International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 5th May 2025. Link

https://ijiemr.org/downloads.php?vol=Volume-14& issue=issue05

DOI: 10.48047/IJIEMR/V14/ISSUE 05/22

Title DRUG DISCOVERY BY AI TOOLS

Volume 14, ISSUE 05, Pages: 237 – 246

Paper Authors
Surendra Babu .L, Ravi Anitha M, Durga Satya Sri. K, Venkata Sai Surya Rekha Sree. B, Kokkeragadda Mohith. K, Ramya Sree. B, Vikash Babu. K

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

DRUG DISCOVERY BY AI TOOLS

Surendra Babu .L*, Ravi Anitha M, Durga Satya Sri. K, Venkata Sai Surya Rekha Sree. B, Kokkeragadda Mohith. K, Ramya Sree. B, Vikash Babu. K
Department of Pharmaceutical Chemistry,
Adikavi Nannaya University College of Pharmaceutical Sciences, Tadepalligudem Campus.
West Godavari, Andhra Pradesh, India.

Abstract
Artificial Intelligence (AI) is revolutionizing the landscape of drug discovery, offering tools that streamline each phase of the process. This mini-review will provide an overview of how the integration of Artificial Intelligence tools in the drug discovery process accelerates the timelines, minimizes failure risks, and lowers overall costs. This manuscript introduces key AI terminology, describes its role in various stages of the drug discovery process, and emphasizes on few popular AI tools used in the discovery process. Additionally, it also discusses the advantages of AI adoption in drug discovery and addresses the challenges faced in its implementation. This review aims to elucidate the current state and prospects of AI-driven drug discovery.

Keywords: Artificial Intelligence, Tools, Drug Discovery, DNA Technology, Machine Learning

Introduction
Artificial intelligence is a generic term that encompasses multiple technologies. In simplified terms, AI is a transformative technology within computer science that helps computers perform tasks by mimicking human cognitive functions to solve real-world problems. Recently, AI has become an integral part of the drug discovery process, including peptide identification, drug design, drug efficacy, and effectiveness assessment, prediction of drug toxicity and physiochemical properties, drug repurposing, Quantitative Structure- Activity Relationship (QSAR), and the design of clinical trials [1]. In this manuscript, to give an overview, we first discussed about conventional drug sdiscovery process, followed by describing various AI technologies. Further, the article describes the AI-based tools employed recently to aid in the drug discovery process. Lastly, the advantages and disadvantages of applying AI in the field of drug discovery are discussed.

Conventional Approaches in Drug Discovery
An investigational compound is known as a New Chemical Entity (NCE) until it gets investigated for a particular condition. Once demonstrates some activity, it is then classified as lead compound, which is further optimized to show better activity and/or fewer side effects. Lead compounds can be obtained through various sources, from

ancient empirical knowledge to structure or computer-based drug design. ssssThe ancient drug discovery process mainly relied on a trial-and-error approach to identify potential medicinal uses of products from plants and animals [2]. Since there was an absence of knowledge about the disease mechanism and molecular targets, although many compounds identified by this process showed efficacy, there were many side effects observed. However, these empirical methods laid the foundation of drug discovery by providing lead compounds. Over decades, with a better understanding of disease processes, lead compounds are emerging by specifically targeting signaling pathways involved in the etiology of the disease.

The modern drug discovery process consists of multiple approaches and stages (Figure 1). It starts with the identification of a lead compound, which could be obtained from the isolation of products from natural sources (such as vincristine and vinblastine from *Catharanthus Roseus*), fermentation coupled with the recombinant DNA technology, synthetic chemistry, and combinatorial chemistry, etc. Lead compounds are then generated by creating and screening the large chemical libraries, employing high-throughput screening methodologies or virtual screening techniques, or utilizing fragment-based lead discovery [3]. Recently, lead compounds are increasingly identified through the use of small molecular weight fragments that bind to target proteins. Once these fragments are discovered, they are further developed by adding chemical groups to create more potent lead molecules [4]. These initial leads undergo optimization, where new analogs will be synthesized to further enhance the efficacy, and bioavailability and reduce the side effects of hits. most promising candidates then enter pre-clinical testing, where their efficacy and safety are evaluated in animal disease models. Compounds that show potential in pre-clinical stages advance to human clinical trials, undergoing extensive testing to further ensure safety and efficacy in humans [5]. Additionally, finding alternative applications of the existing medications through drug repurposing has advantages over conventional drug discovery as it cuts down costs and time duration for the invention of a novel molecule to cure a particular disease. Only after successfully completing these rigorous phases can a drug be considered for market entry. This comprehensive process combines traditional methods with cutting-edge techniques (Figure 1).

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

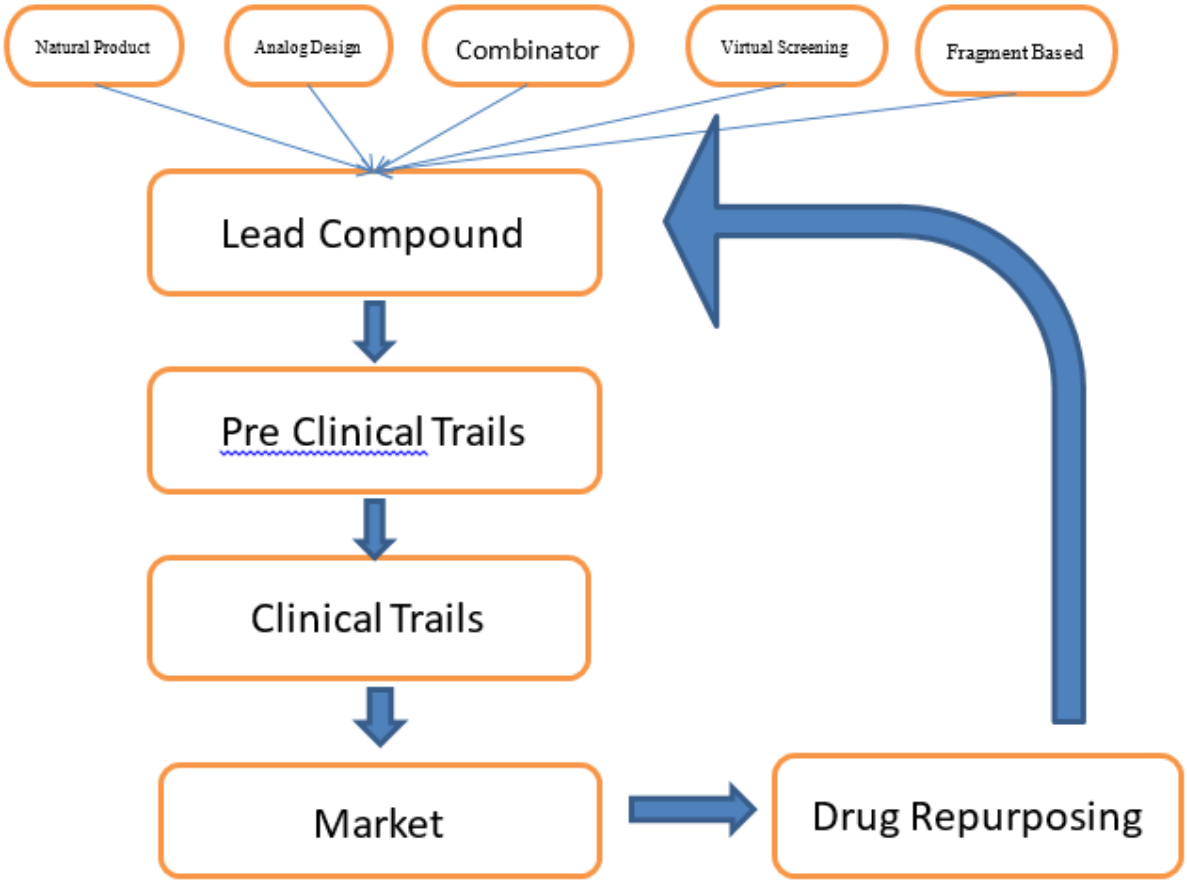

Figure 1: Traditional drug discovery process.

Artificial Intelligence Technologies in Drug Discovery

Artificial Intelligence (AI) is a broad umbrella term used for a variety of technologies designed to mimic human intelligence such as Machine Learning (ML), Deep Learning (DL), and Artificial Neural Networks (ANN). Sometimes the terms for these technologies interchangeably create confusion. Broader definitions of these terms are as below:

A. Artificial Intelligence (AI): AI is a broader term that involves the use of algorithms and data processing to enable machines to simulate human intelligence and cognitive functions and solve complex real-world problems.

B. Machine Learning (ML): A subset of AI that uses algorithms and data analytics to predict the outcomes more accurately.

C. Deep Learning (DL): DL is a further subset of ML that utilizes large data sets and complex algorithms such as image processing and speech recognition for prediction purposes.

D. Artificial Neural Networks (ANN): ANNs are computational models that mimic the functional design of neurons in the brain, which consists of interconnected nodes. Most deep learning algorithms are ANN-based in design; however, ANN can be used for other functions beyond DL.

Artificial intelligence in drug discovery

Recent advancements in GPU technology and deep learning algorithms have led to an explosion of AI applications in the drug

discovery process. These sophisticated algorithms can process large datasets, make automatic decisions, and predict crucial pharmaceutical parameters from diverse information sources without performing costly and time-consuming laboratory-based experiments. Therefore, AI is accelerating the drug discovery process at various stages--drug design, drug screening, polypharmacology, execution of clinical trials, and drug repurposing (Figure 2). In the drug design field, AI can predict the 3D structure of a target protein and its interaction with NCEs. Further, it can also design new drug-like molecules without a starting template, known as de-novo drug design [6]. For example, Protein-Protein Interactions (PPIs) are examined using AI tools like iFitDock, which explores the entire protein surface using small fragment-sized probe molecules to predict interactions with other proteins and small molecules [7]. Further, in the drug screening process, AI can predict various physicochemical properties of the drug, such as log P and solubility, as well as ADME (absorption, distribution, metabolism, and excretion) properties. It can also predict the potential bioactivity and toxicity of the NCEs. Accurately predicting the above critical properties of a drug candidate not only accelerates the drug discovery but also significantly reduces the likelihood of failure in later stages of development. Similarly, deep learning algorithms have been used in designing molecules with multiple targets or predicting multiple targets of new drug molecules (polypharmacology) [8,9]. In later stages of drug discovery, and clinical testing of the molecules, AI could analyze large data from previous trials and patient records and optimize trial design and execution.

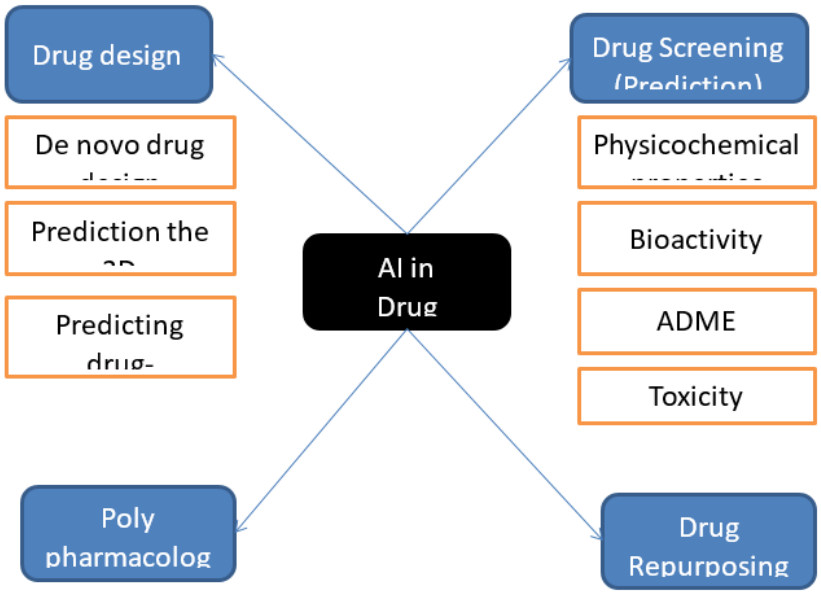

Figure 2: AI applications in drug discovery.

Machine learning in drug discovery

Machine learning is a subset of AI that develops complex algorithms and statistical models that accurately predict drug properties such as chemical, physical, and biological characteristics of NCEs. These ML

tools can be incorporated at any stage of the drug discovery process starting from target identification to clinical trials. Machine Learning is broadly classified into supervised and unsupervised learning. The supervised machine learning tools learn from the data set of known examples. In this method, during the learning process, the algorithm is trained with known input and output data sets to identify patterns. By recognizing these patterns, it can then make predictions about new, unseen data. As an application in drug discovery, this could mean predicting the efficacy or side effects of NCEs based on their structural similarities to existing drugs.
In contrast, in an unsupervised learning model, the algorithm is only provided with input data without any corresponding output information. The tool is expected to discover patterns and structures within this data on its own without training [10]. In the end, the high dimensional data is reduced to a lower dimension to identify the patterns that are easy to interpret. This further includes data visualization, extraction of desired features, and clustering of the data. Using unsupervised ML, the scientists could cluster similar compounds based on identifying structural patterns for potential therapeutic applications or toxicities [11]. For example, a new potential antibiotic, Halicin was discovered by unsupervised learning through a clustering process [12].

Deep learning in drug discovery

Deep learning is a subfield of machine learning that utilizes artificial neural networks with multiple layers to analyze complex patterns in large datasets. Since it has capabilities for image interpretation, structure prediction, and computing large sets of data, it has broader and deeper applications in drug discovery [13]. Its integration in various steps of drug discovery assists in predicting target identification, drug interactions, molecular properties, and biological and toxicological responses.

DL models predict Drug-Target Interactions (DTIs) and Drug- Drug Interactions (DDIs), which are essential for identifying therapeutic candidates and preventing adverse effects. It is used in molecular docking as it predicts the pose of ligands based on interactions observed in protein-ligand complexes. For example, TANKBind, a trigonometry-Aware neural network predicts Drug- Protein binding by segmenting the protein into blocks [14]. They also facilitate personalized medicine by analyzing patient data to forecast individual drug responses and potential side effects. In personalized medicine, Deep learning techniques have shown particular progress in early diagnosis, and identifying cancer that assists in treatment decisions [15]. Recently, deep learning techniques have been applied to generate new chemical structures by using a Variational Autoencoder (VAE). Kadurin et al. [16] used VAE as a molecular descriptor generator which is coupled with a generative adversarial network to generate new molecular structures that were predicted to have anti-neoplastic properties [16]. Blaschke et al. [9], have utilized VAE to generate novel molecules which have predicted activity against dopamine type 2 receptor [17].

Deep Learning (DL) offers several advantages, including improved predictive accuracy, automated feature engineering, and the ability to handle complex, multi-dimensional datasets that are common in drug discovery. Overall, DL is transforming drug development by enhancing both speed and accuracy. However, to fully realize its potential in the pharmaceutical industry, challenges related to data quality, interpretability, and integration need to be addressed. Future research is expected to

focus on building robust datasets and enhancing the explainability of these models [18] (Figure 3).

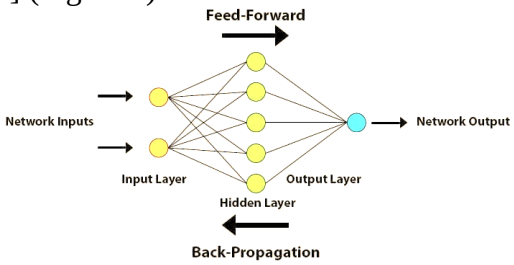

Figure 3: Transformation of neural network in humans to artificial neural network

Artificial Neural Networks

Artificial Neural Networks (ANN) consist of interconnected nodes which process information similar to neurons in humans. A typical ANN structure includes ANN's an input layer that receives the initial data, then one or more hidden layers, where the data is transformed or processed through various weighted connections and activation functions, and finally, the output layer that presents the final decisions. ANNs have different array of network types, each specialized in a particular function. The perceptron is the oldest neural network created by Frank Rosenblatt in 1956. The dual function of ANN, capable of both classification and prediction, has made its way into the AI domain [19].

There are different types of artificial networks including:

Feedforward neural network

FNN is the fundamental neural network that has the simplest construction. FNN is composed of an input layer and an output layer. It might have one hidden layer included if the data set provided is more complex. The input layer receives the values, the hidden layer processes the input, then the output layer produces an output. FNN is formed with the help of a training sample set which is a data set of examples used during the learning process which is used to fit the parameters and backpropagation-gradient descendent algorithm, backpropagation and gradient descendent are used to improve the accuracy of the neural networks by improving the output of the neural network [20].

Multilayer Perceptron (MLP):

MLP is a subpart of a feed-forward neural network with a few more interconnected layers to process complex data. It is comprised of an input layer, a hidden layer, and an output layer with an additional bias term to set a threshold for neuron activation (Figure 4). They are composed of sigmoid neurons, which use sigmoid function. Sigmoid function gives continuous output which would be a number between 0 and 1. It uses the back proportion algorithm, which is a gradient estimation method to train neural network models. It allows adjustment within the network based on the error between the actual and the expected outputs. This error correction helps MLP to make predictions precisely and enhance its accuracy. In drugdiscovery, MLP is used for image recognition, natural language processing, and speech recognition [21].

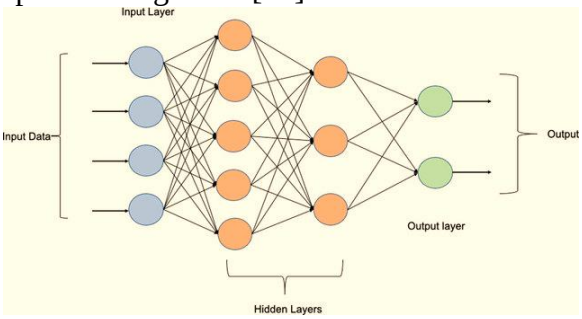

Figure 4: Multilayer perception architecture.

Convolutional Neural Network:

CNN is a type of deep learning algorithm that is mainly designed for image recognition, including image classification, detection, and segmentation. It uses

convolutional layers which have filters to extract features from input data at a large scale to recognize patterns within the image. CNN has subsampling layers which reduce the memory requirement and enhance the learning speed. In drug discovery, they are used to analyze one-dimensional data such as a nucleotide or in QSAR and ligand-based bioactivity predictions. An AI tool, Atom Net which predicts new molecules for different targets is constructed with the help of CNN [22].

Deconvolutional neural network:

DNN reconstructs high-resolution images from lower-resolution inputs or compressed representations. This is achieved through up-sampling which increases the spatial resolution of the feature maps. DNNs consist of two main parts: an encoder and a decoder. The encoder will extract low- and high-level features from the input and then the decoder will reconstruct a high-resolution output image through up-sampling layers [23].

Recurrent neural network:

It is an advanced class of ANN that processes sequential data. In this, the output from the previous step is fed as an input to the current step. The hidden state stores the information of the data in each step. RNN has excelled in its modeling of protein-ligand interactions as it can recognize the sequential relationship between amino acids in proteins and atoms in ligands, which is important for predicting the binding affinity of drug molecules to target the protein of interest.

AI Tools in Drug Discovery

In the rapidly evolving landscape of pharmaceutical research, Artificial Intelligence (AI) tools are being developed and implemented rapidly to revolutionize the drug discovery process. It is important to note that this selection represents only a fraction of the AI innovations currently in the pharmaceutical industry, with new tools and approaches emerging regularly to address specific challenges in drug development.

Software and Applications

S.No.	Software	Significances
1,	Alpha Fold	3D shape of proteins, which is crucial for drug design.
2,	**IBM Watson**	Hidden connections between genes, drugs, and diseases.
3,	EVE Scape	Identifying potential escape mutations can inform our understanding of viral evolution.
4,	Organic	AI can process large amounts of chemical and biological data to identify potential drug candidates.
5,	Addison	Predict the efficacy and safety of potential drug candidates, reducing the need for physical experiments.
6,	Atom Net	Atom Net can accurately predict how small molecules bind to proteins, helping researchers identify potential drug candidates.
7,	Deep tox	Predict the toxicity of small molecules, enabling researchers to prioritize safer candidates.
8,	Seismic therapeutic – IMPACT	Specific disease mechanisms.
9,	Rose TTAFold	Optimizing lead compounds.

10,	Delta vina	Accurately predicting the 3D structure of proteins from their amino acid sequences.
11	Support Vector Machine (SVM)	It supervises machine learning algorithm that classifies data by finding an optimal line or hyperplane that maximizes the distance between each class in an N-dimensional space.

Machine Learning Tools in Drug Discovery

12	Random Forest (RF)	Random Forest's versatility and accuracy make it a valuable tool in drug discovery, enabling researchers to identify promising compounds and optimize their properties.
13	Naive Bayesian (NB)	Naive Bayes can identify potential lead compounds by predicting their binding affinity to a target protein.
14	K-means clustering	Classify protein-ligand interactions based on binding affinity and other properties.

S.No.	Merits	Demerits
01	Reduce human errors	Cost overruns
02	Provides digital assistance	Dearth of talent
03	Helps in lessening repetitive work	Lack of practical products
04	Faster and more accurate decisions	Potential for misuse
05	Better human workflows are discovered	High production cost
06	Employing AI in high –risk circumstances	Making human lazy
07	24/7 availability	Rise in unemployment

Conclusion

Many aspects have influenced the integration of AI and ML into drug discovery particularly in areas like drug design and drug screening. The technological advancements have helped in cutting the time and expense of the research, development, and production and also helped in boosting efficiency. This research paper highlights that AI and Machine Learning can significantly enhance the accuracy of drug discovery and development. They can even replace the clinical trials with stimulations which makes the researchers understand the study of molecules without trials. However, AI and machine learning can develop drug discovery

methods yet several challenges remain. Issues like the cleaning of unstructured and heterogeneous datasets and the occasional limitations of computing devices could hinder progress. Overcoming these barriers will be the key to fully leveraging AI and machine learning, paving the way in the pharmaceutical industry.

References

1. Vijayan R, Jan K, Jason BC, Vasanthanathan P (2022) Enhancing preclinical drug discovery with artificial intelligence. Drug Discovery Today 27(4): 967-984.
2. Dias DA, Urban S, Roessner U (2012) A historical overview of natural products in drug discovery. Metabolites 2(2): 303-336.
3. Erlanson DA, McDowell RS, Brien TO (2004) Fragment-based drug discovery. Journal of Medicinal Chemistry 47(14): 3463-3482.
4. Scott DE, Anthony GC, Sean AH, Chris A (2012) Fragment-based approaches in drug discovery and chemical biology. Biochemistry 51(25): 4990-5003.
5. Khan S, Jaya A, Sunanda P, Nikhat K (2023) Drug repurposing: A futuristic approach in drug discovery. J Pharm Biol Sci 11(1): 66-69.
6. Singh S, Kaur N, Gehlot A (2024) Application of artificial intelligence in drug design: A review. Computers in Biology and Medicine 179: 108810.
7. Abbas M, Abrar R, Fatima K, Rehab A, Maha A (2024) The role of ai in drug discovery. Chembiochem 25(14): e202300816.
8. Anstine DM, Isayev O (2023) Generative models as an emerging paradigm in the chemical sciences. Journal of the American Chemical Society 145(16): 8736-8750.
9. Blaschke T, Bajorath J (2022) Fine-tuning of a generative neural network for designing multi-target compounds. Journal of Computer-Aided Molecular Design 36(5): 363-371.
10. Polanski J (2022) Unsupervised learning in drug design from self-organization to deep chemistry. International Journal of Molecular Sciences 23(5): 2797.
11. Dara S, Swetha D, Surender SJ, Madhu BCh, Mohamed JA (2022) Machine learning in drug discovery: A review. Artificial Intelligence Review 55(3): 1947-1999.
12. Stokes JM, Kevin Y, Kyle S, Wengong J, Andres CR (2020) A deep learning approach to antibiotic discovery. Cell 180(4): 688-702.
13. Liu W, Zhang J, Zhao Y (2022) A comparison of deep learning and traditional machine learning approaches in detecting cognitive impairment using mri scans. IEEE. pp. 998-1001.
14. Lu W, et al. (2022) Tankbind: Trigonometry-aware neural networks for drug-protein binding structure prediction. Advances In Neural Information Processing Systems 525:7236-7249.
15. Liu X, Livia F, Aditya UK, Siegfried KW, Dun JF, et al. (2019) A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. The Lancet Digital Health 1(6): e271-e297.
16. Kadurin A, Nikolenko SI, Khrabrov K, Aliper A, Zhavoronkov A (2017) druGAN: An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico.

Molecular Pharmaceutics 14(9): 3098-3104.

17. Kang S, Morrone JA, Weber JK, Cornell WD (2022) Analysis of training and seed bias in small molecules generated with a conditional graph-based variational autoencoder-insights for practical AI-driven molecule generation. Journal of Chemical Information and Modeling 62(4): 801- 816.
18. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T (2018) The rise of deep learning in drug discovery. Drug Discovery Today 23(6): 1241-1250.
19. Abbas M, Rassam A, Karamshahi F, Abunora R, Abouseada M (2024) The role of AI in drug discovery. Chem Bio Chem 25(14): e202300816.
20. Sazlı MH (2006) A brief review of feed-forward neural networks. Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering 50(1).
21. Popescu MC, Balas VE, Perescu-Popescu L, Mastorakis N (2009) Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems 8(7): 579-588.
22. Wallach I, Dzamba M, Heifets A (2015) AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. arXiv preprint arXiv:1510.02855.
23. Cao F, Yao K, Liang J (2020) Deconvolutional neural network for image super-resolution. Neural Networks 132: 394-404.
24. Neves GF, Chaudron JB, Dion A (2021) Recurrent neural networks analysis for embedded systems. IJCCI 1: 374-383.
25. Patel L, Tripti S, Xiuzhen H, David WU, Shanzhi W (2020) Machine learning methods in drug discovery. Molecules 25(22): 5277.
26. Yadav S, Abhishek S, Rishika S, Jagat PY (2024) Revolutionizing drug discovery: The impact of artificial intelligence on advancements in pharmacology and the pharmaceutical industry. Intelligent Pharmacy 2(3): 367-380.
27. Shiammala PN, Navaneetha KBD, Baskaralingam V, Abdulaziz SA, Esam SAlM, et al. (2023) Exploring the artificial intelligence and machine learning models in the context of drug design difficulties and future potential for the pharmaceutical sectors. Methods 219: 82-94.
28. Cong Y, Endo T (2022) Multi-omics and artificial intelligence-guided drug repositioning: Prospects, challenges, and lessons learned from COVID-19. OMICS 26(7): 361-371.
29. Mukherjee S, Vagha S, Gadkari P (2024) Navigating the future: A comprehensive review of artificial intelligence applications in gastrointestinal cancer. Cureus 16(2): e54467.
30. Terwilliger TC, Dorothee L, Tristan IC, Christopher JW, Airlie JM (2024) AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. Nature Methods 21(1): 110-116.
31. Tiwari PC, Rishi P, Manju JC, Rajendra N (2023) Artificial intelligence revolutionizing drug development: Exploring opportunities and challenges. Drug Development Research 84(8): 1652-1663.