REVIEW ARTICLE

Discovery of new antibiotics using bioinformatics and machine learning methods

Mehmet Nihat Ural*

Pendik Veterinary Control Institute, İstanbul, Türkiye

***Corresponding author**:
Dr. Mehmet Nihat Ural

Address: BatıMah. Yunus Cad. No:2/1 34890 Pendik-İstanbul, Türkiye
Phone: +90544 2736765
ORCID: 0000-0003-1185-6659
E-mail: nihatural@yahoo.com

Original Submission:
10 September 2025
Revised Submission:
02 October 2025
Accepted:
18 November 2025

How to cite this article: Ural MN. 2025. Discovery of new antibiotics using bioinformatics and machine learning methods. Veterinaria, 74(3), 225-36.

ABSTRACT

Antibiotic resistance is a serious global health threat that causes approximately 1.27 million deaths worldwide each year and is expected to reach 10 million by 2050. New antibiotic development is exceptionally challenging, typically requiring 10-15 years and approximately $1.5 billion investment. In this process, genomic and metagenomic analyses play a critical role by revealing the genetic potential of unculturable microorganisms and identifying new antibiotic-producing microorganisms. Additionally, deep learning models analyze molecular structures to identify new compounds with antibacterial activity, and virtual screening techniques analyze large molecular databases to determine potential active compounds. It has been shown that models developed using deep learning can predict antibiotic biosynthesis gene clusters with over 90% accuracy. Alongside these approaches, the identification of antibiotic combinations and the prediction of synergistic effects allow for the development of more effective treatment strategies against multi-drug resistance. These methods contribute to the development of proactive approaches in managing antibiotic resistance and optimize the discovery of new antibiotics and the effective use of existing ones. This review examines the discovery of new antibiotics using bioinformatics and machine learning methods.

Keywords: Bioinformatics, machine learning, new antibiotic discovery

INTRODUCTION

Antibiotic resistance has emerged as one of the most critical global health threats of the 21st century and is defined by the World Health Organization as "one of the greatest threats to global public health, food security, and development" (WHO, 2023). This problem arises from bacteria developing resistance to antibiotics, causing treatable infections to become fatal. Various factors contribute to the rise of antibiotic resistance, including the excessive and inappropriate use of antibiotics, inadequate infection control, and the challenges in developing new antibiotics (Ventola, 2015). According to the World Health Organization (WHO), antimicrobial resistance (AMR) was responsible for an estimated 1.27 million deaths in 2019, based on global statistical modelling across 204 countries (Murray et al., 2022). These estimates include uncertainty intervals (95% UI 0.91–1.71 million deaths) and indicate that, if current trends continue, the annual number of deaths could reach 10 million by 2050. Furthermore, according to World Bank data, global GDP losses of up to 3.8% could occur by 2050 (World Bank, 2017). Addressing this serious threat

doi.org/10.51607/22331360.2025.74.3.225

requires a multifaceted approach, with strategies such as promoting the rational use of antibiotics, developing new antibiotics, strengthening global surveillance systems, adopting the One Health approach, and enhancing international cooperation being of paramount importance (Ajulo and Awosile, 2024).

The rapid spread of antibiotic resistance and the decreasing effectiveness of existing antibiotics have made the development of new antibiotics an urgent global health priority (WHO, 2023). However, the development of new antibiotics is a long, costly, and challenging process. Typically, the period from the discovery of a new antibiotic to its market release ranges from 10 to 15 years, with an average cost of approximately 1.5 billion USD (Luepke et al., 2017; Plackett, 2020). This lengthy process includes basic research, preclinical studies, clinical trials, and regulatory approval stages. Moreover, approximately 90% of drug candidates that enter Phase I–III clinical trials ultimately fail to reach approval, largely due to insufficient efficacy and unmanageable toxicity (Sun et al., 2022). The complexity of the new antibiotic development process, combined with its high cost and low return on investment, has reduced the willingness of pharmaceutical companies to invest in this field, leading to a decline in new antibiotic discoveries (Plackett, 2020). Indeed, while the discovery of new antibiotic classes peaked in the 1940s and 1950s, it has shown a significant decline since the 1960s. The numerical distribution of new antibiotic classes discovered between 1900 and 2009 is shown in Figure 1.

The average annual revenue generated after the market launch of a new antibiotic is approximately 46 million USD (Plackett, 2020). This figure is far from covering the development costs and is considerably lower compared to other types of drugs. In addition, the use of new antibiotics is often restricted to prevent the development of resistance, which further reduces sales and profitability. To address this issue, strategies such as promoting academia–industry collaborations, developing new business models, and increasing public funding have been proposed (Theuretzbacher et al., 2020). New financing models, such as the market entry reward, aim to reduce risk and encourage investment by providing substantial rewards to successful antibiotic developers (Årdal et al., 2020). Furthermore, global initiatives such as the Global Antibiotic Research and Development Partnership (CARB-X) and the Combating Antibiotic-Resistant Bacteria Biopharmaceutical Accelerator (GARDP) are working to accelerate the discovery and development of new antibiotics (Simpkin et al., 2017). Nevertheless, along with the development of new antibiotics, the rational use of existing antibiotics and the prevention of antimicrobial resistance are of great importance (Salam et al., 2023).

Bioinformatics and machine learning methods hold great promise in the fight against antibiotic resistance and have become an important tool in the discovery and development of new antibiotics. These technologies analyze large datasets to rapidly and effectively identify potential antibiotic candidates, offering significant time and cost savings compared to traditional methods (Stokes et al., 2020). Although the applications of bioinformatics

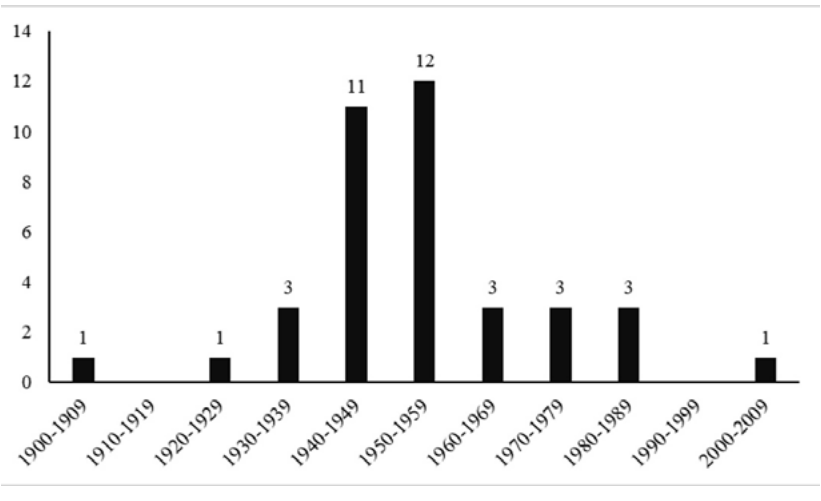

Figure 1 Historical distribution of new antibiotic class discoveries (Stennett et al., 2022)

and machine learning to antibiotic research have been discussed separately in the literature, an integrated perspective that combines both computational fields and their latest methodological advances remains lacking. This narrative review synthesizes recent developments in bioinformatics and machine learning applications for antibiotic discovery. A comprehensive literature search was conducted using PubMed, Web of Science, and Google Scholar databases.

Bioinformatics Methods

Genomic and metagenomic analyses play a crucial role in the discovery and identification of antibiotic-producing microorganisms. These culture-independent approaches are effective in revealing the genetic potential of microorganisms that cannot be cultured (Hover et al., 2018). By overcoming the limitations of traditional culture-based methods, these techniques enable the analysis of genome sequences from microbial communities in natural environments that may contain potential antibiotic-producing organisms. Machine learning algorithms are used to analyze large genomic datasets to predict antibiotic biosynthetic gene clusters (BGCs) and identify new antibiotic candidates (Stokes et al., 2020). Deep learning models can analyze molecular structures to detect compounds with potential antibacterial activity. Compared to conventional high-throughput screening methods, this approach enables the identification of new antibiotic candidates more rapidly and cost-effectively (David et al., 2021). Deep learning tools, such as DeepBGC have demonstrated strong performance in predicting antibiotic biosynthetic gene clusters. This tool achieved an accuracy score of 94.6% AUC in identifying new BGC classes that had not been encountered before (Hannigan et al., 2019). For example, the deep learning model DeepARG, developed by Arango-Argoty et al. (2018), can predict antibiotic resistance genes in metagenomic data with high accuracy. Such models offer higher sensitivity and specificity than traditional approaches. Bioinformatics and machine learning methods are also used for the functional characterization of resistance genes. The Comprehensive Antibiotic Resistance Database (CARD), developed by Alcock et al. (2020), provides a comprehensive catalogue of antibiotic resistance genes and associated phenotypes. This database is used to discover resistance genes present in the genomes of uncultured microorganisms through metagenomic analyses and to predict potential resistance functions of novel genes. A global study conducted by Hendriksen et al. (2019) utilized metagenomic data from wastewater samples to reveal the worldwide distribution and diversity of antibiotic resistance genes. Such studies provide critical information for monitoring and controlling antibiotic resistance at the community level.

The identification of target proteins is a critical step in the discovery of new antibiotics. The deep learning model DeepDrug3D, developed by Pu et al. (2019), can classify drug-binding sites in protein structures with high accuracy by performing three-dimensional analyses. Such models play an important role in identifying and characterizing new antibiotic targets, especially when combined with structural bioinformatics methods. These approaches help identify potential drug-binding regions on proteins, thereby improving the understanding of drug–protein interactions. AlphaFold2, developed by Jumper et al. (2021), enables highly accurate predictions of protein structures, allowing the structural analysis of proteins whose structures have not been determined experimentally. Such tools play a key role in the identification and characterization of potential antibiotic targets. In particular, accurately predicting protein structures is considered a critical step in drug design and in understanding biological processes.

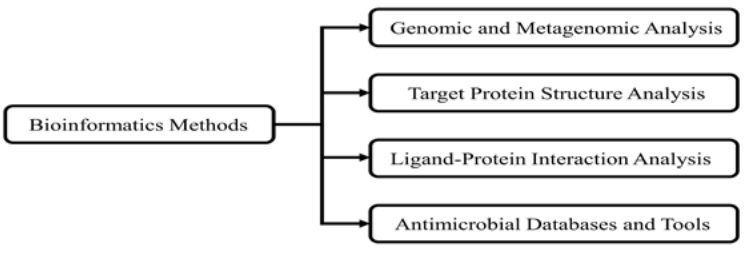

Figure 2 Overview of the Applications of Bioinformatics Methods in Antibiotic Discovery. This schematic illustrates the integrated workflow of bioinformatics approaches in antibiotic research

Modeling ligand–protein interactions is a critical stage in the design and development of new antibiotics, and with the intensive use of bioinformatics and machine learning methods, it has become increasingly sophisticated and effective. For example, molecular docking programs such as AutoDockVina, developed by Trott and Olson (2010), predict possible binding conformations of ligands on proteins, while deep learning models such as DeepDTA, developed by Öztürk et al. (2018), predict drug–target interactions more rapidly and accurately. The GROMACS software, developed by Abraham et al. (2015), offers high performance for conducting molecular dynamics simulations, enabling the investigation of the dynamic behavior of ligand–protein complexes. Pharmacophoremodeling tools such as LigandScout, developed by Wolber and Langer (2005), are used to define the structural features required for the biological activity of ligands through three-dimensional pharmacophore models. These advanced methods make it possible to screen broader chemical spaces more quickly and effectively during the drug discovery process.

Antimicrobial databases, as well as sequence and structure analysis tools, are critical resources in the discovery and development of new antibiotics. These tools provide researchers with access to extensive datasets, facilitating the identification and optimization of potential antibiotic candidates. For example, the Collection of Anti-Microbial Peptides (CAMP) database, introduced by Waghu and Thomas (2020), provides comprehensive information on antimicrobial peptides, enabling the investigation of their structure–activity relationships. Similarly, the Database of Antimicrobial Activity and Structure of Peptides (DBAASP), created by Gogoladze et al. (2014), is used to analyze the structure–activity relationships of antimicrobial peptides. For protein structure analysis, homology modeling tools such as SWISS-MODEL, updated by Waterhouse et al. (2018), help predict protein structures in cases where experimentally determined structures are unavailable. Additionally, molecular visualization and analysis programs such as UCSF Chimera, developed by Pettersen et al. (2004), allow detailed examination of protein structures and ligand–protein interactions. Data mining and text mining techniques are also widely used to extract information about antimicrobial compounds from scientific literature and biological databases. For instance, the DrugBank database, developed by Wishart et al. (2006), provides comprehensive information on approved and experimental drugs, serving as a valuable resource in identifying new antibiotic candidates. These tools and databases accelerate and optimize the process of discovering and developing new antibiotics by providing researchers with a vast pool of information. In the future, the further expansion and integration of these resources will contribute to the development of more effective strategies in combating antibiotic resistance.

Machine Learning Methods

Machine learning methods offer powerful tools for antibiotic discovery in bioinformatics. Supervised and unsupervised learning represent the two main approaches. In supervised learning, models are trained using labeled data; for example, compounds with known antibiotic activity can be used to predict new potential candidates (Stokes et al., 2020). Unsupervised learning, on the other hand, uncovers hidden structures in unlabeled data, which can be important for discovering new classes of antibiotics (Visan and Negut, 2024). Feature selection identifies the most informative features from large molecular datasets, thereby improving model performance and reducing computational load (Saeys et al., 2007). Data preprocessing steps include techniques such as handling missing values, detecting outliers, and eliminating redundant features; these steps enhance the reliability and generalization ability of models (Lee JW, 2022). By enabling the efficient analysis of large-scale biological and chemical datasets, these methods contribute to the faster and more efficient identification of new antibiotic candidates.

Antibiotic activity prediction plays a critical role in the discovery of new and effective antibiotics, with Quantitative Structure–Activity Relationship (QSAR) models and deep learning approaches standing out in this field. QSAR models aim to mathematically describe the relationship between molecular structure and biological activity, offering the ability to rapidly and cost-effectively screen large compound libraries (Cherkasov et al., 2014). These models involve the calculation of molecular descriptors, model construction using statistical or machine learning methods, and model validation. Deep learning approaches, on the other hand, have revolutionized antibiotic activity prediction in recent years, standing out for their ability to learn from complex and large-scale datasets (Stokes et al., 2020). Deep learning models are based on multi-layer artificial neural networks and have the capability to automatically learn complex molecular features. These two approaches play complementary roles: QSAR

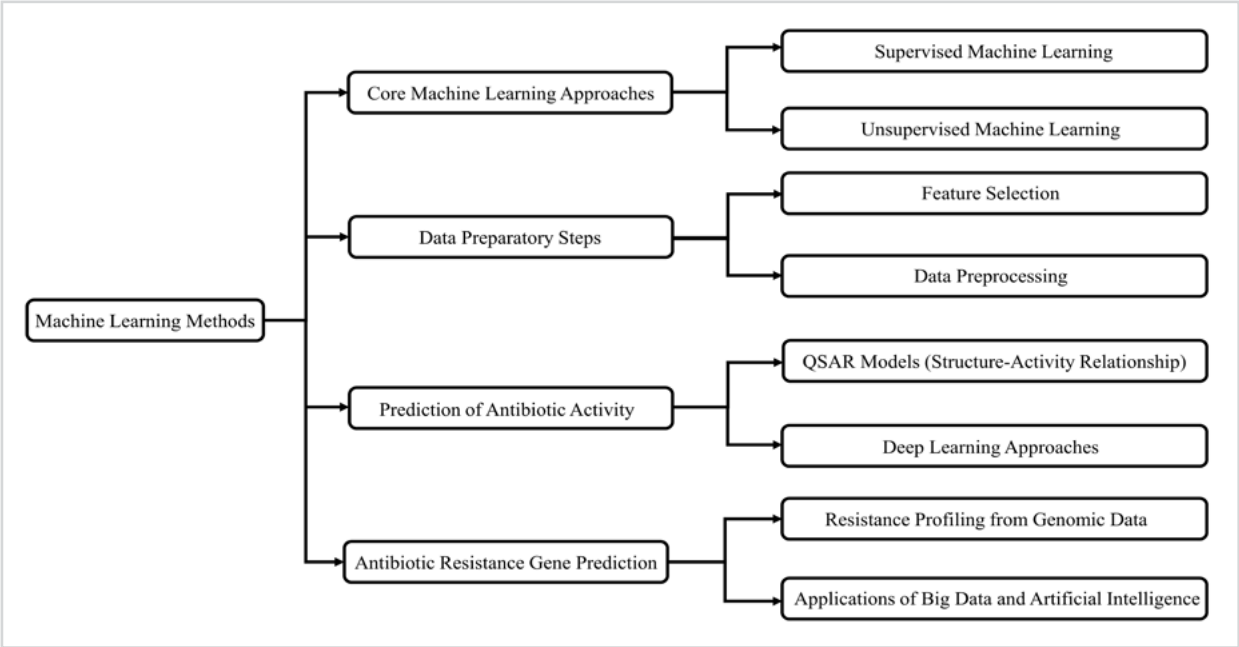

Figure 3 Overview of Machine Learning Applications in Antibiotic Discovery. This diagram depicts the machine learning pipeline for antibiotic research

models are advantageous in terms of interpretability and low computational requirements, whereas deep learning models excel at learning complex patterns and leveraging large datasets.

The prediction of antibiotic resistance genes is of critical importance for developing effective treatment strategies and discovering new antibiotics. In this field, two prominent approaches are the extraction of resistance profiles from genomic data and the application of big data and artificial intelligence.

Resistance profile extraction from genomic data is based on the comprehensive analysis of bacterial genomes, enabling the identification of potential resistance mechanisms. McArthur et al. (2013) developed the Comprehensive Antibiotic Resistance Database (CARD), creating a comprehensive resource that links genomic data with antibiotic resistance. This database provides a powerful tool for identifying and classifying resistance genes. Big data and artificial intelligence applications, on the other hand, integrate genomic data with clinical, epidemiological, and environmental information to produce more comprehensive and accurate predictions. For example, Nguyen et al. (2018) used machine learning algorithms to model the evolution and spread of antibiotic resistance, demonstrating that integrating different data sources can improve prediction accuracy. Similarly, Moradigaravand et al. (2018) applied deep learning methods to predict E. coli antibiotic resistance and showed that this approach achieved higher accuracy than traditional methods. These developments allow for the creation of faster, more accurate, and more comprehensive approaches to predicting antibiotic resistance genes, thereby contributing to the formulation of new strategies in the fight against antimicrobial resistance.

However, despite their high predictive performance, the global generalizability of these models in clinical decision-making and antibiotic discovery remains limited by the current data landscape, which is largely derived from high-income countries. Overcoming this bias will require expanding clinical datasets from diverse geographic and socioeconomic regions and adopting experimental conditions that more closely mimic the biological environment of infection (Nguyen et al., 2018; Peiffer-Smadja et al., 2020; Ayon, 2023).

Applications of Bioinformatics and Machine Learning

Bioinformatics and machine learning methods are driving groundbreaking advancements in the discovery of new antibiotic molecules. In this field, high-throughput screening (HTS), virtual screening techniques, and AI-

assisted drug design stand out. HTS enables the rapid and automated testing of large compound libraries, thereby accelerating the experimental evaluation of potential antibiotic candidates (Ayon, 2023). Virtual screening is a technique that systematically screens large molecular databases using in silico methods during the drug discovery process to predict potential active compounds against a specific biological target. This method serves as an effective pre-screening tool, significantly reducing costs and time in the drug development process before moving on to expensive and time-consuming experimental stages (Oliveira et al., 2023). AI-assisted drug design makes it possible to design new molecules faster and more cost-effectively than with traditional methods. For example, in 2020, James Collins and his team used deep learning models to discover halicin, a broad-spectrum antibiotic. This study is a striking example of the potential of AI in antibiotic discovery (Stokes et al., 2020). Similarly, Zhavoronkov et al. (2019) used generative tensorial reinforcement learning (GENTRL) to design new drug-like molecules, demonstrating that this approach is faster than conventional methods. These developments highlight the critical role of bioinformatics and machine learning methods in the discovery of new antibiotic molecules and pave the way for more effective and faster drug discovery processes in the future. The identification of antibiotic combinations plays a crucial role in combating multidrug resistance and developing more effective treatment strategies. Chandrasekaran et al. (2016) conducted a large-scale antibiotic combination screening study using machine learning algorithms and discovered novel combinations with synergistic effects. This approach reduced the number of experimental trials needed, saving both time and cost. Benefo et al. (2024) utilized genomic data and machine learning algorithms to predict the resistance profiles of pathogens to antibiotic combinations. These advancements demonstrate the advantages that bioinformatics and machine learning methods provide in determining antibiotic combinations and open the door to the development of more effective and targeted combination therapies in the future.

Antibiotic resistance management is of critical importance to global health today, with two key approaches standing out: predicting the development of resistance and establishing effective monitoring and control strategies. The prediction of resistance development has been greatly enhanced through the use of bioinformatics and machine learning methods. For example, Nguyen et al. (2018) developed a system that predicts antibiotic resistance from bacterial genomes using machine learning models such as XGBoost. Feretzakis et al. (2020) employed various machine learning methods to model the relationship between antibiotic use and the development of resistance in hospital settings. In their study, they compared the performance of different methods in predicting antibiotic susceptibility and demonstrated that these approaches could be used to support empirical treatment decisions. The results showed that analyzing microbiological data alongside easily accessible information-such as basic patient details-can enable the early prediction of antibiotic resistance. These developments highlight the advantages that bioinformatics and machine learning methods provide in antibiotic resistance management and pave the way for the development of more effective and proactive resistance management strategies in the future.

The application of bioinformatics and machine learning methods in clinical settings brings not only theoretical advantages but also various practical challenges. Peiffer-Smadja et al. (2020) examined the potential of machine learning-based clinical decision support systems (ML-CDSS) in critical areas such as diagnosis, treatment management, and antibiotic selection in infectious diseases. Their study indicates that these systems can play a significant role in combating antimicrobial resistance by optimizing clinical decision-making processes. However, it also emphasizes that challenges, such as data quality, incompatibilities between healthcare systems, and real-time implementation need to be addressed for effective integration of ML-CDSS. Similarly, Rawson et al. (2018), in their evaluation of artificial intelligence applications within antimicrobial stewardship programs, discussed several challenges encountered in clinical practice. Their work highlights issues such as data integration problems between different hospital systems, the necessity of adapting algorithms to local populations, and the importance of healthcare professionals' trust in these systems. Furthermore, they stress that the successful adoption of AI-based systems requires involving end users in the development process and ensuring transparency. These studies demonstrate that, for bioinformatics and machine learning methods to be successfully implemented in clinical environments, technological advancements must be accompanied by organizational and cultural changes.

Table 1 Summary of representative bioinformatics and machine learning methods applied in antibiotic discovery

Method	Primary Application	Reported Performance (as stated in source)	Main Limitations	Reference
DeepARG	Identification of antibiotic resistance genes (ARGs) in metagenomic samples.	Precision ≈ 0.97; Recall ≈ 0.90 (cross-validation).	Data imbalance for underrepresented ARG classes; uncertainty in novel/rare variants.	Arango-Argoty et al., 2018
DeepDrug3D	Classification of protein–ligand binding pockets (nucleotide vs. heme).	AUROC 0.986 (nucleotide), 0.987 (heme); overall accuracy ≈ 95% (TOUGH-C1 dataset).	Limited to nucleotide/heme classes; depends on voxel representation and 3D structural data availability.	Pu et al., 2019
DeepDTA	Drug–target binding affinity prediction (regression).	CI: 0.878 (Davis), 0.863 (KIBA); MSE: 0.261/0.194; model type: CNN/CNN.	Sequence-only input; lacks 3D structural context; dataset-dependent performance.	Öztürk et al., 2018
AlphaFold2	Highly accurate prediction of three-dimensional protein structure.	In the CASP14 assessment, AlphaFold2 achieved near-experimental accuracy across all targets based on GDT_TS score distributions.	Limited accuracy for multi-domain complexes and flexible/disordered regions; ligand and cofactor positions not directly predicted.	Jumper et al., 2021
AutoDock Vina	Molecular docking and binding pose prediction.	According to the original publication, AutoDock Vina is a molecular docking program that is both much faster and more accurate than its predecessor, AutoDock 4.	Limited correlation with absolute affinities; receptor flexibility and solvent effects underrepresented.	Trott & Olson, 2010
GROMACS (MD simulations)	Molecular dynamics analysis and post-docking refinement.	No predictive accuracy metric; performance determined by force field and setup parameters.	High computational cost; results sensitive to force field and simulation timescale.	Abraham et al., 2015
CAMP	Curation and analysis of antimicrobial peptide (AMP) sequences and activities.	No quantitative metric (database-based resource); regularly updated content.	The ability of CAMPSign to identify peptides according to their families is limited by the number of family signatures it is currently trained on.	Waghu and Thomas, 2020
DBAASP	Antimicrobial peptide (AMP) activity and structure database.	Performance metric not applicable (curated dataset).	Experimental heterogeneity; variation in assay conditions.	Gogoladze et al., 2014
LigandScout	Deriving 3-dimensional (3-D) pharmacophores from protein-bound ligands and using these models as virtual screening filters.	It is fast enough to generate pharmacophores in “a few seconds” and selective enough to identify known targets without error.	Requires known ligand–protein complexes; qualitative and template-dependent.	Wolber and Langer, 2005

Method	Primary Application	Reported Performance (as stated in source)	Main Limitations	Reference
QSAR Models	Predicting biological activity, physicochemical, or toxicological properties of compounds from their molecular structure descriptors.	Performance depends on dataset and validation.	Interpretable and data-efficient, but limited generalizability and extrapolation beyond the training set.	Cherkasov et al., 2014
GENTRL	A deep generative model used to design de novo small molecule drugs by optimizing synthetic feasibility, novelty, and biological activity.	Out of the six compounds designed and synthesized by GENTRL, four exhibited activity in biochemical assays, with IC_{50} values of 10 nM, 21 nM, 278 nM, and 1 μM, respectively.	The generated compounds may require further optimization in terms of selectivity, specificity, and other medicinal chemistry properties.	Zhavoronkov et al., 2019

Case Studies

Bioinformatics and machine learning methods have achieved significant successes in the fields of antibiotic discovery and resistance management. One of the most notable antibiotic discovery projects was conducted by James Collins and colleagues, in which researchers used deep learning models to discover halicin, a broad-spectrum antibiotic. Halicin represents a new class of antibiotics shown to be effective against multidrug-resistant bacteria. This study demonstrated the potential of AI-assisted drug discovery, offering a much faster and more cost-effective discovery process compared to traditional methods (Stokes, 2020). Another antibiotic discovered using machine learning methods is Abaucin, developed against the multidrug-resistant Gram-negative pathogen Acinetobacter baumannii. Researchers screened approximately 7,500 small molecules to identify compounds that inhibit the growth of A. baumannii in vitro. They identified nine effective compounds and highlighted the most effective, Abaucin, as a potential therapeutic candidate. Abaucin is only effective against A. baumannii and acts by disrupting lipoprotein transport via the LolE protein (Liu et al., 2023). Moradigaravand et al. (2018) developed a machine learning model that predicts antibiotic resistance in Escherichia coli with high accuracy using whole-genome sequencing data. In their study, data from 1,936 isolates were used to predict resistance profiles for 11 different antibiotics. This approach was able to predict resistance without prior knowledge of resistance mechanisms and provided an important framework for integrating genomic and epidemiological data into clinical diagnosis. Similarly, Pesesky et al. (2016) evaluated the effectiveness of combining whole-genome sequencing data with machine learning and rule-based algorithms to predict the antibiotic resistance profiles of Gram-negative bacilli. The study showed that genotypic, data-driven predictions could be made more rapidly than phenotypic antibiotic susceptibility testing. This method offers significant potential, particularly in clinical settings, to optimize antibiotic selection and support antimicrobial resistance control strategies. These case studies illustrate the substantial advantages provided by bioinformatics and machine learning methods in antibiotic discovery and resistance management, offering hope for the development of more effective antibiotic therapies in the future.

Trends and Challenges

The applications of bioinformatics and machine learning methods in antibiotic discovery and resistance management offer great potential but also bring various challenges and ethical issues. Chief among these challenges is the processing and integration of large and heterogeneous datasets. While integrating genomic data enables a better understanding of complex biological systems, problems such as data quality and standardization remain significant obstacles (de la Lastra et al., 2024). Moreover, the interpretability and explainability of machine learning models are critically important, especially in clinical applications.

In this context, Rudin (2019) addressed the lack of transparency in AI models used for high-stakes decisions, emphasizing that transparent and interpretable models should be preferred. Future research directions include single-cell genomic analyses and AI-assisted drug design. For example, Zhavoronkov et al. (2019) used generative tensorial reinforcement learning (GENTRL) to design novel drug-like molecules and demonstrated that this approach is much faster and more effective than traditional methods. From an ethical and legal standpoint, the privacy and security of personal genomic data are major concerns. Mittos et al. (2019) discussed ethical issues and legal regulations related to the use of genomic data, drawing attention to the difficulties of balancing data sharing with privacy. Furthermore, the regulation of AI-assisted antibiotic discovery and its application poses new challenges for the pharmaceutical industry and healthcare systems. In light of these developments, it is evident that the effective use of bioinformatics and machine learning methods in antibiotic discovery and resistance management will require interdisciplinary collaboration, the establishment of ethical standards, and continuous technological innovation.

CONCLUSION

Antibiotic resistance is one of the most pressing threats to global health security, and addressing this challenge is critical to safeguarding human health. In this context, the opportunities offered by bioinformatics and machine learning in the fields of antibiotic discovery and resistance management go far beyond traditional approaches, representing a paradigm shift. The in-depth analysis of genomic and metagenomic data enables the identification of antibiotic-producing microorganisms and the mapping of resistance genes, with bioinformatics tools driving significant advances in these areas. Machine learning algorithms contribute by analyzing large datasets to predict antibiotic biosynthetic gene clusters and rapidly identify potential antibiotic candidates, while deep learning techniques greatly accelerate the detection of compounds with antibacterial activity through molecular structure analysis. Furthermore, the integration of virtual screening and high-throughput screening methods speeds up the discovery of potential active compounds and optimizes the experimental validation stage. The identification of antibiotic combinations and the prediction of synergistic effects make it possible to develop more effective treatment strategies against multidrug resistance. At the same time, the prediction and monitoring of resistance genes provide a proactive and dynamic approach to antibiotic resistance management, with the potential to shape global health policies. In conclusion, integrating bioinformatics and machine learning into antibiotic discovery and resistance management is not only accelerating research but also reshaping clinical and public health strategies against infectious diseases. Broader implementation of these approaches will be essential for developing new antibiotics, optimizing existing therapies, and strengthening global preparedness for future health challenges. Achieving this vision will require coordinated efforts across technical, regulatory, and collaborative domains.

ACKNOWLEDGEMENTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest.

REFERENCES

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1-2, 19-25.doi:10.1016/j.softx.2015.06.001

Ajulo S, Awosile B. 2024. Global antimicrobial resistance and use surveillance system (GLASS 2022): Investigating the relationship between antimicrobial resistance and antimicrobial consumption data across the participating countries. PLoS One 19(2), 0297921.doi:10.1371/journal.pone.0297921

Alcock BP, Raphenya AR, Lau TT, Tsang KK, Bouchard M, Edalatmand A, et al. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res, 48(D1), D517-D525. doi:10.1093/nar/gkz935

Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. 2018. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data.

Microbiome, 6(1), 23.doi:10.1186/s40168-018-0401-z

Årdal C, Lacotte Y, PloyMC. 2020. Financing pull mechanisms for antibiotic-related innovation: opportunities for Europe. Clin Infect Dis, 71(8), 1994-9.doi:10.1093/cid/ciaa153

Ayon NJ. 2023. High-Throughput Screening of Natural Product and Synthetic Molecule Libraries for Antibacterial Drug Discovery. Metabolites, 13(5), 625. doi:10.3390/metabo13050625

Benefo EO, Ramachandran P, Pradhan AK. 2024. Genome-based machine learning for predicting antimicrobial resistance in Salmonella isolated from chicken. LWT, 199, 116122. doi:10.1016/j.lwt.2024.116122

Chandrasekaran S, Cokol-Cakmak M, Sahin N, Yilancioglu K, Kazan H, Collins JJ, et al. 2016. Chemogenomics and orthology-based design of antibiotic combination therapies. Molecular systems biology, 12(5), 872.doi:10.15252/msb.20156777

Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. 2014. QSAR modeling: where have you been? Where are you going to?. J Med Chem, 57(12), 4977-5010.doi:10.1021/jm4004285

David L, Brata AM, Mogosan C, Pop C, Czako Z, Muresan L, et al. 2021. Artificial Intelligence and Antibiotic Discovery. Antibiotics (Basel), 10, 10(11), 1376.doi:10.3390/antibiotics10111376.

de la Lastra JMP, Wardell SJT, Pal T, de la Fuente-Nunez C, Pletzer D. 2024. From Data to Decisions: Leveraging Artificial Intelligence and Machine Learning in Combating Antimicrobial Resistance - a Comprehensive Review. J Med Syst, 48(1), 71.doi:10.1007/s10916-024-02089-5.

Feretzakis G, Sakagianni A, Loupelis E, Kalles D, Martsoukou M, Skarmoutsou N, et al. 2020. Using Machine Learning Algorithms to Predict Antimicrobial Resistance and Assist Empirical Treatment. Studies in health technology and informatics, 272, 75-8.doi:10.3233/SHTI200497.

Gogoladze G, Grigolava M, Vishnepolsky B, Chubinidze M, Duroux P, Lefranc MP, et al. 2014. DBAASP: database of antimicrobial activity and structure of peptides. FEMS microbiology letters, 357(1), 63-8.doi:10.1111/1574-6968.12489

Hannigan GD, Prihoda D, Palicka A, Soukup J, Klempir O, Rampula L, et al. 2019. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. Nucleic acids research, 47(18), e110-e110. doi:10.1093/nar/gkz654

Hendriksen RS, Munk P, Njage P, Van Bunnik B, McNally L, Lukjancenko O, et al. 2019. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nature Communications, 10(1), 1124.doi:10.1038/s41467-019-08853-3.

Hover BM, Kim SH, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, et al. 2018. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. Nature Microbiology, 3(4), 415-22.doi:10.1038/s41564-018-0110-1

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-9. doi:10.1038/s41586-021-03819-2

LeeJW, Maria-Solano MA, Vu TNL, Yoon S, Choi S. 2022. Big data and artificial intelligence (AI) methodologies for computer-aided drug design (CADD). Biochemical Society Transactions, 50(1), 241-52.doi:10.1042/BST20211240

Liu G, Catacutan DB, Rathod K, Swanson K, Jin W, Mohammed JC, et al. 2023. Deep learning-guided discovery of an antibiotic targeting *Acinetobacter baumannii*. Nat Chem Biol, 19,1342-50. doi:10.1038/s41589-023-01349-8

Luepke KH, Suda KJ, Boucher H, Russo RL, Bonney MW, Hunt TD, et al. 2017. Past, present, and future of antibacterial economics: increasing bacterial resistance, limited antibiotic pipeline, and societal implications. Pharmacotherapy: The J Human Pharmacol Drug Ther, 37(1), 71-84.doi:10.1002/phar.1868.

McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. 2013. The comprehensive antibiotic resistance database. Antimicrobial agents and chemotherapy, 57(7), 3348-57.doi:10.1128/AAC.00419-13.

Mittos A, Malin B, De Cristofaro E. 2019. Systematizing genome privacy research: A privacy-enhancing technologies perspective. Proceedings on Privacy Enhancing Technologies, 2019(1), 87-107.doi:10.2478/popets-2019-0006

Moradigaravand D, Palm M, Farewell A, Mustonen V, Warringer J, Parts, L. 2018. Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data. PLoS computational biology, 14(12), e1006258.doi:10.1371/journal.pcbi.1006258

Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Aguilar GR, Grayet A, al. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet, 399(10325), 629–55. doi:10.1016/S0140-6736(21)02724-0

Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, et al. 2018. Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. Scientific reports, 8(1), 421.doi:10.1038/s41598-017-18972-w

Oliveira TAD, Silva MPD, Maia EHB, Silva AMD, Taranto AG. 2023. Virtual screening algorithms in drug discovery: A review focused on machine and deep learning methods. Drugs and Drug Candidates, 2(2), 311-34.doi:10.3390/ddc2020017

Öztürk H, Özgür A, Ozkirimli E. 2018. DeepDTA: deep drug-target binding affinity prediction. Bioinformatics, 34(17), i821-i829.doi:10.1093/bioinformatics/bty593

Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Georgiou P, Lescure FX, et al. 2020. Machine learning for

clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect, 26(5), 584-95.doi:10.1016/j.cmi.2019.09.009

Pesesky MW, Hussain T, Wallace M, Patel S, Andleeb S, Burnham CA, et al. 2016. Evaluation of machine learning and rules-based approaches for predicting antimicrobial resistance profiles in gram-negative bacilli from whole genome sequence data. Frontiers in microbiology, 7, 1887.doi:10.3389/fmicb.2016.01887.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J Computat Chemist, 25(13), 1605-12.doi:10.1002/jcc.20084.

Plackett B. 2020. Why big pharma has abandoned antibiotics. Nature, 586(7830), S50-S52.

Pu L, Govindaraj RG, Lemoine JM, WuHC, Brylinski M. 2019. DeepDrug3D: classification of ligand-binding pockets in proteins with a convolutional neural network. PLoS computational biology, 15(2), e1006718.doi:10.1371/journal.pcbi.1006718

Rawson TM, Ahmad R, Toumazou C, Georgiou P, Holmes AH. 2018. Artificial intelligence can improve decision-making in infection management. Nature Human Behaviour, 2(8), 452-4. doi:10.1038/s41562-019-0583-9

Rudin C. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215. doi:10.1038/s42256-019-0048-x

Saeys Y, Inza I, Larranaga P. 2007. A review of feature selection techniques in bioinformatics. bioinformatics, 23(19), 2507-2517.Doi:10.1093/bioinformatics/btm344

Salam MA, Al-Amin MY, Salam MT, Pawar JS, Akhter N, Rabaan AA. 2023. Antimicrobial Resistance: A Growing Serious Threat for Global Public Health. Healthcare (Basel). 11(13), 1946. doi:10.3390/healthcare11131946.

Simpkin VL, Renwick MJ, Kelly R, Mossialos E. 2017. Incentivising innovation in antibiotic drug discovery and development: progress, challenges and next steps. J Antibiot,70(12),1087-96. doi:10.1038/ja.2017.124.

Stennett HL, BackCR, Race PR. 2022. Derivation of a precise and consistent timeline for antibiotic development. Antibiotics, 11(9), 1237.doi:10.3390/antibiotics11091237

Stokes JM, Yang K, Swanson K, JinW, Cubillos-Ruiz A, Donghia NM, et al. 2020. A deep learning approach to antibiotic discovery. Cell, 180(4), 688-702.doi:10.1016/j.cell.2020.01.021

Sun D, Gao W, Hu H, Zhou S. 2022. Why 90% of clinical drug development fails and how to improve it? Acta Pharmaceutica Sinica B. 12(7), 3049-62. doi:10.1016/j.apsb.2022.02.002.

Theuretzbacher U, Outterson K, Engel A, Karlén A. 2020. The global preclinical antibacterial pipeline. Nature Rev Microbiol, 18(5),275-85.doi:10.1038/s41579-019-0288-0.

Trott O, Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Computat Chemist, 31(2), 455-61.doi:10.1002/jcc.21334

Ventola CL. 2015. The antibiotic resistance crisis: part 1: causes and threats. Pharmacy and Therapeutics, 40(4), 277-83.

Visan AI, Negut I. 2024. Integrating Artificial Intelligence for Drug Discovery in the Context of Revolutionizing Drug Delivery. Life, 14(2), 233.doi:10.3390/life14020233.

Waghu FH, Idicula-Thomas S. 2020. Collection of antimicrobial peptides database and its derivatives: Applications and beyond. Protein Sci, 29(1), 36-42. doi:10.1002/pro.3714

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res, 46(W1), W296-W303.doi:10.1093/nar/gky427.

Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. 2006. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res, 1;34(Database issue):D668-72. doi:10.1093/nar/gkj067.

Wolber G, Langer T. 2005. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Cheml Inform Model, 45(1), 160-9. doi:10.1021/ci049885e.

World Bank. 2017. Drug-Resistant Infections: A Threat to Our Economic Future. https://www.worldbank.org/en/topic/health/publication/drug-resistant-infections-a-threat-to-our-economic-future. (accessed04.07.2024).

World Health Organization (WHO) 2023. Antimicrobial resistance. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance (accessed04.07.2024).

Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya, AV, et al. 2019. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature biotechnology, 37(9), 1038-40.doi:10.1038/s41587-019-0224-x

Otkrivanje novih antibiotika korištenjem bioinformatike i metoda mašinskog učenja

SAŽETAK

Otpornost na antibiotike predstavlja ozbiljnu globalnu zdravstvenu prijetnju koja svake godine uzrokuje oko 1,27 miliona smrtnih slučajeva, a smatra se da će taj broj do 2050. godine doseći 10 miliona. Razvoj novih antibiotika je izrazito zahtjevan proces koji obično traje 10-15 godina uz investiciju od oko 1,5 milijarde američkih dolara. U ovom procesu genomske i metagenomske analize igraju odlučujuću ulogu u otkrivanju genetskog potencijala mikroorganizama koji se ne mogu kultivirati, kao i u identificiranju novih mikroorganizama koji proizvode antibiotike. Osim toga, modeli dubokog učenja analiziraju molekularne strukture s ciljem identifikacije novih spojeva s antibakterijskom aktivnošću, dok virtualne tehnike skrininga analiziraju velike molekularne baze podataka s ciljem određivanja potencijalno aktivnih spojeva. Dokazano je da modeli razvijeni korištenjem dubokog učenja mogu predvidjeti genske klastere za biosintezu antibiotika s preciznošću od preko 90%. Osim ovakvih pristupa, identifikacija antibiotskih kombinacija i predviđanje sinergističkih učinaka omogućavaju razvoj efektivnijih terapijskih strategija u borbi protiv multirezistentnosti na lijekove. Ovakve metode doprinose razvoju proaktivnih pristupa upravljanju antibiotskom rezistencijom i optimiziranju otkrivanja novih antibiotika i učinkovitijoj primjeni postojećih. Ovaj rad ispituje otkrivanje novih antibiotika korištenjem bioinformatike i metoda mašinskog učenja.

Ključne riječi: Bioinformatika, mašinsko učenje, otkrivanje novih antibiotika