The Future of Pharmaceutical: Artificial Intelligence in Drug Discovery

Ms. Sakshi P. Malekar[1] , Ms. Shruti P. Kamble[2] , Ms. Pornima P. Mangure[3] , Ms. Sneha S. Kadam[4]

[1,2,3,4] Department of Pharmacy / Ashokrao Mane College of Pharmacy , Peth-Vadgaon / Shivaji University 416112, Maharashtra, India.

Corresponding author: sk6832138@gmail,com
Doi: https://doi.org/10.5281/zenodo.17934319
Received: 03 December 2025
Accepted: 16 December 2025

Abstract

Artificial intelligence (AI), specifically machine learning and deep learning, has revolutionized drug discovery by significantly accelerating and mitigating risks throughout the pipeline, from target identification to clinical candidate selection. AI integrates and analyzes multidimensional data related to biology (genomics, proteomics, transcriptomics, and real-world clinical datasets) in order to clarify disease mechanisms, deduce causal gene-disease correlations, prioritize druggable targets, and uncover predictive biomarkers for precision medicine. In the earliest research, generative AI models (e.g., variational autoencoders, GANs, and diffusion models) combined with reliable protein structure prediction (AlphaFold2/3) allow for de novo drug design and virtual screening of billions of compounds. Graph neural networks and physics-informed neural networks improve polypharmacology profiling, binding affinity prediction, and ADMET property forecasting, significantly lowering attrition rates in hit-to-lead and lead optimization stages. Automation powered by AI simplifies retrosynthetic planning, high-throughput screening, and multi-objective strength, selection, and safety profile tuning. As a result, development schedules have been shortened from ten to fifteen years to possibly three to five years, all the while increasing success rates. Despite these developments, there are still issues that need to be resolved, such as the need for impartial, high-quality training data; the difficulty of interpreting deep learning models; ethical issues with privacy of data and algorithmic bias; and the possibility of overly optimistic extrapolated forecasts. For AI to be sustainably integrated into pharmaceutical development and research, these issues must be addressed through strong data collection, explainable artificial intelligence frameworks, and hybrid human–AI processes.

Introduction

Even while the conventional drug discovery approach has produced many life-saving medications, there are issues with it that cause patients' access to novel cancer treatments to be delayed. For example, it typically takes 10 to 15 years and costs about $2.6 billion to bring a new medicine to market, taking into account expenses that are both direct and indirect[1,2]. Furthermore, the success rate of bringing medication candidates to market is just under 10

percent, indicating a high failure rate. Given these figures, there is an unfulfilled need to find methods for streamlining the entire process of developing drugs. In fact, safety and lack of activity are linked to the majority of clinical development failures, suggesting that preclinical procedures like drug discovery and target selection/identification need to be improved[2]. In recent years, artificial intelligence (AI) has become increasingly popular in the biomedical profession because it makes it possible to analyze and interpret vast amounts of data that are not possible using conventional statistical techniques[4].

For both pharmaceutical corporations and academic organizations, moving a therapeutic candidate from rigorous preclinical optimization to phase-I clinical development is seen as a major milestone. Lead chemical success rates in clinical trials have been increased by the use of large-scale computer screening and docking. Nevertheless, there are drawbacks to these approaches, including inaccuracy and inefficiency. Deep learning, also called DL, and algorithms for machine learning (ML), considered to be subsets of artificial intelligence, have been recognized as potential answers to these problems[5,6]. These AI technologies have cheap processing costs and can accurately anticipate macrosystem features. Because of this, chemical and biological experts are increasingly using AI algorithms in the drug discovery process. Drug development makes substantial use of machine learning (ML), including methods like deep learning (DL), support vector machines (SVM), random forests (RF), Bayesian networks (BN), and clustering[7]. Large volumes of data are processed and analyzed by DL models for tasks including bioactivity predictions, virtual screening (VS), and clinical imaging. BNs forecast patients' reactions to therapy as well as toxicity or bioactivity[8]. While clustering finds patterns or links within data, RF models are utilized for feature selection and molecular target identification[9]. Predicting pharmacokinetic characteristics, VS, and toxicity are just a few uses for SVM, a supervised learning system that categorizes data[10].

Objectives

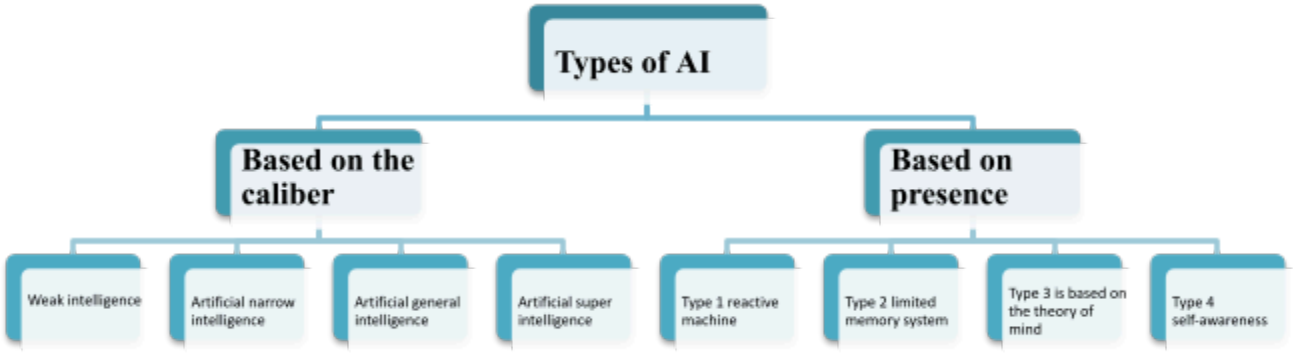
1. The literature review summarizes present AI/ML utilization and highlights approaches that impact current paradigms in order to critically examine applications of AI/ML in drug discovery.
2. To determine important AI/ML methods and how to apply them in various stages of drug development;
3. To evaluate the effectiveness of those approaches, solve their shortcomings and difficulties, and present new avenues for optimization.
4. To assess new trends, such as the use of huge language models, in relation to data availability, ethics, bias, and legal constraints[11].

Scope

The focus of the scope is on cloud-based implementations and contemporary model architectures that are especially pertinent to applications in the pharmaceutical business. Target Identifying, Hit/Lead Detection & Optimizing, and Early Clinical Analysis & Safety are the main areas of focus for this study, which examines English-language publications worldwide and includes research from several multinational teams. The focus is on computational and in silico approaches using AI for target selection and design and the study of chemical and biological characteristics. Studies that only address formulation, automation, robotics, or high throughput are not included unless they directly use AI[11].

- **AI and It's Types:**

AI can be classified in two different ways[3]:



- **Target Identification**

AI has recently made great progress in the diagnosis of infectious and non-communicable diseases. AI techniques have significantly improved infectious illness diagnosis. The creation of AI models for the efficient diagnosis of the COVID-19 pandemic was given special attention. An effective indirect technique for diagnosing COVID-19 is a chest X-ray. It was employed to identify pneumonia linked to this viral infection. Numerous machine learning algorithms have been created to forecast whether patterns would appear in X-ray radiographs or not. For instance, Narin et al. suggested an automated diagnostic model based on convolutional neural networks (CNNs) to identify coronavirus-induced pneumonia. Other infectious disorders, such as urinary tract infections (UTIs), that are frequently linked to high rates of diagnostic errors, can also be diagnosed with the help of machine learning models. A prospective cohort analysis of about 80,387 persons who visited the emergency room with UTI symptoms was

published by Taylor et al. The prevention and diagnosis of HIV, a virus that causes increasing immune system failure in humans and promotes cancer, has been greatly aided by machine learning models[14].

In order to avoid HIV infection, Xiang et al. created an ensemble method that combines GCN (graph convolutional network) with LR and RF to identify high-risk individuals. A neurodegenerative condition affecting the brain is Alzheimer's disease. Age and genetics are the two main risk factors for AD. Although it has aided in the diagnosis of AD, new research suggests that environmental and lifestyle variables may also play a role. Promising outcomes in the early diagnosis and prediction of AD have been demonstrated using AI-based systems. AI systems have shown encouraging results in the diagnosis, prognosis, and treatment of non-communicable diseases such as cancer, diabetes, and Alzheimer's disease[.

AI-based algorithms can help identify these illnesses early, allowing for prompt interventions and individualized treatment regimens that will ultimately improve patient outcomes. Healthcare workers may be able to lessen the burden of NCDs, improve the standard of treatment, and maximize healthcare resources by utilizing AI. Functional investigations supported whole genome analysis by identifying novel oncogenic vulnerabilities. High-throughput screening (HTS) techniques like CRISPR-Cas9 and gene knockout research have proved crucial in identifying possible targets[12,13]. Certain target screening strategies have been used in situations where tumors were more reliant on a particular genetic mutation or were impacted by modifications in the cancer's microenvironment. Finding vulnerabilities like synthetic lethality interactions has been made possible by these techniques. The strong genomic dependence between PRMT5 inhibition and MTAP (methylthioadenosine phosphorylase) deletion in a number of cancers is one recent example[14].

But not every protein can be drugged. A protein must have certain qualities that make it a suitable target for therapeutic treatments, such as tiny molecules or biologics like antibodies, in order for it to be druggable. One of these features is a distinct binding site, or "pocket," where tiny molecules can physically attach[15,16]. A medication should be able to engage with this binding site with significant affinity and modify the activity of the protein without disrupting other proteins if it is sufficiently accessible and selective. Accordingly, for a medicine to bind properly, the protein needs to be sufficiently stable to retain an appropriate shape.

Therefore, undruggables are proteins that might not have a stable structure or a drug-binding pocket[17,18]. Certain proteins that are involved in large protein-protein interactions have large, flat surfaces that make it difficult for small molecules to bind to them, while other proteins may have highly flexible or dynamic structures[17]. These proteins include scaffolding proteins and transcription factors, which control several cellular functions but do not have readily accessible drug-binding sites[19]. see the Fig.1[39].

Drug Discovery

Computer-aided drug design (CADD) has historically used computational methods to forecast how medications will interact with biological targets.

AI in virtual screening :-A critical stage in the drug development process is virtual screening, in which molecules that are likely to react with a particular biological target are found by computationally analyzing enormous chemical libraries[20]. In order to forecast how various chemicals would attach to the chosen pocket, structure-based virtual

screening (SBVS) uses the target's three-dimensional structure. The target's binding location must be thoroughly understood in order to use this strategy[21]. For many years, ML models have been utilized in ligand-based virtual screening (LBVS) techniques, which use predictive models called Quantitative Structure-Activity Relationships (QSAR) to investigate new candidates based on the characteristics (or descriptors) of known ligands for a particular target[22]. With the use of novel atomic representation and DL structures, the revolution caused by artificial intelligence in drug discovery has just recently been applied to QSAR. Therefore, more effective screenings of extremely big compound libraries are made possible by so-called deep QSAR, which can be coupled with virtual screening methods like molecular docking or pharmacophore modeling. The last method has been extensively employed in SBVS techniques, which find possible inhibitor chemicals by using information on the target protein's and compounds' three-dimensional structures. AI models in this sector have improved scoring functions to assess ligand-protein binding affinity, binding pocket finding, and classification techniques[23].

De Novo drug design with AI : De novo drug design, in which AI models are utilized to create completely novel molecule structures that have not been synthesized previously, is one of the most revolutionary uses of artificial intelligence in drug discovery. As previously indicated, conventional drug discovery techniques rely on pre-existing chemical libraries; however, AI enables the synthesis of new molecules that are tailored for particular biological characteristics[24]. In de novo drug design, the RL technique is used to repeatedly enhance molecular design by getting feedback on how well each interaction meets specific requirements like binding affinity, stability, or posture[25,26]. By learning from existing data, generative models such as variational autoencoders (VAEs) and GANs are used to create novel chemical structures[27].

- By forecasting how changes to a chemical structure would impact its overall drug-like qualities, AI can aid with lead optimization:

1. ***Molecular dynamics (MD) simulations with AI.***- By offering more precise predictions of how molecules would behave in various physiological contexts, AI improves MD simulations. Simulations of MD are used to forecast how a medication will react with membranes in biological systems, enzymes, or transporters, as well as how stable it will be while bound to its target[28]. These simulations are accelerated by AI-based models, enabling the simulation of intricate biological interactions in just a few seconds of the time required by conventional techniques[29,30].
2. ***Predicting drug-drug interactions and toxicity.***. Predicting a compound's ADMET qualities in conventional drug discovery is expensive and time-consuming. By examining the structures of molecules and biological pathways, AI models can forecast the possibility of drug-drug interactions and bad effects. AI models can discover substances that are susceptible to causing toxicity or other negative consequences in patients by combining information about chemical structure, biological activities, and known drug interactions[31,32,33]. Furthermore, it can make it possible to stop developing possibilities with unfavorable profiles earlier, which lowers the likelihood that early clinical trials would fail owing to toxicity.

3. ***Use of multimodal models and QSAR in drug discovery***: For a more thorough understanding, multimodal models combine many data types (such as chemical, biological, and several other characteristics) into a single framework[34]. QSAR models can incorporate multimodal models as well. The models mentioned above are used in the drug discovery process to forecast a chemical compound's biological action based on its structure. In QSAR modeling, structural characteristics of molecules, such as polarity, molecular weight, or functional groups, are correlated with their reported biological activity[35]. AI seeks to greatly improve QSAR techniques so that more precise and predictive models may be created.

- **Challenges and Limitation**

The availability of appropriate data is one of the main obstacles. For training, AI-based methods usually need a lot of data. The accuracy and dependability of the results may be impacted by the fact that there is frequently a limited amount of data available or that the data is inconsistent or of poor quality[36]. Ethical issues provide another difficulty because AI-based methods may give rise to questions regarding prejudice and fairness (see next section). For instance, predictions made by an ML system may be unfair or erroneous if the data that was used to train it is biased or unrepresentative. One crucial issue that needs to be addressed is guaranteeing the ethical and equitable application of AI for the creation of novel medicinal molecules. The challenges AI faces in the field of chemical medicine can be addressed in a number of ways. Using data augmentation, which entails creating synthetic data to complement preexisting datasets, is one strategy. This can boost the amount and variety of data available for ML algorithm training, enhancing the precision and dependability of the outcomes[37,38].

Another strategy is the application of explainable AI (XAI) techniques, which seek to offer clear and understandable justifications for the predictions generated by machine learning algorithms. In addition to offering an improved awareness of the fundamental mechanisms and presumptions behind the predictions, this can help allay worries about prejudice and equity in AI-based approaches.

However, current AI-based approaches cannot take the place of conventional experimental methods, nor can they replace the expertise and knowledge of human researchers. AI can only make assumptions based on the given data; human researchers must then verify and interpret the findings. However, the drug development process can potentially be improved by combining AI with conventional experimental techniques. It is feasible to streamline the process of drug discovery and hasten the creation of new drugs by fusing the predictive capabilities of AI with the skills and expertise of human researchers[2].

- **References**

1. Accelerating biopharmaceutical development while reducing costs| McKinsey. https://www.mckinsey.com/industries/life-sciences/our-insights/the-pursuit-of-excellence-in-new-drug-development

2. Wouters, O. J., Mckee, M., & Luyten, J. (2009). Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA*, *323*, 844–853.
3. Raza, M. A., Aziz, S., Noreen, M., Saeed, A., Anjum, I., Ahmed, M., & Raza, S. M. (2022). Artificial intelligence (AI) in pharmacy: an overview of innovations. *INNOVATIONS in pharmacy*, *13*(2), 10-24926.
4. Murmu, A., & Győrffy, B. (2024). Artificial intelligence methods available for cancer research. *Front Med*, *18*, 778–797.
5. Paul, S. M. (2010). How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*, *9*, 203–214.
    a. CrossrefGoogle Scholar
6. Bolcer, J. D., & Hermann, R. B. (n.d.). *The development of computational chemistry in the United States*.
7. Y. Ding, J.H. Sohn, M.G. Kawczynski, *et al.*A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brainRadiology, 290 (2019), pp. 456-464
8. C. Berzuini, R. Bellazzi, S. Quaglini, *et al.*Bayesian networks for patient monitoring Artif. Intell. Med., 4 (1992), pp. 243-260
9. S. Kapsiani, B.J. Howlin Random forest classification for predicting lifespan-extending chemical compounds Sci. Rep., 11 (2021), pp. 1-13.
10. Rehman, A. U., Li, M., Wu, B., Ali, Y., Rasheed, S., Shaheen, S., & Zhang. (2024). *Role of artificial intelligence in revolutionizing drug discovery*.
11. Ferreira, F. J., & Carneiro, A. S. (2025). AI-Driven Drug Discovery: A Comprehensive Review. *ACS Omega*.
12. Vazquez, F., & Sellers, W. R. (2021). Are CRISPR screens providing the next generation of therapeutic targets?? *Cancer Res*, *81*, 5806–5809. Article CAS PubMed PubMed Central Google Scholar
13. Wang, T. (2015). Identification and characterization of essential genes in the human genome. *Science*, *350*, 1096–1101.
14. Rehman, A. U., Li, M., Wu, B., Ali, Y., Rasheed, S., Shaheen, S., & Zhang. (2024). *Role of artificial intelligence in revolutionizing drug discovery*.
15. Rask-Andersen, M., Almén, M. S., & Schiöth, H. B. (2011). Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov*, *10*, 579–590.
16. Imming, P., Sinning, C., & Meyer, A. (2006). Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov*, *5*, 821–834.
17. Arkin, M. R., & Wells, J. A. (2004). Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov*, *3*, 301–317.
18. Nevola, L., & Giralt, E. (2015). Modulating protein-protein interactions: the potential of peptides. *Chem Commun*, *51*, 3302–3315.

19. Ocana, A., Pandiella, A., Siu, L. L., & Tannock, I. F. (2011). Preclinical development of molecular-targeted agents for cancer. *Nat Rev Clin Oncol*, *8*, 200–209.
20. Jorgensen, W. L. (2004). The many roles of computation in drug discovery. *Science*, *303*, 1813–1818.
21. Lionta, E., Spyrou, G., Vassilatis, D. K., & Cournia, Z. (n.d.). Structure-Based virtual screening for drug discovery: principles, applications and recent advances. *Curr Top Med Chem*, *14*, 1923–1938.
22. Meli, R., Morris, G. M., & Biggin, P. C. (2022). Scoring functions for Protein-Ligand binding affinity prediction using Structure-based deep learning: A review. *Front Bioinform*, *2*.
23. Crampon, K., Giorkallos, A., Deldossi, M., Baud, S., & Steffenel, L. A. (2022). Machine-learning methods for ligand-protein molecular Docking. *Drug Discovery Today*, *27*, 151–164.
24. Schneider G. Automating drug discovery. Nat Rev Drug Discov. 2018;17:97–113.
25. Popova, M., Isayev, O., & Tropsha, A. (2018). Deep reinforcement learning for de Novo drug design. *Sci Adv*, *4*.
26. Fang, Y., Pan, X., & Shen, H.-B. (2023). De Novo drug design by iterative multiobjective deep reinforcement learning with graph-based molecular quality assessment. *Bioinformatics*, *39*.
27. Sanchez-Lengeling, B., & Aspuru-Guzik, A. (2018). Inverse molecular design using machine learning: generative models for matter engineering. *Science*, *361*, 360–365.
28. Noé, F., Tkatchenko, A., Müller, K.-R., & Clementi, C. (2020). Machine learning for molecular simulation. *Annu Rev Phys Chem*, *71*, 361–390.
29. Chmiela, S. (2017). Machine learning of accurate energy-conserving molecular force fields. *Sci Adv*, *3*.
30. Zhang, L., Han, J., Wang, H., & Car, R. (2018). Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys Rev Lett*, *120*.
31. Bhhatarai, B., Walters, W. P., Hop, C., Lanza, G., & Ekins, S. (2019). Opportunities and challenges using artificial intelligence in ADME/Tox. *Nat Mater*, *18*, 418–422.
32. Ferreira, L., & Andricopulo, A. D. (2019). ADMET modeling approaches in drug discovery. *Drug Discovery Today*, *24*, 1157–1165.
33. Lin, Z., & Chou, W.-C. (2022). Machine learning and artificial intelligence in toxicological sciences. *Toxicol Sci*, *189*, 7–19.
34. Kearnes, S., Mccloskey, K., Berndl, M., Pande, V., & Riley, P. (2016). Molecular graph convolutions: moving beyond fingerprints. *J Comput Aided Mol Des*, *30*, 595–608.
35. Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Mol Inf*, *29*, 476–488.
36. Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., … Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, *4*(2), 268–276. doi:10.1021/acscentsci.7b00572
37. Gawehn, E., Hiss, J. A., & Schneider, G. (2016). Deep learning in drug discovery. *Molecular Informatics*, *35*(1), 3–14. doi:10.1002/minf.201501008

38. Lysenko, A., Sharma, A., Boroevich, K. A., & Tsunoda, T. (2018). An integrative machine learning approach for prediction of toxicity-related drug safety. *Life Science Alliance*, *1*(6), e201800098. doi:10.26508/lsa.201800098.
39. https://www.linkedin.com/posts/mdhasnain_pharmastudents-drugdiscovery-ai-activity-7396976060002652160-1UIU?utm_source=li_share&utm_content=feedcontent&utm_medium=g_dt_web&utm_campaign=copyeferences.