Progress in Microbes and Molecular Biology

HH PUBLISHER

Review Article

Unleashing the Power of Artificial Intelligence-Driven Drug Discovery in *Streptomyces*

Wei-Shan Ang[1], Jodi Woan-Fei Law[2], Yatinesh Kumari[3], Vengadesh Letchumanan[4], Loh Teng-Hern Tan[2]*

Article History

Received: 21 September 2024;

Received in Revised Form: 28 October 2024;

Accepted: 20 November 2024;

Available Online: 25 November 2024

[1]Novel Bacteria and Drug Discovery Research Group (NBDD), Microbiome and Bioresource Research Strength (MBRS), Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Bandar Sunway 47500, Selangor Darul Ehsan, Malaysia; wei.ang@monash.edu (W-SA)

[2]Microbiome Research Group, Research Centre for Life Science and Healthcare, Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute (CBI), University of Nottingham Ningbo China, Ningbo 315000, China; jodi-woan-fei.law@nottingham.edu.cn (JW-FL)

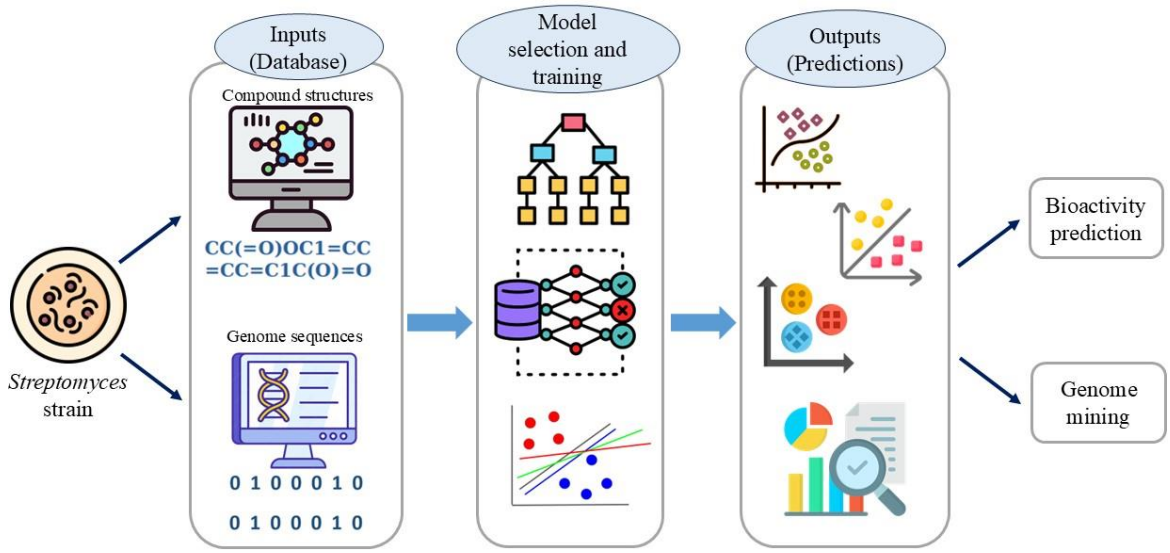
[3]Neurological disorder and aging research group (NDA), Neuroscience Research Strength (NRS), Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Selangor 47500, Malaysia; yatinesh.kumari@monash.edu (YK)

[4]Pathogen Resistome Virulome and Diagnostic Research Group (PathRiD), Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Bandar Sunway 47500, Selangor Darul Ehsan, Malaysia; vengadesh.letchumanan1@monash.edu (VL)

*Corresponding author: Loh Teng-Hern Tan; Microbiome Research Group, Research Centre for Life Science and Healthcare, Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute (CBI), University of Nottingham Ningbo China, Ningbo 315000, China Email: loh-teng-hern.tan@nottingham.edu.cn (LT-HT)

Abstract: The rise of antibiotic resistance has created an urgent need for the discovery of new antibiotic compounds. Streptomycin, the first antibiotic isolated from *Streptomyces* sp., paved the way for discovering other antibiotics for combating bacterial infections. By exploring the genome-based biosynthetic potential of various *Streptomyces* species, a vast array of secondary metabolites with potential therapeutic applications can be identified, contributing a transformative impact on the field of medicine. However, conventional screening approaches on novel natural products (NPs) from *Streptomyces* sp. have entered a bottleneck due to inefficiency. Fortunately, artificial intelligence (AI) and machine learning (ML) models enable rapid exploration and prediction of potential antibiotic compounds, increasing the probability of discovering new antibacterial compounds. AI-driven drug discovery in *Streptomyces* sp. represents a paradigm shift in the future quest for novel pharmaceutical agents. Various ML models have been developed and applied in different practical applications. Overall, the ML model is trained using input data and generates

outcomes based on prediction output. This review discusses the continued potential of *Streptomyces* sp. as a source of novel NPs, along with the application of ML throughout the NP drug discovery pipeline involving genome mining, biological activities prediction, and optimization compound production in *Streptomyces* microbial systems.



Graphical abstract. The role of machine learning in drug discovery from *Streptomyces.*

Keywords: *Streptomyces*; Secondary metabolites; Natural products; Drug discovery; Machine Learning; Artificial Intelligence; SDG 3 Good health and well-being

1. Introduction

Bioprospecting has played an important role in natural product-based drug discovery through the exploration, extraction, and screening of new natural compounds derived from plants, microorganisms, and animals for commercially valuable applications, particularly in harnessing their therapeutic effects in the pharmaceutical industry [1-5]. During the 'Golden Age' of antibiotic discovery (1940 to 1970), this approach was used in identifying 23 classes of antibiotics that are currently in clinical settings to prevent and treat human diseases [6]. Major antibiotic classes include aminoglycosides, cephalosporins, fluoroquinolones, macrolides, tetracyclines, and β-lactam. The introduction of antibiotics has remarkably changed the therapeutic paradigm and saved millions of lives from a wide range of bacterial infections [7]. Most of the new drug classes were identified originating from natural resources such as microbes and screened based on molecular, species, and genetic levels [8]. Whole-cell screen strategies have been implemented extensively on microorganisms for exploring potential biotherapeutic activity within the cellular context [9].

Over two-thirds of the clinically used antibiotics are natural products produced by the genus *Streptomyces* which have been considered as a bio-factory of a diverse range of natural compounds with antagonistic or pharmacological properties [10]. This high biosynthetic potential of *Streptomyces* was explored actively due to its large genome of 8-10 Mbp with high GC contents and large biosynthetic gene clusters (BGCs) that are able to synthesize a large variety of secondary metabolites [11, 12].

Nowadays, widespread antibiotic resistance has raised alarms in healthcare systems across the globe [13, 14]. Antibiotic resistance emerges when bacteria, viruses, fungi and parasites evolve different mechanisms to evade the therapeutic effects of antibiotics [15, 16]. One of the most widespread infections worldwide is Methicillin-resistant *Staphylococcus aureus* (MRSA) infection where *S. aureus* becomes resistant to methicillin and other β-lactam antibiotics in hospital settings, leading to significant morbidity, and mortality [17, 18]. The excessive and inappropriate use of antibiotics in both human medicine and agriculture has significantly accelerated the development of antibiotic resistance, making it one of the top 10 global public health threats as recognized by the World Health Organization (WHO) [15]. Discovering new antibiotics has been an alternative approach to combat this growing serious health threat. However, nature-driven drug discovery research on microbes has taken a back seat and evolved over the years. Traditional drug discovery approaches often face technical difficulties, particularly in screening programs, separation and isolation techniques of natural products produced from the primary or secondary metabolism of bacterial species under laboratory conditions. To address these issues, several technologies, such as combinatorial chemistry, high throughput screening (HTS), computational modeling, and artificial intelligence (AI), have been developed to accelerate the drug discovery process for natural products. [19]. The application of AI and machine learning (ML) incorporated with algorithms marks a revolutionary shift in drug discovery and development. AI refers to the development of machine learning ML models that simulate human-like intelligence and perform tasks adaptively [20]. In natural product discovery, ML techniques have been applied throughout the process involving compound screening, detecting biosynthetic gene clusters (BGCs), drug target interaction, compound optimisation and compound dereplication [21].

Herein, this review highlights the continued potential of *Streptomyces* as a source of novel natural products with the current state-of-the-art technology in antimicrobial natural product drug discovery and the utility of ML approaches in advancing microbial natural product discovery focusing on *Streptomyces sp*. Employing ML approaches in *Streptomyces* natural products discovery is still relatively nascent and needs to be explored more.

2. *Streptomyces* as a Source of Valuable Compounds

Presently, two-thirds of commercial and therapeutical antibiotics are derived from actinomycetes, nature's topmost antibiotic producers, and almost exclusively from *Streptomyces* sp. [22, 23]. Several previous studies have extensively reported the morphology, taxonomy, and genetics of *Streptomyces* as well as various metabolic pathways and enzymatic functions [24-26]. *Streptomyces* is a Gram-positive bacteria and the largest genus in the *Actinobacteria* phylum living in a wide range of environments, such as harsh, underexplored

habitats, terrestrial, marine regions and mangroves [22, 27]. It undergoes a complex life cycle that includes vegetative growth, sporulation, and antibiotic production. Besides, *Streptomyces* undergo multiple levels of morphological differentiation in response to the growing environments [28]. When a typical *Streptomyces* spore encounters favourable conditions, it forms filamentous branching structures of vegetative mycelium, further differentiates into the reproductive aerial mycelium and eventually leads to the formation of sporulation septa under nutrient depletion or stressed environment [29]. The mature spore is often associated with the production of secondary metabolites, including antibiotics as a result of nutrient limitation. Unlike other bacteria, *Streptomyces* have a large genome size of 8-10 Mb with an exceptionally high G+C content of > 70% and multiple biosynthetic gene clusters (BGCs) [30].

Remarkably, *Streptomyces* account for over 70% of commercially useful antibiotics, major types of antibiotics such as aminoglycosides, anthracyclines, glycopeptides, β-lactams, macrolides, ansamycins, nucleosides, peptides, polyenes, polyesters, and tetracyclines (Table 1). The diverse range of natural products with high structural diversity exhibits broad-spectrum activity against both Gram-positive and Gram-negative bacteria, antiviral, antifungal, cytotoxic, antitumor, anti-protozoal, anti-hypertensive, immunosuppressive, insecticide, and antioxidative properties [31-33]. Bioactive compounds with different biological activities isolated from different *Streptomyces* are represented in Table 2. In 1943, the first aminoglycoside antibiotic, streptomycin was discovered by Albert Schatz and Selman Waksman [34]. This antibiotic was isolated from *Streptomyces griseus* and contributed significantly to the treatment of various bacterial infections, including tuberculosis, plague, tularemia, and brucellosis. In recognition of his achievements in the discovery, Selman Waksman was awarded the Nobel Prize for Medicine in 1952 [35]. Besides, erythromycin is an antibiotic produced by *Streptomyces erythreus* that is used to treat a variety of bacterial infections, particularly respiratory tract and skin infections [36]. *Streptomyces* have profoundly revolutionized medicine with the discovery of antibiotics leading to the antibiotic era and have significantly reduced live mortality. Exploration of bioactive compounds from *Streptomyces* remains a valuable ally in the search for new solutions to overcome antibiotic resistance [37]. The unique and diverse range of bioactive compounds synthesised by *Streptomycetes* have high versatility and broad-spectrum antagonistic activity against both Gram-positive and Gram-negative bacteria [38-40].

3. Secondary Metabolite Biosynthetic Gene Clusters (smBGCs) in *Streptomyces*

Secondary metabolite production in *Streptomyces* was triggered during the stationary phase in response to environmental stress or lack of nutrients. Secondary metabolites are natural products that, different from primary metabolites involve normal growth, development, and reproduction. In contrast, secondary metabolites are primarily involved in the defence system [61]. In *Streptomyces*, the biosynthesis of secondary metabolites is mainly regulated by nonribosomal polyketide synthetase (NRPS) pathways and polyketide synthetase (PKS) [62]. Polyketides are secondary metabolites or natural products produced by *Streptomyces* and synthesized through sequential reactions catalyzed by a set of enzyme

complexes known as polyketide synthases (PKSs) [63]. *Streptomyces*-derived bioactive natural products are produced by means of complex 'secondary metabolic' pathways encoded by the secondary metabolite biosynthetic gene clusters (smBGCs) [64]. smBGCs is a grouping of genes in genome-sequenced bacteria that encode the enzymes and proteins involved in the pathways of precursor biosynthesis, assembly, modification, transport, and regulation of a particular secondary metabolite. Different modular structures within the gene clusters are responsible for distinct steps in the secondary metabolite biosynthetic pathway [65]. Each smBGC contains a set of core biosynthetic genes, accessory genes, and regulatory elements that work synergistically to synthesize secondary metabolites. The expression of these gene clusters is tightly controlled by molecular mechanisms in complex regulatory networks in response to environmental stresses found in the bacteria's native habitats [66]. Advances in sequencing technology have demonstrated that a typical *Streptomyces* genome encodes around 25–50 BGCs. However, approximately 90% of them are cryptic or silent under laboratory fermentation conditions, limiting the synthesis of secondary metabolites [67]. Therefore, to maximize secondary metabolite production by discovering the unexplored biosynthetic potential of *Streptomyces*, methods to activate silent BGCs are crucial to current natural product-derived drug discovery research.

Table 1. Lists of classes of antibiotics and their examples from *Streptomyces*.

Class	Antibiotics	References
Aminoglycosides	gentamicin streptomycin tobramycin neomycin kanamycin	[41]
Anthracyclines	doxorubicin	[42]
β-lactams	monobactams, cephalosporin carbapenems	[43]
Macrolides	clarithromycin erythromycin azithromycin	[44]
Ansamycins	rifamycin	[45]

Table 2. Bioactive compounds isolated from *Streptomyces* species with biological activities.

Bioactive molecules	Bioactivities	Species	References
Bleomycin	Anticancer	*S. verticillus*	[46]
Chloramphenicol	Antibiotic	*S. venezuelae*	[47]
Clavulanic acid	β-lactamase inhibitor	*S. clavuligerus*	[48]
Clindamycin	Antibiotic	*S. lincolnensis*	[49]
Daptomycin	Antibiotic	*S. roseosporus*	[50]
Daunomycin	Antitumor	*S. peucetius*	[51]
Erythromycin	Antibiotic	*S. erythraeus*	[36]
FK506 (Tacrolimus)	Immunosuppressant	*S. tsukubaensis*	[52]
Ivermectin	Antiparasitic	*S. avermitilis*	[53]

Kanamycin	Antibiotic	*S. kanamyceticus*	[54]
Lincomycin	Antibiotic	*S. lincolnensis*	[55]
Nystatin	Antifungal	*S. noursei*	[56]
Streptomycin	Antibiotic	*S. griseus*	[57]
Tetracycline	Antibiotic	*S. aureofaciens and S. rimosus*	[58, 59]
Vancomycin	Antibiotic	*S. orientalis*	[60]

4. Traditional approaches for drug discovery in *Streptomyces*

In 1941, Selman Waksman known as the Father of Antibiotics discovered streptomycin isolated from *Streptomyces griseus* through *in vitro* screening tests against pathogenic bacteria, including tuberculosis-causing mycobacteria Woodruff [34]. During the golden era of antibiotic discovery in the 1940s to 1960s, his systematic screening approach known as the Waksman Platform was widely adopted and implemented by researchers to identify antimicrobial agents produced by actinomycetes, particularly *Streptomyces* [68]. Traditionally, the identification and discovery of secondary metabolites and their associated biosynthetic gene clusters in *Streptomyces* were achieved through a combination of classical microbiological and molecular biology techniques [69]. To get the first insight into the biological activities of soil-borne *Streptomyces sp*. against pathogens, plenty of *Streptomyces* species were isolated from various environments and screened for the production of secondary metabolites through phenotypic screening approaches based on physical observations such as pigment production or inhibition zones and biochemical assays [70]. The isolation of secondary metabolites by *Streptomyces* is first performed by fermentation under various culture conditions, fractional extraction, and a series of purification steps involving chromatography techniques such as column chromatography, thin-layer chromatography (TLC), or high-performance liquid chromatography (HPLC). To identify and characterize the natural products, the purified compounds are then subjected to structural elucidation using spectroscopic methods such as nuclear magnetic resonance (NMR) and mass spectrometry [71]. Bioassay-guided fractionation is also used to investigate the bioactivity of fractions of a crude extract and isolate the active fractions indicating the specific secondary metabolites [72, 73]. However, these methods are time-consuming and laborious, and a large fraction of secondary metabolites are not expressed actively and cannot be identified under laboratory culture conditions.

In the early 2000s, classical molecular technologies such as genome mining and bioinformatic analysis revealed enormous numbers of BGCs in the *Streptomyces* genomes [74]. Unlike the traditional bioactivity-guided isolation of NPs, the genome sequencing and bioinformatic analysis of sequenced *Streptomyces* genomes lead to the exploration and prediction of the cryptic or silent smBGCs for novel NPs, which are silent under standard laboratory conditions and are potential targets to be activated [75]. To identify smBGCs within the *Streptomyces* genome sequences, several bioinformatics tools have been developed such as antiSMASH [76], PRISM [77], NP.searcher [78], ClustScan [79], BLAST [80] and BAGEL [81] integrated with databases such antiSMASH database, Minimum Information about a Biosynthetic Gene cluster (MIBiG) [82], the biosynthetic gene cluster families database (BIG-

FAM) [83] or the Integrated Microbial Genomes Atlas of Biosynthetic gene Clusters (IMG-ABC) [84]. The urgent need to discover novel secondary metabolites has resulted in the development of genomic engineering to activate these cryptic or silent gene clusters [85]. Silent smBGCs can be activated through heterologous expression and *in situ* activation. Heterologous expression was established by cloning the target silent gene cluster and heterologous expressed in a heterologous host. Besides, the manipulation of smBGCs for unlocking silent or cryptic gene clusters can be achieved by introducing constitutive promoters, regulating the transcription factors (TFs), and modifying ribosomes through targeted mutagenesis involving knock-in or knock-out techniques [67, 86]. In the past studies reported by Ochi [85], the CRISPR-Cas9 system was applied to promote the effective activation of silent BGCs in five *Streptomyces* species [85]. In addition, overexpression of activator gene under promoter such as *bldA* successfully triggered the expression of cryptic biosynthetic gene clusters for the production of the antibiotics actinorhodin, undecylprodigiosin, and methylenomycin [87]. Although all these techniques have shown efficacy, there is a high investment and low return rate in silent BGC activation. Nonetheless, the application of these methods in studies on NPs discovery provides better insights into the hidden potential of microorganisms (Figure 1). By leveraging techniques such as genome mining, predictive analysis, and targeted genetic manipulation, new bioactive compounds or biosynthetic pathways of NPs can be identified and elucidated from metabolomics which involve metabolite profiling.

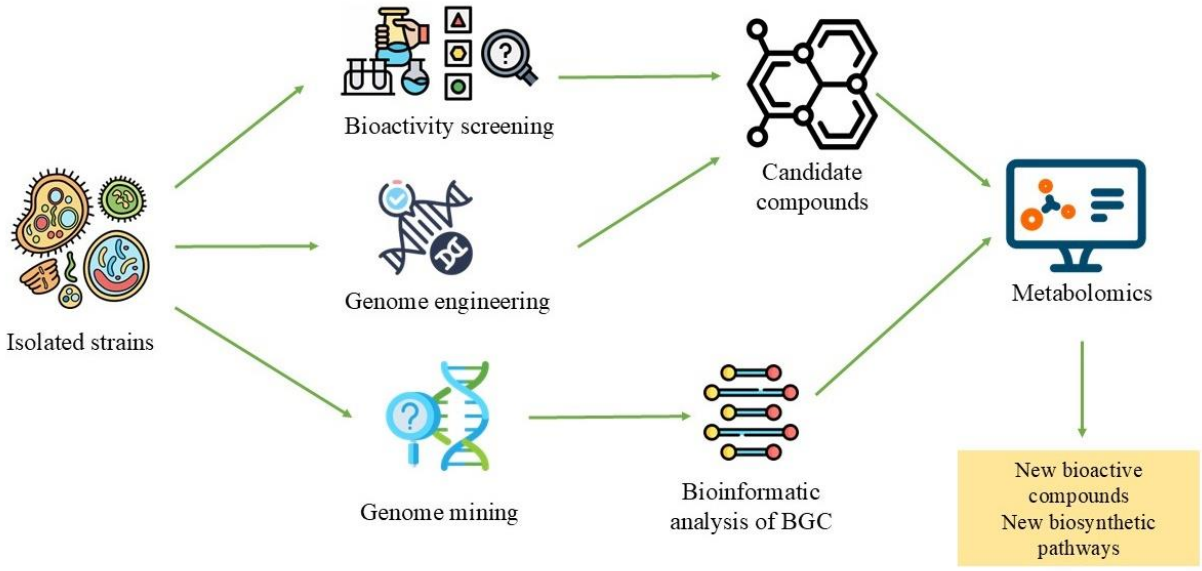


Figure 1. Summary workflows of natural product discovery in *Streptomyces.* The icons in the figure are generated by Freepik from https://www.flaticon.com/.

5. Application of artificial intelligence in natural product-derived drug discovery research

In 2020, group researchers at MIT's Jameel Clinic discovered the first powerful antibiotic using AI named Halicin and it has made a ground-breaking drug discovery [88]. Halicin was initially used for diabetes treatment and was accidentally rediscovered to exhibit broad-spectrum antibacterial activity, particularly against multidrug-resistant pathogens like *Clostridiodes difficile*, *Acinetobacter baumannii*, and *Mycobacterium tuberculosis* [89]. This discovery demonstrated the potential of AI to repurpose existing drugs for new therapeutic applications. The researchers employed an ML model and trained using a deep learning algorithm with a diverse dataset of about 2,500 FDA-approved drugs and natural products. The ML model was trained to recognize patterns of different compounds associated with various chemical structures and further correlate with their antibacterial properties [90]. The algorithm was designed to predict the effectiveness of potential antibiotics. Remarkably, there is a study that highlights the efficacy of Halicin in both *in vitro* and *in vivo* settings, with particular success in murine models infected with *Clostridioides difficile* and pan-resistant *Acinetobacter baumannii* [89]. While the research was in its early stages, the discovery of Halicin highlighted the potential of AI in accelerating the identification of novel antibiotic compounds. It raised hopes to combat the rising threat of antibiotic resistance. AI-based drug discovery also showed another breakthrough in discovering a new antibacterial molecule, namely abaucin that targets *Acinetobacter baumannii*, which causes blood, urinary tract, and lung infections [65, 91]. The convergence of AI and drug discovery has catalyzed a paradigm shift in the pharmaceutical industry to speed up the discovery of new antimicrobial drugs. Exploring new NPs with biological activity has been a cornerstone of drug discovery research. However, conventional drug discovery in *Streptomyces* sp. namely, the Waksman Platform and other approaches are usually characterized by time-consuming processes, high costs and low success rates. Therefore, AI and machine learning (ML) approaches have been utilized to bypass the limitations and challenges of traditional approaches hence accelerating the drug discovery process in a better efficient, more cost-effective and time-effective way. These advanced technologies enable researchers to analyze large number of datasets, predict molecular properties, and identify potential drug candidates with greater precision and speed. As a result, AI and ML have emerged as invaluable tools in the quest for novel therapeutics from natural sources.

Machine learning (ML) is a subset of AI, which mimics the human cognitive processes to interpret information. It is a mathematical model that learns from data, understands the patterns and makes predictions or decisions [92]. ML techniques can be classified into supervised, unsupervised and semi-supervised learning. In supervised learning, a given dataset with known class labels is used to train the algorithm for the classification or regression tasks. For unsupervised learning algorithms, the model learns from unlabelled data without any explicit guidance or predefined outputs. It is used for the three main tasks including clustering, association and dimensionality reduction. Semi-supervised learning is a hybrid model that includes both supervised and unsupervised learning [93]. Supervised learning algorithms, such as Random Forest (RF) [94], Support Vector

Machines (SVM) [95], Naive Bayes (NB) [96], decision tree (DT) [97] and linear regression are the most commonly used supervised algorithms for NP discovery. Unsupervised learning algorithms like Hierarchical Clustering [98] and Chemical Space Mapping [99] also facilitate NP discovery research. The choice of ML algorithm depends on several factors, such as the size and quality of the data, the type of machine learning task, and the interpretability of outputs. In general, constructing a machine learning model consists of multiple steps involving data collection and preprocessing, model selection and training, evaluation and validation of the model's performance and eventually ends with model deployment and continuous monitoring (Figure 2). Initially, the workflow of machine learning starts with data collection from various sources and data preprocessing to remove duplicates, handling missing values, and normalizing or scaling features. Once the data is pre-processed, a model architecture such as decision trees and neural networks with optimal hyperparameters such as learning rate, regularization strength, and batch size is selected and trained on the data [100]. The model learns patterns and relationships from the input features to make predictions. After the model is trained, the model's performance metrics, such as precision, accuracy, or F1 score are evaluated and validated. After evaluating and validating the model, it can be deployed for real-world application or production. To evaluate the model's performance consistency across different data partitions, cross-validation is utilized by randomly dividing the original datasets into multiple training and testing sets, typically referred to as "folds." K-fold cross-validation is employed to compare different models, evaluate efficient models with hyperparameters, and subsequently obtain greater model performance with high reliability and robustness [101]. This approach enhances the generalization ability of a learning algorithm on unseen data so that it can be used to make predictions on new data, resulting in more accurate and dependable results in practical applications. In natural product-derived drug discovery, ML tools could play a vital role in processes involving detecting BGCs, drug target identification, lead compound prioritization and optimization as well as compound screening and drug design [102].

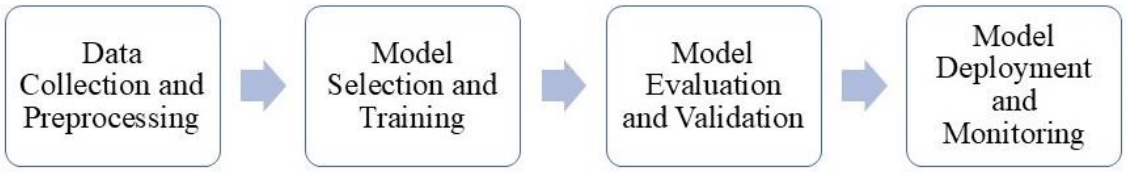


Figure 2. Machine learning workflow.

5.1. Genome mining

Genome mining refers to the analysis of genetic information within the genome of sequenced microorganisms to identify or characterise bioactive compounds such as natural products (NPs) or secondary metabolites [103]. Especially in the case of *Streptomyces*, a genus known for its prolific production of bioactive compounds, genome mining is indispensable for identifying novel drug candidates. Utilising NGS platforms like MiSeq (short-read sequencing) or PacBio (long-read sequencing), numerous studies have sequenced the whole genome of *Streptomyces* isolated from various environments [104-106]. These advanced

approaches have significantly contributed to drug discovery research and expanded our understanding of their biosynthetic capabilities. To enhance the efficiency of discovering NPs and characterization of their bioactivity, various approaches such as computational tools, bioinformatics techniques, and experimental validation methods have been employed to establish the gene–metabolite link by identifying biosynthetic genes that encode enzymes involved in the biosynthesis of bioactive metabolites [65, 107].

The biosynthetic machinery of *Streptomyces* is a highly regulated system, where the biosynthetic gene cluster (BGCs) responsible for the production of various NPs such as polyketides, nonribosomally synthesized peptides (NRPs), ribosomally synthesized and posttranslationally modified peptides (RiPPs), along with alkaloids, and terpene. A novel pentacyclic polyketide named formicamycin has been discovered and identified by using antiSMASH, which employs profile-hidden Markov models (pHMMs) to identify the BGC [108]. In another study, the RiPPER genome mining tool enabled the isolation of novel thioamidated RiPPs [109]. RiPPs differ from other natural products like polyketides and NRPs, which are typically synthesized by multi-modular enzyme complexes. Unlike other classes of NP, RiPPs exhibit unique biosynthetic pathways that lack universal signature biosynthetic genes across all RiPP families, making it challenging to develop universal bioinformatics tools or predictive models for RiPP genome mining [110].

To fully reveal the biosynthetic potential of *Streptomyces*, genome mining tools that incorporate machine learning approaches could be used to detect different classes of NPs. To aid in the identification of BGCs for all major NP classes, DeepBGC, Deep-BGCpred and BIGCARP were introduced by Hannigan et al. [111], Yang et al. [112], and Rios-Martinez et al. [113], respectively. Each genome mining tools implement a deep learning approach such as neural networks combined with vector representations of protein family (Pfam) to predict BGC boundaries and annotate BGC function associated with the biosynthesis of secondary metabolites. It also adopted word embedding techniques such as Word2vec, a natural language processing (NLP) algorithm to analyse literature, patents and databases to extract information about gene functions, gene-drug associations, natural products, and their biosynthesis pathways facilitating the understanding of drug mechanisms and identification of potential therapeutic targets [114]. This has streamlined the target selection process, accelerating the process of identifying novel targets for drug development. Furthermore, NRPSpredictor2 and SANDPUMA (Specificity of Adenylation Domain Prediction Using Multiple Algorithms) were developed to predict NRPS adenylation domain specificity [115, 116]. Both of the tools employ SVM algorithms or other machine learning algorithms to identify NRPS BGCs. Deep learning models such as DeepRiPP, Data-driven Exploratory Class-independent RiPP TrackER (decRiPPter) and NeuRiPP have been developed to tackle these challenges for mining RiPP BGCs [117-119]. A recent study reported that a new class of lanthipeptides (termed “class V”) and 42 new RiPP family candidates were identified with the help of decRiPPter on genome mining of 1295 *Streptomyces* genomes. Unlike traditional methods, decRiPPter does not rely on prior knowledge of core enzymatic machinery or specific modifications. It employs an SVM classifier trained on 175 known RiPP precursors regardless of RiPP subclasses. This approach incorporates pan-genomic analysis to identify

putative precursor genes located within specialized genomic regions. These genomic regions contain multiple enzyme-coding genes and are part of the accessory genome of a genus [120]. By employing decRiPPter, researchers have the potential to unlock a treasure trove of previously undiscovered natural products of RiPP, paving the way for groundbreaking advancements in drug discovery.

Over the years, other ML-assisted genome mining tools such as the hidden Markov model-based method ClusterFinder [121], GECCO17 [122], RiPPMiner-Genome [123], Pytorch [124] and SanntiS [125] have been developed and utilized to detect and annotate potential BGCs that encoded putative bioactive compounds different classes of NPs. Integrating AI technologies with classical approaches might hold tremendous potential for accelerating the discovery process of discovering novel drugs from the vast genomic reservoir of *Streptomyces* sp. AI systems help to integrate various data types such as genomics, transcriptomics, proteomics, metabolomics, structural data, and bioactivity data. This integration enables the discovery of the complex relationships between features and supports the development of finely tuned hypothesis [126].

5.2. Biological activities prediction

The predictive power of AI extends beyond gene cluster identification. The incorporation of automation and AI-powered systems into in silico screening of natural products has also revolutionized the initial phase of drug discovery. ML plays a crucial role in advancing molecular property prediction (MPP) and chemical reaction prediction (CRP) [127] contributing to the discovery of lead compounds. This enables the rapid screening of compound libraries against specific drug targets to identify potential bioactive compounds with therapeutic activity in *Streptomyces* strains. The virtual screening of natural products (NPs) derived from *Streptomyces* sp. utilises datasets from the StreptomeDB 2.0 database which includes about 2,877 NPs originating from *Streptomyces* [128]. In ML technology, molecular featurization is implemented to digitize chemical structures of novel molecules from natural products into a machine-readable format [127]. Molecules have been featured through various techniques such as molecular representations, descriptors, fingerprints and latent vectors derived from molecular embedding [129]. During the molecular featurization process, the chemical structures of the molecules are represented as SMILES (Simplified Molecular Input Line Entry System) [130] or international chemical identifier (InChES) [131] annotations, images, strings or molecular graphs to serve as input information and datasets for machine learning models including various neural network, SVM, multilayer perceptron (MLP) and random forest (RF) [132], subsequently generate outputs to predict the biological activities of molecules such as antibacterial, antifungal, antiviral, antitumor, or immunomodulatory activity (Figure 3).

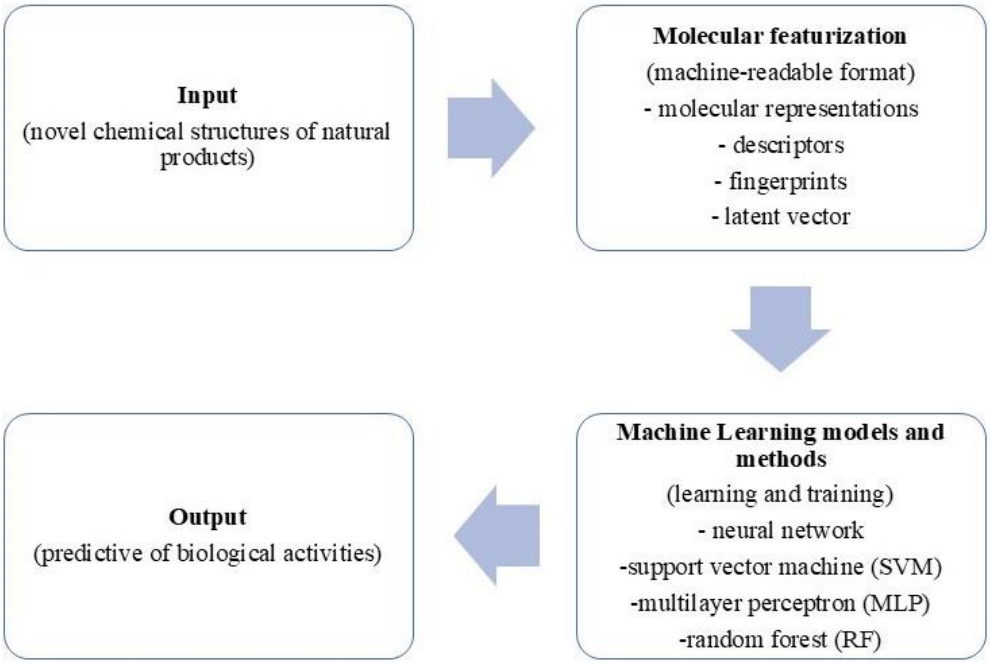


Figure 3. General overview of ML approaches in natural products drug discovery.

In the context of structural elucidation, computational prediction of spectroscopic data, mainly NMR, gas/liquid chromatography (GC/LC) or mass spectrometry (MS) could be integrated with improved performance of deep learning models for more accurate prediction of chemical structures of natural products and theoretical NMR correlation data [133]. Deep learning models employ deep neural networks such as Artificial Neural Network (ANN), Convolutional Neural Network (CNN) and Graph Neural Network (GNN) can be applied in ADMET (Adsorption, Distribution, Metabolism, Excretion, and Toxicity) modelling and Quantitative Structure-Activity relationship (QSAR) modelling involving various aspects of drug discovery including drug target identification, drug target interaction prediction, pharmacokinetics and toxicity of compound [134-137]. This approach could computationally predict the biological activity of *Streptomyces*-derived compounds to target proteins based on chemical structures and select targeted compounds with favourable pharmacokinetic properties and reduced toxicity risks [138]. The interpretation of results could help identify potential compounds, prioritize targeted compounds for testing, and reduce the experimental workload in screening [139]. Moreover, AI algorithms such as Random Forest and SVM speed up multi-omics data processing [140] and hence dereplicate known compounds in natural extracts efficiently [141]. This speeds up the process of identifying novel compounds by eliminating redundancy and focusing on unique chemical entities.

5.3. Optimisation of bioactive compound production

The natural fermentation process has been employed for drug discovery in microbes where cultivated microbes synthesise naturally occurring bioactive compounds through fermentation technology. Some major challenges of microbial fermentation are low-cost efficiency of raw material inventory, low quality of biomass concentration and low yield of fermentation products. The drug discovery process could be enhanced by improving fermentation strategies for the production of secondary metabolites in *Streptomyces* sp. [142]. Therefore, fermentation optimization in terms of incubation time, medium compositions, and environmental factors such as temperature and pH are crucial steps to optimize the conditions for growing *Streptomyces* cultures maximize the valuable secondary metabolite yields, and determine and minimize the input variables. Conventionally, non-statistical techniques such as One-Factor-at-Time (OFAT) and statistical methods such as response surface methodology (RSM) have been practised widely for medium optimization [143, 144]. OFAT is a traditional experimental design method where one independent variable is varied at a time while all other variables remain unchanged. In the context of fermentation optimization, OFAT involves altering one fermentation parameter while keeping all other parameters constant and observing the resulting changes in fermentation yield. This traditional approach could be time-consuming and may overlook synergistic effects that impact overall outcomes due to the difficulty in estimating interactions between multiple variables from the experiments. RSM is a mathematical approach that employs a few experimental designs such as Box-Behnken design (BBD) or Central Composite Design (CCD) to model the relationship between multiple explanatory variables and one or more response variables. The experimental data is then analysed using statistical analysis techniques including multiple regression analysis and analysis of variance (ANOVA) to fit mathematical models such as higher-order polynomial equations, to predict the response surface and identify the optimal combination of fermentation conditions that maximizes the production of desired metabolites or bioactive compounds. Even though RSM is extensively used, some associated limitations are the assumption of linearity and high dependence on the experimental design that might yield biased or unreliable results. Fermentation is a complex system where the output results can be influenced by multiple variables [145, 146]. Compared to conventional methods, AI-driven approaches can analyse complex interactions between variables and adaptively refine solutions over time. ML-based predictive models namely ANNs and statistical models specifically RSM can be coupled with four evolutionary algorithms (EAs) such as GA, DE, simulated annealing algorithm, and particle swarm optimization to optimize the fermentation parameters [147-150]. This coupling of machine learning model and EAs signifies powerful, hybrid optimisation techniques leveraging the strengths of both paradigms, where machine learning models provide predictions while EAs fine-tune and optimize parameters more precisely [151].

Introducing AI or ML-based approaches in genome mining, compound screening, identification of metabolites or bioactive compounds and optimisation of metabolite expression have contributed significant impacts in *Streptomyces*-related drug discovery. Research studies involving the application of AI algorithms or models related to

Streptomyces-related drug discovery are presented in Table 3. Despite the transformative potential, AI systems might have key challenges such as limited data quality, availability, heterogeneity, dataset size, and data privacy [152]. During predictive model construction, unbalanced active to inactive compound datasets with limited coverage of inactive compounds might influence the accuracy of prediction results. Additionally, AI models might lack explainability, interpretability, reproducibility and validation on diverse datasets due to their underlying complexity. Besides, overfitting might occur when the ML model fits the training data too precisely, resulting in poor generalization of new test data [153]. Biases might be present in training data including systemic biases, selection bias, and automation bias, which could potentially lead to biased predictions and ethical considerations [154]. Hence, the accuracy of the prediction may be uncertain when biased data are utilized. Other significant challenges in developing and implementing AI in drug discovery are the need for significant computational power, expertise, and financial investment.

Table 3. Research studies implementations of AI algorithms or models in *Streptomyces sp*.

Aim of the study	AI algorithms or models	*Streptomyces* strains	Findings	Ref.
Genome mining	decRiPPter (Data-driven Exploratory Class-independent RiPP TrackER) combines a Support Vector Machine (SVM)	*Streptomyces pristinaespiralis* ATCC 25468	• Identify 42 novel Ribosomally synthesized and post-translationally modified peptides (RiPP) families • Discover novel family of lanthipeptides within the RiPP biosynthetic gene cluster	[120]
	DeepT2, DeepBGC and antiSMASH model with four machine learning algorithms (random forest, XGBoost, SVM, and MLP)	37 selected *Streptomyces* isolates	DeepT2 outperforms both DeepBGC and antiSMASH in the prediction of type II polyketides (T2PK) using only KS_{β} sequences	[155]
	genetic algorithm (GA)	*Streptomyces coelicolor*	• Identify 11 secondary metabolite gene clusters of the antibiotic-producing eubacterium *Streptomyces coelicolor* • identified gene regulator based on transcriptomic and expression data.	[156]
	SVM-based model	*Streptomyces coelicolor*	Predict and verify the operon structure using different binary classifiers.	[157]
	NRPSpredictor2	*Streptomyces lincolnensis*	Characterized a new nonribosomal peptide namely cysteoamide, 1 and identified its NRPs biosynthetic gene cluster	[158]
Strain identification	artificial neural network (ANN)	three putatively novel *Streptomyces* species	Identify members of the three target streptomycete taxa.	[159]

Optimisation of compound production	MSHub/GNPS (Global Natural products Social Molecular Networking)	*Streptomyces Volatilomes* (37 selected isolates)	• Detect and annotate more volatile organic compounds (VOCs) than using the conventional method. • Remove the volatilome variability between media and isolates.	[160]
	ANN	*Streptomyces flavolimosus*	• Predict the optimal conditions that maximize the biosynthesis of AuNPs using the cell-free supernatant of *Streptomyces flavolimosus* at high efficacy compared to mathematical models, central composite design. • Demonstrate antitumor properties in-vitro (MCF-7 human breast cancer and Hela carcinoma cell lines) and in vivo against Ehrlich ascites carcinoma	[161]
	Response Surface Methodology-Genetic Algorithm (RSM-GA)	*Streptomyces rimosus* MTCC 10792	This combination approach optimizes the medium components for extracellular cholesterol oxidase (COD) production in *Streptomyces* (3.6 folds higher compared to un-optimized medium)	[162]
	ANN coupled with GA and Nelder-Mead downhill simplex (NMDS)	*Streptomyces sindenensis* MTCC 8122	ANN-NMDS optimization was found to be more efficacious compared to the ANN-GA optimization maximum antibiotic production where 197 microgram/ml was obtained in ANN-NMDS optimization; 176 microgram/ml was obtained in ANN-GA optimization	[163]
	ANN/GA	*Streptomyces triostinicus*	ANN/GA prediction model shows better optimal performance in actinomycin V yield of 36.7% higher than RSM model.	[164]
	ANN/GA	*Streptomyces* sp. NICM 5500	ANN/GA prediction model shows better optimal performance with 60% higher COD concentration than RSM model.	[165]
	GA	*Streptomyces hygroscopicus*	GA predicts optimal cultivation parameters which contribute maximum antifungal activity	[166]

ANN	*Streptomyces microflavus* strain NEAE-83	ANN prediction model shows better accuracy than CCD model and aligns closely with the validation experimental in optimization of the chitosan nanoparticles biosynthesis	[167]
ANN	*Streptomyces noursei*	ANN prediction model shows better accuracy than RSM model and aligns closely with the validation experimental in fermentation optimisation	[168]
ANN	*mutant Streptomyces durhamensis GC23*	ANN optimization shows higher predictive efficiency than RSM and aligns closely with the validation experimental for cellulase production.	[169]

6. Conclusion

The global spread of multidrug-resistant bacterial pathogens has become a major threat to the healthcare system and hence driven the urgent need to discover new sources of natural products or antibiotics. In the past few decades, *Streptomyces* sp. has been the largest bio-factory for the production of secondary metabolites with a wide range of biological activities. However, *Streptomyces* sp. possesses large smBGCs in which numerous cryptic BGCs are silenced under laboratory culture conditions with only a limited number of secondary metabolites that have been expressed actively, rendering a largely untapped source of drugs. Drug discovery was traditionally performed through random screening and this traditional compound screening might be challenging. Besides, optimisation of cultivation parameters for bacteria is crucial for enhanced production of secondary metabolites, which is significant for the drug discovery process. Indeed, AI tools are increasingly used to enhance drug discovery efficacy by offsetting the limitations of ineffective traditional approaches in drug discovery. Therefore, applying AI integrated with multidisciplinary techniques is critical to unlocking hidden pathways and discovering new natural products in *Streptomyces* sp.

Author Contribution: Investigation, Writing – original draft, review & editing, WSA; Writing – review & editing, JW-FL, VL, YK, LT-HL; Conceptualization, supervision, funding acquisition, LT-HL.

Funding: This research was funded by the Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute (CBI), University of Nottingham Ningbo China, Life Science and Healthcare Centre research seed fund number (I01240300006); Jeffrey Cheah School of Medicine and Health Sciences (JCSMHS) Strategic Grant 2022 (Grant Code: STG-000107).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aware C and Jadhav J. *Chapter 10 - Bioprospecting potential of microbes for the therapeutic application*, in *Bioprospecting of Microbial Diversity*, P. Verma and M.P. Shah, Editors. 2022, Elsevier. p. 223-255.

2. Tan LT-H, Chan K-G, Lee L-H*, et al*. Streptomyces Bacteria as Potential Probiotics in Aquaculture. Front Microbiol 2016; 7.

3. Tang C, Hoo PC-X, Tan LT-H*, et al*. Golden Needle Mushroom: A Culinary Medicine with Evidenced-Based Biological Activities and Health Promoting Properties. Front Pharmacol 2016; 7.

4. Tay K-C, Tan LT-H, Chan CK*, et al*. Formononetin: A Review of Its Anticancer Potentials and Mechanisms. Front Pharmacol 2019; 10.

5. Tan LTH, Lee LH, Yin WF*, et al*. Traditional Uses, Phytochemistry, and Bioactivities of Cananga odorata (Ylang-Ylang). Evid Based Complement Alternat Med 2015; 2015(1): 896314.

6. Aminov RI. A brief history of the antibiotic era: lessons learned and challenges for the future. Front Microbiol 2010; 1: 134.

7. Kemung HM, Tan LT-H, Khan TM*, et al*. Streptomyces as a Prominent Resource of Future Anti-MRSA Drugs. Front Microbiol 2018; 9.

8. Atanasov AG, Zotchev SB, Dirsch VM*, et al*. Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov 2021; 20(3): 200-216.

9. Bonnett SA, Ollinger J, Chandrasekera S*, et al*. A Target-Based Whole Cell Screen Approach To Identify Potential Inhibitors of Mycobacterium tuberculosis Signal Peptidase. ACS Infect Dis 2016; 2(12): 893-902.

10. Beroigui O and Errachidi F. Streptomyces at the heart of several sectors to support practical and sustainable applications: A review. Prog Microbes Mol Biol 2023; 6(1).

11. Law JW, Ser HL, Ab Mutalib NS*, et al*. Streptomyces monashensis sp. nov., a novel mangrove soil actinobacterium from East Malaysia with antioxidative potential. Sci Rep 2019; 9(1): 3056.

12. Lee N, Kim W, Hwang S*, et al*. Thirty complete Streptomyces genome sequences for mining novel secondary metabolite biosynthetic gene clusters. Sci Data 2020; 7(1): 55.

13. Ghimire U, Kandel R, Neupane M*, et al*. Biofilm Formation and blaOXA Genes Detection Among Acinetobacter baumannii from Clinical Isolates in a Tertiary Care Kirtipur Hospital, Nepal. Prog Microbes Mol Biol 2021; 4(1).

14. Rizki LP, Murni IK, Aman AT*, et al*. Environmental Metagenomic Analysis of "ESKAPE" Pathogens in the Pediatric Intensive Care Unit of General Hospital Yogyakarta Indonesia. Prog Microbes Mol Biol 2024; 7(1).

15. World Health Organization. Global action plan on antimicrobial resistance. 2015 [Accessed 12 January 2024]; Available from: https://iris.who.int/bitstream/handle/10665/193736/9789241509763_eng.pdf.

16. Chong KJ, Feng H, Letchumanan V*, et al*. Tackling Microbial Resistance and Emerging Pathogens with Next-Generation Antibiotics. Prog Microbes Mol Biol 2024; 7(1).

17. Prestinaci F, Pezzotti P, and Pantosti A. Antimicrobial resistance: a global multifaceted phenomenon. Pathog Glob Health 2015; 109(7): 309-18.

18. Kuai YH, Law JW-F, Ong YS*, et al*. Role of SarA in Staphylococcus aureus: A Virulence Target For Therapeutic Strategies. Prog Microbes Mol Biol 2024; 7(1).

19. Baranova AA, Alferova VA, Korshun VA, *et al*. Modern Trends in Natural Antibiotic Discovery. Life 2023; 13(5): 1073.

20. Makridakis S. The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. Futures 2017; 90: 46-60.

21. Singh S, Kumar R, Payra S, *et al*. Artificial Intelligence and Machine Learning in Pharmacological Research: Bridging the Gap Between Data and Drug Discovery. Cureus 2023; 15(8): e44359.

22. Alam K, Mazumder A, Sikdar S, *et al*. Streptomyces: The biofactory of secondary metabolites. Front Microbiol 2022; 13: 968053.

23. Thye AY-K, Letchumanan V, Tan LT-H, *et al*. Malaysia's breakthrough in Modern Actinobacteria (MOD-ACTINO) drug discovery research. Prog Microbes Mol Biol 2022; 5(1): a0000275.

24. Tan LT-H, Chan K-G, Khan TM, *et al*. Streptomyces sp. MUM212 as a Source of Antioxidants with Radical Scavenging and Metal Chelating Properties. Front Pharmacol 2017; 8.

25. Tan LT-H, Chan K-G, Pusparajah P, *et al*. Mangrove derived Streptomyces sp. MUM265 as a potential source of antioxidant and anticolon-cancer agents. BMC Microbiol 2019; 19(1): 38.

26. Tan LT-H, Ser H-L, Yin W-F, *et al*. Investigation of Antioxidative and Anticancer Potentials of Streptomyces sp. MUM256 Isolated from Malaysia Mangrove Soil. Front Microbiol 2015; 6.

27. Law JW-F, Tan LT-H, Letchumanan V, *et al*. Streptomyces griseiviridis sp. nov., a novel "modern actinobacteria" isolated from Malaysia mangrove soil. Prog Microbes Mol Biol 2023; 6(1): a0000270.

28. McCormick JR and Flärdh K. Signals and regulators that govern Streptomyces development. FEMS Microbiol Rev 2012; 36(1): 206-31.

29. Yagüe P, López-García MT, Rioseras B, *et al*. Pre-sporulation stages of Streptomyces differentiation: state-of-the-art and future perspectives. FEMS Microbiol Lett 2013; 342(2): 79-88.

30. Hopwood D. Highlights of Streptomyces genetics. Heredity (Edinb) 2019; 123: 23-32.

31. Tan LT-H, Chan K-G, Chan CK, *et al*. Antioxidative Potential of a Streptomyces sp. MUM292 Isolated from Mangrove Soil. BioMed Res Int 2018; 2018(1): 4823126.

32. Pusparajah P, Letchumanan V, Law JW-F, *et al*. Streptomyces sp.—A Treasure Trove of Weapons to Combat Methicillin-Resistant Staphylococcus aureus Biofilm Associated with Biomedical Devices. Int J Mol Sci 2021; 22(17): 9360.

33. Goh JXH, Tan LT-H, Law JW-F, *et al*. Streptomyces sp. MUM 195J: A Promising Probiotic for Controlling Vibrio parahaemolyticus Infection in Aquaculture. Prog Microbes Mol Biol 2024; 7(1).

34. Woodruff HB. Selman A. Waksman, winner of the 1952 Nobel Prize for physiology or medicine. Appl Environ Microbiol 2014; 80(1): 2-8.

35. Schatz A, Bugie E, and Waksman SA. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. 1944. Clin Orthop Relat Res 2005(437): 3-6.

36. Weber JM, Wierman CK, and Hutchinson CR. Genetic analysis of erythromycin production in Streptomyces erythreus. J Bacteriol 1985; 164(1): 425-33.

37. Loo K-Y, Tan LT-H, Law JW-F, *et al*. Vibrio parahaemolyticus: Exploring its Incidence in Malaysia and the Potential of Streptomyces sp. as an Anti-Vibrio Agent. Prog Microbes Mol Biol 2023; 6(1).

38. Mangzira Kemung H, Tan LT-H, Chan K-G*, et al*. Streptomyces sp. Strain MUSC 125 from Mangrove Soil in Malaysia with Anti-MRSA, Anti-Biofilm and Antioxidant Activities. Molecules 2020; 25(15): 3545.

39. Loo K-Y, Tan LT-H, Law JW-F*, et al*. Detection of multidrug resistant Vibrio parahaemolyticus and anti-Vibrio Streptomyces sp. MUM 178J. Prog Microbes Mol Biol 2023; 6(1).

40. Elsalam RM, Goh KW, Mahadi M*, et al*. The Antibacterial Activities of Secondary Metabolites Derived from Streptomyces sp. Prog Microbes Mol Biol 2022; 5(1).

41. Krause KM, Serio AW, Kane TR*, et al*. Aminoglycosides: An Overview. Cold Spring Harb Perspect Med 2016; 6(6).

42. Venkatesh P and Kasi A. *Anthracyclines*, in *StatPearls*. 2023, StatPearls Publishing Copyright © 2023, StatPearls Publishing LLC.: Treasure Island (FL).

43. Tahlan K and Jensen SE. Origins of the β-lactam rings in natural products. J Antibiot 2013; 66(7): 401-410.

44. Ma Z, Ginsberg AM, and Spigelman M. *7.24 - Antimycobacterium Agents*, in *Comprehensive Medicinal Chemistry II*, J.B. Taylor and D.J. Triggle, Editors. 2007, Elsevier: Oxford. p. 699-730.

45. Morimitsu Y and Hirota A. Ansamycin antibiotics as free radical scavengers isolated from Streptomyces by using the bactericidal action of the hydroxyl radical. Biosci Biotechnol Biochem 1996; 60(9): 1507-1509.

46. Matsuo H, Mochizuki H, Davies J*, et al*. Production of bleomycin N-acetyltransferase in Escherichia coli and Streptomyces verticillus. FEMS Microbiol Lett 1997; 153(1): 83-88.

47. Shaw WV and Hopwood DA. Chloramphenicol Acetylation in Streptomyces. Microbiology 1976; 94(1): 159-166.

48. Howarth TT, Brown AG, and King TJ. Clavulanic acid, a novel β-lactam isolated from Streptomyces clavuligerus; X-ray crystal structure analysis. J Chem Soc Chem Commun 1976(7): 266b-267.

49. Stevens HA. Clindamycin. Prim Care Update Ob Gyns 1997; 4(6): 251-253.

50. Huber FM, Pieper RL, and Tietz AJ. The formation of daptomycin by supplying decanoic acid to Streptomyces roseosporus cultures producing the antibiotic complex A21978C. J Biotechnol 1988; 7(4): 283-292.

51. Ye J, Dickens ML, Plater R*, et al*. Isolation and sequence analysis of polyketide synthase genes from the daunomycin-producing Streptomyces sp. strain C5. J Bacteriol 1994; 176(20): 6270-6280.

52. Wollenberg A and Bieber T. *FK-506/Tacrolimus*, in *Strategies for Immunointerventions in Dermatology*, G. Burg and R.G. Dummer, Editors. 1997, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 53-57.

53. Campbell WC, Fisher MH, Stapley EO*, et al*. Ivermectin: A Potent New Antiparasitic Agent. Science 1983; 221(4613): 823-828.

54. Kharel MK, Subba B, Basnet DB*, et al*. A gene cluster for biosynthesis of kanamycin from Streptomyces kanamyceticus: comparison with gentamicin biosynthetic gene cluster. Arch Biochem Biophys 2004; 429(2): 204-214.

55. Peschke U, Schmidt H, Zhang HZ*, et al*. Molecular characterization of the lincomycin‐production gene cluster of Streptomyces lincolnensis 78‐11. Mol Microbiol 1995; 16(6): 1137-1156.

56. Zotchev S, Haugan K, Sekurova O*, et al*. Identification of a gene cluster for antibacterial polyketide-derived antibiotic biosynthesis in the nystatin producer Streptomyces noursei ATCC 11455. Microbiology 2000; 146(3): 611-619.

57. Waksman SA, Reilly HC, and Johnstone DB. Isolation of streptomycin-producing strains of Streptomyces griseus. J Bacteriol 1946; 52(3): 393-397.

58. Darken MA, Berenson H, Shirk RJ*, et al*. Production of tetracycline by Streptomyces aureofaciens in synthetic media. J Appl Microbiol 1960; 8(1): 46-51.

59. Petkovic H, Cullum J, Hranueli D*, et al*. Genetics of Streptomyces rimosus, the oxytetracycline producer. Microbiol Mol Biol Rev 2006; 70(3): 704-728.

60. Griffith RS. Vancomycin use—an historical review. J Antimicrob Chemother 1984; 14(suppl_D): 1-5.

61. Durairaj T, Alagappan C, Suresh SSR*, et al*. *An Introductory Chapter: Secondary Metabolites*, in *Secondary Metabolites*, V. Ramasamy and S.S.R. Suresh, Editors. 2018, IntechOpen: Rijeka. p. Ch. 1.

62. Mohammed H, Hamdi B, Miloud B*, et al*. *Streptomyces Secondary Metabolites*, in *Basic Biology and Applications of Actinobacteria*, E. Shymaa, Editor. 2018, IntechOpen: Rijeka. p. Ch. 6.

63. Risdian C, Mozef T, and Wink J. Biosynthesis of Polyketides in Streptomyces. Microorganisms 2019; 7(5).

64. Seipke RF. Strain-level diversity of secondary metabolism in Streptomyces albus. PLoS One 2015; 10(1): e0116457.

65. Lee N, Hwang S, Kim J*, et al*. Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in Streptomyces. Comput Struct Biotechnol J 2020; 18: 1548-1556.

66. Craney A, Ahmed S, and Nodwell J. Towards a new science of secondary metabolism. J Antibiot 2013; 66(7): 387-400.

67. Liu Z, Zhao Y, Huang C*, et al*. Recent Advances in Silent Gene Cluster Activation in Streptomyces. Front Bioeng Biotechnol 2021; 9: 632230.

68. Lewis K. The Science of Antibiotic Discovery. Cell 2020; 181(1): 29-45.

69. Amin DH, Abdallah NA, Abolmaaty A*, et al*. Microbiological and molecular insights on rare Actinobacteria harboring bioactive prospective. Bull Natl Res Cent 2020; 44(1): 5.

70. Newman D. Screening and identification of novel biologically active natural compounds. F1000Research 2017; 6: 783.

71. Kavitha A, Prabhakar P, Vijayalakshmi M*, et al*. Purification and biological evaluation of the metabolites produced by Streptomyces sp. TK-VL_333. Res Microbiol 2010; 161(5): 335-345.

72. Nothias L-F, Nothias-Esposito M, da Silva R*, et al*. Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. J Nat Prod 2018; 81(4): 758-767.

73. Tan LT-H, Chan C-K, Chan K-G*, et al*. Streptomyces sp. MUM256: A Source for Apoptosis Inducing and Cell Cycle-Arresting Bioactive Compounds against Colon Cancer Cells. Cancers (Basel) 2019; 11(11): 1742.

74. Ziemert N, Alanjary M, and Weber T. The evolution of genome mining in microbes–a review. Nat Prod Rep 2016; 33(8): 988-1005.

75. Alam K, Hao J, Zhong L*, et al*. Complete genome sequencing and in silico genome mining reveal the promising metabolic potential in Streptomyces strain CS-7. Front Microbiol 2022; 13: 939919.

76. Blin K, Shaw S, Augustijn HE*, et al*. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res 2023; 51(W1): W46-W50.

77. Wenger AM, Clarke SL, Guturu H*, et al*. PRISM offers a comprehensive genomic approach to transcription factor function prediction. Genome Res 2013; 23(5): 889-904.

78. Li MH, Ung PM, Zajkowski J*, et al*. Automated genome mining for natural products. BMC Bioinformatics 2009; 10: 185.

79. Starcevic A, Zucko J, Simunkovic J*, et al*. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. Nucleic Acids Res 2008; 36(21): 6882-6892.

80. Altschul SF, Madden TL, Schäffer AA*, et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25(17): 3389-3402.

81. de Jong A, van Hijum SA, Bijlsma JJ*, et al*. BAGEL: a web-based bacteriocin genome mining tool. Nucleic Acids Res 2006; 34(suppl_2): W273-W279.

82. Terlouw BR, Blin K, Navarro-Muñoz JC*, et al*. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucleic Acids Res 2022; 51(D1): D603-D610.

83. Kautsar SA, Blin K, Shaw S*, et al*. BiG-FAM: the biosynthetic gene cluster families database. Nucleic Acids Res 2020; 49(D1): D490-D497.

84. Palaniappan K, Chen IA, Chu K*, et al*. IMG-ABC v.5.0: an update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. Nucleic Acids Res 2020; 48(D1): D422-d430.

85. Ochi K. Insights into microbial cryptic gene activation and strain improvement: principle, application and technical aspects. J Antibiot 2017; 70(1): 25-40.

86. Yu G, Duan Q, Cui T*, et al*. Development of a bacterial gene transcription activating strategy based on transcriptional activator positive feedback. J Adv Res 2023.

87. Hackl S and Bechthold A. The gene bldA, a regulator of morphological differentiation and antibiotic production in Streptomyces. Arch Pharm 2015; 348(7): 455-462.

88. Trafton A *Artificial intelligence yields new antibiotic*. 2020, Massachusetts Institute of Technology.

89. Booq RY, Tawfik EA, Alfassam HA*, et al*. Assessment of the Antibacterial Efficacy of Halicin against Pathogenic Bacteria. Antibiotics (Basel) 2021; 10(12).

90. Walker AS and Clardy J. A Machine Learning Bioinformatics Method to Predict Biological Activity from Biosynthetic Gene Clusters. J Chem Inf Model 2021; 61(6): 2560-2571.

91. Liu G, Catacutan DB, Rathod K*, et al*. Deep learning-guided discovery of an antibiotic targeting Acinetobacter baumannii. Nat Chem Biol 2023; 19(11): 1342-1350.

92. Jordan MI and Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015; 349(6245): 255-260.

93. Mahesh B. Machine learning algorithms-a review. International Journal of Science and Research 2020; 9(1): 381-386.

94. Zhang C and Ma Y. *Random forest for bioinformatics*. Ensemble machine learning: Methods and applications. 2012. 307-323.

95. Noble WS. What is a support vector machine? Nat Biotechnol 2006; 24(12): 1565-1567.

96. Zhang H. *The Optimality of Naive Bayes*. in *The Florida AI Research Society*. 2004.

97. Hyafil L and Rivest RL. Constructing optimal binary decision trees is NP-complete. Inf Process Lett 1976; 5(1): 15-17.

98. Böcker A, Derksen S, Schmidt E*, et al*. A hierarchical clustering approach for large compound libraries. J Chem Inf Model 2005; 45(4): 807-815.

99. Medina-Franco JL, Martínez-Mayorga K, Giulianotti MA*, et al*. Visualization of the chemical space in drug discovery. Curr Comput Aided Drug Des 2008; 4(4): 322-333.

100. Yang L and Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 2020; 415: 295-316.

101. Fushiki T. Estimation of prediction error by using K-fold cross-validation. Stat Comput 2011; 21: 137-146.

102. Qureshi R, Irfan M, Gondal TM*, et al*. AI in drug discovery and its clinical relevance. Heliyon 2023; 9(7): e17575.

103. Bauman KD, Butler KS, Moore BS*, et al*. Genome mining methods to discover bioactive natural products. Nat Prod Rep 2021; 38(11): 2100-2129.

104. Pusparajah P, Law JW-F, Chan K-G*, et al*. Whole-genome sequence of Streptomyces pluripotens strain MUM 16J, a potential resource of glycopeptide antibiotic and biocontrol agent against biofilm-forming bacteria. Prog Microbes Mol Biol 2023; 6(1): a0000330.

105. Loo K-Y, Tan LT-H, Chua K-O*, et al*. Complete Whole-Genome Sequence of Streptomyces sp. MUM 178J, a Potential Anti-Vibrio Agent. Prog Microbes Mol Biol 2024; 7(1): a0000399.

106. Law JW-F, Letchumanan V, Hong K-W*, et al*. Streptomyces learnhanii sp. nov., unveiling a Mangrove-Derived Novel "Modern Actinobacteria" in Malaysia. Prog Microbes Mol Biol 2023; 6(1): a0000342.

107. Kjærbølling I, Mortensen UH, Vesth T*, et al*. Strategies to establish the link between biosynthetic gene clusters and secondary metabolites. Fungal Genet Biol 2019; 130: 107-121.

108. Qin Z, Munnoch JT, Devine R*, et al*. Formicamycins, antibacterial polyketides produced by Streptomyces formicae isolated from African Tetraponera plant-ants. Chem Sci 2017; 8(4): 3218-3227.

109. Santos-Aberturas J, Chandra G, Frattaruolo L*, et al*. Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool. Nucleic Acids Res 2019; 47(9): 4624-4637.

110. Russell AH and Truman AW. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. Comput Struct Biotechnol J 2020; 18: 1838-1851.

111. Hannigan GD, Prihoda D, Palicka A*, et al*. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. Nucleic Acids Res 2019; 47(18): e110.

112. Yang Z, Liao B, Hsieh C*, et al*. Deep-BGCpred: A unified deep learning genome-mining framework for biosynthetic gene cluster prediction. bioRxiv 2021: 2021.11. 15.468547.

113. Rios-Martinez C, Bhattacharya N, Amini AP*, et al*. Deep self-supervised learning for biosynthetic gene cluster detection and product classification. PLoS Comput Biol 2023; 19(5): e1011162.

114. Saldívar-González F, Aldas-Bulos V, Medina-Franco J*, et al*. Natural product drug discovery in the artificial intelligence era. Chem Sci 2022; 13(6): 1526-1546.

115. Chevrette MG, Aicheler F, Kohlbacher O, *et al*. SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria. Bioinformatics 2017; 33(20): 3202-3210.

116. Röttig M, Medema MH, Blin K, *et al*. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res 2011; 39(suppl_2): W362-W367.

117. Merwin NJ, Mousa WK, Dejong CA, *et al*. DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. Proc Natl Acad Sci 2020; 117(1): 371-380.

118. de Los Santos EL. NeuRiPP: Neural network identification of RiPP precursor peptides. Sci Rep 2019; 9(1): 13406.

119. Kloosterman AM, Cimermancic P, Elsayed SS, *et al*. Integration of machine learning and pan-genomics expands the biosynthetic landscape of RiPP natural products. bioRxiv 2020: 2020.05. 19.104752.

120. Kloosterman AM, Cimermancic P, Elsayed SS, *et al*. Expansion of RiPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lanthipeptides. PLoS Biol 2020; 18(12): e3001026.

121. Cimermancic P, Medema MH, Claesen J, *et al*. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 2014; 158(2): 412-421.

122. Tietz J, Schwalen C, Patel P, *et al*. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. Nat Chem Biol 2017; 13.

123. Agrawal P, Amir S, Deepak, *et al*. RiPPMiner-Genome: A Web Resource for Automated Prediction of Crosslinked Chemical Structures of RiPPs by Genome Mining. J Mol Biol 2021; 433(11): 166887.

124. Melnikov AD, Tsentalovich YP, and Yanshole VV. Deep learning for the precise peak detection in high-resolution LC–MS data. Anal Chem 2019; 92(1): 588-592.

125. Fragoso J, Rogers J, Rogers A, *et al*. Expansion of novel biosynthetic gene clusters from diverse environments using SanntiS. bioRxiv 2023: 540769.

126. Yagin FH. *Machine Learning Approaches for Multi-omics Data Integration in Medicine*, in *Machine Learning Methods for Multi-Omics Data Integration* L.R. Abedalrhman Alkhateeb, Editor. 2023, Springer International Publishing: New York City. p. 23-38.

127. Shilpa S, Kashyap G, and Sunoj RB. Recent Applications of Machine Learning in Molecular Property and Chemical Reaction Outcome Predictions. J Phys Chem A 2023; 127(40): 8253-8271.

128. Klementz D, Döring K, Lucas X, *et al*. StreptomeDB 2.0—an extended resource of natural products produced by streptomycetes. Nucleic Acids Res 2015; 44(D1): D509-D514.

129. Raghunathan S and Priyakumar UD. Molecular representations for machine learning applications in chemistry. Int J Quantum Chem 2022; 122(7): e26870.

130. Bjerrum E, Rastemo T, Irwin R, *et al*. PySMILESUtils–enabling deep learning with the SMILES chemical language. Chemrxiv 2021.

131. Heller SR, McNaught A, Pletnev I, *et al*. InChI, the IUPAC international chemical identifier. J Cheminform 2015; 7(1): 1-34.

132. Jeon J, Kang S, and Kim HU. Predicting biochemical and physiological effects of natural products from molecular structures using machine learning. Nat Prod Rep 2021; 38(11): 1954-1966.

133. Martínez-Treviño SH, Uc-Cetina V, Fernández-Herrera MA*, et al*. Prediction of Natural Product Classes Using Machine Learning and 13C NMR Spectroscopic Data. J Chem Inf Model 2020; 60(7): 3376-3386.

134. Han R, Yoon H, Kim G*, et al*. Revolutionizing Medicinal Chemistry: The Application of Artificial Intelligence (AI) in Early Drug Discovery. Pharmaceuticals 2023; 16(9): 1259.

135. Scarselli F, Gori M, Tsoi AC*, et al*. The Graph Neural Network Model. IEEE Trans Neural Netw 2009; 20(1): 61-80.

136. Agatonovic-Kustrin S and Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal 2000; 22(5): 717-727.

137. Albawi S, Mohammed TA, and Al-Zawi S. *Understanding of a convolutional neural network*. in *2017 international conference on engineering and technology (ICET)*. 2017. Ieee.

138. Miethke M, Pieroni M, Weber T*, et al*. Towards the sustainable discovery and development of new antibiotics. Nat Rev Chem 2021; 5(10): 726-749.

139. Lu M, Yin J, Zhu Q*, et al*. Artificial Intelligence in Pharmaceutical Sciences. Engineering 2023.

140. Azman NS, A Samah A, Lin JT*, et al*. Support Vector Machine – Recursive Feature Elimination for Feature Selection on Multi-omics Lung Cancer Data. Prog Microbes Mol Biol 2023; 6(1).

141. Caesar LK, Montaser R, Keller NP*, et al*. Metabolomics and genomics in natural products research: complementary tools for targeting new chemical entities. Nat Prod Rep 2021; 38(11): 2041-2065.

142. Adeyemo J and Enitian A. Optimization of fermentation processes using evolutionary algorithms-A review. Sci Res Essays 2011; 6(7): 1464-1472.

143. Bezerra MA, Santelli RE, Oliveira EP*, et al*. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. Talanta 2008; 76(5): 965-977.

144. Frey DD and Wang H. Adaptive one-factor-at-a-time experimentation and expected value of improvement. Technometrics 2006; 48(3): 418-431.

145. Tan LT-H, Lee L-H, and Goh B-H. Critical review of fermentation and extraction of anti-Vibrio compounds from Streptomyces. Prog Microbes Mol Biol 2020; 3(1): a0000051.

146. Singh V, Haque S, Niwas R*, et al*. Strategies for fermentation medium optimization: an in-depth review. Front Microbiol 2017; 7: 2087.

147. Dong Z, Wu L, Wang L*, et al*. Optimization of Fracturing Parameters with Machine-Learning and Evolutionary Algorithm Methods. Energies 2022; 15(16): 6063.

148. Chiou J-P and Wang F-S. Hybrid method of evolutionary algorithms for static and dynamic optimization problems with application to a fed-batch fermentation process. Comput Chem Eng 1999; 23(9): 1277-1291.

149. Price KV. *Differential evolution*, in *Handbook of optimization: From classical to modern approach*. 2013, Springer. p. 187-214.

150. Kumar K, Shah H, and Moholkar VS. *Genetic algorithm for optimization of fermentation processes of various enzyme productions*, in *Optimization of sustainable enzymes production*. 2022, Chapman and Hall/CRC. p. 121-144.

151. Nagata Y and Chu KH. Optimization of a fermentation medium using neural networks and genetic algorithms. Biotechnol Lett 2003; 25(21): 1837-1842.

152. Vora LK, Gholap AD, Jetha K*, et al*. Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design. Pharmaceutics 2023; 15(7).

153. Ying X. *An overview of overfitting and its solutions*. in *Journal of physics: Conference series*. 2019. IOP Publishing.

154. Blanco-González A, Cabezón A, Seco-González A*, et al*. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. Pharmaceuticals (Basel) 2023; 16(6).

155. Huang J, Gao Q, Tang Y*, et al*. A deep learning model for type II polyketide natural product prediction without sequence alignment. Digit Discov 2023; 2(5): 1484-1493.

156. To CC and Vohradsky J. A parallel genetic algorithm for single class pattern classification and its application for gene expression profiling in Streptomyces coelicolor. BMC Genomics 2007; 8(1): 49.

157. Charaniya S, Mehra S, Lian W*, et al*. Transcriptome dynamics-based operon prediction and verification in Streptomyces coelicolor. Nucleic Acids Res 2007; 35(21): 7222-7236.

158. Wang M, Chen D, Zhao Q*, et al*. Isolation, Structure Elucidation, and Biosynthesis of a Cysteate-Containing Nonribosomal Peptide in Streptomyces lincolnensis. J Org Chem 2018; 83(13): 7102-7108.

159. Chun J, Atalan E, Kim S-B*, et al*. Rapid identification of streptomycetes by artificial neural network analysis of pyrolysis mass spectra. FEMS Microbiol Lett 1993; 114(1): 115-119.

160. Liu J, Clarke J-A, McCann S*, et al*. Analysis of Streptomyces Volatilomes Using Global Molecular Networking Reveals the Presence of Metabolites with Diverse Biological Activities. Microbiol Spectr 2022; 10(4): e00552-22.

161. El-Naggar N, Rabei N, Elmansy M*, et al*. Artificial neural network approach for prediction of AuNPs biosynthesis by Streptomyces flavolimosus, characterization, antitumor potency in-vitro and in-vivo against Ehrlich ascites carcinoma Sci Rep 2023; 13(1).

162. Srivastava A, Singh V, Haque S*, et al*. Response Surface Methodology-Genetic Algorithm Based Medium Optimization, Purification, and Characterization of Cholesterol Oxidase from Streptomyces rimosus. Sci Rep 2018; 8(1): 10913.

163. Tripathi CK, Khan M, Praveen V*, et al*. Enhanced antibiotic production by Streptomyces sindenensis using artificial neural networks coupled with genetic algorithm and Nelder-Mead downhill simplex. J Microbiol Biotechnol 2012; 22(7): 939-46.

164. Singh V, Khan M, Khan S*, et al*. Optimization of actinomycin V production by Streptomyces triostinicus using artificial neural network and genetic algorithm. Appl Microbiol Biotechnol 2009; 82(2): 379-385.

165. Pathak L, Singh V, Niwas R*, et al*. Artificial Intelligence versus Statistical Modeling and Optimization of Cholesterol Oxidase Production by using Streptomyces Sp. PLoS One 2015; 10(9): e0137268.

166. Mitrović I, Lukić N, Grahovac M*, et al*. Optimization of Streptomyces hygroscopicus Cultivation Parameters in a Lab‐scale Bioreactor. Chem Eng Technol 2021; 44(2): 349-358.

167. El-Naggar N-E, Bashir S, Rabei N*, et al*. Innovative biosynthesis, artificial intelligence-based optimization, and characterization of chitosan nanoparticles by Streptomyces microflavus and their inhibitory potential against Pectobacterium carotovorum. Sci Rep 2022; 12(1).

168. Bankar S, Dhumal V, Bhotmange D*, et al*. Empirical predictive modelling of poly-ε-lysine biosynthesis in resting cells of Streptomyces noursei. Food Sci Biotechnol 2014; 23: 201-207.

169. Lakshmi ES, Rao MRN, and Sudhamani M. Response surface methodology-artificial neural network based optimization and strain improvement of cellulase production by Streptomyces sp. Bioscience 2020; 36(4).