American Academic Scientific Research Journal for Engineering, Technology, and Sciences
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402
https://asrjetsjournal.org/index.php/American_Scientific_Journal/index

# Application of Artificial Intelligence and Machine Learning in Drug Development for Infectious Diseases

Harmony Sunny Ka-Alenebari[a]*,Peter Oluwasayo Adigun[b], Paris MIkem Yambali[c],Michael A. Adekanye[d],Qurat-ul-Ain[e]

[a]*Department of Biological Science, New Mexico Highlands University, 1005 Diamond St, Las Vegas, New Mexico, USA*

[b]*Department of Computer Science, New Mexico Highlands University, 1005 Diamond St, Las Vegas, New Mexico, USA*

[c]*Department of Chemistry, New Mexico Highlands University, 1005 Diamond St, Las Vegas, New Mexico, USA*

[d]*Miba Group, Dr.-Mitterbauer-Str. 3; A-4663 Laakirchen, Austria*

[e]*Pir Mehr Ali Shah Arid Agriculture University Rawalpindi islamabad Shamsabad, Muree Road, Punjab 46000, Pakistan*

[a]*Email: hsunny@live.nmhu.edu*,[b]*Email: poadigun@nmhu.edu*,[c]*Email: mparis@live.nmhu.edu*

[d]*Email: adekanye_michael@yahoo.com*,[e]*Email: aineenwz@gmail.com*

**Abstract**

The persistent threat of infectious diseases and the growing challenge of antimicrobial resistance have validated the need for innovative approaches in drug discovery and development. This review explores the transformative role of artificial intelligence (AI), and machine learning (ML), in addressing these challenges across the drug development pipeline for infectious diseases. AI models and ML algorithms excel at integrating and analyzing diverse datasets, including genomic, proteomic, chemical, clinical, and epidemiological data, to accelerate target identification, optimize compound screening, and predict drug efficacy and safety. The review synthesizes recent advances, highlighting the application of supervised and deep learning models in pathogen identification, antimicrobial susceptibility testing, and the rational design of novel therapeutics. It also examines the use of generative AI for precision medicine, drug repurposing, and real-time outbreak prediction. Despite these advances, the review identifies ongoing challenges related to data quality, model interpretability, and ethical considerations. Positively, AI and ML stand as a catalyst for more efficient, cost-effective, and personalized therapeutic strategies, with the promise of significantly improving global public health outcomes.

***Keywords:*** Artificial Intelligence (AI); Machine Learning (ML); Deep Learning (DL); Drug Discovery; Infectious Diseases; Precision Medicine; Antimicrobial Resistance.

---

*Received: 6/1/2026*
*Accepted: 8/1/2026*
*Published: 8/10/2026*

---

* *Corresponding author.*

## 1. Introduction

Infectious diseases persist as a significant global health issue, particularly with the confrontation of the ongoing dual challenges of emerging pathogens and the escalating problem of antimicrobial resistance (AMR). Traditional drug discovery and development processes are competing to keep up with the pathological evolution, due to the complexity of pathogen evolution and the extensive chemical space of potential therapeutics [1, 2]. Recently, machine learning (ML) and artificial intelligence (AI) have emerged as transformative tools in tackling these challenges. They allow for the analysis of large, complex datasets and help identify patterns that inform every stage of the drug development process [3, 4].

AI and Machine learning algorithms (MLA) are particularly effective at handling various types of data, such as genomic sequences, chemical structures, clinical records, and epidemiological trends, which makes them especially useful for addressing infectious diseases [3, 5]. In the field of clinical microbiology, AI and ML are transforming the identification of pathogens and testing for antimicrobial susceptibility. For instance, models that are trained using mass spectrometry and genomic data can swiftly and accurately forecast resistance profiles, enhancing antibiotic stewardship and infection control efforts [2, 3].

AI and ML are significantly accelerating the discovery of new anti-infective agents beyond traditional diagnostics. Additionally, high-throughput screening methods are now being enhanced by ML-driven virtual screening and AI molecular modelling, which efficiently evaluate millions of candidate compounds in silico, predict their efficacy, and optimize their chemical properties [6, 7]. Generative AIs further facilitate the rational design of new molecules, potentially broadening the range of available therapeutics (Hanna and his colleagues, 2025). More vital applications of AI and ML are the roles they play in drug repurposing by identifying new uses for existing medications through the analysis of patterns in drug-target interactions and clinical outcomes [7]. In public health, AI and ML are employed to forecast outbreaks, track disease spread, and inform resource allocation during epidemics [8, 9].

Despite these advancements, several challenges remain. Issues related to data quality, model interpretability, and ethical considerations are critical to ensuring the safe and effective integration of ML into clinical and research settings [1, 2]. Continued research and interdisciplinary collaboration are essential to fully realize the potential of ML in combating infectious diseases and developing the next generation of anti-infective drugs (Winkler, 2022).

Hence, this review assesses the application of AI and ML in drug development for infectious diseases. The objectives are:

i. To review types of AI models and ML algorithms primarily applied in the development of drugs for infectious diseases.
ii. To review areas and types of infectious diseases in which AI and ML have been applied.
iii. To draw out the future application of AI and ML for the development of drugs for infectious diseases.

## 2. Narrative Review Approach and Method

This narrative review was conducted to synthesize current evidence on the application of AI and ML applications in the development of drugs for infectious diseases. The review followed the SANRA (Scale for the Assessment of Narrative Review Articles) framework to ensure methodological standards and validity. A total of 15 peer-reviewed articles published between 2020 and 2025 were selected based on relevance, diversity of perspectives, and contribution to the intersection of AI and ML toward health interventions and drug development for infectious diseases. These sources included scoping reviews, experimental studies, meta-analyses, and conceptual papers drawn from journals ranked Q1-Q2 such as *Annals of Medicine, Antimicrobials and Resistance, Briefings in Bioinformatics, Frontiers in Chemistry, Frontiers in Cellular and Infection Microbiology, Clinical Microbiology and Infection, Pharmaceuticals, Scientific Reports*, etc.

The literature was identified through a purposive sampling strategy, targeting publications that addressed AI-based models, ML algorithms for addressing issues for drug development, and diverse datasets, including genomic, proteomic, chemical, clinical, and epidemiological data, to accelerate target identification, optimize compound screening, and predict drug efficacy and safety. No formal database search protocol was applied, as the goal was to capture a broad conceptual landscape rather than empirical outcomes. Each article was reviewed for its contribution to one or more of the following thematic domains: (1) types of AI models and ML algorithms, (2) type of infectious disease, (3) pharmacological impact and clinical efficacy of AI models and ML algorithms, (3) ethical and privacy considerations and implications, and (4) future directions for AI and ML in drug development for infectious diseases.

Scientific reasoning was applied to compare findings across studies, highlight methodological strengths and limitations, and identify gaps in current knowledge. Also, referencing was conducted using the IEEE style, and all sources were critically appraised for relevance and credibility.

## 3. Applications of AI and ML in Drug Development for Infectious Diseases

Machine learning (ML) and artificial intelligence (AI) have significantly transformed the fields of infectious disease research and drug development. They offer innovative solutions to ongoing challenges, such as antimicrobial resistance, the emergence of new pathogens, and the demand for faster and more effective therapeutic interventions [1-4]. Al and ML algorithms are built to learn from vast and complex datasets, which play a crucial role in analyzing genomic, proteomic, clinical, and epidemiological data. These models extract patterns, insights, and facts that inform every stage of the drug development process [3, 5, 6].

Cesaro and colleagues conducted an overview of the obstacles and uses of AI in infectious diseases and resistance to antimicrobials [2]. The authors conduct a thorough examination of various AI architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) models, emphasizing their usefulness in both clinical and laboratory settings. The review demonstrates how AI accelerates pathogen detection by analyzing images of Gram stains, interpreting mass spectrometry data, and conducting genomic sequencing, which allows for the rapid identification of resistant strains. Additionally, Cesaro

and his colleagues explore the application of AI in predicting antimicrobial susceptibility, an essential factor for determining effective treatment plans. They highlight real-world applications such as predicting sepsis in intensive care units, showcasing how AI models can anticipate patient deterioration and facilitate timely interventions. The authors also address challenges related to data quality, model transparency, and ethical considerations, advocating for the development of explainable AI systems that clinicians can rely on.

Hanna and colleagues examined the future of AI and ML in transforming trends in pathology and the field of medicine [10]. They observed that health care organizations are beginning to establish management strategies for integrating AI and ML toolsets. These platforms harness the computational power of advanced algorithms to analyze data, leading to better insights that enhance clinical decision-making and improve patient outcomes. They noted that the implementation of AI and ML technologies in pathology and medicine is reshaping the field by offering innovative solutions that enhance diagnostic accuracy, improve operational workflows, and support clinical decision-making. Hanna and colleagues noted that these tools are increasingly valuable in pathology research, contributing to automated image analysis, biomarker discovery, drug development, clinical trials, and predictive analytics. They also identified several related trends, including the adoption of ML operations to manage models in clinical settings, the application of multimodal and multiagent AI to utilize diverse data sources, expedited translational research, and virtual education for training and simulation. The authors emphasized that future deployments of AI in health care will likely be subject to evolving regulations as risks to patients become clearer.

Shiwlani and colleagues carried out a comprehensive review on the interpretation of immune biomarkers by generative artificial intelligence (GAI) regarding precision medicine [11]. The study focused on GAI applications, which include machine learning (ML), generative adversarial networks (GANs), transformers, and large language models (LLMs). The findings indicate that generative AI significantly enhances real-time decision-making in the customization of diagnostics and therapies, particularly for autoimmune and infectious diseases such as lupus, rheumatoid arthritis, sepsis, and COVID-19. However, the review also points out several challenges, including data inconsistency, ethical concerns, and limited interpretability of AI. The authors stressed the urgent need for improved data standardization, interdisciplinary collaboration, and systematic integration of AI to leverage the potential of generative AI in precision medicine fully. [11].

Vaghasiya and colleagues performed a meta-analysis to study the role of AI and ML in managing health, biotechnology, and vaccine projects [12]. Their analysis included 44 peer-reviewed studies found in databases like PubMed, IEEE Xplore, Scopus, Web of Science, EMBASE, and Google Scholar. They discovered that using AI and ML significantly enhanced planning, optimized resources, managed risks, and relied on data in decision-making. Also, AI and ML played a key role in selecting vaccines for trials and enhanced the speed of improvements in both trial optimization and safety reviews. Their analyses showed that the biggest effect sizes were found in controlling infectious diseases (1.2). Their study found that key decision-making for colon cancer treatment involves as many as 95% CI: 0.85–1.50) of patients, followed by those with medical imaging (85%; 95% CI: 0.75–0.95). Studies that use AI-based approaches reported a combined effect size of 0.83 (95% CI: 0.78–1.08). Even with a lot of variability among the studies ($I^2 = 99.04\%$) and moderate-to-high risks of bias, the main outcomes remained robust after sensitivity analyses.

Al Meslamani and colleagues examined the growing role of AI, ML, and big data analytics in the global fight against infectious diseases [1]. The authors emphasized the potential of ML algorithms to analyze complex and diverse datasets for various applications, including outbreak prediction, pathogen identification, drug discovery, and personalized medicine. They also addressed key limitations that impede the widespread implementation of these technologies, such as data quality issues, integration complexities, and the interpretability of ML outcomes. To overcome these challenges, the authors proposed strategic uses of ML for identifying biomolecular targets, developing effective therapeutics, and utilizing vaccinomics, a method that leverages big data analytics to discover and characterize protective vaccine antigens. Furthermore, they highlighted the importance of preparing for catastrophic evolutionary events. By using intelligent analytical tools, they aim to proactively reduce infection risks.

Kwofie and his colleagues established a perspective on the use of AI, ML, and Big Data in the discovery of drugs for the Ebola virus [7]. They seek to provide insights that complement existing research on the Ebola virus (EBOV) by discussing the role of ML techniques in predicting small-molecule inhibitors of EBOV. They noted that various ML algorithms, including Bayesian, support vector machines, and random forests, have been employed to predict anti-EBOV compounds, demonstrating strong models with credible outcomes. Kwofie and colleagues observed that the use of deep learning (DL) models for predicting anti-EBOV molecules is still underutilized. Therefore, they explored how such models could be harnessed to develop fast, efficient, robust, and innovative algorithms to assist in the discovery of anti-EBOV drugs. Additionally, they highlighted deep neural networks as a promising ML algorithm for predicting anti-EBOV compounds. They also summarized the extensive data sources required for ML predictions, presenting high-dimensional data systematically and comprehensively. They reported that with ongoing efforts to eradicate Ebola virus disease (EVD), the application of AI and ML in EBOV drug discovery research will facilitate data-driven decision-making and may help reduce the high attrition rates of compounds in the drug development pipeline. [7]

Parvatikar and his colleagues examined the use of AI along with ML techniques for examining extensive databases and facilitating drug discovery [13]. According to the authors, finding drugs for diseases became challenging during outbreaks due to rapid changes in the genetic code of infectious organisms. There is an emphasis on virtual screening as it is often used to find promising drug targets in large chemical collections. It explained that integrating deep learning and neural network algorithms changed the field, allowing for achievements in multiple drug discovery steps. Examples included the creation of peptides by synthesis, organizing databases by ligands and structures, checking for toxicity, following drugs throughout their lifecycle, timed gradual drug release, creating models for drug activity, studying relationships between structures and how they affect drugs, repurposing existing drugs, examining polypharmacology effects and studying the physical and chemical properties of drugs. They concluded that the identification and optimization of lead compounds from large chemical libraries is a very crucial step in drug discovery, which is very time-consuming and involves a lot of investment. Also, the lead compound should possess all possible pharmacodynamic and pharmacokinetic properties and should have high-affinity binding and specificity for a target associated with a disease, which can be used as a drug target [13].Hu and associates presented a comprehensive review of ML applications in the study and management of protozoal pathogens and related infectious diseases [8]. Their work emphasizes how ML has evolved into an essential tool not only for academic research but also for practical applications in the real world.

This includes areas such as pathogen detection, public health surveillance, analysis of host-pathogen interactions, and the discovery of drug and vaccine candidates. The authors focus specifically on protozoal pathogens, including *Plasmodium*, *Trypanosoma*, *Toxoplasma*, *Cryptosporidium*, and *Giardia*, which pose significant threats to both human and animal health. With the increasing availability of high-throughput biological data and decreasing computational costs, integrating ML into parasitology and infectious disease research has become more feasible and impactful. The authors provide foundational knowledge on ML workflows and commonly used algorithms, such as support vector machine (SVM), random forests (RF), and neural networks (NN), as well as methods for feature selection and model evaluation.

Mishra and his colleagues assessed the application of MLA in diagnosing a range of infectious diseases, highlighting their potential in enhancing early detection and containment [14]. The authors reviewed various supervised and ensemble ML techniques, including logistic regression, SVM, Naive Bayes, decision trees, RF, K-nearest neighbors, artificial neural networks (ANNs), and CNNs. They evaluated these models on datasets related to diseases such as tuberculosis (TB), influenza, HIV, dengue fever, COVID-19, cystitis, and nonspecific urethritis. Their findings indicated that ensemble approaches outperform individual classifiers by combining multiple models, which enhances diagnostic precision. However, the study also highlighted ongoing challenges in ML-based disease diagnosis, particularly regarding data availability, quality, heterogeneity, and temporal dynamics. The authors emphasize that developing reliable diagnostic models requires diverse, high-quality datasets that accurately reflect real-world clinical variability. Their review supports the growing consensus that when AI and ML are carefully integrated with domain-specific knowledge and a large data infrastructure, it will play a crucial role in improving diagnostics for infectious diseases and enhancing public health outcomes. [14]

Chiu and colleagues investigated the use of ML techniques to address the clinical and logistical challenges posed by emerging infectious diseases (EIDs), with a particular focus on optimizing patient triage and resource allocation during pandemic scenarios such as COVID-19 [5]. The methodology of this study focuses on two ML-based prediction models: decision tree (DT) models and deep neural network (DNN) models. The authors highlight that DT models appeal to clinicians because they provide clear decision rules. However, these models rely on univariate analysis and do not incorporate linear or non-linear transformations. In contrast, the framework of DNN models generally offers superior prediction performance compared to other model types. Conclusively, these models are particularly useful in situations where laboratory diagnostics may be inaccessible or limited, as they reliably identify low-risk cases based on clinical and demographic features alone. Overall, the study suggests that machine ML is a scalable and effective approach to support real-time decision-making and enhance preventive medical strategies during outbreaks of emerging infectious diseases (EID) [5].

Winkler outlines the increasing importance of ML in finding and creating drugs to fight tuberculosis (TB), which continues to be one of the deadliest infectious diseases in the world [4]. He established that early quantitative structure-activity relationship (QSAR) methods that assumed a simple linear relationship between the structure of a chemical and how active it is. Often, the first AI models were not versatile because they used small and limited kinds of data. Because of progress in computer science, ML in chemistry now relies more on nonlinear methods such as RFs, SVMs, NNs, and Bayesian classifiers that can model intricate, multi-dimensional connections in chemicals. One example is the use of Bayesian models to detect inhibitors of *Mycobacterium tuberculosis*

*topoisomerase I*, and another is the virtual search for possible inhibitors using CoMFA/COMSIA models from large compound libraries. Winkler points out that the growing amount of data, improved automation in reviewing, and the rise of open-source ML tools have greatly increased the effectiveness of these approaches. Even so, the review mentions several challenges, such as using classification more than regression, not testing enough, and relying on datasets without enough different types of chemicals.

Gaudelet and colleagues reported a detailed overview of how Graph machine learning (GML) is applied in drug development, highlighting its power in dealing with issues that are not addressed by traditional ML [3]. They assert that biological and chemical data, such as molecular shapes, interactions between proteins, and networks of genes, are best represented through graphs since they are relation-based. Graph neural networks (GNNs) are given attention as effective options for learning from complex data in chemistry and biology. Their review shows that GML can adapt to different types of data and has a positive impact on crucial drug discovery stages such as finding targets, optimizing leads, and repurposing drugs. The authors also highlighted developments related to making models more scalable and easier to understand, supported by explainable AI and large-scale graph architectures. Yet, their report points out issues, including variations in data accuracy, the restricted use of models in different applications, and the importance of having agreed-upon ways of benchmarking. Gaudelet and his colleagues believe that advances in each of these three areas are vital for leveraging GML to speed up and reduce risks in modern drug discovery [3].

Trans and co-researchers explored the utilization of AI and ML in infectious disease testing [15]. Their study showed that the ML models have demonstrated high levels of accuracy in discriminating COVID-19 cases based on MALDI–TOF–MS spectra, with a sensitivity and specificity ratio of 100% and 96%, respectively. They also observed the same level of improvements in testing efficiency in malaria, Lyme disease, hepatitis, meningitis, and sepsis. As reported by the authors, the use of supervised learning (SL) methods, such as support vector machines (SVMs), random forests (RFs), and neural networks (NNs), is frequently employed due to their ability to process complex clinical and laboratory data. Some of the mentioned tools for auto-machine learning (Auto-ML) also play a role in expediting the development of models by simplifying the processes of selecting and validating features.

Winkler explored the harnessing of AI and ML in the quest to uncover new treatments for neglected tropical diseases (NTDs) [6]. They emphasized that key technological innovations, which have advanced drug discovery in developed countries, have not been equally leveraged for NTDs. Factors such as unsustainable pharmaceutical business models, the high cost of drug development, limited patient access, and restricted availability of technologies for scientists in affected regions have contributed to this disparity. However, they noted that the growing adoption of computational approaches, particularly ML, is beginning to transform the landscape of drug discovery for NTDs. The increased availability of large-scale datasets facilitates this shift, improves predictive accuracy of ML models, reduces barriers to entry for researchers, and provides widespread access to open-source ML tools. Winkler provided a comprehensive overview of the current and projected applications of ML in modeling biological activity and identifying drug candidates for NTDs. They discussed the developmental trajectory of ML and other AI technologies in this context, along with existing obstacles and the potential for synergistic advancements to accelerate progress soon [6].

Peiffer-Smadja and his colleagues conducted a narrative review that maps the landscape of ML applications in clinical decision support systems (CDSS) for infectious diseases [9]. Drawing from 60 ML-CDSS studies, the review revealed a predominant focus on bacterial infections (62%), followed by viral infections (17%), tuberculosis (15%), and a smaller fraction (7%) addressing general infections. The systems served various clinical purposes, including infection diagnosis (33%), sepsis prediction and stratification (30%), treatment response forecasting (22%), and antimicrobial guidance, such as predicting resistance (7%), selecting regimens (5%), and optimizing antiretroviral combinations (3%). These tools were primarily developed for intensive care units (40%), infectious disease consultations (25%), hospital wards (20%), emergency departments (7%), primary care (5%), and antimicrobial stewardship settings (2%). Notably, 88% of the systems were trained using datasets from high-income countries, with only 12% incorporating data from low- and middle-income countries (LMICs), underscoring a critical disparity in data representation. Moreover, the evaluation of these ML-CDSS was mostly confined to internal performance metrics, like sensitivity and specificity, with a mere 5% tested in real-world clinical settings. The authors advocate for enhanced integration of diverse patient data from underrepresented healthcare contexts and more rigorous real-world validation to optimize the applicability and reliability of ML-CDSS in global infectious disease management [9].

**Table 1:** List of AI Models and ML Algorithms used for Drug Development of ID

| Author(s) | Type of Application | Area | Contribution |
|---|---|---|---|
| Al Meslamani and his colleagues (2024) | ML, Big Data analytics | Outbreak prediction, drug discovery, vaccinomics | Addressed key ML use cases and barriers in infectious disease management; proposed strategies like biomolecular target prediction and vaccinomics. |
| Cesaro and his colleagues (2025) | CNNs, RNNs, LSTM | Antimicrobial resistance, pathogen detection | Applied AI to image interpretation, genomic sequencing, and predicting antimicrobial susceptibility. |
| Chiu and his colleagues (2022) | Decision Trees, Deep Neural Networks (DNN) | Triage during pandemics | Demonstrated effectiveness of DNNs in predicting the severity of EIDs using clinical and demographic data. |
| Gaudelet and his colleagues (2021) | Graph ML (GNNs) | Target identification, drug repurposing | Highlighted how GML surpasses conventional ML by modeling relationships in biological data. |
| Hanna and his colleagues (2025) | AI/ML platforms, multimodal/multiagent AI | Pathology, diagnostic workflows | Showed how AI is revolutionizing diagnosis, biomarker discovery, and clinical trials; discussed MLOps and virtual education. |
| Hu and his colleagues (2022) | SVM, RF, neural networks | Protozoal infections (e.g., Plasmodium, Giardia) | Reviewed ML tools in detecting pathogens, host-pathogen interaction, and vaccine candidate identification. |
| Kwofie and his colleagues (2023) | Bayesian, SVM, RF, Deep Neural Networks | Ebola virus drug discovery | Identified promising ML tools for predicting small-molecule inhibitors; emphasized underutilized deep learning. |
| Mishra and his colleagues (2022) | Logistic regression, SVM, RF, CNN, ensemble methods | Diagnosis of TB, HIV, dengue, COVID-19, etc. | Found that ensemble methods outperform single classifiers; emphasized data quality issues. |
| Parvatikar and his colleagues | Deep learning, Neural networks, Virtual screening | Broad drug discovery (focus on antivirals) | Documented how AI/ML streamlined lead compound identification, repurposing, toxicity prediction, and pharmacodynamics modeling. |

| Author(s) | Type of Application | Area | Contribution |
|---|---|---|---|
| (2023) | | | |
| Peiffer-Smadja and his colleagues (2020) | ML-CDSS (SVM, RF, etc.) | Clinical decision support for ID | Reviewed 60 ML-CDSS systems across ICUs, primary care, and ID clinics; evaluated for infection diagnosis, sepsis prediction, etc. |
| Shiwlani and his colleagues (2025) | Generative AI (GANs, LLMs, transformers) | Immune biomarker analysis, EHR interpretation | Enhanced real-time risk prediction and treatment customization for autoimmune/infectious diseases. |
| Tran and his colleagues (2021) | Supervised ML (SVM, RF, neural networks); AutoML | Infectious disease testing (COVID-19, malaria, etc.) | Demonstrated high diagnostic accuracy using ML with MALDI–TOF–MS spectra and proposed data fusion with wearable/EMR systems. |
| Vaghasiya and his colleagues (2025) | Meta-analysis (AI/ML tools) | Project management in biotech and vaccine dev. | Found AI/ML improved planning, resource use, and trial speed; largest effects in infectious disease control. |
| Winkler (2021) | ML (QSAR, SVM, RF) | Neglected tropical diseases | Showed how ML opens new drug discovery paths for diseases lacking commercial incentives; stressed access to public ML tools. |
| Winkler (2022) | QSAR, Bayesian, SVM, RF | Tuberculosis drug discovery | Reviewed the evolution from linear QSAR to complex ML for anti-TB drug screening; emphasized data size and model applicability. |

## 4. Future Development of Machine Learning in Drug Development for Infectious Diseases

AI and ML have emerged as a driving force in pharmaceutical innovation, particularly in the domain of infectious disease drug discovery. Its ability to process high-dimensional biomedical data, predict compound efficacy, and uncover novel therapeutic targets is increasingly evident in recent literature. As infectious diseases continue to pose global threats, the future of drug development will likely be shaped by increasingly sophisticated AI and ML tools, integrated platforms, and interdisciplinary collaboration.

Recent contributions reflect a growing trend toward diverse ML techniques in infectious disease contexts. For instance, Kwofie and his colleagues employed Bayesian models, SVM, RFs, and DNNs in drug discovery for the Ebola virus. Their study emphasized ML's capacity to identify small-molecule inhibitors, and notably called

attention to the underutilization of DL in this field, an area that future research will expand upon to enhance prediction coverage and chemical diversity [7]. Similarly, Winkler examined tuberculosis drug discovery using QSAR models, Bayesian inference, SVM, and RF. This work traced a clear evolution from linear models to nonlinear ML techniques for anti-TB compound screening, validating the importance of larger, heterogeneous datasets for optimizing model performance and generalizability [4, 6].

Generative AI and graph-based models represent two of the most promising advancements in this field. Shiwlani and colleagues applied GANs, LLMs, and transformers to interpret immune biomarker data from electronic health records (EHRs). These tools enabled real-time risk prediction and treatment customization for autoimmune and infectious diseases. Going forward, generative AI can facilitate de novo drug design and dynamic patient-specific therapy planning, key tenets of precision medicine [11]. In parallel, Gaudelet and his colleagues introduced Graph Machine Learning (GML), particularly Graph Neural Networks (GNNs), in drug discovery. Their work demonstrated how graph-based representations outperform traditional ML models in capturing complex biological interactions during target identification and drug repurposing [3].

A recurring theme in the literature is the integration of ML with large-scale, heterogeneous datasets. Hu and his colleagues reviewed ML tools such as SVM, RF, and neural networks, used for protozoal infections, including malaria and Giardia [8]. They demonstrated the utility of ML in detecting pathogens, modeling host-pathogen interactions, and identifying vaccine candidates. Future efforts should aim to incorporate multi-omic data (genomics, proteomics, metabolomics) into ML workflows to provide a more holistic view of pathogenesis and therapeutic targets. Likewise, Mishra and associates emphasized ensemble methods, combinations of logistic regression, SVM, RF, and CNN as superior to single models in diagnosing diseases like tuberculosis, HIV, and dengue. Their findings also underscored the critical importance of data quality [14]. Thus, future development must not only refine algorithmic architectures but also improve the curation and interoperability of biomedical datasets.

As ML applications advance, concerns around transparency, ethics, and regulatory integration become more pressing. Hanna and his colleagues discussed the use of AI/ML platforms and multimodal/multiagent systems in pathology and diagnostics [10]. Their study highlighted the rise of machine learning operations (MLOps) and virtual diagnostics, stressing the importance of navigating regulatory and ethical frameworks. Al Meslamani and his colleagues similarly addressed ML's applications in outbreak prediction, drug discovery, and vaccinomics, and identified critical limitations such as poor model interpretability and data fragmentation [1]. Future systems will require robust auditing mechanisms and explainable AI methods to gain clinical and regulatory trust. One of ML's most promising applications lies in real-time outbreak response and strategic drug development. Chiu and his colleagues demonstrated the use of Decision Trees and Deep Neural Networks (DNNs) to support triage during pandemics, particularly when laboratory testing is delayed or unavailable [5]. Their work validates the potential of ML in creating scalable and adaptive triage systems for public health emergencies. Furthermore, Vaghasiya and his colleagues conducted a meta-analysis showing how AI and ML applied in project management improved planning, resource allocation, and clinical trial speed in vaccine development for emerging viral threats [12]. Their findings suggest that, in the future, ML will be critical not only in lab research but also in the broader orchestration of health innovation projects. Parvatikar and his colleagues reinforced this by documenting how deep learning

and neural networks have streamlined compound screening, repurposing, and toxicity prediction across a wide spectrum of antiviral drug targets [13].

The future of machine learning in drug development for infectious diseases is multidimensional and rapidly evolving. From compound screening and biomarker discovery to patient stratification and pandemic planning, ML is becoming indispensable across the therapeutic pipeline. Authors such as Kwofie and his colleagues, Shiwlani and his colleagues, Gaudelet and his colleagues, Winkler, and Vaghasiya and his colleagues have laid the groundwork by applying a variety of algorithms from classic supervised learning to graph-based and generative AI in different infectious diseases [3, 6, 7, 11, 12].

## 5. Conclusion

In conclusion, AI and ML is deemed to revolutionize the development of drugs for infectious diseases by integrating multi-omics data, explainable models, real-time clinical support, generative design, and robust ethical governance. By addressing the existing challenges related to data quality, interpretability, and equity, these advancements promise to deliver safer, faster, and more accessible therapies for global health challenges. Also, the incorporation of AI and ML into drug development represents a paradigm shift in the biomedical sciences, offering an expediting process, precision, and adaptability in discovering and optimizing new therapeutics.

ML-driven approaches have proven valuable in nearly every aspect of the drug development pipeline, ranging from target identification and compound screening to resistance prediction, clinical trial optimization, and personalized therapy [2,7]. These advancements have been enabled by the exponential increase in high-quality biomedical data and the development of sophisticated ML algorithms capable of extracting actionable insights from complex, multidimensional datasets [1,5].

## 6. List of Abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| AMR | Antimicrobial Resistance |
| Auto-ML | Auto-Machine Learning (Automated Machine Learning) |
| CDSS | Clinical Decision Support Systems |
| CNN | Convolutional Neural Network |
| COVID-19 | Coronavirus Disease 2019 |
| DT | Decision Trees |
| DL | Deep Learning |
| EID | Emerging Infectious Diseases |
| EMR | Electronic Medical Record |
| EHR | Electronic Health Record |
| GAN | Generative Adversarial Network |
| GAI | Generative Artificial Intelligence |
| GML | Graph Machine Learning |

| | |
|---|---|
| GNN | Generative Neural Network |
| LSTM | Long Short-Term Memory (a type of recurrent neural network) |
| EBOV | Ebola Virus |
| EVD | Ebola Virus Disease |
| GANs | Generative Adversarial Networks |
| HIV | Human Immunodeficiency Virus |
| LLMs | Large Language Models |
| LMICs | Low-Middle Income Countries |
| MALDI–TOF–MS | Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry |
| ML | Machine Learning |
| MLA | Machine Learning Algorithm |
| ML-CDSS | Machine Learning-based Clinical Decision Support Systems |
| MLOps | Machine Learning Operations |
| NN | Neural Network |
| NTDs | Neglected Tropical Diseases |
| RF | Random Forest |
| RNN | Recurrent Neural Network |
| SVM | Support Vector Machine |
| TB | Tuberculosis |

**References**

[1] A. Z. Al Meslamani, S. Isidro, and J. and de la Fuente, "Machine learning in infectious diseases: potential applications and limitations," *Annals of Medicine,* vol. 56, no. 1, p. 2362869, 2024/12/31 2024, doi: 10.1080/07853890.2024.2362869.

[2] A. Cesaro, S. C. Hoffman, P. Das, and C. de la Fuente-Nunez, "Challenges and applications of artificial intelligence in infectious diseases and antimicrobial resistance," *npj Antimicrobials and Resistance,* vol. 3, no. 1, p. 2, 2025/01/07 2025, doi: 10.1038/s44259-024-00068-x.

[3] T. Gaudelet *et al.*, "Utilizing graph machine learning within drug discovery and development," *Briefings in Bioinformatics,* vol. 22, no. 6, 2021, doi: 10.1093/bib/bbab159.

[4] D. A. Winkler, "The impact of machine learning on future tuberculosis drug discovery," *Expert Opinion on Drug Discovery,* vol. 17, no. 9, pp. 925–927, 2022/09/02 2022, doi: 10.1080/17460441.2022.2108785.

[5] H.-Y. R. Chiu *et al.*, "Machine learning for emerging infectious disease field responses," *Scientific Reports,* vol. 12, no. 1, p. 328, 2022/01/10 2022, doi: 10.1038/s41598-021-03687-w.

[6] D. A. Winkler, "Use of Artificial Intelligence and Machine Learning for Discovery of Drugs for Neglected Tropical Diseases," (in English), *Frontiers in Chemistry,* Review vol. Volume 9 - 2021, 2021–

March–15 2021, doi: 10.3389/fchem.2021.614073.

[7] S. K. Kwofie *et al.*, "Artificial Intelligence, Machine Learning, and Big Data for Ebola Virus Drug Discovery," *Pharmaceuticals*, vol. 16, no. 3, p. 332, 2023. [Online]. Available: https://www.mdpi.com/1424-8247/16/3/332.

[8] R.-S. Hu, A. E.-L. Hesham, and Q. Zou, "Machine Learning and Its Applications for Protozoal Pathogens and Protozoal Infectious Diseases," (in English), *Frontiers in Cellular and Infection Microbiology*, Review vol. Volume 12 - 2022, 2022–April–28 2022, doi: 10.3389/fcimb.2022.882995.

[9] N. Peiffer-Smadja *et al.*, "Machine learning for clinical decision support in infectious diseases: a narrative review of current applications," *Clinical Microbiology and Infection*, vol. 26, no. 5, pp. 584–595, 2020/05/01/ 2020, doi: https://doi.org/10.1016/j.cmi.2019.09.009.

[10] M. G. Hanna *et al.*, "Future of Artificial Intelligence—Machine Learning Trends in Pathology and Medicine," *Modern Pathology*, vol. 38, no. 4, p. 100705, 2025/04/01/ 2025, doi: https://doi.org/10.1016/j.modpat.2025.100705.

[11] A. Shiwlani, S. Kumar, and H. A. Qureshi, "Leveraging Generative AI for Precision Medicine: Interpreting Immune Biomarker Data from EHRs in Autoimmune and Infectious Diseases," *Annals of Human and Social Sciences*, vol. 6, no. 1, pp. 244–260, 02/20 2025, doi: 10.35484/ahss.2025(6-I)22.

[12] J. Vaghasiya, M. Khan, and T. Milan Bakhda, "A meta-analysis of AI and machine learning in project management: Optimizing vaccine development for emerging viral threats in biotechnology," *International Journal of Medical Informatics*, vol. 195, p. 105768, 2025/03/01/ 2025, doi: https://doi.org/10.1016/j.ijmedinf.2024.105768.

[13] P. P. Parvatikar *et al.*, "Artificial intelligence: Machine learning approach for screening large database and drug discovery," *Antiviral Research*, vol. 220, p. 105740, 2023/12/01/ 2023, doi: https://doi.org/10.1016/j.antiviral.2023.105740.

[14] S. Mishra, R. Kumar, S. K. Tiwari, and P. Ranjan, "Machine learning approaches in the diagnosis of infectious diseases: a review," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3509–3520, 2022.

[15] N. K. Tran *et al.*, "Evolving Applications of Artificial Intelligence and Machine Learning in Infectious Diseases Testing," *Clinical Chemistry*, vol. 68, no. 1, pp. 125–133, 2021, doi: 10.1093/clinchem/hvab239.