

## Review study

# Predictive Computational Approaches in Pharmaceutical Microbiology: Machine Learning and *In Silico* Integration: A Review Study

Alhadi Wajiej<sup>id</sup>, Safinaz Aburagaegah\*<sup>id</sup>

Department of Microbiology and Immunology, Faculty of Pharmacy, University of Elmergib, Al-Khoms City, Libya

Corresponding email. [safinazjuma@elmergib.edu.ly](mailto:safinazjuma@elmergib.edu.ly)**Abstract**

Employing *in-silico* programs and Machine Learning (ML) in microbiology, making computational microbiology (dry lab), is changing the way the data is investigated, visualized, and used. Complicated biological datasets are increasing over time due to discoveries. ML models such as Decision Trees (DT) and Artificial Neural Networks (ANN) have predictive algorithms for many applications in microbiology, such as antimicrobial resistance (AMR), pathogen detection, and microbiome characterization. Moreover, ML can help in identifying biomarkers, enhancing diagnosis, and visualizing microbial interactions. Although this technology has significant advantages, it still presents some issues, including ethical and legal concerns, missing analyzed data, and the model's efficiency. Besides ML models, *in-silico* programs such as digital plate reading (DPR), statistical, and genomic comparison programs improve the processing and visualization of the data. Regardless of the challenges, this technology is always progressing, and the increased data accessibility is used to support dependable AI microbiological applications and public health.

**Keywords.** Pharmaceutical Microbiology, Computational Biology, *In-Silico*, Machine Learning.

**Introduction**

Microorganisms are among the most primitive organisms to exist on Earth, having first appeared about 3.5 billion years ago [1]. In a microbial community, several species influence one another's development to create symbiosis. These ecosystem networks control the structural and functional characteristics of a community [2]. Conventional microbiological tests typically provide qualitative or semi-quantitative documentation, needing a high level of skill and a longer time for interpreting and analyzing the results, and until laboratory tests can be conducted, predictions based on biological data might not be confirmed [3]. A fundamental change toward algorithmic methodologies for effective data interpretation is required as we enter this era due to the enormous amount of biological data obtained from high-throughput technology [4]. In microbiology, machine learning (ML) is an important subject under artificial intelligence (AI), and it is one of the most rapidly increasing methods. [2]. Sophisticated algorithms are used in machine learning to allow the examination of large, complicated datasets for pattern detection and knowledge discovery [5]. The best way to analyze and comprehend this vast and increasingly complicated data, sometimes referred to as big data, frequently surpasses human capabilities and calls for the deployment of tools like ML [6]. The growing number of ML applications makes it impossible to even quantify them. Growing applications are an example of ML's significance to microbiology. The main goal of machine learning in microbiology is to identify the model that best matches the data. To forecast future data, we fit a model to the historical data. Additionally, unsupervised machine learning algorithms aim to uncover hidden connections between data and structural objects through similarities or contrasts [7]. In this mini-review, we discussed the potential of machine learning, including topics such as pathology, rapid diagnostics, microbiome research, precision medicine, and minimizing antibiotic resistance (AMR). This could transform our understanding of microbial ecosystems, diagnosis, and potentially treatment. It will involve a thorough review of recent literature, as well as the identification and assessment of machine learning techniques employed in computational microbiological research.

**Types of Machine Learning Algorithms Used**

The two primary learning modes in machine learning are predictive, which is supervised learning, that uses training data to anticipate future events, and descriptive is unsupervised learning, which is experimental and lacks data for training, a target, or a result [8]. Additionally, to machine learning, there are deep learning (DL), and reinforcement learning (RL). Every field of microbiological studies, including bacteriology, parasitology, virology, and mycology, has employed machine learning [2]. ML algorithms can generate unique rules by analyzing the data, in contrast to expert systems that are dependent on programming a set of rules [6]. Black-box models are models that don't explicitly explain how or why they make a certain prediction, yet they produce the most accurate machine-learning predictions. On the other hand, predictions that are made by the white-box models are significantly less accurate but simpler to understand [9]. As it was mentioned before, there are supervised ML and unsupervised ML algorithms, for example, types of algorithms used in supervised ML are Decision Trees (DT), ridge regression, eXtreme Gradient Boosting (XGBoost), Artificial Neural Networks (ANN), Convolutional Neural Network (CNN), Logistic Regression (LR), Naive Bayes (NB), K-Nearest Neighbors (K-NN), Support Vector Machines (SVM) and, random forests are

used for predictive modeling. As for unsupervised ML: Clustering, dimensionality reduction, and association analysis [10,11]. Standardization methods are used for preprocessing, while LASSO is used for single feature selection, and Schipper Egli Sandoz (SES) is used for multiple feature selection. For model assessment, there is Generalized cross-validation used for Performance estimation, Grid search with heuristics is used for configuration space search, lastly, BBC-CV is used for performance correction [12]. The majority of unsupervised machine learning applications in microbiology use clustering or dimensionality reduction/ordination techniques. A popular unsupervised learning method for spotting patterns in high-dimensional data is clustering, which finds data clusters with the most closely related data in each cluster [2]. As for supervised learning, it is the most used in microbial research. For instance, a microbiologist can label photomicrographs as they can be holding or not holding a particular microbe of interest. The images can then be used as a source to create a machine-learning algorithm that can identify the microbe [13].

### **In-Silico Platforms for Microbial Prediction**

It quickly became clear that true in silico experiments and using computers should be developed, starting with data analysis and storage, as in vivo and in vitro experiments were the norm. [14]. For microbiologists and other scientific researchers, the increasing need for digital microbiological data offers an exceptional opportunity [15]. The development of digital microbiology could have a big impact on monitoring infections and public health [16]. As clinical microbiology labs continue to move toward increased automation, digital plate reading (DPR) computer vision software is used to provide the next efficiency boost and is already making progress in these labs [17]. The statistical test ANOVA was applied to evaluate the differences in *A. hydrophila* bacterial shelf life and microbiological load in sardines stored at various temperatures [18]. Additionally, comparison methods are essential for determining the microbial phylogenomics of pathogenic isolates and creating complex networks that illustrate how different patients spread during epidemics [19].

### **Role of Machine Learning in Microbial Prediction**

#### **Predicting Antimicrobial Resistance**

Microbiology provides the potential answers for the prevention and treatment of diseases. Antimicrobial resistance (AMR), which develops when microbes adapt to decrease or completely eradicate the effectiveness of antimicrobial medications, is one of the biggest healthcare issues [2]. Antibiotic-resistant bacteria are the primary cause of antibiotic resistance, which occurs due to the consumption of antibiotics, causing bacteria to change their genetic makeup [20]. With little assistance from humans, AI increased the precision of predicting antimicrobial resistance (AMR) trends and enhanced the overall effectiveness of treatments [3]. ML and AI are essential to the healthcare system because they protect individuals against multi-resistant microbes and antimicrobial resistance (AMR) while also promoting the health of plants, animals, and ecosystems [21]. Additionally, combining AI with technologies like matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has benefited AMR profiling and microbiological identification, providing quick and affordable solutions [22,23]. An ML-based model was developed, which is a diagnostic model that was based on a random forest algorithm. The 44 criteria available at the time of admission to the hospital were taken into consideration. The model anticipated a 30% decrease in the use of antibiotics in cases of nonpathogenic SIRS and took into consideration the early detection of all sepsis illnesses [24]. Therefore, AI and ML could be used as a tool to fight high AMR rates [11]. Additionally, ML has been used to predict AMR in organisms such as nontyphoidal *Salmonella* and *M. tuberculosis*, with known resistance pathways [25]. Instead of accurately predicting resistance, the study's objective was to extract important insights from AMR genomic data by using machine learning (ML) models to identify known AMR-implicated genes of *M. tuberculosis* and to identify new acknowledged AMR determinants and resolve interactions driving AMR evolution.[26]. Nevertheless, any mistake made in any of the previously described phases leads to incorrect identification and diagnosis, which is then linked to a poor prognosis, unsuccessful treatments, and the development of antimicrobial resistance [27].

### **Pathogen Detection and Classification**

Infectious agent identification is a multilevel procedure that requires careful sample handling and collection before appropriate laboratory culturing and staining methods, imaging, and molecular analysis [28]. Predicting disease outbreaks and accurately identifying and characterizing pathogens are essential stages in resisting the threat [2]. AI can anticipate the diagnosis and prognosis of diseases and assist in the creation of customized treatment plans. It can help identify epidemics and develop efficient infection control measures, which can greatly aid in infection prevention and control [29,4]. The research studies involve viruses, bacteria, protozoa, fungi, and helminths in the fields of epidemiology, clinical applications, antibiotic resistance, medicine and vaccine development, climate change, microorganisms in plants, microbiomes, and taxonomy [2]. Moreover, to enhance the automated recovery and prediction of bacteriophages, several computational methods based on machine-learning techniques have been created [5]. Annotation and sequence signature data from previously found bacteriophages are used by MARVEL (Metagenomic Analysis and Retrieval of Viral Elements), an ML-based algorithm, to identify double-stranded bacteriophages in metagenomic bins [30]. Fourier-transform infrared spectroscopy (FTIR) features were combined with

machine learning (ML) approaches to demonstrate the viability of tracking microbes in foods under variable storage conditions [10].

In comparison with bacteriological examination, specimens for Mycobacteria Identification with the help of Deep Learning in Pathology improve accuracy in early pulmonary tuberculosis diagnosis [31]. In addition, it was created and verified by researchers a DL algorithm to check for mycobacteria in sections of tissue on digital acid-fast-stained (AFS) slides [32]. To anticipate and view bacterial infection, a study developed a deep neural network called AI-based bacterial infection detection using whole Slide Images, which uses hematoxylin and eosin-stained pathology slides. This is the first AI-based model to examine bacterial infections in pathology images [33]. Many pathology labs have established digital pathology systems that use artificial intelligence (AI) software to enhance the detection of pathogens on cytological and histological slides [34]. When it comes to plants, machine learning (ML) helps with early field-level identification of plant viral infections by evaluating hyperspectral images. ML can also help us predict the subcellular location as well as the structure of the viral proteins. Finding inhibitors of these effector proteins may be made easier with the correct structural prediction of virus-encoded proteins [35].

### **Microbiome Profiling and Interaction Networks**

Microbiomes, or communities of microorganisms inhabiting a host organism's ecosystems, are capable of communicating with their host in complicated and meaningful ways that affect the host's behavior, immunological response, metabolism, and digestion [36]. Identifying the distinct qualities of the microbiome may help to recognize and identify the disease-associated microbiome because a healthy microbiome can be detected through a number of shared traits with non-healthy people [37]. The sophisticated structure of microbiome data shows that species may engage in a variety of interactions with one another, such as rivalry, parasitism, commensalism, and mutualism [12]. Microorganisms and their genomes, commonly referred to as the microbiome, are the subject of increased investigation due to advancements in metagenomics and genome sequencing platforms [2]. In microbiome studies, the most frequent ML tasks that are used are those that deal with disease diagnosis, prediction, or treatment response [38]. Identifying diagnostic or predictive biomarkers in the microbiome—that is, the informative characteristics (genes, taxa, or functions) most closely linked to a disease, phenotype, environmental factor, or response to treatment—is possible with machine learning. [12].

Establishing a prediction algorithm, a performance evaluation process, a model selection protocol, and a performance optimization metric are all part of the modeling activity. The models demonstrated a high degree of accuracy in distinguishing between multiple kinds of cancer and between cancer and normal tissue, indicating that microbial signatures are specific to each type of cancer. However, they fell short in distinguishing between distinct stages of the same cancer type, indicating that microbial signatures may not be correlated with cancer phases for all types of cancer [39]. The suggested ML approach could potentially be used as a cancer diagnostic tool based on the microbiome. A demonstration of how to use RF to anticipate microbial interactions, knowing each community organism's characteristic identifiers [40]. Selecting a predicting algorithm that is naturally comprehensible, such as decision trees, logistic regression, or linear SVMs, is one method to guarantee interpretability. Understanding the connections between the microbiome features and the result is made easier by these models' intuitive link between the input and the output [12].

### **Challenges and Limitations**

Predictive ML models also can present some drawbacks, for instance, it may be difficult to detect and reduce bias in AI models used for microbiological applications due to possible bias in training data that affects results. Low-quality microbiological data availability and insufficiently structured datasets for machine learning model training [10]. Data management issues include gathering, ensuring quality, storing, and protecting data as a result of the digitization of microbiology diagnostic procedures [41]. In terms of AI and ML in microbiology, the future will bring more thorough discussions and possibly new rules around the privacy of data, ethical issues, and patient authorization [42]. Moreover, in fields like gene editing and synthetic biology, legislation and regulations are falling behind technological developments, thereby creating threats and uncertainty [10]. Many of these technologies need an external dataset to test their particular algorithms in a clinical setting, in addition to the limited case studies [35].

### **Conclusion and Future Perspectives**

Given the promise to improve research, diagnosis, and treatment, artificial intelligence and machine learning have been included in microbiology more and more. A critical viewpoint, however, highlights both the benefits and drawbacks of this integration. Scientists are visualizing a time when microbiological research and practice are heavily reliant on artificial intelligence and machine learning, necessitating the need for more readily available databases. For AI models to be reliable and generalizable, they need to be thoroughly validated using a variety of high-quality datasets [10]. Furthermore, the majority of ML, DL, and AI technologies used up to this point have been either supervised or semi-supervised, requiring specialized



professionals to name and analyze the results [35]. It is essential to create self-updating, flexible AI models that can operate in a variety of healthcare settings, assuring comfort and privacy.

**Conflict of interest.** Nil

## References

1. Nannipieri P, Ascher J, Ceccherini MT, Landi L, Pietramellara G, Renella G. Microbial diversity and soil functions. *Eur J Soil Sci.* 2003;54(4):655–670. doi:10.1046/j.1351-0754.2003.0556.x.
2. Goodswen SJ, Barratt JLN, Kennedy PJ, Kaufer A, Calarco L, Ellis JT. Machine learning and applications in microbiology. *FEMS Microbiol Rev.* 2021;45(5):fuab015. doi:10.1093/femsre/fuab015.
3. Khalaf WS, Morgan RN, Elkhatib WF. Clinical microbiology and artificial intelligence: Different applications, challenges, and future prospects. *J Microbiol Methods.* 2025;232–234:107125. doi:10.1016/j.mimet.2025.107125.
4. Tsitou VM, Rallis D, Tsekova M, Yanev N. Microbiology in the era of artificial intelligence: Transforming medical and pharmaceutical microbiology. *Biotechnol Biotechnol Equip.* 2024;38(1):2349587. doi:10.1080/13102818.2024.2349587.
5. Nami Y, Imeni N, Panahi B. Application of machine learning in bacteriophage research. *BMC Microbiol.* 2021;21(1):193. doi:10.1186/s12866-021-02256-5.
6. Peiffer-Smadja N, Delli  re S, Rodriguez C, Birgand G, Lescure FX, Fourati S, et al. Machine learning in the clinical microbiology laboratory: Has the time come for routine practice? *Clin Microbiol Infect.* 2020;26(10):1300–1309. doi:10.1016/j.cmi.2020.02.006.
7. Cieslak MC, Castelfranco AM, Roncalli V, Lenz PH, Hartline DK. Distributed Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis. *Mar Genomics.* 2020;51:100723. doi:10.1016/j.margen.2019.100723.
8. Mitchell TM. *Machine Learning.* New York: McGraw-Hill; 1997.
9. London AJ. Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Cent Rep.* 2019;49(1):15–21. doi:10.1002/hast.973.
10. Mohseni P, Ghorbani A. Exploring the synergy of artificial intelligence in microbiology: Advancements, challenges, and future prospects. *Comput Struct Biotechnol Rep.* 2024;1:100005. doi:10.1016/j.csbr.2024.100005.
11. Rabaan AA, Alhumaid S, Mutair AA, Garout M, Abulhamayel Y, Halwani MA, et al. Application of Artificial Intelligence in Combating High Antimicrobial Resistance Rates. *Antibiotics.* 2022;11(6):784. doi:10.3390/antibiotics11060784.
12. Papoutsoglou G, Tarazona S, Lopes MB, Klammsteiner T, Ibrahimi E, Eckenberger J, et al. Machine learning approaches in microbiome research: Challenges and best practices. *Front Microbiol.* 2023;14:1261889. doi:10.3389/fmicb.2023.1261889.
13. Burns BL, Rhoads DD, Misra A. The Use of Machine Learning for Image Analysis Artificial Intelligence in Clinical Microbiology. *J Clin Microbiol.* 2023;61(9):e02336-21. doi:10.1128/jcm.02336-21.
14. Danchin A. In vivo, in vitro and in silico: An open space for the development of microbe-based applications of synthetic biology. *Microb Biotechnol.* 2022;15(1):42–64. doi:10.1111/1751-7915.13937.
15. Gruson D, Helleputte T, Rousseau P, Gruson D. Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation. *Clin Biochem.* 2019;69:1–7. doi:10.1016/j.clinbiochem.2019.04.013.
16. Basak SC, Majumdar S, Nandy A, Roy P, Dutta T, Vracko M, et al. Computer-Assisted and Data Driven Approaches for Surveillance, Drug Discovery, and Vaccine Design for the Zika Virus. *Pharmaceuticals.* 2019;12(4):157. doi:10.3390/ph12040157.
17. Smith KP, Wang H, Durant TJS, Mathison BA, Sharp SE, Kirby JE, et al. Applications of Artificial Intelligence in Clinical Microbiology Diagnostic Testing. *Clin Microbiol Newsl.* 2020;42(8):61–70. doi:10.1016/j.clinmicnews.2020.03.006.
18. Bulat FN, Kili   B, Atalay SD. Microbial ecology of different sardine parts stored at different temperatures and the development of prediction models. *Food Biosci.* 2020;38:100770. doi:10.1016/j.fbio.2020.100770.
19. Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, et al. Whole-Genome Sequencing of Bacterial Pathogens: The Future of Nosocomial Outbreak Analysis. *Clin Microbiol Rev.* 2017;30(4):1015–1063. doi:10.1128/cmr.00016-17.
20. Ahmed N, Khan M, Saleem W, Karobari MI, Mohamed RN, Heboyan A, et al. Evaluation of Bi-Lateral Co-Infections and Antibiotic Resistance Rates among COVID-19 Patients. *Antibiotics.* 2022;11(2):276. doi:10.3390/antibiotics11020276.
21. Pinto-de-S   R, Sousa-Pinto B, Costa-de-Oliveira S. Brave New World of Artificial Intelligence: Its Use in Antimicrobial Stewardship—A Systematic Review. *Antibiotics.* 2024;13(4):307. doi:10.3390/antibiotics13040307.
22. Signoroni A, Ferrari A, Lombardi S, Savardi M, Fontana S, Culbreath K. Hierarchical AI enables global interpretation of culture plates in the era of digital microbiology. *Nat Commun.* 2023;14(1):6874. doi:10.1038/s41467-023-42563-1.
23. Wilson S, Steele S, Adeli K. Innovative technological advancements in laboratory medicine: Predicting the lab of the future. *Biotechnol Biotechnol Equip.* 2022;36(S1):S9–S21. doi:10.1080/13102818.2021.2011413.

24. Rodríguez-González A, Zanin M, Menasalvas-Ruiz E. Public Health and Epidemiology Informatics: Can Artificial Intelligence Help Future Global Challenges? An Overview of Antimicrobial Resistance and Impact of Climate Change in Disease Epidemiology. *Yearb Med Inform.* 2019;28:224–231. doi:10.1055/s-0039-1677910.
25. Kim JI, Maguire F, Tsang KK, Gouliouris T, Peacock SJ, McAllister TA, et al. Machine Learning for Antimicrobial Resistance Prediction: Current Practice, Limitations, and Clinical Perspective. *Clin Microbiol Rev.* 2022;35(3):e00179–21. doi:10.1128/cmr.00179-21.
26. Kavvas ES, Catoiu E, Mih N, Yurkovich JT, Seif Y, Dillon N, et al. Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance. *Nat Commun.* 2018;9(1):4306. doi:10.1038/s41467-018-06634-y.
27. Smith K. Chapter 5—Artificial intelligence in clinical microbiology. In: Krittanawong C, editor. *Artificial Intelligence in Clinical Practice*. Academic Press; 2024. p. 47–50. doi:10.1016/B978-0-443-15688-5.00008-5.
28. Shelke YP, Badge AK, Bankar NJ, Shelke YP, Badge A, Bankar NJ. Applications of Artificial Intelligence in Microbial Diagnosis. *Cureus.* 2023;15(11):e49366. doi:10.7759/cureus.49366.
29. Karalis VD. The Integration of Artificial Intelligence into Clinical Practice. *Appl Biosci.* 2024;3(1):2. doi:10.3390/applbiosci3010002.
30. Amgarten D, Braga LPP, da Silva AM, Setubal JC. MARVEL, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins. *Front Genet.* 2018;9:304. doi:10.3389/fgene.2018.00304.
31. Zaizen Y, Kanahori Y, Ishijima S, Kitamura Y, Yoon HS, Ozasa M, et al. Deep-Learning-Aided Detection of Mycobacteria in Pathology Specimens Increases the Sensitivity in Early Diagnosis of Pulmonary Tuberculosis Compared with Bacteriology Tests. *Diagnostics.* 2022;12(3):709. doi:10.3390/diagnostics12030709.
32. Pantanowitz L, Wu U, Seigh L, LoPresti E, Yeh FC, Salgia P, et al. Artificial Intelligence–Based Screening for Mycobacteria in Whole-Slide Images of Tissue Samples. *Am J Clin Pathol.* 2021;156(1):117–128. doi:10.1093/ajcp/aqaa215.
33. Zhu M, Guo M, Liu CQ, Songyang Z, Dou WX, Xiong Y. Deep-learning model AIBISI predicts bacterial infection across cancer types based on pathological images. *Heliyon.* 2023;9(4):e15400. doi:10.1016/j.heliyon.2023.e15400.
34. Marletta S, L'Imperio V, Eccher A, Antonini P, Santonicco N, Girolami I, et al. Artificial intelligence-based tools applied to pathological diagnosis of microbiological diseases. *Pathol Res Pract.* 2023;243:154362. doi:10.1016/j.prp.2023.154362.
35. Ghosh D, Chakraborty S, Kodamana H, Chakraborty S. Application of machine learning in understanding plant virus pathogenesis: Trends and perspectives on emergence, diagnosis, host-virus interplay and management. *Virol J.* 2022;19(1):42. doi:10.1186/s12985-022-01767-5.
36. Marchesi JR, Ravel J. The vocabulary of microbiome research: A proposal. *Microbiome.* 2015;3(1):31. doi:10.1186/s40168-015-0094-5.
37. Behrouzi A, Nafari AH, Siadat SD. The significance of microbiome in personalized medicine. *Clin Transl Med.* 2019;8(1):16. doi:10.1186/s40169-019-0232-y.
38. Brouillette M. Cancer debugged. *Nat Biotechnol.* 2023;41(3):310–313. doi:10.1038/s41587-023-01677-z.
39. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Retracted: Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature.* 2020;579(7800):567–574. doi:10.1038/s41586-020-2095-1.
40. DiMucci D, Kon M, Segrè D. Machine Learning Reveals Missing Edges and Putative Interaction Mechanisms in Microbial Ecosystem Networks. *mSystems.* 2018;3(5):e00181–18. doi:10.1128/msystems.00181-18.
41. Henry KE, Kornfield R, Sridharan A, Linton RC, Groh C, Wang T, et al. Human-machine teaming is key to AI adoption: Clinicians' experiences with a deployed machine learning system. *NPJ Digit Med.* 2022;5(1):97. doi:10.1038/s41746-022-00597-7.
42. Miller S, Chiu C, Rodino KG, Miller MB. Point-Counterpoint: Should We Be Performing Metagenomic Next-Generation Sequencing for Infectious Disease Diagnosis in the Clinical Laboratory? *J Clin Microbiol.* 2020;58(3):e01739–19. doi:10.1128/jcm.01739-19.