AI in Drug Discovery and Development: A Review on Current Trends and Tools.

Ishwar Bhikule[1], Prajwal Bansode[2],Rohan Ghige[3], Shraddha Dingare[4*]

[1,2,3]Student, [4]Assistant professor, Department of Pharmaceutical Chemistry, TMV's Lokmanya Tilak Institute of Pharmaceutical Sciences, Pune, India 411037

Corresponding author: Shraddha Dingare[4]
Email: shraddhadespande2@gmail.com
Doi: 10.5281/zenodo.15711110
Received: 25 April 2025
Accepted: 21 May 2025

Abstract

AI in drug development defines as the application of artificial intelligence (AI) technologies like deep learning (DL), machine learning (ML), and natural language processing (NLP) to enhance the process drug development by discovering, designing, testing, and launching new active drug to market. AI accelerates drug development by analysing vast biological, chemical, and clinical datasets to identify potential drug candidates. AI increases newly synthesized drug safety and efficacy and reduce its toxicity.

This review aims to cover an overview of current tools, technologies, and trends in AI for drug discovery and development and introduce the power of AI in a new era for drug discovery and development, accelerated research functioning, by powering new tools. Simplifying new drug discovery by explaining AI powered tools in each stage of drug development. As the field continues to evolve, ongoing collaboration between scientists, technologists, and regulatory bodies will be essential to harness these advances for the benefit of global health.

Keywords:

Artificial intelligence (AI), Deep learning (DL), Machine learning (ML), Natural language processing (NLP)

Introduction:

In recent decades, continuous new drug introduction and rapid attrition of older drugs have been seen. Traditionally, new drug development is a highly complex, tedious, competitive, costly, and commercially risky process. From the synthesis/identification of the molecule to marketing, a new drug development takes at least 10 years and costs 500–1000 million US$. As such, the invention and development of new drugs is now possible through artificial intelligence tools. [1] AI is shaping the structure of drug discovery and development, offering solutions to challenges in pharmaceutical industry. AI stands at the first line of this reconfiguration, promising to accelerate timelines, reduce costs, and enable more precise targeting of therapies.[2]

Table1.1: AI contribution in acceleration in drug development [1] [3]

No.	Stages	Traditional Duration	With AI Acceleration	AI Contributions
1.	Synthesis/Isolatio n of Compound	1–2 years	Months to 1 year	AI-developed de novo drug design, predictive modelling, virtual compound generation
2.	Preclinical Studies	2–4 years	1–2 years	In silicon screening, automated toxicity prediction, AI-animal model correlation
3.	Clinical Trial Approval	3–6 months	3–6 months	No major time reduction
4.	Pharmaceutical Formulation & Standardization	0.5–1 year	0.5–1 year	AI optimizes formulations and bioassay standardization
5.	Clinical Trials	3–10 years	1.5–6 years	AI in patient selection, adaptive trials, digital biomarkers,real-time monitoring
6.	Marketing Permission	0.5–2 years	0.5–1 year	AI-assisted submission, better organized clinical data for faster approval
7.	Post marketing Surveillance (Phase IV)	Ongoing	Ongoing (enhanced)	Real-timepharmacovigilance using AI & NLP from EHRs, social media, registries

The combination of AI into drug discovery is listed several key trends. Generative AI, leveraging deep learning and advanced algorithms, is capable of simulating complex biological systems, designing novel molecules, and predicting their properties with accuracy. These technologies identify the drug candidates, streamline target validation, and support the virtual screening of millions of compounds, reducing traditional labor-intensive methods. Tools such as DeepMind's AlphaFold have revolutionized protein structure prediction, while generative adversarial networks (GANs) and other machine learning models enable the creation of chemical structures built for efficacy and safety.[3] [4]

AI tools deal with the vast, multimodal biological datasets powered to generate and analyze giant volumes of data, which support drug development that can reach clinical stages faster. The technology firms and pharmaceutical giants' businesses are evolving proportionally, with a nurturing of AI-driven startups and collaborations between them. For example, companies like Atomwise utilize deep learning for structure-based drug design, enabling the rapid exploration of vast chemical spaces and the identification of novel drug candidates. While AI holds promise, ensuring the responsible development and deployment of AI tools will be critical to realizing their full potential in delivering more effective, accessible, and affordable therapies to patients worldwide. [3]

CURRENT TRENDS OF AI IN DRUG DISCOVERY AND DEVLOPMENT:

AI technologies are trending in various stages of pharmaceutical research and new drug development with the goal of improving monitoring, accelerating quality and patient safety.

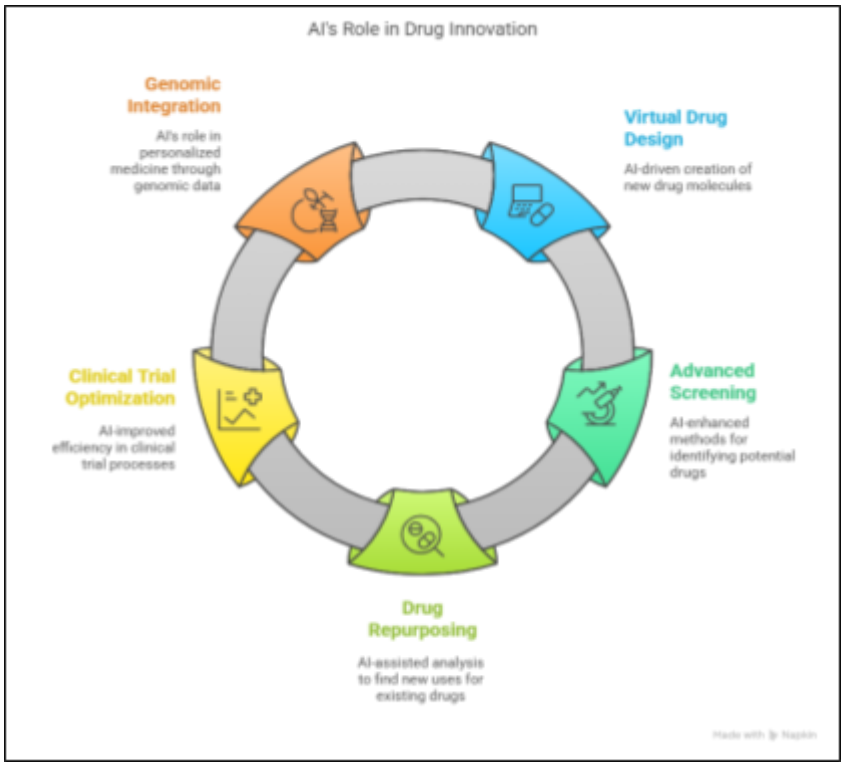

Fig. No.1.1 Current Role if AI in Drug Development [5]

1. **Virtual Drug Design:**

Virtual drug design accelerates the development of therapeutic compounds by unlocking active compound physical, chemical, and biological activity by computer-aided drug design (CADD) methodologies to predict and optimize drug-target interactions. Machine learning predicts reaction outcomes, optimizes conditions, and designs novel pathways.[6]

a. Structure-Based Methods

- Molecular Docking: Predict molecule binding to target proteins using tools like AutoDock Vina, Glide, and RosettaVS. Advanced platforms like RosettaVS incorporate receptor flexibility, improving pose and affinity predictions.
- Molecular Dynamics (MD) Simulations: Analyzes time-dependent molecular behavior with tools such as Gromacs and OpenMM.

b. Ligand-Based Methods

Pharmacophore Modeling: Identifies essential molecular features (e.g., hydrogen bonds, hydrophobic regions) using tools like ROCS.

- Shape/Field-Based Screening: Matches molecular shapes or electrostatic fields to known active compounds. [6][7]

2. Advanced screening and predictive modelling

Machine learning classifies compounds and analyse their active/inactive descriptors, then identify structurally similar compound by AI tool Random Forests (RF) and Support Vector Machines.Deep learning analyzes protein-ligand binding poses, capturing spatial interactions with 3D-CNNs. ADMET prediction through Graph Neural Networks (GNNs) AI Multi-task learning models define absorption, toxicity, and metabolic stability simultaneously, trained on datasets like PubChem etc. [8]

3. Drug repurposing and combination analysis

Drug repurposing is the searching a new therapeutic use of previously existing drug by Literature Mining, ML-based Prediction, Network Analysis, Matrix Factorization, Deep Learning renew understanding chemical, genomic features with tools like Matrix Decomposition, Collaborative Filtering, Encoder-Decoder, CNNs, RNNs etc. [9]

Clinical trial optimization

Al tools like consertAI CTO 2.0 assist in clinical trials by identifying participating volunteer disease conditions. The huge data of clinical trials are maintained and updated in the database through AI. Medidata AI emphasizes protocol design using historical trial data and also patient matching via AI algorithms for recruitment. Aicure was developed as a mobile application to measure medication adherence in a phase II trial. AI prediction of relevant biomarkers of disease allows specific recruitment of patient populations. This predictive modeling in the selection of patient populations enhances the rate of clinical trials. [10]

4. Combination of genomic and personalized medicine

BioGPT and MedPaLM, advanced AI models that change genomic data into personalized medicine, supporting clinicians with building effective treatment strategies, and treating patients. They support diagnosis, treatment planning, and patient communication, and care through real-time analysis, predictive modelling, etc. [10]

MECHANISM OF DRUG DISCOVERY AND DEVLOPMENT PROCESS THROUGH AI:

Artificial Intelligence (AI) plays a major role in modern drug development and research studies, ensuring drug efficacy, safety, and minimal toxicity to humans. From target identification to post-marketing studies, AI tools are transforming every stage of the drug development.[10]

AI tools have reduced the time span of drug discovery and enhanced the therapeutic effectiveness of new drugs. Primery-stage research, which traditionally took 2–3 years, can now be completed within 3–6 months, streamlining the entire development process. AI has revolutionized data management across clinical trials, chemical reports, and in vivo studies, making the organization, analysis, and complex datasets faster and more accurate.[2]

Table No.1.2: Drug Discovery Stages and their AI tools/Platforms.[3]

Stages	AI Role	Tools/Platforms
1. Target Identification	- Pattern recognition in omics data- Network biology analysis	DeepMind's AlphaFold
2. Biomarker Discovery	- Feature selection- Predictive modelling	SOPHiA GENETICS, Tempus, BioXcel
3. Lead Compound Identification	- QSAR modelling- ML-based compound prioritization	AtomNet (Atomwise), Chemprop, DeepChem
4. Virtual Screening & Optimization	- Molecular docking simulations- Reinforcement learning for scoring functions	Schrödinger, DeepDock, AutoDock Vina, DeepScreen
5. Chemical Structure Modification	- Generative models for structure optimization	REINVENT, GENTRL, MolGAN, ChemAxon
6. De Novo Drug Design *(Optional)*	- Generative AI (VAEs, GANs)- Reinforcement learning	Insilico Medicine, Deep Genomics, MolGAN, Junction Tree VAE
7. Preclinical Development	- Toxicity prediction- PK/PD simulation	ADMET Predictor, pkCSM, DeepTox, CytoReason

8. Clinical Trial Design	- Patient stratification- Trial simulation & prediction	Unlearn.AI, Trials.ai, QuantHealth, Deep 6 AI
9. Clinical Trials (Phases I-III)	- Anomaly detection- NLP on trial reports- Predictive analytics	Medidata, Flatiron Health, AiCure, DeepTrial
10. Regulatory Submission & FDA Approval	- Automated data extraction- Standardization & documentation	IQVIA, Accenture INTIENT, Genpact AI for Pharma Compliance
11. Post-Market Surveillance & RWE	- NLP for adverse event detection- Signal detection from real-world data	Aetion, Saama, Evidation Health, OpenAI Codex (for data parsing)

AI TOOLS USE IN DRUG DISCOVERY AND DEVLOPMENT:

Sr. No.	**Tools**	**Feature**	**Method**	**URL link**	**Reference**
1.	ChEMBL	Chemical database containing bioactive and drug molecules.	Machine learning	https://www.ebi.ac.uk/chembl/	[11]
2.	PubChem	database containing drug and their activity against biological targets.	Machine learning	https://pubchem.ncbi.nlm.nih.gov/	[12]
3.	AutoGrow 4	De novo drug design and lead optimization.	Genetic algorithm	https://durrantlab.pitt.edu/autogrow4/	[13]
4.	DEEPScreen	High-performance drug- target interaction prediction.	Convolutional	https://github.com/cansyl/DEEPscreen	[13]

			neural networks		
5.	QSAR modeling	Open-source toolkit for multi-target QSAR modeling.	Machine learning and classificati on model	https://github .com/ncordei rfcup/QSAR-Co-X	[13]
6.	ChemSAR	An online pipelining platform for molecular SAR modeling.	RDKit or ChemoPy package, scikit- learn package	http://chemsa r.scbdd.com/	[13]
7.	DeepConv -DTI	Prediction of drug-target interactions via protein sequences.	Deep learning	https://github .com/GIST-CSBL/DeepC onv-DTI	[13]
8.	ADMETla b	A platform for ADMET evaluation base on collected ADMET database.	Django framewor k in Python	http://admet.s cbdd.com/	[14]
9.	SMPDB 2.0	database for visualizing human metabolic, drug action, drug metabolism, physiological activity and metabolic disease pathways.	Machine Learning	http://www.s mpdb.ca/	[14]
10.	Ambit-SMIRKS	A software module for reaction representation, reaction search and structure transformation.	The Chemistry Developm ent Kit	http://ambit.s ourceforge.ne t/smirks	[14]

11.	TargetNet	A web service for predicting potential drug-target interaction profiling via multi-target SAR models.	Naïve Bayes models	http://targetnet.scbdd.com/	[14]
12.	DeepDrug	A general graph-based deep learning framework for drug relation prediction.	Graph convolutional networks	https://github.com/wanwenzeng/deepdrug	[14]
13.	IDDkin	Prediction of kinase inhibitors.	Deep diffusion model	https://github.com/CS-BIO/IDDkin	[14]
14.	MDeePred	Novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery.	Deep learning	https://github.com/cansyl/MDeePred	[14]
15.	Vienna LiverTox	Prediction of interactions profiles of small molecules with transporters relevant for regulatory agencies.	Machine learning classification model	https://livertox.univie.ac.at/	[14]
16.	DTI-CDF	A cascade deep forest model toward the prediction of drug-target interactions.	Deep forest model	https://github.com/a96123155/DTI- CDF	[14]
17.	DeepPurpose	Library for drug-target interaction prediction.	Deep learning	https://github.com/kexinhuang12345/DeepPurpose	[12]

18.	Neg Stacking	Drug-target interaction prediction.	Ensemble learning and logistic regression	https://github.com/Open-ss/NegStacking	[14]

19.	DrugNet	Network-based drug-disease prioritization by integrating heterogeneous data.	Machine learning	http://genome2.ugr.es/drug net/	[14]
20.	BioGRID	Database information about protein, genetic and chemical interactions	Machine Learning	https://thebiogrid.org/	[13]
21.	BioCyc	Database containing information of organism specific Pathway/ Genome Databases.	Machine Learning	https://biocyc.org/	[13]
22.	BRENDA	Database containing information of enzyme function data.	Machine Learning	https://www.brenda-enzymes.org/	[13]
23.	Reactome	database containing information of metabolic, protein trafficking and signaling pathways.	Machine Learning	https://reactome.org/	[13]
24	KEGG	Database containing information of genomic, chemical and functional information.	Machine Learning	https://www.genome.jp/ke gg/	[13]
25.	DrugMatri x	Database of toxicogenomic reference resource.	Machine Learning	https://ntp.niehs.nih.gov/data/drugmatrix/	[13]

APPLICATION/USES OF AI IN DRUG DISCOVERY AND DEVELOPMENT: [15] [16][10]

a) Property Prediction:

The number of molecules potentially synthesizable in a drug discovery program is huge. Some have put the number of possible organic molecules at 10^{60}, greater than the number of stars in the heavens. One way to restrict the number of molecules to be investigated is

to estimate the biological activity and physical characteristics of a molecule before it is synthesized. The predicted properties are then used to guide the priority of follow- up experiments.

b) Molecule Generation:

In the past few years, we have witnessed the advent of a new discipline called generative molecular design. In this method, a machine learning algorithm learns the rules and syntax of Chemistry from a given set of molecules and applies this knowledge to create new, novel molecules.

Generative models are more advanced than earlier de novo design methods that needed hard-coded rules and heuristics to extend and transform existing chemical structures.

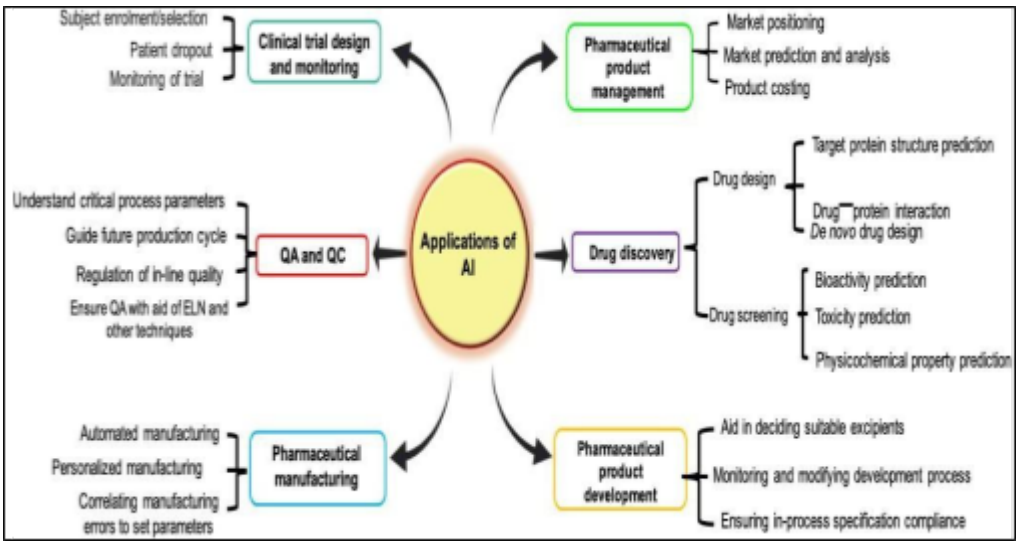

Figure No.1.2: Applications of AI in various fields such as Clinical trial design and monitoring, QA and QC, Pharmaceutical manufacturing, product management, and Drug discovery. [10]

c) Image Analysis:

In the past few years, we have witnessed the advent of a new discipline called generative molecular design. In this method, a machine learning algorithm learns the rules and syntax of Chemistry from a given set of molecules and applies this knowledge to create new, novel molecules. Generative models are more advanced than earlier de novo design methods that

needed hard-coded rules and heuristics to extend and transform existing chemical structures.

d) Organic Synthesis Planning:

The way that drugs and other organic molecules are constructed from chemical building blocks has always been as much an art as a science. Earlier efforts to automate route design from available reagents to complex molecules have been only partially successful. Deep learning is now being applied to search large databases of chemical reactions and suggest feasible routes for the synthesis of complex molecules.

e) Predicting drug–protein interactions:

Drug–protein interactions play a crucial part in the success of a treatment. Prediction of the interaction between a drug and a receptor or protein is important to determine its efficacy andthe effectiveness, facilitates drug repurposing and avoids poly- pharmacology. Several AI approaches have been helpful in the correct prediction of ligand–protein interactions, leading to improved therapeutic efficacy. The role of AI in the denovo molecular design advantage industry due to its numerous benefits, including offer ing learningconcurrent optimization.

Prediction of bioactivity:

The effectiveness of drug molecules is based on their affinity towards the target protein or receptor. Drug molecules that lack any interaction or affinity towards the target protein will be unable to produce the therapeutic effect. In a few cases, it may also be possible that drug molecules developed interact with non-target proteins or receptors, which may cause toxicity. Therefore, drug target binding affinity (DTBA) is critical to forecast drug–target inter actions. AI-based approaches are able to calculate the binding affinity of a drug based on either the features or similarities of the drug and its target. Feature-based interactions identify the chemical moieties of the drug and of the target in order to find the feature vectors. Conversely, in similarity-based interaction, similarity between drug and target is taken into account, and it is presumed that similar drugs interact with similar targets.

f) Prediction of toxicity:

Toxicity prediction of any drug molecule is critical to prevent toxicities. Cell-based in vitro assays are usually utilized as initial studies, and then animal studies to determine the toxicity of a compound, making drug discovery more costly. A number of web-available tools, including LimTox, pkCSM, admetSAR, and Toxtree, are available to make the process cost-effective. Sophisticated AI-driven methods search for similarities between compounds or predict the toxicity of the compound on the basis of input features. The National Institutes of Health, Environmental Protection Agency (EPA), and US Food and Drug Administration (FDA) hosted the Tox21 Data Challenge, which was an effort to compare various computational methods to predict the toxicity of 12,707 environmental chemicals and drugs; an ML method called DeepTox surpassed all approaches by finding static and dynamic features in the chemical descriptors of the molecules like molecular weight (MW) and Van der Waals volume, and was able to accurately predict the toxicity of a molecule on the basis of pre-defined 2500 toxicophore features. The different AI tools used in drug discovery are listed in Table No.:

g) Target identification and validation:

The drug discovery process begins with molecule searching. Specific small molecule databases can match given health issues. Ranges of docking software are designed to examine molecular-bioactivity. Scientists also employ analogues

of previously known molecules.

h) Biomarker discovery and validation:

During the age of molecular medicine, the process of drug discovery is aided by biomarker,discovery. Discovery of a biomarker necessitates the gathering and proper e xamination of a high volume of samples uniformly. Validation proves that the marker is reliable, reproducible, and the sensitivity and specificity of the marker are adequate. AI may be used in this step. In the drug development scenario, biomarkers serve as an outcome measure in clinical trials to assist with the identification and validation of drug targets. Therefore, the appropriate treatment for each patient according to the biomarkers measured would be determined.

i) Pre-clinical and clinical development:

Prediction of potential responses to a drug is a key step within a drug design pipeline. Both similarity or feature-based machine learning approaches can predict the response of a drug

to individual cells as well as drug-target interaction effectiveness by binding affinity or free energy of binding. Similarity approaches have the hypothesis that comparable drugs affect similar targets, and feature-based approaches identify a unique feature of the drug and the target and give the classifier the drug-target feature vector. One example is a deep learning- based approach DeepConv-DTI that learns drug and target embeddings from convolutional structure and attention mechanism. AI-based methods can help in choosing potential patients for pre-clinical trials by finding relevant human-disease bio markers and predicting possible toxic or unwanted side effects and by filtering a high dimensional set of clinical variables to choose a cohort of patients. AI can also aid in predicting the result of clinical trials well before the actual trial reducing the likelihood of any adverse effect on patients.

CHALLENGES AND FUTURE DIRECTION:

data generation has increased due to the introduction of technology and advancements in processing power, creating a number of massive compound data sources. This required many artificial intelligence and machine learning techniques into practice. In different phases of drug research in order to extract medicinal knowledge from massive amounts of "big" data. [13]

discovering new compounds and further making them has been triggered these AI algorithms in big data. This method has advanced the drug development process by automating and personalizing it and confirming the importance of big data. In addition to drug screening, artificial intelligence (AI) techniques are used at various phases of the drug discovery cycle, including patient recruitment and surveillance, protein structure prediction, ADME characteristics, toxicity, bioactivity, and physical property prediction. Aside from the wide range of uses for AI-based technology, there are certain restrictions and difficulties to be resolved.[9]

The success of AI-based technologies depends on how frequently and easily consumers can get data its better data management and curation as well as user-friendly online portals. Thus, to extract meaningful information, trustworthy and well-curated data is necessary. The absence of sufficient curated data and data accessibility can prove to be a barrier. quickly and consistently updating the software that is now accessible in accordance with the format of the data that is generated and the newly created algorithm. Furthermore, there is a shortage of qualified workers for the full-fledged operation of AI-based drug discovery applications. [10]

Protein binding affinity with therapeutic molecule and target that may be not investigated properly. The length and quality of the data are the only limitations of this technique because deep learning necessitates large amounts of data. [7]

CONCLUSION:

AI help in determining the best drug dosage form, produce least toxic drugs with reliable outcomes, assist with clinical trial safety and efficacy, and conduct an extended market analyzation and studies. This AI-aided algorithms reduce the time of drug to reach the market, and increase the quality of the product in the market, management of resource for production, cost management. This AI and its tools give chance to pharmaceutical companies for new drug development, Drug repurposing and combination analysis this increase start-up from this field. As we improve AI tools to overcome challenges, we can expect more from AI in drug development.

REFERENCES:

1. Tripathi, K. (2019). *Essentials of Medical Pharmacology* (8th Edition ed.). Jaypee Brothers Medical Publishers (P) LTD, New Delhi.

2. Bharadwaj, S. D. (2024). Exploring the Artificial Intelligence and Its Impact in Pharmaceutical Sciences: Insights Toward the Horizons Where Technology Meets Tradition. *Chemical Biology & Drug Design, 104*(4), 1-23. doi:https://doi.org/10.1111/cbdd.14639

3. Rajagopalan, R. &. (2024). A Role of Artificial Intelligences in Drug Discovery and Drug Development – A Critical Review. *INTERNATIONAL JOURNAL OF PHARMACEUTICAL QUALITY ASSURANCE, 15*(3), 1714-1722. doi: 10.25258/ijpqa.15.3.94

4. U.S. Food and Drug Administration. (2023). *Artificial Intelligence in Drug Manufacturing, Drug Development, and Drug Use: Discussion Paper.* U.S. Department of Health and Human Services, united states. Retrieved from https://www.fda.gov/media/167973/download

5. Mohapatra M, S. C. (2024). Trends of Artificial Intelligence (AI) Use in Drug Targets, Discovery and Development: Current Status and Future Perspectives. Curr Drug Targets. doi:10.2174/0113894501322734241008163304

6. Niazi, S. K. (2024). Computer-Aided Drug Design and Drug Discovery: A Prospective Analysis. *Pharmaceuticals, 17*(1), 22. doi:https://doi.org/10.3390/ph17010022

7. Zhou, G. R. (2024). An artificial intelligence accelerated virtual screening platform for drug discovery. *Nat Commun, 15*. doi:https://doi.org/10.1038/s41467-024-52061-7

8. Carpenter, K. A. (2018). Machine Learning-based Virtual Screening and Its Applications to Alzheimer's Drug Discovery: A Review. . *Current pharmaceutical design, 24*(28), 3347–3358. doi: https://doi.org/10.2174/1381612824666180607124038

9. Daniele Parisi, M. F. (2020). Drug repositioning or target repositioning: A structural perspective of drug-target-indication relationship for available repurposed drugs,. *Computational and Structural Biotechnology Journal, 18*, 1043-1055. doi:https://doi.org/10.1016/j.csbj.2020.04.004.

10. Mak, K. K. (2019). Artificial intelligence in drug development: present status and future prospects. *Drug discovery today, 24*(3), 773–780. doi: https://doi.org/10.1016/j.drudis.2018.11.014

11. Wei Chen, X. L. (2023). Artificial intelligence for drug discovery: Resources, methods, and applications. *Molecular Therapy: Nucleic Acids Vol. 31*, 693.

12. Rizwan Qureshi, M. I. (2023). AI in drug discovery and its clinical relevance. *Heliyon 9*, 5-6.

13. Manish Kumar Tripathi, A. N. (2021). Evolving scenario of big data and Artificial Intelligence (AI) in drug . *Molecular Diversity*, 1443-1454.

14. Rohan Gupta, D. S. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular Diversity*, 18-24.

15. Maria Kokudeva, M. V. (2024). Artificial intelligence as a tool in drug discovery and development. *World Journal of Experimental Medicine*, 96042.

16. Debleena Paul, G. S. (2021). Artificial intelligence in drug discovery and development.

17. *Drug Discovery Today*, 80-90.