoS Pathogens. 2020;**16**(10):1-18

[22] Chow NA, de Groot T, Badali H, Abastabar M, Chiller TM, Meis JF. Potential fifth clade of *Candida auris*, Iran, 2018. Emerging Infectious Diseases. 2019;**25**(9):1780-1781

[23] Osei SJ. *Candida auris*: A systematic review and meta-analysis of current updates on an emerging multidrug-resistant pathogen. Microbiology. 2018;**7**(4):1-29

[24] Magiorakos A-P, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. Clinical Microbiology and Infection. 2012;**18**(3):268-281

[25] Lyman M, Forsberg K, Reuben J, Dang T, Free R, Seagle EE, et al. Notes from the field: Transmission of pan-resistant and Echinocandin-resistant *Candida auris* in health care facilities—Texas and the District of Columbia, January–April 2021. MMWR. Morbidity and Mortality Weekly Report. 2021;**70**(29):1022-1023

[26] Verweij PE, Lucas JA, Arendrup MC, Bowyer P, Brinkmann AJF, Denning DW, et al. The one health problem of azole resistance in *Aspergillus fumigatus*: Current insights and future research agenda. Fungal Biology Reviews. 2020;**34**(4):202-214

[27] Rudramurthy SM, Shankarnarayan SA, Dogra S, Shaw D, Mushtaq K, Paul RA, et al. Mutation in the squalene epoxidase gene of *Trichophyton interdigitale* and *Trichophyton rubrum* associated with Allylamine resistance. Antimicrobial Agents and Chemotherapy. May 2018;**62**(5):1-9

[28] Kano R, Kimura U, Kakurai M, Hiruma J, Kamata H, Suga Y, et al. *Trichophyton indotineae* sp. nov.: A new highly terbinafine-resistant anthropophilic dermatophyte species. Mycopathologia. 2020;**185**(6):947-958

[29] Laxminarayan R, Heymann DL. Challenges of drug resistance in the developing world. BMJ. 2012;**344**(7852):3-6

[30] Laxminarayan R, Duse A, Wattal C, Zaidi AKM, Wertheim HFL,

Sumpradit N, et al. Antibiotic resistance-the need for global solutions. The Lancet Infectious Diseases. 2013;**13**(12):1057-1098

[31] Huang AM, Newton D, Kunapuli A, Gandhi TN, Washer LL, Isip J, et al. Impact of rapid organism identification via matrix-assisted laser desorption/ionization time-of-flight combined with antimicrobial stewardship team intervention in adult patients with bacteremia and candidemia. Clinical Infectious Diseases. 2013;**57**(9):1237-1245

[32] Burnham CAD, Leeds J, Nordmann P, O'Grady J, Patel J. Diagnosing antimicrobial resistance. Nature Reviews. Microbiology. 2017;**15**(11):697-703

[33] Moult J, Fidelis K, Kryshtafovych A, Schwede T, Topf M. Critical Assessment of Techniques for Protein Structure Prediction, Fourteenth Round. 2020. pp. 1-344

[34] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;**596**(7873):583-589

[35] Mikolov T, Deoras A, Povey D, Burget L, Černocký J. Strategies for training large scale neural network language models. In: 2011 IEEE Work Autom Speech Recognit Understanding, ASRU 2011, Proc. 2011. pp. 196-201

[36] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017;**60**(6):84-90

[37] Samuel AL. Some studies in machine learning using the game of checkers. IBM Journal of Research and Development. 1959;**3**:210-229

[38] Awad M, Khanna R. Machine learning. In: Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Berkeley, CA: Apress; 2015. pp. 1-18

[39] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of go with deep neural networks and tree search. Nature. 2016;**529**(7587):484-489

[40] Degrave J, Felici F, Buchli J, Neunert M, Tracey B, Carpanese F, et al. Magnetic control of tokamak plasmas through deep reinforcement learning. Nature. 2022;**602**(7897):414-419

[41] Weissler EH, Naumann T, Andersson T, Ranganath R, Elemento O, Luo Y, et al. The role of machine learning in clinical research: Transforming the future of evidence generation. Trials. 2021;**22**(1):537

[42] Ripoli A, Sozio E, Sbrana F, Bertolino G, Pallotto C, Cardinali G, et al. Personalized machine learning approach to predict candidemia in medical wards. Infection. 2020;**48**(5):749-759

[43] Jaroszewisz A, Ernst J. An integrative approach for fine-mapping chromatin interactions. Bioinformatics. 2020;**36**(6):1704-1711

[44] Movva R, Greenside P, Marinov GK, Nair S, Shrikumar A, Kundaje A. Deciphering regulatory DNA sequences and noncoding genetic variants using neural network models of massively parallel reporter assays. PLoS One. 2019;**14**(6):1-20

[45] Rozenwald MB, Galitsyna AA, Sapunov GV, Khrameeva EE, Gelfand MS. A machine learning framework for the prediction of chromatin folding in Drosophila using epigenetic features. PeerJ Computer Science. 2020;**6**:2-21

[46] Talukder A, Barham C, Li X, Hu H. Interpretation of deep learning in genomics and epigenomics. Briefings in Bioinformatics. 2021;**22**(3):1-16

[47] Chen X, Clarence Yan C, Luo C, Ji W, Zhang Y, Dai Q. Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity. Scientific Reports. 2015;**5**(June):1-12

[48] Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Pantelis G, Lescure FX, et al. Machine learning for clinical decision support in infectious diseases: A narrative review of current applications. Clinical Microbiology and Infection. 2020;**26**(5):584-595

[49] Martinez JL. General principles of antibiotic resistance in bacteria. Drug Discovery Today: Technologies. 2014;**11**:33-39

[50] Zhang G, Feng J. The intrinsic resistance of bacteria. Yi chuan = Hered. 2016;**38**(10):872-880

[51] Brauner A, Fridman O, Gefen O, Balaban NQ. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. Nature Reviews. Microbiology. 2016;**14**(5):320-330

[52] Berman J, Krysan DJ. Drug resistance and tolerance in fungi. Nature Reviews. Microbiology. 2020;**18**(6):319-331

[53] Wood TK, Knabel SJ, Kwan BW. Bacterial persister cell formation and dormancy. Applied and Environmental Microbiology. 2013;**79**:7116-7121

[54] Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. Bacterial persistence as a phenotypic switch. Science (80-). 2004;**305**(5690):1622-1625

[55] Moyed HS, Bertrand KP. hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. Journal of Bacteriology. 1983;**155**(2):768-775

[56] Balaban NQ, Helaine S, Lewis K, Ackermann M, Aldridge B, Andersson DI, et al. Definitions and guidelines for research on antibiotic persistence. Nature Reviews. Microbiology. 2019;**17**(7):441-448

[57] Hammoud MS, Al-Taiar A, Fouad M, Raina A, Khan Z. Persistent candidemia in neonatal care units: Risk factors and clinical significance. International Journal of Infectious Diseases. 2013;**17**(8):e624-e628

[58] Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;**297**(5584):1183-1186

[59] Adam M, Murali B, Glenn NO, Potter SS. Epigenetic inheritance based evolution of antibiotic resistance in bacteria. BMC Evolutionary Biology. 2008;(8):52

[60] Farquhar KS, Rasouli Koohi S, Charlebois DA. Does transcriptional heterogeneity facilitate the development of genetic drug resistance? BioEssays. 2021;**43**(8):1-7

[61] Reygaert WC. An overview of the antimicrobial resistance mechanisms of bacteria. AIMS Microbiology. 2018;**4**(3):482-501

[62] Chancey ST, Zähner D, Stephens DS. Acquired inducible antimicrobial resistance in Gram-positive bacteria. Future Microbiology. 2012;**7**(8):959-978

[63] Lambert PA. Cellular impermeability and uptake of biocides and antibiotics in

gram-positive bacteria and mycobacteria. Symposium Series (Society for Applied Microbiology). 2002;**31**:46S-54S

[64] Blair JMA, Richmond GE, Piddock LJV. Multidrug efflux pumps in Gram-negative bacteria and their role in antibiotic resistance. Future Microbiology. 2014;**9**(10):1165-1177

[65] Gill MJ, Simjee S, Al-Hattawi K, Robertson BD, Easmon CS, Ison CA. Gonococcal resistance to beta-lactams and tetracycline involves mutation in loop 3 of the porin encoded at the penB locus. Antimicrobial Agents and Chemotherapy. 1998;**42**(11):2799-2803

[66] Mah T-F. Biofilm-specific antibiotic resistance. Future Microbiology. 2012;**7**(9):1061-1072

[67] Reygaert W. Methicillin-resistant *Staphylococcus aureus* (MRSA): Molecular aspects of antimicrobial resistance and virulence. Clinical Laboratory Science. 2009;**22**(2):115-119

[68] Cox G, Wright GD. Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions. International Journal of Medical Microbiology. 2013;**303**:287-292

[69] Roberts MC. Resistance to macrolide, lincosamide, streptogramin, ketolide, and oxazolidinone antibiotics. Applied Biochemistry and Biotechnology—Part B Molecular Biotechnology. 2004;**28**:47-62

[70] Redgrave LS, Sutton SB, Webber MA, Piddock LJV. Fluoroquinolone resistance: Mechanisms, impact on bacteria, and role in evolutionary success. Trends in Microbiology. 2014;**22**(8):438-445

[71] Huovinen P, Sundström L, Swedberg G, Sköld O. Trimethoprim and sulfonamide resistance. Antimicrobial Agents and Chemotherapy. 1995;**39**(2):279-289

[72] Blair JMA, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJV. Molecular mechanisms of antibiotic resistance. Nature Reviews. Microbiology. 2015;**13**(1):42-51

[73] Kumar A, Schweizer HP. Bacterial resistance to antibiotics: Active efflux and reduced uptake. Advanced Drug Delivery Reviews. 2005;**57**(10):1486-1513

[74] Beck-Sagué C, Jarvis WR. Secular trends in the epidemiology of nosocomial fungal infections in the United States, 1980-1990. National Nosocomial Infections Surveillance System. The Journal of Infectious Diseases. 1993;**167**(5):1247-1251

[75] White TC, Holleman S, Dy F, Mirels LF, Stevens DA. Resistance mechanisms in clinical isolates of *Candida albicans*. Antimicrobial Agents and Chemotherapy. 2002;**46**(6):1704-1713

[76] White TC. Increased mRNA levels of ERG16, CDR, and MDR1 correlate with increases in azole resistance in Candida albicans isolates from a patient infected with human immunodeficiency virus. Antimicrobial Agents and Chemotherapy. 1997;**41**(7):1482-1487

[77] Franz R, Kelly SL, Lamb DC, Kelly DE, Ruhnke M, Morschhäuser J. Multiple molecular mechanisms contribute to a stepwise development of fluconazole resistance in clinical *Candida albicans* strains. Antimicrobial Agents and Chemotherapy. 1998;**42**(12):3065-3072

[78] Braun BR, van het Hoog M, d'Enfert C, Martchenko M, Dungan J, Kuo A, et al. A human-curated annotation of the Candida albicans genome. PLoS Genetics. 2005;**1**:0036-0057

[79] Sanglard D, Coste A, Ferrari S. Antifungal drug resistance mechanisms in fungal pathogens from the perspective of transcriptional gene regulation. FEMS Yeast Research. 2009;**9**(7):1029-1050

[80] Flowers SA, Barker KS, Berkow EL, Toner G, Chadwick SG, Gygax SE, et al. Gain-of-function mutations in UPC2 are a frequent cause of ERG11 upregulation in azole-resistant clinical isolates of *Candida albicans*. Eukaryotic Cell. 2012;**11**(10):1289-1299

[81] Sanglard D. Diagnosis of antifungal drug resistance mechanisms in fungal pathogens: Transcriptional gene regulation. Current Fungal Infection Reports. 2011;**5**(3):157-167

[82] Selmecki A, Forche A, Berman J. Genomic plasticity of the human fungal pathogen *Candida albicans*. Eukaryotic Cell. 2010;**9**(7):991-1008

[83] Gulshan K, Moye-Rowley WS. Multidrug resistance in fungi. Eukaryotic Cell. 2007;**6**(11):1933-1942

[84] Cowen LE, Steinbach WJ. Stress, drugs, and evolution: The role of cellular signaling in fungal drug resistance. Eukaryotic Cell. 2008;**7**(5):747-764

[85] Perlin DS. Current perspectives on echinocandin class drugs. Future Microbiology. 2011;**6**(4):441-457

[86] Katiyar S, Pfaller M, Edlind T. *Candida albicans* and *Candida glabrata* clinical isolates exhibiting reduced Echinocandin susceptibility. Antimicrobial Agents and Chemotherapy. 2006;**50**(8):2892-2894

[87] Munro CA, Selvaggini S, de Bruijn I, Walker L, Lenardon MD, Gerssen B, et al. The PKC, HOG and Ca2+ signalling pathways co-ordinately regulate chitin

synthesis in *Candida albicans*. Molecular Microbiology. 2007;**63**(5):1399-1413

[88] Walker LA, Munro CA, de Bruijn I, Lenardon MD, McKinnon A, Gow NAR. Stimulation of chitin synthesis rescues *Candida albicans* from Echinocandins. Cormack BP, editor. PLoS Pathogens. 2008;**4**(4):e1000040

[89] Loo AS, Muhsin SA, Walsh TJ. Toxicokinetic and mechanistic basis for the safety and tolerability of liposomal amphotericin B. Expert Opinion on Drug Safety. 2013;**12**(6):881-895

[90] Vanden Bossche H, Marichal P, Odds FC. Molecular mechanisms of drug resistance in fungi. Trends in Microbiology. 1994;**2**(10):393-400

[91] Nolte FS, Parkinson T, Falconer DJ, Dix S, Williams J, Gilmore C, et al. Isolation and characterization of fluconazole- and amphotericin B-resistant *Candida albicans* from blood of two patients with leukemia. Antimicrobial Agents and Chemotherapy. 1997;**41**(1):196-199

[92] Blum G, Hörtnagl C, Jukic E, Erbeznik T, Pümpel T, Dietrich H, et al. New insight into amphotericin B resistance in *Aspergillus terreus*. Antimicrobial Agents and Chemotherapy. 2013;**57**(4):1583-1588

[93] Eddouzi J, Parker JE, Vale-Silva LA, Coste A, Ischer F, Kelly S, et al. Molecular mechanisms of drug resistance in clinical Candida species isolated from Tunisian hospitals. Antimicrobial Agents and Chemotherapy. 2013;**57**(7):3182-3193

[94] CLSI. Reference Method for Broth Dilution Antifungal Susceptibility Testing of Filamentous Fungi; Approved Standard—CLSI Document M38-A2. Vol. 28. Clinical and Laboratory Standards Institute (CLSI); 2008. p. 52

[95] European Committee on Antimicrobial Susceptibility Testing—EUCAST. EUCAST reading guide for broth microdilution. Read Guid broth microdilution. 2020;**1.0**(March):17

[96] McEwen SA, Collignon PJ. Antimicrobial resistance: A one health colloquium. Microbiology Spectrum. 2018;**6**(2):1-26

[97] Pulido MR, García-Quintanilla M, Martín-Peña R, Cisneros JM, McConnell MJ. Progress on the development of rapid methods for antimicrobial susceptibility testing. The Journal of Antimicrobial Chemotherapy. 2013;**68**(12):2710-2717

[98] Vasala A, Hytönen VP, Laitinen OH. Modern tools for rapid diagnostics of antimicrobial resistance. Frontiers in Cellular and Infection Microbiology. 2020;**10**:308

[99] Boyle D. Unitaid TB Diagnostics—NAAT for Microscopy Stations [Internet]. 2017. Available from: http://unitaid.org/assets/2017-Unitaid-TB-Diagnostics-Technology-Landscape.pdf

[100] Peytavi R, Raymond FR, Gagné D, Picard FJ, Jia G, Zoval J, et al. Microfluidic device for rapid (<15 min) automated microarray hybridization. Clinical Chemistry. 2005;**51**(10):1836-1844

[101] Dougherty K, Smith BA, Moore AF, Maitland S, Fanger C, Murillo R, et al. Multiple phenotypic changes associated with large-scale horizontal gene transfer. PLoS One. 2014;**9**(7):e102170

[102] Weis C, Cuénod A, Rieck B, Dubuis O, Graf S, Lang C, et al. Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. Nature Medicine. 2022;**28**(1):164-174

[103] Mitchell TM. Machine Learning. McGraw Hill; 1997. p. 414

[104] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of go without human knowledge. Nature. 2017;**550**(7676):354-359

[105] Davies A, Veličković P, Buesing L, Blackwell S, Zheng D, Tomašev N, et al. Advancing mathematics by guiding human intuition with AI. Nature. 2021;**600**(7887):70-74

[106] Russell S, Norvig P. Artificial Intelligence: A Modern Approach. New Jersey: Pearson; 2010

[107] Trevor H, Jerome F, Robert T. The elements of statistical learning data mining, inference, and prediction. The Elements of Statistical Learning. 2009;**27**:83-85

[108] Gareth J, Daniela W, Hastie T, Robert T. An Introduction to Statistical Learning with Applications in R. 2nd ed. New York: Springer Text in Statistics; 2011. 110p

[109] Wang Q, Feng Y, Huang J, Wang T, Cheng G. A novel framework for the identification of drug target proteins: Combining stacked auto-encoders with a biased support vector machine. PLoS One. 2017;**12**(4):e0176486

[110] Gupta S, Chaudhary K, Kumar R, Gautam A, Nanda JS, Dhanda SK, et al. Prioritization of anticancer drugs against a cancer using genomic features of cancer cells: A step towards personalized medicine. Scientific Reports. 2016;**6**(1):23857

[111] Lemos P, Jeffrey N, Cranmer M, Ho S, Battaglia P. Rediscovering orbital mechanics with machine learning. arXiv. 2022

[112] Cortes C, Vapnik V, Saitta L. Support-vector networks. Machine Learning. 1995;**20**(3):273-297

[113] Breiman L. Random forests. Machine Learning. 2001;**45**(1):5-32

[114] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;**521**:436-444

[115] Hinton G, Sejnowski T. Unsupervised learning: Foundations of neural computation. Computers & Mathematics with Applications. 1999;**38**(5-6):256

[116] Steinwart I, Gov D, Gov J. A classification framework for anomaly detection Don hush Clint Scovel. Journal of Machine Learning Research. 2005;**6**:211-232

[117] Shon T, Moon J. A hybrid machine learning approach to network anomaly detection. Information Sciences. 2007;**177**(18):3799-3821

[118] Tenenbaum JB, De Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science (80-). 2000;**290**(5500):2319-2323

[119] Van Der Maaten L, Postma E, Van den Herik J. Dimensionality reduction: A comparative review. Journal of Machine Learning Research. 2009;**10**:66-71

[120] Chapelle O, Schölkopf B, Zien A. Semi-supervised learning. 2010;**508**:373-440

[121] van Engelen JE, Hoos HH. A survey on semi-supervised learning. Machine Learning. 2020;**109**(2):373-440

[122] Hartigan JA, Wong MA. Algorithm AS 136: A K-means clustering algorithm. Applied Statistics. 1979;**28**(1):100

[123] Likas A, Vlassis N, J. Verbeek J. The global k-means clustering algorithm. Pattern Recognition. 2003;**36**(2):451-461

[124] Johnson SC. Hierarchical clustering schemes. Psychom. 1967;**32**(3):241-254

[125] Murtagh F, Contreras P. Algorithms for hierarchical clustering: An overview, II. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2017;**7**(6):e1219

[126] Birant D, Kut A. ST-DBSCAN: An algorithm for clustering spatial–temporal data. Data & Knowledge Engineering. 2007;**60**(1):208-221

[127] Liu FT, Ting KM, Zhou ZH. Isolation forest. In: Proc—IEEE Int Conf Data Mining. ICDM; 2008. pp. 413-422

[128] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems. 1987;**2**(1-3):37-52

[129] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science (80-). 2000;**290**(5500):2323-2326

[130] Moon TK. The expectation-maximization algorithm. IEEE Signal Processing Magazine. 1996;**13**(6):47-60

[131] Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: A survey. Journal of Artificial Intelligence Research. 1996;**4**:237-285

[132] Sutton RS, Barto AG. Reinforcement learning. In: An Introduction. 2nd ed. United States: MIT Press; 2018. pp. 1-3

[133] Carkli Yavuz B, Yurtay N, Ozkan O. Prediction of protein secondary structure with clonal selection algorithm and multilayer perceptron. IEEE Access. 2018;**6**:45256-45261

[134] Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. In: Shawe-Taylor J,

Zemel R, Bartlett P, Pereira F, Weinberger KQ, editors. Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2011

[135] Refaeilzadeh P, Tang L, Liu H. Cross-validation. In: Encyclopedia of Database Systems. 2016. pp. 1-7

[136] García S, Luengo J, Herrera F. Dealing with missing values. IntelligentSystems Reference Library. 2015;**72**:59-105

[137] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. {TensorFlow}: A system for {large-scale} machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16). Savannah, GA: USENIX Association; 2016. pp. 265-283

[138] Mousavizadegan M, Mohabatkar H. Computational prediction of antifungal peptides via Chou's PseAAC and SVM. Journal of Bioinformatics and Computational Biology. 2018;**16**(4):1850016

[139] Fabian P, Michel V, Varoquaux G, Thirion B, Dubourg V, Passos A, et al. Scikit-learn: Machine learning in python. Journal of Machine Learning Research. 2011;**12**:2825-2830

[140] Paszke A, Gross S, Massa F, Lerer A, Bradbury Google J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Conference proceedings: Advances in Neural Information Processing Systems: 2019

[141] Hicks AL, Wheeler N, Sánchez-Busó L, Rakeman JL, Harris SR, Grad YH. Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome

sequencing data. PLoS Computational Biology. 2019;**15**(9):e1007349

[142] Li D, Wang Y, Hu W, Chen F, Zhao J, Chen X, et al. Application of machine learning classifier to *Candida auris* drug resistance analysis. Frontiers in Cellular and Infection Microbiology. 2021;**11**:742062

[143] Delavy M, Cerutti L, Croxatto A, Prod'hom G, Sanglard D, Greub G, et al. Machine learning approach for *Candida albicans* fluconazole resistance detection using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Frontiers in Microbiology. 2020;**10**(January):3000

[144] Liu Z, Deng D, Lu H, Sun J, Lv L, Li S, et al. Evaluation of machine learning models for predicting antimicrobial resistance of *Actinobacillus pleuropneumoniae* from whole genome sequences. Frontiers in Microbiology. 2020;**11**(February):1-7

[145] Davis JJ, Boisvert S, Brettin T, Kenyon RW, Mao C, Olson R, et al. Antimicrobial resistance prediction in PATRIC and RAST. Scientific Reports. 2016;**6**:27930

[146] Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, et al. Developing an in silico minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. Scientific Reports. 2018;**8**(1):421

[147] Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. Nature Communications. 2015;**6**:10063

[148] Her HL, Wu YW. A pan-genome-based machine learning approach for

predicting antimicrobial resistance activities of the *Escherichia coli* strains. Bioinformatics. 2018;**34**(13):i89-i95

[149] Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, et al. Prediction of staphylococcus aureus antimicrobial resistance by whole-genome sequencing. Journal of Clinical Microbiology. 2014;**52**(4):1182-1191

[150] Drouin A, Letarte G, Raymond F, Marchand M, Corbeil J, Laviolette F. Interpretable genotype-to-phenotype classifiers with performance guarantees. Scientific Reports. 2019;**9**(1):4071

[151] Feretzakis G, Sakagianni A, Loupelis E, Kalles D, Skarmoutsou N, Martsoukou M, et al. Machine learning for antibiotic resistance prediction: A prototype using off-the-shelf techniques and entry-level data to guide empiric antimicrobial therapy. Healthcare Informatics Research. 2021;**27**(3): 214-221

[152] Lewin-Epstein O, Baruch S, Hadany L, Stein GY, Obolski U. Predicting antibiotic resistance in hospitalized patients by applying machine learning to electronic medical records. Clinical Infectious Diseases. 2021;**72**(11):e848-e855

[153] Didelot X, Pouwels KB. Machine-learning-assisted selection of antibiotic prescription. Nature Medicine. 2019;**25**(7):1033-1034

[154] Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics. 2019;**20**(2):273-286

[155] Vohora D, Singh G. Pharmaceutical Medicine and Translational Clinical Research. Elsevier; 2018. pp. 1-497

[156] Jeon J, Nim S, Teyra J, Datti A, Wrana JL, Sidhu SS, et al. A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. Genome Medicine. 2014;**6**(7):57

[157] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economics. 2016;**47**:20-33

[158] Turner JR. New Drug Development. New York, NY: Springer New York; 2010

[159] Xia Z, Wu L-Y, Zhou X, Wong STC. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. BMC Systems Biology. 2010;**4**(Suppl 2):S6

[160] Costa PR, Acencio ML, Lemke N. A machine learning approach for genome-wide prediction of morbid and druggable human genes based on systems-level data. BMC Genomics. 2010;**11**(Suppl. 5):S9

[161] Ament SA, Pearl JR, Cantle JP, Bragg RM, Skene PJ, Coffey SR, et al. Transcriptional regulatory networks underlying gene expression changes in Huntington's disease. Molecular Systems Biology. 2018;**14**(3):e7435

[162] Wang C, Kurgan L. Survey of similarity-based prediction of drug-protein interactions. Current Medicinal Chemistry. 2020;**27**(35):5856-5886

[163] Matsumoto A, Aoki S, Ohwada H. Comparison of random forest and SVM for raw data in drug discovery: Prediction of radiation protection and toxicity case study. International Journal of Machine Learning and Computing. 2016;**6**(2):145-148

[164] Li L, Wang B, Meroueh SO. Support vector regression scoring of receptor-ligand complexes for rank-ordering and virtual screening of chemical libraries.

Journal of Chemical Information and Modeling. 2011;**51**(9):2132-2138

[165] Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M. Combining global and local measures for structure-based druggability predictions. Journal of Chemical Information and Modeling. 2012;**52**(2):360-372

[166] Bundela S, Sharma A, Bisen PS. Potential compounds for oral cancer treatment: Resveratrol, nimbolide, lovastatin, bortezomib, vorinostat, berberine, pterostilbene, deguelin, andrographolide, and colchicine. PLoS One. 2015;**10**(11):e0141719

[167] Maltarollo VG, Kronenberger T, Espinoza GZ, Oliveira PR, Honorio KM. Advances with support vector machines for novel drug discovery. Expert Opinion on Drug Discovery. 2019;**14**:23-33

[168] Schneider G, Hartenfeller M, Proschak E. De novo drug design. Lead Generation Approaches in Drug Discovery. 2010. pp. 165-185

[169] Kinnings SL, Liu N, Tonge PJ, Jackson RM, Xie L, Bourne PE. A machine learning-based method to improve docking scoring functions and its application to drug repurposing. Journal of Chemical Information and Modeling. 2011;**51**(2):408-419

[170] Samigulina G, Zarina S. Immune network technology on the basis of random forest algorithm for computer-aided drug design. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Cham: Springer; 2017. pp. 50-61

[171] Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Central Science. 2018;**4**(2):268-276

[172] Ramsundar B, Liu B, Wu Z, Verras A, Tudor M, Sheridan RP, et al. Is multitask deep learning practical for pharma? Journal of Chemical Information and Modeling. 2017;**57**(8):2068-2076

[173] Leelananda SP, Lindert S. Computational methods in drug discovery. Beilstein Journal of Organic Chemistry. 2016;**12**:2694-2718

[174] Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G. KDEEP: Protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. Journal of Chemical Information and Modeling. 2018;**58**(2):287-296

[175] Kadurin A, Aliper A, Kazennov A, Mamoshina P, Vanhaelen Q, Khrabrov K, et al. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. Oncotarget. 2017;**8**(7):10883-10890

[176] Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature Biotechnology. 2019;**37**(9):1038-1040

[177] Kingma DP, Welling M. An introduction to variational autoencoders. Foundations and Trends in Machine Learning. 2019;**12**:307-392

[178] Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. Cell. 2020;**180**(4):688-702.e13

[179] Mohimani H, Kersten RD, Liu WT, Wang M, Purvine SO, Wu S, et al.

Automated genome mining of ribosomal peptide natural products. ACS Chemical Biology. 2014;**9**(7):1545-1551

[180] Cao L, Gurevich A, Alexander KL, Naman CB, Leão T, Glukhov E, et al. MetaMiner: A scalable peptidogenomics approach for discovery of ribosomal peptide natural products with blind modifications from microbial communities. Cell Systems. 2019;**9**(6):600-608.e4

[181] Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Research. 2017;**45**(D1):D945-D954

[182] Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Research. 2016;**44**(D1):D1045-D1053

[183] Farrell LJ, Lo R, Wanford JJ, Jenkins A, Maxwell A, Piddock LJV. Revitalizing the drug pipeline: AntibioticDB, an open access database to aid antibacterial research and development. The Journal of Antimicrobial Chemotherapy. 2018;**73**(9):2284-2297

[184] Xiao X, Wang P, Lin W-Z, Jia J-H, Chou K-C. iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types. Analytical Biochemistry. 2013;**436**(2):168-177

[185] Wang L, Le X, Li L, Ju Y, Lin Z, Gu Q, et al. Discovering new agents active against methicillin-resistant *Staphylococcus aureus* with ligand-based approaches. Journal of Chemical Information and Modeling. 2014;**54**(11):3186-3197

[186] Li L, Le X, Wang L, Gu Q, Zhou H, Xu J. Discovering new DNA gyrase inhibitors using machine learning approaches. RSC Advances. 2015;**5**(128):105600-105608

[187] Fu C, Zhang X, Veri AO, Iyer KR, Lash E, Xue A, et al. Leveraging machine learning essentiality predictions and chemogenomic interactions to identify antifungal targets. Nature Communications. 2021;**12**(1):6497

[188] Singh V, Shrivastava S, Kumar Singh S, Kumar A, Saxena S. Accelerating the discovery of antifungal peptides using deep temporal convolutional networks. Briefings in Bioinformatics. 2022;**23**(2):bbac008

[189] Sharma R, Shrivastava S, Kumar Singh S, Kumar A, Saxena S, Kumar SR. Deep-AFPpred: Identifying novel antifungal peptides using pretrained embeddings from seq2vec with 1DCNN-BiLSTM. Briefings in Bioinformatics. 2022;**23**(1):1-16

[190] Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable ai: A review of machine learning interpretability methods. Entropy. 2021;**23**(1):1-45