OPEN ACCESS

KJMR

Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 7 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

REVOLUTIONIZING DRUG DISCOVERY: THE TRANSFORMATIVE ROLE OF ARTIFICIAL INTELLIGENCE IN MODERN MEDICINE

[1]Khubaib Ahmad, [2]Aima Ali, [3]Muneeba Siddique, [4]Ahsan Ali, [5]Iqra Jamil*

[1, 3]Student, BS Medical Laboratory Technology, University of Central Punjab, Lahore, Pakistan.

[2]BS MLT student, University of Central Punjab, Pakistan.

[4]Lecturer, University Institute of Medical Laboratory Technology, University of Lahore, Pakistan.

[5]Lecturer, Department of Microbiology, University of Central Punjab, Lahore.

Corresponding author: iqrajameel@hotmail.com

Article Info

Abstract

Application of artificial intelligence (AI), particularly machine learning (ML), in medicine has significantly advanced drug discovery. When it comes to bridging the knowledge gap between possible therapeutic molecules and illness understanding, artificial intelligence is a potent catalyst. The most recent developments in artificial intelligence and its use in drug discovery are comprehensively summarized in this review. Starting with disease identification and covering diagnostics, target identification, screening, and lead discovery, different phases of the drug discovery process have been briefly discussed. In these phases, artificial intelligence's capacity to examine large datasets and identify trends is crucial, improving forecasts and productivity in the areas of disease detection, medication development, and clinical trial administration. It is made clear how AI may speed up drug development through the analysis of large amounts of data, which would cut down on the time and expense of introducing new drugs to the market. Significance of algorithm instruction, data quality, and ethical issues, particularly when managing information about patients in clinical trials is also of greater concern. AI has the potential to revolutionize medication development by taking these aspects into account, with significant benefits for both patients and society.

Keywords:

Artificial Intelligence; Machine Learning; Deep Learning; Support Vector Machine; Artificial Neural Networks; Computer-Aided Drug Design; Mass Spectrometry

INTRODUCTION

Each authorized drug can take up to 15 years and cost between $1 and $2 billion, making the traditional drug discovery process a difficult and expensive undertaking [1]. The main causes of this are longer clinical trial durations and increased attrition rates [2]. Even after progressing to the phase-I clinical trial, about 90% of possible treatment candidates fail, despite significant resource investment [3]. Pharmaceutical corporations and academic institutions alike view the advancement of a drug candidate to a phase-I clinical trial following thorough preclinical optimization as a major milestone [4]. To increase the success rate of lead drugs in clinical trials, extensive computational screening and docking have been used [5]. Nevertheless, these approaches have disadvantaged such inaccuracy and inefficiency [6]. Deep learning (DL) and machine learning (ML) algorithms, which are subsets of artificial intelligence, have been recognized as viable solutions to these problems. These artificial intelligence (AI) techniques can accurately forecast the features of macro systems at cheap computational costs. As a result, chemical and biological experts are increasingly using AI algorithms in the drug discovery process. Drug development makes substantial use of machine learning (ML), which uses methods like support vector machines (SVM), clustering, random forests (RF), Bayesian networks (BN), and deep learning (DL). Fig. 1 illustrates the broad classification of ML. Large volumes of data are processed and analyzed by DL models in tasks like bioactivity predictions, virtual screening (VS), and clinical imaging [7-8]. BNs forecast a patient's reaction to treatment as well as toxicity or bioactivity [9]. Clustering finds correlations or patterns in data, while RF models are utilized for feature selection and molecular target identification [10]. SVM is a form of supervised learning that is used to categorize data and has applications in toxicity prediction, VS, and pharmacological property prediction [11].

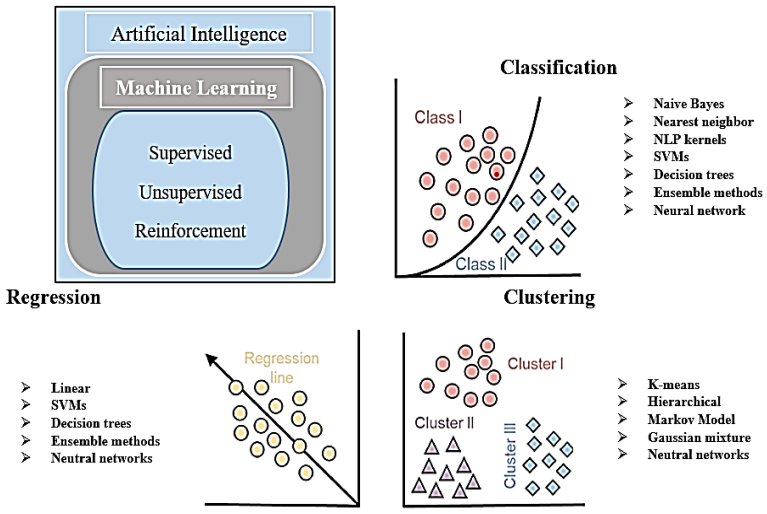

Figure 1: Specific varieties of artificial intelligence. Machine learning can be further classified into three primary categories: supervised, unsupervised, and reinforcement learning. These subcategories significantly expand the range of drug discovery applications and are essential to the architecture of machine learning algorithms.

The toxicity issues brought on by off-target interactions have been resolved and lead compounds from chemical libraries with over 106 million compounds may now be found more quickly thanks to machine learning techniques [12]. Additionally, ML-based tools like Alpha Fold have simplified the process of

estimating a target protein's three-dimensional structure, which is essential in the drug discovery process. Google's DeepMind recently created the AI-based tool Alpha Fold. AI has also been utilized by researchers to find new peptides for medicinal uses.AI is being used more and more to calculate the right dosage of drugs [13].

whereas the impediment in protein-protein interactions is predicted using machine and statistical learning techniques such as k-nearest neighbor (kNN), Naive Bayesian, SVM, ANN, decision trees (DT), and RF [14]. In computer-aided drug design (CADD), which entails selecting a promising therapeutic chemical from a pool of compounds, VS is an effective technique.

With a filtered dataset, machine learning (ML) may be applied to visual search (VS) using techniques like SVM, RF, and DT. Once the correctness of the trained model has been confirmed, it is applied to fresh data sets to find compounds with the appropriate activity against a target. Before starting clinical trials, the nominated drugs go through a number of bioassays and ADMET (absorption, distribution, metabolism, excretion, and toxicity) analyses. VS can be increased, boosted, and even have less false positives thanks to machine learning. For LBVS, the algorithms PARASHIFT, HEX, USR, and Shape have been developed. Several tools and models have been developed for both LBVS and structure-based virtual screening (SBVS) in recent years due to the rise of AI algorithms in the healthcare and pharmaceutical industries [15-16].

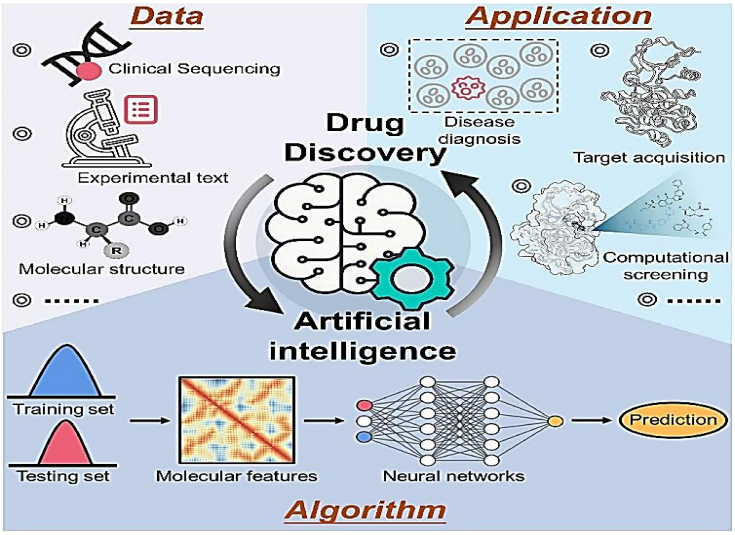

Figure 2: Schematic representation of the integration of artificial intelligence (AI) into the drug development process is presented in this picture. The diagram shows the progression of data collecting, which includes molecular structural analysis, experimental text, and clinical sequencing, to the use of AI algorithms and neural networks.

Disease Identification Using AI:

Artificial intelligence (AI) has demonstrated significant promise in the detection, diagnosis, and treatment of infectious diseases and non-communicable diseases (NCDs). AI technologies, such as machine learning (ML) and deep learning (DL), enable more accurate diagnostics, early disease detection, and personalized treatment plans. By leveraging vast amounts of data, AI can identify high-risk individuals, predict disease

spread, and improve patient outcomes. This article reviews key applications of AI in healthcare, particularly in the diagnosis of infectious diseases, HIV, diabetes, and diabetic retinopathy [17].

AI's role in infectious disease detection has been transformative, especially in the context of epidemics like COVID-19. During the pandemic, AI models were developed to analyze chest X-rays and CT scans for pneumonia caused by COVID-19, achieving up to 96% accuracy in categorization. For instance, Narin [18]. created a convolutional neural network (CNN)-based diagnostic model that analyzed X-ray images of healthy patients, COVID-19 patients, and patients with viral or bacterial pneumonia. These models were able to identify specific patterns associated with COVID-19-induced pneumonia, enhancing diagnostic accuracy [19]. In addition to COVID-19, AI has also shown potential in diagnosing other infectious diseases. For example, Taylor et al. published a study on urinary tract infections (UTIs) where six AI algorithms, including support vector machine (SVM) and random forest (RF), were developed to diagnose UTIs based on symptoms and laboratory results. The models achieved an area under the curve (AUC) of 0.88 and 0.90 for full and reduced datasets, respectively, with XGBoost algorithms outperforming other models in terms of accuracy [20]. This demonstrates AI's ability to reduce diagnostic errors and improve the detection of common infections.

AI has significantly advanced HIV detection, prevention, and counseling. Xiang et al. developed an ensemble technique that combined logistic regression (LR), RF, and graph convolutional networks (GCN) to identify individuals at high risk for HIV infection [21]. The model showed promising results, with the GCN+LR combination achieving 93.4% accuracy and an F1 score of 88.4%, and the GCN+RF combination reaching 96.6% accuracy and an F1 score of 94.6% [22]. These models can be used to predict individuals at risk for HIV and guide preventative measures. In addition to predictive models, AI has enabled the creation of virtual HIV counselors [23]. Heerden et al. developed a conversational agent using natural language processing (NLP) to guide users through HIV counseling and testing. This virtual assistant encourages users to communicate freely with the system, offering a more anonymous and confidential experience compared to previous agents [24]. Human evaluations confirmed the effectiveness of this system in improving user engagement and ensuring privacy [25].

AI has also made significant contributions to the diagnosis and management of diabetes, particularly in the detection of diabetic retinopathy, a common complication of diabetes that can lead to blindness [26]. Created a deep CNN model for retinal image analysis, achieving sensitivity of 97.5% and specificity of 93.4% in detecting diabetic retinopathy [27]. This model has shown superior performance compared to human ophthalmologists, offering a more efficient and accurate method for diagnosing this condition [28]. In 2018, the U.S. Food and Drug Administration (FDA) approved IDx-DR, the first AI-based medical device for the autonomous detection of diabetic retinopathy [29]. The device analyzes retinal images and classifies them as either "more than mild diabetic retinopathy detected: refer to an eye care professional" or "negative for more than mild diabetic retinopathy; rescreen in 12 months." This approval marked a significant milestone in AI's integration into medical practice [30]. Furthermore, AI has been applied to predict type 2 diabetes using speech recognition technology. Developed an interactive AI model that functions as a virtual doctor. Using the CMUS phinx open-source technology and the Vox Forge dataset, the model can recognize and synthesize speech, allowing it to interact with patients in German [31]. This

virtual doctor achieved an area under the curve (AUC) of 0.84 in predicting type-2 diabetes, demonstrating the potential of AI in diagnosing and managing chronic diseases [32].

AI is revolutionizing healthcare by enhancing the detection, diagnosis, and management of infectious and non-communicable diseases. Through the development of predictive models and diagnostic tools, AI is improving early disease detection, reducing diagnostic errors, and enabling personalized treatment options [33-34]. From HIV and diabetes to sepsis and cancer, AI technologies are transforming medical practice, leading to better patient outcomes and more efficient healthcare systems. However, further research and regulatory frameworks are needed to address challenges related to data quality, bias, and ethical considerations. As AI continues to evolve, it holds immense potential to improve healthcare delivery and outcomes worldwide.

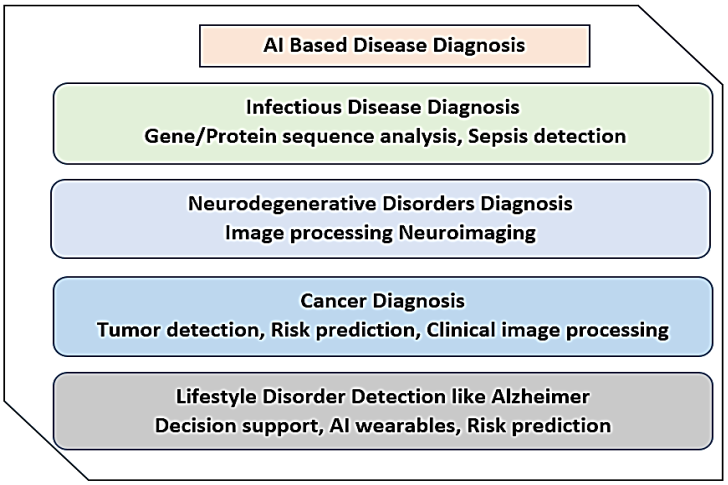

Figure 3: An outline of AI-based disease detection. Infectious illness diagnosis, cancer diagnosis, lifestyle disorders, and neurodegenerative disease diagnosis are the four categories of AI-based disease identification.

IDENTIFICATION OF THE TARGET

One of the most important phases in the drug discovery process is target identification. The conventional method uses expensive and time-consuming experimental techniques including X ray crystallography and high-throughput screening (HTS). However, the application of AI has transformed this subject by making it possible to identify possible targets using computational techniques [35].

AI-based target identification has the potential to revolutionize the drug discovery process by identifying previously overlooked or unknown targets. Using machine learning (ML) algorithms, large datasets from various sources can be analyzed to uncover hidden patterns and correlations that traditional methods might miss [36]. This approach can lead to the discovery of new biological pathways and potential therapeutic targets, thereby transforming drug discovery. Computational techniques allow for the identification of targets in a much more efficient manner compared to experimental methods, which are often resource-intensive, laborious, and time-consuming [37].

One major challenge in the past was the difficulty in studying proteins and protein complexes. However, in the 1980s, cryo-electron microscopy (cryo-EM) revolutionized the field.Cryo-ET, for example, has been particularly useful for studying protein complexes in their natural environments [38]. Using

computational algorithms, images from cryo-EM can be combined to form 3D representations of the protein structures, providing deeper insights into their functions.

In proteomics, mass spectrometry (MS) has been widely used to analyze proteins [39]. One key challenge in proteomics is predicting how proteases, such as trypsin, will cleave proteins during digestion [40]. To improve this prediction, Deep Digest, a deep learning (DL)-based algorithm, was developed to forecast the proteolytic cleavage sites of eight different protease enzymes. This tool was trained using a vast dataset of samples from various species and has demonstrated significant predictive power in identifying cleavage sites, thereby improving the accuracy of protein identification [41].

Advancements in computational tools, high-performance computing, and ML algorithms have paved the way for new drug discovery methods. A significant hurdle in drug discovery is the need to predict the 3D structure of proteins, particularly when experimental methods, like X-ray diffraction, are limited to crystallizable materials [42]. For many years, computational methods were used to predict protein structures in the absence of experimental data. One notable advancement in this area is the development of Alpha Fold, a deep learning algorithm created by DeepMind [43]. Alpha Fold is able to predict the 3D structure of proteins from their amino acid sequences with high accuracy, and its success is considered a breakthrough in drug discovery. By solving the structures of approximately 200 million proteins, including 98.5% of the human proteome, Alpha Fold has opened new avenues for drug development. In 2021, the Alpha Fold database was made publicly available, providing detailed 3D structures of the human proteome. However, Alpha Fold has limitations, such as its inability to account for the impact of mutations on protein folding and its lack of consideration for the dynamic nature of proteins, which can exist in multiple states [44-45].

In addition to Alpha Fold, other methods, like ESMFold and EMBER3D, have emerged as promising tools for predicting protein structures. ESMFold, created by Lin 2022, utilizes a masked transformer protein language model that understands biological features. ESM Fold outperforms Alpha Fold in terms of prediction speed and has shown significant accuracy for structures predicted with high confidence [46]. It is able to predict over 617 million structures from metagenomic datasets, with millions of them predicted with high confidence. EMBER3D, although less accurate than Alpha Fold, is known for its faster processing speed. It also performs better than Alpha Fold in predicting how single amino acid mutations affect protein structure. EMBER3D's predictions align closely with experimental results, making it a valuable tool for studying mutations and their impact on protein function [47].

A notable development in this area is the Passer tool, created by Tian. Passer uses an ensemble learning approach to predict allosteric sites on proteins. It combines XGBoost and graph convolutional networks (GCN) to enhance prediction accuracy [48]. With an accuracy of 0.97, Passer is able to identify allosteric sites with high precision, making it an invaluable tool in drug discovery. This method is able to identify potential targets for therapeutic intervention by studying the structure and physical properties of protein pockets [49].

Finally, BANDIT, a Bayesian machine learning platform for drug target prediction, integrates data from various sources to identify potential drug targets [50]. It uses over 20 million data points, including bioassay results and post-treatment reactions, to make accurate predictions. BANDIT has been successfully tested on thousands of molecules and has identified promising new targets for cancer therapy [51]. This platform holds significant promise for accelerating drug development by prioritizing targets with unique mechanisms of action [52].

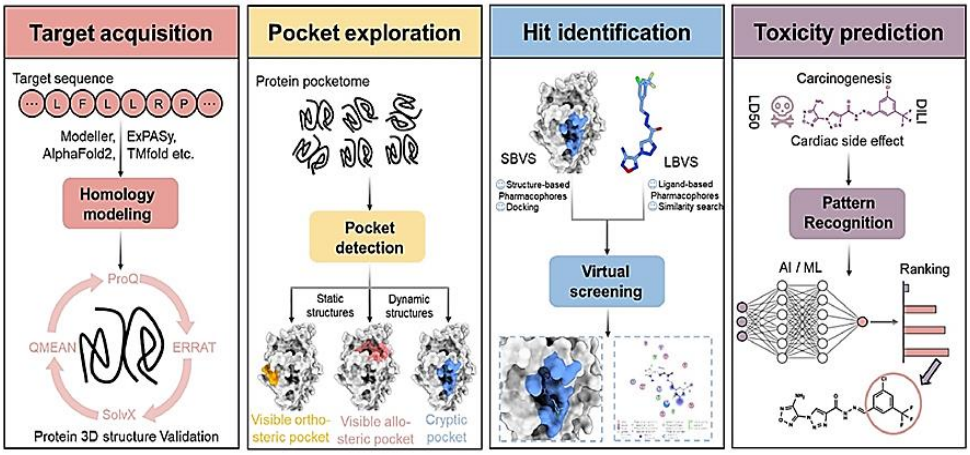

Figure 4: shows a summary of the AI-based method for locating and assessing tiny compounds. First comes target acquisition, then protein 3D structural validation with ProQ and SolvX, and finally homology modeling. Programs such as Modeller and AlphaFold2 are used to model protein sequences. The next stage is pocket exploration, which finds orthosteric and allosteric binding sites across the protein proteome using both static and dynamic structures.

AI-ENABLED VIRTUAL SCREENING IN DRUG DISCOVERY: OPPORTUNITIES AND CHALLENGES

In the first stage of drug discovery, many compounds are typically computationally screened to find those with the desired cellular or biochemical effects; new techniques are continually being developed to speed up, improve the efficiency, and reduce the cost of this process. Primary "hit" compounds are identified by a positive response in the first round of screening in a biochemical assay; further screening is then conducted to determine whether the hit compounds' physicochemical and pharmacological properties are suitable. With any luck, a lead may eventually receive drug approval, a process that could take 12 to 15 years [53]. Drug development is still a costly and time-consuming procedure, even with major improvements in medicinal chemistry and drug discovery technology. In order to find hit compounds that may be improved to lead compounds with desired attributes including higher potency, solubility, and reduced toxicity and off-target effects, hundreds of compounds are subjected to in vitro tests as part of the current standard procedure, which is known as HTS [54].

Thousands of compounds must be synthesized and tested as part of the traditional drug discovery process, which is expensive and time-consuming. It also requires a lot of protein supply and well-established laboratory techniques for bioactivity testing [55]. On the other hand, VS has become a viable and affordable method for screening millions of commercially accessible chemicals and identifying those that should be further tested, synthesized, or purchased techniques are divided into two groups: ligand-based techniques and structure-based techniques [56]. But even with the potential advantages, developing a single medication still takes an average of 10 to 15 years and more than $2 billion [57].

STRUCTURE-BASED METHODS

Structure-based virtual screening (SBVS) relies heavily on obtaining structural data of protein complexes. Molecular docking, a key technique in SBVS, estimates the ligand's binding affinity by generating various binding poses for the ligand in the protein’s binding site. These poses are ranked using scoring functions (SFs), which have seen significant advancements with the introduction of machine learning (ML) and deep learning (DL)-based SFs. Unlike SBVS, ligand-based virtual screening (LBVS) methods such as

quantitative structure-activity relationship (QSAR) modeling, molecular similarity searches, and ligand-based pharmacophore approaches require only the ligand information and are well-established techniques for drug discovery [58-59].

To improve the efficiency of the SBVS process, AI-enabled tools have been developed. One notable advancement is Deep Docking, an open-source protocol designed by Gentile, which accelerates SBVS by 100-fold. Deep Docking integrates both molecular docking and machine learning to screen large molecule libraries [60]. Initially, a small subset of the library undergoes molecular docking, and the results are used by a deep neural network (DNN) to infer the ranking of the remaining molecules [61]. This method allows the efficient screening of billions of compounds while minimizing computational costs. Deep Docking can be paired with various docking software, including Glide, Autodock-GPU, and OpenEye's FRED. Despite its advantages, the approach faces challenges, such as the reliance on the availability of graphics processing units (GPUs) [62].

Several machine learning approaches are commonly employed for virtual screening (VS), including artificial neural networks (ANNs), k-nearest neighbors (kNN), support vector machines (SVMs), random forests (RFs), and Naive Bayes (NB) classifiers. For example, KNN is simple to build and can be used for multi-task learning (MTL), whereas NB classifiers are particularly effective for identifying scaffold pieces beneficial for drug discovery. Additionally random forests can be parallelized and boosted, enhancing performance. Combining multiple machine learning models in an ensemble is often recommended to improve prediction accuracy [63].

In ligand-based virtual screening (LBVS), NB classifiers are frequently used due to their strong performance. demonstrated the effectiveness of NB classifiers in identifying inhibitors for methicillin-resistant Staphylococcus aureus. They also developed an improved ensemble model incorporating both NB and SVMs, which identified potent neuraminidase inhibitors [64].

SVMs, first introduced by Vapnik, are highly effective in virtual screening when supervised learning is applied. SVMs create feature vectors for input data, mapping them into a higher-dimensional space for classification. Chandra and colleagues demonstrated the power of SVMs in identifying inhibitors for PTP1B, a potential target for Type 2 diabetes. Their SVM model successfully identified five inhibitory compounds, two of which showed significant activity in vitro. Similarly, Deshmukh et al. used an SVM model to identify both known and novel inhibitors of FEN1, a target involved in DNA repair. Furthermore, SVM models with regression have proven to be more effective than RF models in predicting a compound's skin permeability [65].

AI techniques, especially those employing machine learning algorithms, have revolutionized virtual screening methods for drug discovery. By combining structure-based and ligand-based approaches with AI tools, researchers can significantly improve the efficiency and accuracy of the drug discovery process [66].

LIGAND-BASED VIRTUAL SCREENING (LBVS)

Ligand-based virtual screening (LBVS) forms a crucial part of the drug discovery process by identifying compounds that share structural similarities with known active ligands. One of the main techniques within LBVS is pharmacophore-based virtual screening. This approach involves creating two-dimensional fingerprints of active ligands based on chemical descriptors, such as hydrogen-bond donors, hydrogen-bond acceptors, and aromatic rings . By utilizing these chemical descriptors, LBVS can screen large databases effectively, even without structural data of the target protein [67].

In addition to fingerprint-based approaches, machine learning (ML) plays a significant role in LBVS. ML models are employed to explore the relationship between the biological activity of ligands and their molecular or atomic descriptors. For example, Quantitative Structure-Activity Relationship (QSAR) is an area where ML is frequently used to predict the biological activity of compounds based on their molecular features demonstrated this in their study of hybrid inhibitors for cancer treatment. The study achieved area under the curve (AUC) values of 0.98 and 0.94 for the prediction of inhibitors targeting BCR-ABL showing the potential of these hybrid compounds in cancer therapy [68]. In another example, Dhamodharan. developed three AI models to predict dual inhibitors for acetylcholinesterase and beta-secretase 1 (BACE1) in Alzheimer's disease, with the artificial neural network (ANN) model providing the highest prediction accuracy. Furthermore, a significant development in LBVS is the introduction of AtomNet, a convolutional neural network (CNN) model created by Atomwise, Inc. AtomNet departs from traditional LBVS techniques by using a structure-based approach to predict ligand binding. This model utilizes convolutional layers to recognize binding features from chemical functions, achieving AUC values greater than 0.74 across various datasets. AtomNet's ability to identify binding features greatly improves the accuracy of virtual screening [69].

The complexity of diseases like Alzheimer's has led to the use of polypharmacology in drug discovery. Polypharmacology focuses on designing multi-target directed ligands (MTDLs) that can modulate multiple targets associated with complex diseases. Fang et al. used LBVS methods, including Naïve Bayes and recursive partitioning classifiers, to identify drugs binding to multiple targets involved in Alzheimer's disease. These methods provided insight into potential multi-target inhibitors and contributed to a better understanding of polypharmacology in AD drug discovery [70].

Recent advancements in computational tools have greatly enhanced the efficiency of virtual screening. Supercomputing platforms such as Open Eye, Giga Docking and Virtual Flow have been employed to dock massive chemical libraries, screening billions of molecules within a short time frame using thousands of CPUs/GPUs. However, these platforms require significant computational resources, making Deep Docking (DD) a more desirable alternative for large-scale virtual screening. Deep Docking uses less computational power compared to other docking methods and speeds up the screening process by focusing on compounds with the highest docking scores, making it an effective tool for high-throughput screening of large molecule libraries. Drug repurposing is another emerging field that benefits from these advanced screening techniques. By identifying new therapeutic uses for existing drugs, drug repurposing can bypass pre-clinical research and lead optimization, accelerating the drug development process. A novel method for predicting the molecular targets of approved drugs was developed by Reker and colleagues. Their self-organizing map-based prediction of drug equivalence relationships (Spider) has shown impressive predictive power, with ROC values between 0.86 and 0.93 in cross-validation studies.

Prediction of Drug Toxicity with AI

One essential step in lowering the failure rate and increasing the effectiveness of drug discovery is predicting drug toxicity during the preclinical phases. The use of crude models and small datasets in traditional drug toxicity prediction techniques limits their applicability. However, by utilizing vast and varied data sources biological processes and clinical data, AI-based methods have surfaced as viable substitutes. AI-based methods can increase the precision and effectiveness of forecasting the potentially harmful effects of novel compounds by employing machine learning (ML) algorithms. This can assist to lower the risks involved in clinical trials, lower the expenses of medication development, and ultimately improve patient outcomes [71].

Recently, there has been an increase in the application of AI-based computer models to predict medication toxicity. Numerous researches have used machine learning (ML) and deep learning (DL) methods,

including neural networks, to evaluate large drug and toxicity data sets in order to find possible hazardous effects during drug development. These models can expedite the discovery of novel medications by detecting toxicity early on. AI-based toxicity prediction models can also identify novel therapeutic targets and toxicity pathways, as well as rank drugs for testing. Numerous reviews have addressed the prediction of AI toxicity. Numerous toxicity features in the broad field of AI-based toxicity prediction make a single review difficult [72].

CARDIAC SIDE EFFECT PREDICTION

Machine learning (ML)-based techniques are widely used to predict hERG toxicity, which is critical for assessing cardiac side effects in drug development. Several ML models, including Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), Support Vector Regression (SVR), k-Nearest Neighbors (kNN), Decision Trees (DT), Gradient Boosting (GB), Partial Least Squares (PLS), and extreme Gradient Boosting (XGB), are commonly applied for this purpose.

Venkatraman developed an RF model based on Extended Connectivity Fingerprint (ECFP6) descriptors to predict hERG toxicity. This model used 7,889 compounds with known experimental hERG blocking bioactivities from four assays. Compounds with a blocking value of ≤10 μM were considered positive, totaling 4,355 compounds, while 3,534 compounds were negative. The RF model, which builds multiple decision trees, achieved an accuracy of 80% and an ROC-AUC of 88%, demonstrating robust prediction performance for hERG toxicity. Other studies have also demonstrated the power of RF models utilized RF for hERG prediction with high accuracy. In a different approach, Ogura et al. used 72 NSGA-II-selected descriptors combined with ECFP_4 structural fingerprints to build an SVM model, achieving a kappa statistic of 0.733 and an accuracy of 98.4%. This model outperformed others in predicting hERG toxicity [73].

Konda developed consensus models using RF, Sequential Minimal Optimization (SMO), and Multilayer Perceptron (MLP), incorporating 2D descriptors. Their consensus model outperformed others, reaching an accuracy of 92%. More advanced models have also been introduced, such as Directed Message-Passing Neural Networks (DMPNN) created a DMPNN model using moe206 descriptors, achieving an accuracy of 80%, surpassing other models. Furthermore, Zhang developed HergSPred, which achieved an impressive accuracy of 98.3%. Other notable models include Deep Hit, created by Ryu et al., which excelled in terms of accuracy, Matthews Correlation Coefficient (MCC), and sensitivity (SE), and Interpretable-ADMET by Wei which showed a high accuracy of 91.9% on 8,672 chemicals. ADMET Lab 2.0 [74].

LD50 PREDICTION

Toxicology research frequently uses the median lethal dose (LD50) to assess a substance's toxicity. LD50 represents the dose of a chemical that causes death in 50% of a population within a set timeframe, typically in animal models. It is an essential measure in drug screening. However, due to ethical concerns and variability between species, conventional LD50 testing is being phased out in favor of in silico prediction models and cell culture-based toxicity assays. These alternative methods aim to reduce animal testing and provide more accurate, efficient ways to predict toxicity across different species. One common approach to predicting LD50 involves the use of binary classification models. Substances are classified into two categories: hazardous (LD50 = 2,000 mg/kg) or benign (LD50 = 50 mg/kg). Several computational tools, such as FP-ADMET and Interpretable-ADMET, have incorporated LD50 prediction algorithms to aid in toxicity assessments. Ballabio developed binary fingerprint models using Naïve Bayes (NB), N-Nearest Neighbors, and Extended Connectivity to predict LD50 for 8,992 compounds. Their model achieved 84% sensitivity and 81% specificity, demonstrating its utility in toxicity prediction. Further advancements in

LD50 prediction have involved machine learning (ML) and deep learning (DL) models [93]. For example, Gadaleta utilized Random Forest (RF), regression techniques, and other ML methods to create a QSAR model for 8,448 chemicals. Their integrated approach achieved over 70% accuracy for multiclass endpoints and 80% for binary endpoints, with an RMSE of 0.477%. In another study, created a DL consensus model combining RF, Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), and Graph Convolutional Networks (GCNN) for 80,081 chemicals, achieving an RMSE of 0.65 and R2 of 0.5[75].

These advancements in computational methods, such as RF, DNN, and CNN, have significantly improved the prediction of LD50 values, allowing for safer and more efficient drug development while minimizing the need for animal testing.

AI-BASED MODELING FOR PERSONALIZED DRUG DOSING

Traditionally, clinical practice has followed a "one-size-fits-all" approach to therapy, where drugs are prescribed uniformly. However, individual variations in genetic profiles can lead to differences in drug metabolism, causing treatments to be more effective in some patients and less so, or even harmful, in others. These variations highlight the importance of precision medicine, a more personalized approach to treatment based on a patient's genetic makeup. The goal is to minimize side effects and optimize treatment outcomes by tailoring treatments and dosages to individual patients or groups with similar genetic profiles [76].

Artificial intelligence (AI) has significantly advanced the field of precision medicine. For instance, CURATE.AI, an AI-driven platform, uses patient data to predict treatment outcomes and recommend optimal dosages. It creates a personalized patient profile that is adjusted dynamically as the patient's disease progresses or recedes. This approach is particularly beneficial as drug regimens become more complex, often involving combination therapies that target multiple aspects of a disease, such as cancer.AI applications have also been used in oncology to predict patient responses to treatments. Kureshi developed an AI decision tree to predict treatment effectiveness for non-small cell lung cancer (NSCLC) patients, achieving an accuracy of 76.6%. AI-based tools like IBM Watson for Oncology have further enhanced personalized treatment plans, drawing on vast databases of clinical trials and medical literature to make precise treatment recommendations [77].

THE ROLE OF AI IN RARE DISEASE RESEARCH

Nearly 10% of the US population suffers from rare diseases (RDs), making diagnosis and treatment challenging due to their complexity and rarity. Delays in diagnosis, which can take up to seven years, lead to significant treatment setbacks. AI offers potential solutions for improving RD diagnosis and therapy. Techniques like Naive Bayes (NB), Random Forest (RF), XGBoost, Convolutional Neural Networks (CNN), and Autoencoders (AE) are being employed to detect and treat rare diseases. For example, developed a deep learning model using InceptionV3 CNN to identify tubers in MRI images, achieving 95% accuracy in diagnosing a rare neurological disorder. Similarly, Founta created a classifier based on XGBoost and RF for diagnosing Amyotrophic Lateral Sclerosis (ALS), achieving an accuracy of 88.89%.AI-based PET scans also show promise in early RD detection [78]. However, ethical, legal, and societal implications of AI in healthcare need careful consideration. Collaborations with patient advocacy groups and diverse datasets are essential to ensure the safety, efficacy, and relevance of AI-based medical devices [79].

CONCLUSION

AI technology has significantly impacted drug design, enhancing the efficiency and accuracy of various stages in drug development, from disease diagnosis to post-market surveillance. In early disease prediction, AI models can analyze patient data and predict potential diseases, aiding in personalized medicine and optimizing drug dosages for individual needs. Furthermore, AI accelerates target and lead identification by predicting protein structures and assessing the biological activities of small molecules, reducing the reliance on experimental validation.

Clinical trials benefit from AI by assisting in patient stratification, recruitment, and monitoring, ensuring more effective trials with fewer failures. It also aids in regulatory approval processes, including with the FDA, and contributes to pharmacovigilance, reducing the chances of adverse drug reactions post-market. AI's role in drug development has led to faster discoveries, cost reductions, and minimized reliance on animal testing, addressing ethical concerns. However challenges remain, such as ensuring model explain ability, maintaining high-quality training data, and preventing bias and overfitting. AI’s ethical implications, particularly concerning data privacy, also need careful attention. Despite these challenges, the integration of AI into healthcare is progressing, with emerging fields like the development of virtual human models to simulate molecular interactions, predicting drug behaviors and side effects with unprecedented accuracy. This approach could revolutionize drug discovery and precision medicine, offering new possibilities for therapeutic safety and efficacy prediction.

REFERENCES

1. Gao N, Li J, Wei T, Cheng Q, Gao S, Hu Y. Discovery and exploration of adaptive immune response-related drug targets in diabetic retinopathy by Mendelian randomization. Diabetes Res Clin Prac. 2025 Feb 1;220:111987.
2. Kenéz B, Koplányi G, Decsi B, Molnár Z, Horváth P, Katona G, Balogh GT, Balogh-Weiser D. Development of a Novel Human Serum Albumin-Based Tool for Effective Drug Discovery: The Investigation of Protein Quality and Immobilization. J Med Chem. 2025 Jan 15;68(3):2840-8.
3. Parveen S, Shahbaz L, Shafiq N, Rashid M, Mohany M, Zhu M. An integrated theoretical study on natural alkaloids as SARS-CoV-2 main protease inhibitors: a step toward discovery of potential drug candidates with anti-COVID-19 activity. RSC advances. 2025;15(3):2045-65.
4. Saha P, Kumar M, Sharma DK. Potential of Mycobacterium tuberculosis Type II NADH-Dehydrogenase in Antitubercular Drug Discovery. ACS Infectious Diseases. 2025 Jan 15;11(2):398-412.
5. Shukla S, Kushwah S, Mani A. Targeting Malaria's Achilles' Heels: A Review of Plasmodium Life Cycle Vulnerabilities for Drug Discovery. Curr Topics Med Chem. 2025 Jan 10.
6. Sobral PS, Carvalho T, Izadi S, Castilho A, Silva Z, Videira PA. Advancements in drug discovery: integrating CADD tools and drug repurposing for PD-1/PD-L1 axis inhibition. RSC Adv. 15(4):2298–316.
7. Sone N, Gotoh S. Micro-physiological system of human lung: The current status and application to drug discovery. Drug Metab Pharmacokinet. 2025 Jan 9:101050.
8. Tan D, Chen Y, Ilboudo Y, Liang KY, Butler-Laporte G, Richards JB. Caution when using network partners for target identification in drug discovery. Human Genet Gen Adv. 2025 Apr 10;6(2).
9. Fatima K, Haroon H, Laraib I, Jabbar E, Yousaf F, Jamil I. Integrating Rapid Diagnostics, AI Technologies, and Omics Approaches for Sepsis Management and Antimicrobial Stewardship. J Med Health Sci Rev. 2025 Jun 18;2(2).
10. Ameen MU, Ashfaq M, Shahbaz F, Dilbar M, Ali A, Jamil I. Transforming Histopathology with Artificial Intelligence: Enhancing Diagnosis, Prognosis, and Personalized Care. MSRA. 2025 Jul 17;3(3):64-76.
11. Wang D, Jin J, Shi G, Bao J, Wang Z, Li S, Pan P, Li D, Kang Y, Hou T. ADMET evaluation in drug discovery: 21. Application and industrial validation of machine learning algorithms for Caco-2 permeability prediction. J Cheminform. 2025 Jan 10;17(1):3.
12. Wang H, Ma X, Sun L, Bi T, Yang W. Applications of innovative synthetic strategies in anticancer drug discovery: The driving force of new chemical reactions. Bioorg Med Chem Lett. 2025 Jan 9:130096.
13. Wang R, Qin Y, Jiang X, Bai H, Liu Y, Gao X, Zhao L. Cell membrane biomimetic magnetic fluorescent bifunctional nanoplatform for drug lead discovery. Analy\Chim Acta. 2025 Feb 8;1338:343583.
14. Win R, Minto W, Mah IK, Boyd K. Integration of Chromogenic RNAscope In Situ Hybridization for Target Validation in Drug Discovery. Toxicol Pathol. 2025 Jan;53(1):21-30.
15. Zhou Y, Zhang Y, Zhang Z, Zhou Z, Zhu F. AI comes to the Nobel Prize and drug discovery. J Pharma Anal. 2024 Nov 1;14(11):101160.

16. Zheng B, Duan M, Huang Y, Wang S, Qiu J, Lu Z, Liu L, Tang G, Cheng L, Zheng P. Discovery of a heparan sulfate binding domain in monkeypox virus H3 as an anti-poxviral drug target combining AI and MD simulations. Elife. 2025 Jan 16;13:RP100545.
17. Zhang Y, Liu C, Liu M, Liu T, Lin H, Huang CB, Ning L. Attention is all you need: utilizing attention in AI-enabled drug discovery. Brief bioinform. 2023 Nov 22;25(1).
18. Zhang Z, Bian Y, Xie A, Han P, Zhou S. Can Pretrained Models Really Learn Better Molecular Representations for AI-Aided Drug Discovery? J Chem Inf Model. 64(7):2921–30.
19. Zhang C, Xie L, Lu X, Mao R, Xu L, Xu X. Developing an Improved Cycle Architecture for AI-Based Generation of New Structures Aimed at Drug Discovery. Molecules. 29(7).
20. Yoo J, Kim TY, Joung I, Song SO. Industrializing AI/ML during the end-to-end drug discovery process. Curr Opin Struct Biol. 2023 Apr 1;79:102528.
21. Yildirim Z, Swanson K, Wu X, Zou J, Wu J. Next-Gen Therapeutics: Pioneering Drug Discovery with iPSCs, Genomics, AI, and Clinical Trials in a Dish. Annu Rev Pharmacol Toxicol. 65(1):71–90.
22. Yang K, Xie Z, Li Z, Qian X, Sun N, He T. MolProphet: A One-Stop, General Purpose, and AI-Based Platform for the Early Stages of Drug Discovery. J Chem Inf Model. 64(8):2941–7.
23. Wu K, Li X, Zhou Z, Zhao Y, Su M, Cheng Z, Wu X, Huang Z, Jin X, Li J, Zhang M. Predicting pharmacodynamic effects through early drug discovery with artificial intelligence-physiologically based pharmacokinetic (AI-PBPK) modelling. Front Pharmacol. 2024 Feb 16;15:1330855.
24. Workman P, Antolin AA, Al-Lazikani B. Transforming cancer drug discovery with Big Data and AI. Expert Opin Drug Discov. 14(11):1089–95.
25. Wenteler A, Cabrera CP, Wei W, Neduva V, Barnes MR. AI approaches for the discovery and validation of drug targets. Camb Prisms: Prec Med. 2024 Jan;2:e7.
26. Wang L, Wang S, Yang H, Li S, Wang X, Zhou Y, Tian S, Liu L, Bai F. Conformational space profiling enhances generic molecular representation for AI-powered ligand-based drug discovery. Adv Sci. 2024 Oct;11(40):2403998.
27. Walters WP, Barzilay R. Critical assessment of AI in drug discovery. Expert Opin Drug Discov. 16(9):937–47.
28. Vert JP. How will generative AI disrupt data science in drug discovery? Nat Biotechnol. 41(6):750–1.
29. Vashishat A, Gupta GD, Kurmi BD. Revolutionizing Drug Discovery: The Role of AI and Machine Learning. Curr Pharm Des. 29(39):3087–8.
30. Vemula D, Jayasurya P, Sushmitha V, Kumar YN, Bhandari V. CADD, AI and ML in drug discovery: A comprehensive review. Eur J Pharma Sci. 2023 Feb 1;181:106324.
31. Urbina F, Lentzos F, Invernizzi C, Ekins S. AI in drug discovery: A wake-up call. Drug Discov Today. 28(1).
32. Turon G, Hlozek J, Woodland JG, Kumar A, Chibale K, Duran-Frigola M. First fully-automated AI/ML virtual screening cascade implemented at a drug discovery centre in Africa. Nat Commun. 14(1).
33. Tripathi MK, Nath A, Singh TP, Ethayathulla AS, Kaur P. Evolving scenario of big data and Artificial Intelligence (AI) in drug discovery. Mol Divers. 25(3):1439–60.
34. Trepte P, Secker C, Olivet J, Blavier J, Kostova S, Maseko SB. AI-guided pipeline for protein-protein interaction drug discovery identifies a SARS-CoV-2 inhibitor. Mol Syst Biol. 20(4):428–57.

35. Trepte P, Secker C, Olivet J, Blavier J, Kostova S, Maseko SB, Minia I, Silva Ramos E, Cassonnet P, Golusik S, Zenkner M. AI-guided pipeline for protein–protein interaction drug discovery identifies a SARS-CoV-2 inhibitor. Molecular systems biology. 2024 Apr 2;20(4):428-57.
36. Thuault S. Drug discovery by AI trained on aging biology. Nat Aging. 4(4).
37. Smith JS, Roitberg AE, Isayev O. Transforming Computational Drug Discovery with Machine Learning and AI. ACS Med Chem Lett. 9(11):1065–9.
38. Smalley E. AI-powered drug discovery captures pharma interest. Nat Biotechnol. 35(7):604–5.
39. Shen C, Song J, Hsieh CY, Cao D, Kang Y, Ye W. DrugFlow: An AI-Driven One-Stop Platform for Innovative Drug Discovery. J Chem Inf Model. 64(14):5381–91.
40. Serrano DR, Luciano FC, Anaya BJ, Ongoren B, Kara A, Molina G. Artificial Intelligence (AI. Applications in Drug Discovery and Drug Delivery: Revolutionizing Personalized Medicine Pharmaceutics. 16(10).
41. Schauperl M, Denny RA. AI-Based Protein Structure Prediction in Drug Discovery: Impacts and Challenges. J Chem Inf Model. 62(13):3142–56.
42. Rusinko A, Rezaei M, Friedrich L, Buchstaller HP, Kuhn D, Ghogare A. AIDDISON: Empowering Drug Discovery with AI/ML and CADD Tools in a Secure. Web-Based SaaS Platform J Chem Inf Model. 64(1):3–8.
43. Roy K. AI and ML for small molecule drug discovery in the big data era II. Mol Divers. 28(4):1847–8.
44. Roy K. Special issue of molecular diversity on "AI and ML for small molecule drug discovery in the big data era." Mol Divers. 25(3):1259–60.
45. Rajaei F, Minoccheri C, Wittrup E, Wilson RC, Athey BD, Omenn GS, Najarian K. AI-based computational methods in early drug discovery and post market drug assessment: a survey. IEEE/ACM Trans Comput Biol Bioinform. 2024 Nov 6.
46. Qureshi R, Irfan M, Gondal TM, Khan S, Wu J, Hadi MU, Heymach J, Le X, Yan H, Alam T. AI in drug discovery and its clinical relevance. Heliyon. 2023 Jul 1;9(7).
47. Quinn TR, Giblin KA, Thomson C, Boerth JA, Bommakanti G, Braybrooke E. Accelerated Discovery of Carbamate Cbl-b Inhibitors Using Generative AI Models and Structure-Based Drug Design. J Med Chem. 67(16):14210–33.
48. Qiu X, Li H, Ver Steeg G, Godzik A. Advances in AI for Protein Structure Prediction: Implications for Cancer Drug Discovery and Development. Biomolecules. 14(3).
49. Pitt WR, Bentley J, Boldron C, Colliandre L, Esposito C, Frush EH. Real-World Applications and Experiences of AI/ML Deployment for Drug Discovery. J Med Chem. 68(2):851–9.
50. Ouyang B, Shan C, Shen S, Dai X, Chen Q, Su X, Cao Y, Qin X, He Y, Wang S, Xu R. AI-powered omics-based drug pair discovery for pyroptosis therapy targeting triple-negative breast cancer. Nat Commun. 2024 Aug 30;15(1):7560.
51. Nozawa K, Tsunoyama K. Drug discovery research using AI and data science: analyzing literature information to understand pathological mechanisms. Nihon Yakurigaku Zasshi. 158(1):3–9.
52. Nishan MN. AI-powered drug discovery for neglected diseases: accelerating public health solutions in the developing world. Journal]Glob Health. 2025 Jan 10;15:03002.
53. Nierengarten MB. Using AI to tailor cancer care and drug discovery. Cancer. 130(18).

54. Niazi SK. The Coming of Age of AI/ML in Drug Discovery, Development, Clinical Testing, and Manufacturing: The FDA Perspectives. Drug Des Devel Ther. 17:2691–725.
55. Mullard A. When can AI deliver the drug discovery hits? Nat Rev Drug Discov. 23(3):159–61.
56. Mullard A. Creating an AI-first drug discovery engine. Nat Rev Drug Discov. 23(10):734–5.
57. Moingeon P. Harnessing the power of AI-based models to accelerate drug discovery against immune diseases. Expert Rev Clin Immunol. 20(10):1135–8.
58. Mahapatra M, Sahu C, Mohapatra S. Trends of artificial intelligence (AI) use in drug targets, discovery and development: current status and future perspectives. Curr Drug Targets. 2024 Oct 29.
59. Mock M, Edavettal S, Langmead C, Russell A. AI can help to speed up drug discovery - but only if we give it the right data. Nature. 621(7979):467–70.
60. Mirakhori F, Niazi SK. Harnessing the AI/ML in drug and biological products discovery and development: the regulatory perspective. Pharmaceuticals. 2025 Jan 3;18(1):47.
61. Luo Y, Liu XY, Yang K, Huang K, Hong M, Zhang J, Wu Y, Nie Z. Toward unified ai drug discovery with multimodal knowledge. Health Data Science. 2024 Feb 23;4:0113.
62. Mak KK, Balijepalli MK, Pichika MR. Success stories of AI in drug discovery-where do things stand?. Expert opinion on drug discovery. 2022 Jan 2;17(1):79-92.
63. Luo Y, Liu XY, Yang K, Huang K, Hong M, Zhang J, Wu Y, Nie Z. Toward unified ai drug discovery with multimodal knowledge. Health Data Science. 2024 Feb 23;4:0113.
64. Yap TA, Smith AD, Ferraldeschi R, Al-Lazikani B, Workman P, De Bono JS. Drug discovery in advanced prostate cancer: translating biology into therapy. Nature reviews Drug discovery. 2016 Oct;15(10):699-718.
65. Liu Z, Roberts RA, Lal-Nag M, Chen X, Huang R, Tong W. AI-based language models powering drug discovery and development. Drug Discovery Today. 2021 Nov 1;26(11):2593-607.
66. Liu J, Du H, Huang L, Xie W, Liu K, Zhang X, Chen S, Zhang Y, Li D, Pan H. AI-powered microfluidics: shaping the future of phenotypic drug discovery. ACS Applied Materials & Interfaces. 2024 Jul 17;16(30):38832-51.
67. Liu H, Chen P, Zhai X, Huo KG, Zhou S, Han L, Fan G. PPB-Affinity: Protein-Protein Binding Affinity dataset for AI-based protein drug discovery. Scientific data. 2024 Dec 3;11(1):1316.
68. Le MH, Nguyen PK, Nguyen TP, Nguyen HQ, Tam DN, Huynh HH, Huynh PK, Le NQ. An in-depth review of AI-powered advancements in cancer drug discovery. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease. 2025 Jan 19:167680.
69. Kumar M, Ashok AK, Bhat T, Ballamoole K, Gollapalli P. Computational-driven discovery of AI-2 quorum sensing inhibitor targeting the 5′-methylthioadenosine/S-adenosylhomocysteine nucleosidase (MTAN) to combat drug-resistant Helicobacter pylori. Computers in Biology and Medicine. 2025 Jan 1;184:109409.
70. Klaus B. AI models and drug discovery within pharmaceutical drug market. Delaware Journal of Public Health. 2023 Nov 30;9(4):52.
71. Kint S, Dolfsma W, Robinson D. Strategic partnerships for AI-driven drug discovery: The role of relational dynamics. Drug Discovery Today. 2024 Dec 1;29(12):104242.
72. Khan SR, Al Rijjal D, Piro A, Wheeler MB. Integration of AI and traditional medicine in drug discovery. Drug discovery today. 2021 Apr 1;26(4):982-92.

73. Khan MK, Raza M, Shahbaz M, Hussain I, Khan MF, Xie Z, Shah SS, Tareen AK, Bashir Z, Khan K. The recent advances in the approach of artificial intelligence (AI) towards drug discovery. Frontiers in Chemistry. 2024 May 31;12:1408740.

74. Kargbo RB. AI in pharma: transforming drug discovery and strategic management with MYC-modulating compounds and BET protein inhibitors. ACS Medicinal Chemistry Letters. 2024 Feb 13;15(3):334-6.

75. Jayatunga MK, Xie W, Ruder L, Schulze U, Meier C. AI in small-molecule drug discovery: a coming wave. Nat. Rev. Drug Discov. 2022 Mar 1;21(3):175-6.

76. Iwata H. Transforming drug discovery: the impact of AI and molecular simulation on R&D efficiency. Bioanalysis. 2024 Dec 16;16(23-24):1211-7.

77. Hutson M. AI for drug discovery is booming, but who owns the patents?.

78. Hervey M. Harnessing AI in drug discovery without losing patent protection. Drug discovery today. 2020 Jun;25(6):949-50.

79. Han R, Yoon H, Kim G, Lee H, Lee Y. Revolutionizing medicinal chemistry: the application of artificial intelligence (AI) in early drug discovery. Pharmaceuticals. 2023 Sep 6;16(9):1259.