Journal of Army Engineering University of PLA ISSN: 2097-0730
陆军工程大学学报

AN INTELLIGENT APPROACH FOR DE-NOVO DRUG DISCOVERY: A SYSTEMATIC REVIEW

Muzaffar Ahmad Sofi[1], Dhanpratap Singh[2], Tawseef Ahmed Teli[3*]

[1,2] LPU, Jalandhar Punjab,
INDIA
[1]muzaffarsofi.g@gmail.com
[2]dhanpratap.25706@lpu.co.in
[3*] Higher Education Department, JK, INDIA
mtawseef805@gmail.com

Abstract. The drug is a substance that when introduced into the body brings physiological effects in the body. Drug discovery is the process comprising many stages of creating such substances to bring physiological effects for particular target diseases. Techniques like machine learning (ML) and deep learning (DL) have been successful in streamlining this difficult, expensive, and time-consuming drug discovery process. Due to the availability of abundant and quality data, ML and DL methods have been implemented at the different phases of drug discovery that include; target identification, target validation, drug-target interaction, lead optimization, etc., and have reduced manifolds the overall time drug introduction in the market. The de-novo method creates fresh chemical structures out of basic building blocks without the use of past knowledge or connections. By employing computational growth techniques, de-novo describes the creation of unique chemical structures that adhere to a group of restrictions. De-novo drug design has a number of benefits, such as the ability to search a larger chemical space, the creation of molecules with unique desirable features, and the quick and economical production of therapeutic candidates.

Keywords: De novo, Physiological, Machine Learning, Deep Learning, Cancer, Drug Discovery.

1. Introduction

The drug is a substance that when introduced into the body brings physiological effects in the body and drug discovery is a pipelined process for identifying such elements. It is a profoundly complex, costly, cumbersome, and multi-factor dependence task. On average, a new drug molecule costs around $2.6 billion for identification and development [1]. The main aim of a drug, usually a protein, is to get attached to the target protein in the body whose modification can bring a change in the target disease. Despite taking an enormous time and money the success rate of discovering a new drug is modicum, not satisfactory. The overall pipeline has largely been unfruitful. On the other hand, living beings (humans, more specifically) are continuously fighting diseases, which makes discovering new drugs a vital task for the survival of humans and has compelled scientists to go for the optimization of this drug discovery. In this vast field, which has been mostly considered as barren as far as the process's efforts (time, cost and output) are concerned, traces of fertility are seen after the

application of data-drivenmethods like ML and DL. The trend of the articles published in drug discovery using ML and DL approaches is shown in Fig. 1.

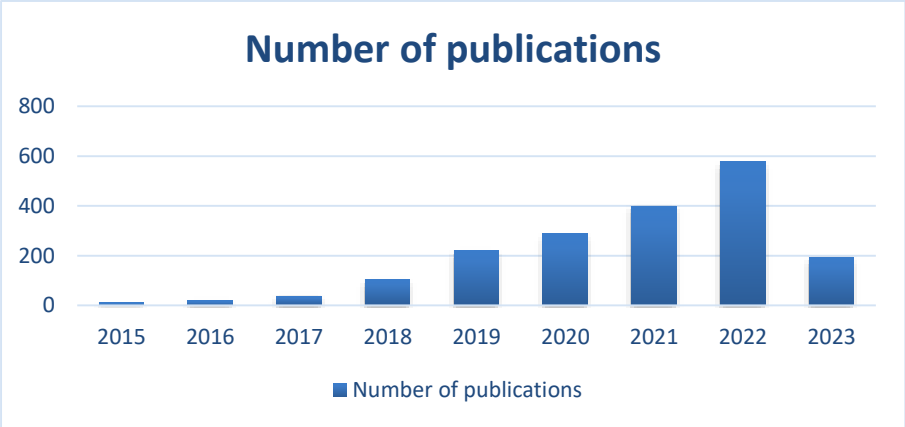

Fig 1. Papers published over the years

2. ML Approaches in drug discovery

Great benefits are witnessed by pharmaceutical companies through the implementation of ML techniques in development of new drugs. A large number of ML-based models were developed for predicting the features which may be, chemical, biological, or physical, of various compounds that aid in drug discovery [5–13]. All stagesof drug discovery incorporate ML models. These techniques have been extensively used for drug repurposing, to analyze and successfully predict drug-protein interactions and discover and establish drug efficacy etc [14–18].

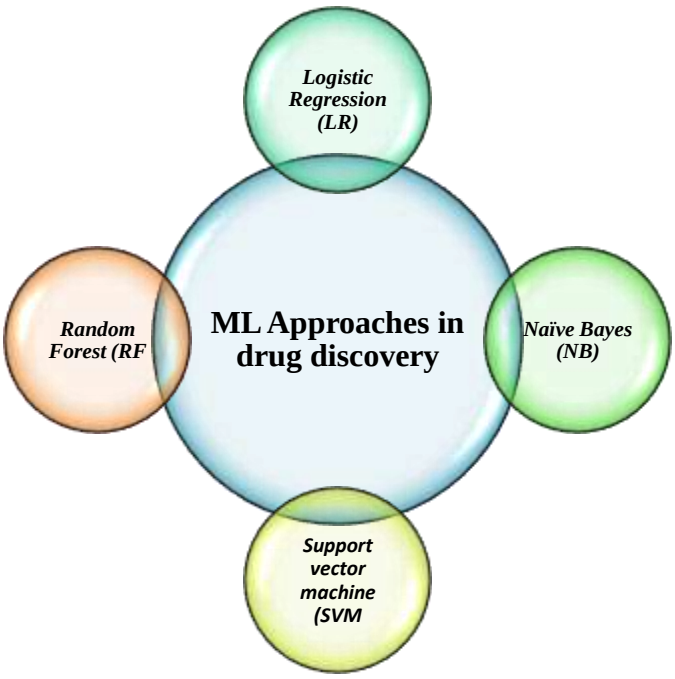

Fig 2. AI and drug discovery phases.

Various ML techniques used in drug discovery include Random Forest,Naïve Bayes, SVM, etc. as shown in Fig 2.

Random Forest is a reworking of decision trees [19]. It is a set of different classifiers returning the mode of the outputs of each decision tree (flow-chart type classification algorithm [20]). Random forest prevents over-fitting by reducing estimator variances. Random Forest is considered to be one of the revolutionary algorithms in ML. The random forest has been used in drug-target interaction prediction, used with lasso [21]. The random forest has also been used in response prediction, and improvement in scoring function performance in binding affinity [22]. A random forest-based method SMRF (scalable map-reduced random forest) [23] has been proposed for big data learning. Initialization, generation and voting are the three steps of the algorithm's operation. To explain attributions, the initialization process creates a file that acts as a descriptor. The generating step entails randomly splitting the original dataset into numerous subgroups and using the bagging algorithm and bootstrap samples for building the random forest which is created from the training dataset. In the voting step, the decision trees vote on the outcome.

The question that makes RF significant to drug discovery pertains to some of the following reasons:

- Quicker training process
- The use of a reduced number of parameters
- Easy handling of missing data [24]

Another variation of RFs known as multivariate random forest assists in reducing error. The data containing genetic information is used to analyse and extrapolate the mean and confidence interval of drug responses, which is a significant quality for studying any medicine that will be used in clinical trials [25].

The Bayes' theorem is used to create a probabilistic classifier called Nave Base. NB performs so well in machine learning applications because of dependence distribution [26]. NB methods can help predict ligand-target interactions, a significant accomplishment that could eventually help in drug discovery [27]. Scientists around the world recently combined various NB approaches into a variety of drug discovery applications. Researchers employed NB models and other methodologies to classify drugs in breast cancer [28]. Complementary Nave Bayes [29] is a variation of the NB classifier with no generative interpretation. It enhances the performance of Naïve Bayesclassification by utilizing the data from all groups but the one that is being concentrated.It's also utilized in huge datasets [30].

The SVM model is used to define a collection of hyperplanes in high dimensions,using kernel functions like RBF, Linear, poly, etc., this is very often used for classification tasks [31]. It is a supervised algorithm which is used in drug-target interaction [32], anticancer drug classification [28], quantifying anti-cancer cell properties [33] and many more.

Logistic Regression (LR) classifier is used for predicting the likelihood of an event occurring. It employs several predictor variables which might be numerical or categorical. It extends thelogistical function. LR accepts a vector of variables as input. It also includes weights associated with each of the input variables. LR has been widely used in many drug

discovery applications recently with high performance and cost-effective properties [46].

3. DL approaches to drug discovery

Deep Learning based approaches have solved various complex problems easily in the recent past. DL-based models have achieved higher performance rates while keeping the margin for error very low. Image and object detection etc., can all benefit from the deployment ofCNNs. RNNs and their successors, such as LSTMs, and GRUs are great tools whensolving challenges like language translation and voice recognition etc. The field of drug discovery can be greatly enhanced by deep learning. DL could be seen as a rapid and cost-effective way of drug discovery, which is a cumbersome andcomputationally expensive procedure. DL has been used in drug discovery for threemain purposes.

- Drug properties prediction
- De Novo Drug Design
- Drug target interaction

Various deep learning techniques include DNN, CNN and GAN as shown in Fig 3.

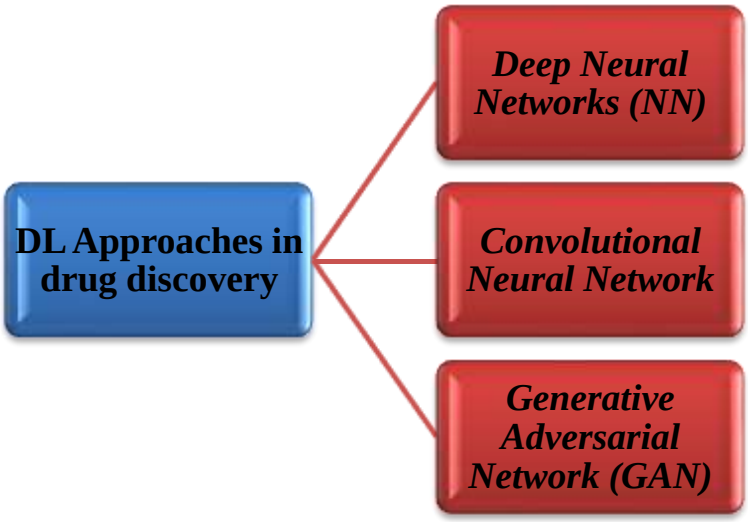

Fig 3. DL Approaches to Drug Discovery

A neural network consists of nodes organized in many layers, where neurons/nodes are connectedbetween many layers. The layered structure itself makes it diverse and it has been proved that simple 3-layer NN can act as a universal approximator, i.e., it can approximate any complex function [47]. The general structure is an input layer, many hidden layers and an output layer. Multiple hidden layers make the network deep which is called a Deep neural network (DNN) [48]. DNNs have heavily been studiedfor the drug discovery process, and one of the important deep learning-based platforms is DeepChem [49]. A Multilayer perceptron is a variation of the typical linear perceptron [50- 51].

Deep CNN is yet another type of NN which provides promising results in the drug discovery process including feature extraction, DTI, etc. The main power of CNN lies in its

convolutional filters followed by pooling layers for spatial reduction of layers. It has been already discussed throughout the survey and is considered the most important deep learning architecture.

Generative Adversarial Network (GAN) [52] is another kind of deep latent variable model. It is considered as the top influencing algorithm in the generative and prediction process. The idea is to develop a NN that will apply more focus on the areas where its predictions are weak. Mainly it has two networks; Discriminator and Generator, the former is simply a NN classifier which tries to segregate the real and fake data samples, the generator creates fake data such that the discriminator should not be able to differentiate between the real data samples from fake ones, while discriminator is trained to minimize the loss for classifying them. Both are trained in an adversarial setting, i.e., training in a competitive way. After being successfully trained, the generator generates samples that are indistinguishable from real ones. A generator can be easily used to generate new samples as it samples inputs from a random standard distribution. This model architecture is used for designing new drug molecules [25] and target interaction prediction [52-54].

4. Tools and databases for drug discovery

4.1 ML and DL-Based Tools for Drug Discovery

There has been extensive research on drug discovery using a plethora of machine learning and deep learning-based tools using different features and techniques. Table 1 gives a summary of tools developed for drug discovery.

Table 1. ML and DL-based tools for drug discovery.

Tools	Details	References
Deep chem.	MLP model to search for an appropriate candidate	[34]
DeepTox	Predicts toxicity of drugs	[35]
DeepNeuralNetQSA	Detection of the molecular activity	[36]
Organic	Creation of molecules having	[37]

	a specific property .	
PotentialNet	Predicts binding affinity	[38]
Hit Dexter	Prediction based on the response to biochemical assays	[39]
DeltaVina	Scoring function	[40]
Neural graph Fingerprint	Prediction of properties of new molecules	[41]
AlphaFold	Predicts 3D structures of proteins	[42]
Chemputer	chemical synthesis	[43]
GoPubMed	A search engine	[44]
Textpresso	Full-text engine	[44]
BioRAT	Full-text search engine	[44]
ABNER	Text analyzer	[44]
PPICurator	Mining of protein-protein interaction	[44]

4.2 ML and DL-Based Databases for Drug Discovery

Drug discovery using ML and DL-based tools require differentdatasets/databases. Table 2 gives a summary of datasets used for drug discovery.

Table 2. ML and DL-based databases for drug discovery.

Datasets	Details	References
BRENDA	Enzyme information dataset	[55]
KEGG	Genomic information	[55]
PubChem	Dataset on chemicals and	[55]

	biological activities	
TTD	Dataset on the DRM and GE etc.	[55]
DrugBank	Dataset on drug data and drug-target	[55]
SuperTarget	Drug-related databases	[55]
TDR targets	chemogenomic dataset on neglected tropical diseases	[55]
STITCH	Chemical-Protein interaction datasets	[55]
SMD	Raw microarray datasets	[39]
Gene Expression Omnibus	Raw microarray datasets	[39]
caArray	Cancer microarray datasets	[39]
CGAP database	Cancer microarray datasets	[39]
Oncomine	Cancer microarray datasets	[39]
UniHI	Human molecular interaction networks	[39]
Pathguide	resources and molecular interactions	[39]

UniProt	Protein information dataset	[39]
InterPro	Protein domain dataset	[39]

5. DL-based tools for cancer drug discovery

DL-based methods have been predominantly been used for drug response prediction. Deep Neural Network performs way better than any feed-forwardneural network since the ratio s:m, where s is sample count and m represents measurements per sample, does not prove good for feed-forward neural networks. These architectures are also prone to overfitting. But new DNN-based models have shown some good results. Many such tools/models used in drug discovery are shown in Table 3.

Table 3. Cancer drug discovery tools.

Name	Details	Reference
HNMDRP	Drug–target interaction and PPI	[57]
KRL	Gene expression	[58]
CDRscan	DNN, somatic mutations and drug compound fingerprints	[59]
Dr.VAE	Gene expression autoencoder (drug perturbation)	[60]
CancerDP	SVM (mutations, CNVs, expression levels)	[61]
BMTMKL	Bayesian Multiview, multitask model	[62]

The search count of different tools used in different research articles, Table 4, is shown as under:

Table 3. Cancer drug discovery tools.

Tool	**Search Count**
Deepchem	64
Alphafold	1371
Chemputer	30

Deeptox	19
Potentialnet	15
Deltavina	06
Gopubmed	113
Textpresso	83
Biorat	16
Abner	1664
Ppicurator Carscan	2
Hnmdrp	13
Bmtmkl	05
Cancer Dp	03
Divae	2

6. The De-Novo Era

It takes time, money, and risk to create a chemical entity, test it, assess it, and then give it the all-clear to be marketed as a medicine. Only 5 out of 5000 therapeutic concepts are expected to get to the human testing stage after undergoing preclinical testing. It's important to note that only one medicine reaches the market. Big data and AI together are regarded as the fourth industrial revolution, and they have the power to change how scientific research is carried out dramatically. AI is transforming the fields of pathology, radiology, and other medical specialities. DL technologies are beginning to be applied in the drug development process in molecular docking, transcriptomics, reaction mechanism elucidation, and molecular energy prediction, among others. Using only the building blocks of atoms, the de-novo approach generates new chemical structures without reference to prior work or relationships. Traditional methods rely on the active binders or active site properties of a biological target. Nowadays, drug development uses cutting-edge tools including computer-aided drug design (CADD) methods. These methods include ligand-based design methods like pharmacophore modelling and quantitative structure-activity relationships (QSAR), as well as structure-based design methods like molecular docking and dynamics. Along with current, rapid, and affordable hardware, the expansion of biological targets' X-ray, NMR, and electron microscopy structures has accelerated the development of more precise computational approaches. New chemical entities have been discovered more quickly as a result. According to the meaning of "de novo," which means "from the beginning," this technique enables the production of novel molecular entities without the need for a starting template [8].

6.1 De Novo Methodology for Drug Design

The de novo drug creation strategy to creating novel chemical entities uses just a biological target (receptor) or its known active binders (ligands determined to show effective binding or inhibitory action against the receptor). It primarily involves modeling the ligand or active site of the receptor, synthesizing the molecules (sampling), and evaluating the chemicals that are created. Various de-novo techniques are shown in Fig 4.

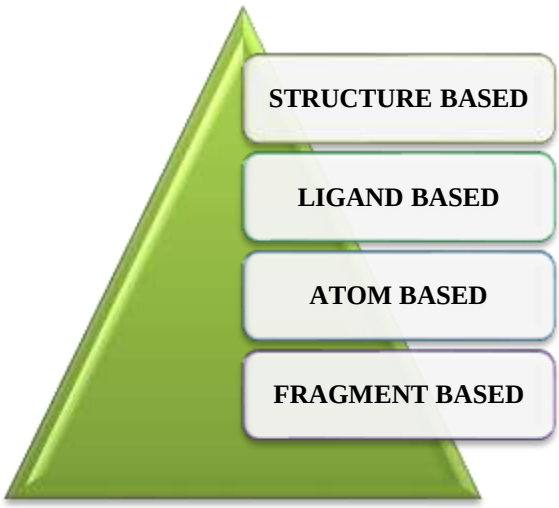

Fig 4. Methodologies Used for De-novo Drug Designing

7. Deep learning models for de novo drug design

Various DL-based models for de-novo drug design are shown in Fig. 5

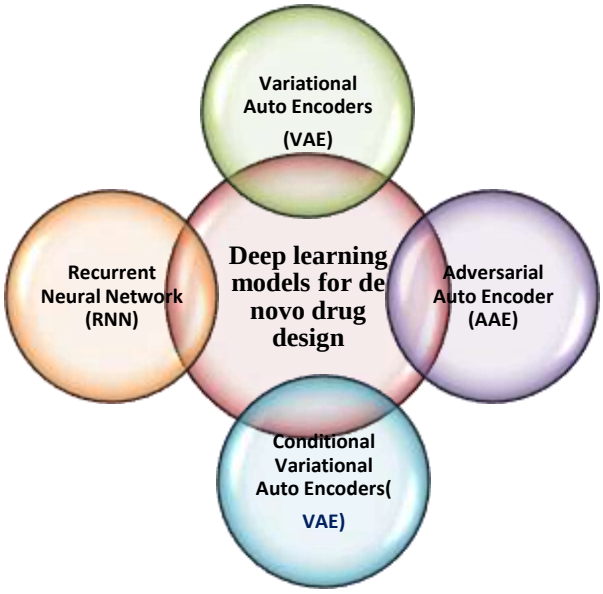

Fig 5. DL Models for De-novo Drug Designing

7.1 Recurrent Neural Network (RNN)

The researchers in the study [31] used a long short-term memory (LSTM) model to generate new drug-like compounds. The generative RNN model is comprised of two LSTM layers, each with a hidden dimension of 256 and dropout regularization. Then, a dense output layer with softmax activation was applied. The model was trained on SMILES strings (from ChEMBL221) by first converting them to numerical representation by one-hot encoding.

After training for 22 epochs, the model was able to generate token-wise compounds with 58% validity. However, they folded this model and used it to grow user-defined molecule fragments. This formulation increased accuracy, hence generating molecules with better validity.

7.2 Variational Auto Encoders (VAE)

The researchers have shown how to map molecular structures to continuous latent space using various autoencoders. They have demonstrated that latent space preserves molecular similarity, allowing us to generate new molecules with modified properties. They have evaluated proposed models, compared the performance, and found that a variational autoencoder with an additional discriminator to force the output of the encoder to follow a user-defined (target) distribution is significantly better than a simple VAE. The model that follows Gaussian distribution proved to generate 77.4% valid molecules, while the model with uniform distribution generated 78.3% valid variational autoencoder with an additional discriminator to force the output of the encoder to follow a user-defined (target) distribution is significantly better than a simple VAE. The model that follows the Gaussian distribution proved to generate 77.4% valid molecules, while the model with uniform distribution generated 78.3% valid molecules. The results were obtained on the ChEMBL version 22.34 dataset [32].

7.3 Adversarial Auto Encoder (AAE)

Adversarial autoencoders (AAE) are used to formulate novel drug molecules with anti-cancer properties. The researchers used a 7-layer AAE model. They have introduced a neuron responsible for growth inhibition percentage in latent space. The proposed method was tested on NCI-60 cell line data2and converted the SMILES to binary 166-bit Molecular ACCess System (MACCS) chemical fingerprints. Researchers trained a proposed model on 6525 compounds and sampled 640 vectors from latent space. After feeding them to a decoder with log concentration, they found the generated compounds had anti-cancer properties [33]

7.4 Conditional Variational Auto Encoders (VAE)

Researchers used a conditional variational autoencoder (CVAE) in this work to generate novel drug candidates with the desired properties. The CVAE is essentially an extended version of a variational autoencoder in which the encoder and decoder are conditioned to specify the target properties of generated compounds. The model was trained on the NCI-60 dataset with 166-bit MACCS fingerprints and one-hot-encoding of normalized G150 (growth inhibition by 50%) as a condition vector. After training, they found that feeding the conditional vector and latent vector (sampled) to the decoder of CVAE generates molecules

with anti-cancer properties. They also verified the characteristics of the generated molecules by comparing them with the drugs approved by the FDA for breast cancer. On computing Tanimoto similarity coefficients, they found that the generated molecules have high similarity. They show how to use the proposed method for searching for similar molecules from public datasets [34]. Other ML and DL-based applications for data sciences include [63-74].

8. Conclusion

Machine Learning has tremendously optimized and expedited the process of drug discovery. Deep Learning, in particular, can cope with heterogeneous and huge amounts of data. Deep Learning requires no human interventionand deals with immensely complex data which makes it profoundly useful for applications in drug discovery. Some bottlenecks like the non-availability of pharmaceutical data and complex biological associations for interpretation etc., limit theapplications and the applicability of deep learning techniques in drug discovery. In thispaper, various ML and DL techniques were discussed that are used for drug discovery.Various DL-based tools for cancer drug discovery were also discussed. The tools and databases that are available for drug discovery were also discussed. Pertinent to mention that DL-based approaches provide promising ways to discover drugs efficiently and accurately.

References

1. Kiriiri, G.K., Njogu, P.M. & Mwangi, A.N. Exploring different approaches to improve the success of drug discovery and development projects: a review. Futur J Pharm Sci 6, 27 (2020). https://doi.org/10.1186/s43094-020-00047-9.
2. Duch, W. et al. (2007) Artificial intelligence approaches for rational drug design and discovery. Current Pharmaceutical Design 13, 1497–1508.
3. Blasiak, A. et al. (2020) CURATE. AI: optimizing personalized medicine with artificial intelligence. SLAS Technol. 25, 95–105
4. Baronzio, G. et al. (2015) Overview of methods for overcoming hindrance to drug delivery to tumors, with special attention to tumor interstitial fluid. Frontiers in Oncology 5, 165
5. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. Nat. Rev. Drug Discov.2019, [6], 463–477.
6. Zo_mann, S.; Vercruysse, M.; Benmansour, F.; Maunz, A.; Wolf, L.; Marti, R.B.; Heckel, T.; Ding, H.;Truong, H.H.; Prummer, M.; et al. Machine learning-powered antibiotics phenotypic drug discovery. Sci. Rep. 2019, 9, 5013.
7. Ekins, S.; Puhl, A.C.; Zorn, K.M.; Lane, T.R.; Russo, D.P.; Klein, J.J.; Hickey, A.J.; Clark, A.M. Exploiting machine learning for end-to-end drug discovery and development. Nat. Mater. 2019, [12], 435–441.
8. Agarwal, S.; Dugar, D.; Sengupta, S. Ranking chemical structures for drug discovery: A new machine learning approach. J. Chem. Inf. Model. 2010, 50, 716–731.
9. Leelananda, S.P.; Lindert, S. Computational methods in drug discovery. Beilstein J. Org. Chem. 2016, [6] 2694–2718.
10. Gao, D.; Chen, Q.; Zeng, Y.; Jiang, M.; Zhang, Y. Application of Machine Learning on

Drug Target Discovery. Curr. Drug Metab. 2020.
11. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. Nat. Rev. Drug Discov. 2019, 18, 463–477.
12. Zo_mann, S.; Vercruysse, M.; Benmansour, F.; Maunz, A.; Wolf, L.; Marti, R.B.; Heckel, T.; Ding, H.; Truong, H.H.; Prummer, M.; et al. Machine learning-powered antibiotics phenotypic drug discovery. Sci. Rep. 2019, 9, 5013.
13. Ekins, S.; Puhl, A.C.; Zorn, K.M.; Lane, T.R.; Russo, D.P.; Klein, J.J.; Hickey, A.J.; Clark, A.M. Exploiting machine learning for end-to-end drug discovery and development. Nat. Mater. 2019, 18, 435–441.
14. Sarica, A.; Cerasa, A.; Quattrone, A. Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review. Front. Aging Neurosci. 2017, 9, 329.
15. Yang, Y.; Adelstein, S.J.; Kassis, A.I. Target discovery from data mining approaches. DrugDiscov. Today 2009, 147–154.
16. Maia, E.H.B.; Assis, L.C.; de Oliveira, T.A.; da Silva, A.M.; Taranto, A.G. Structure-Based Virtual Screening From Classical to Artificial Intelligence. Front. Chem. 2020, 8, 343.
17. Jiawei Hanl, Yanheng Liul, Xin Sunl, "A Scalable Random Forest Algorithm Based on MapReduce", 4th IEEE nternational Conference on Software Engineering and Service Science (ICSESS), Bejin, May 2013.
18. Cano, G.; Garcia-Rodriguez, J.; Garcia-Garcia, A.; Perez-Sanchez, H.; Benediktsson, J.; Thapa, A.; Barr, A. Automatic selection of molecular descriptors using random forest: Application to drug discovery. Expert Syst. Appl. 2017, 72, 151–159.
19. Kadurin A, Nikolenko S, Khrabrov K, Aliper A, Zhavoronkov A. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. Mol Pharm. 2017 Sep 5;14(9):3098-3104. doi: 10.1021/acs.molpharmaceut.7b00346. Epub 2017 Aug 4. PMID: 28703000.
20. Chan, H.S. et al. (2019) Advancing drug discovery via artificial intelligence. Trends in Pharmacological Sciences 40(8), 592–604
21. Zhu, H. (2020) Big data and artificial intelligence modeling for drug discovery. Annual Review of Pharmacology and Toxicology 60, 573–589.
22. Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, 2, 359-366.
23. Chen, R.; Liu, X.; Jin, S.; Lin, J.; Liu, J. Machine Learning for Drug-Target Interaction Prediction. Molecules 2018, 23, 2208.
24. Brown, N. (2015) In Silico Medicinal Chemistry: Computational Methods to Support DrugDesign, Royal Society of Chemistry
25. Pereira, J.C. et al. (2016) Boosting docking-based virtual screening with deep learning. Journal of Chemical Information and Modeling 56, 2495–2506.
26. Rifaioglu, A.S.; Atas, H.; Martin, M.J.; Cetin-Atalay, R.; Atalay, V.; Dogˇ an, T. Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools and databases. Brief Bioinform. 2019, 14, 1878–1912.
27. Nigsch, F.; Bender, A.; Jenkins, J.L.; Mitchell, J.B.O. Ligand-Target Prediction Using

Winnow and Naïve Bayesian Algorithms and the Implications of Overall Performance Statistics. J. Chem. Inf. Model. **2008**, 48, 2313–2325.
28. Pang, X.; Fu, W.; Wang, J.; Kang, D.; Xu, L.; Zhao, Y.; Liu, A.L.; Du, G.H. Identification of Estrogen Receptor Antagonists from Natural Products via In Vitro and In Silico Approaches. Oxid. Med. Cell. Longev. 2018, 6040149.
29. J. D. M. Rennie, L. Shih, J. Teevan, and D. R. Karger. Tackling the poor assumptions of Naive Bayes classifiers. In Proceedings of International Conference on Machine Learning, pages 616–623, 2003.
30. S. Owen, R. Anil, T. Dunning, E. Friedman, “Mahout in Action”, page 275, 2012.
31. Loc Nguyen, Academic Network. (2016). Tutorial on Support Vector Machine. Special Issue “Some Novel Algorithms for Global Optimization and Relevant Subjects”, Applied and Computational Mathematics (ACM). 6. 1-15. 10.11648/j.acm.s.2017060401.11.
32. Najm, Matthieu & Azencott, Chloe-Agathe & Playe, Benoit & STOVEN, Veronique. (2021). Target identification of drug candidates with machine-learning algorithms: how to choose negative examples for training. 10.1101/2021.04.06.438561.
33. Chen R, Liu X, Jin S, Lin J, Liu J. Machine Learning for Drug-Target Interaction Prediction. Molecules. 2018 Aug 31;23(9):2208. doi: 10.3390/molecules23092208. PMID: 30200333; PMCID: PMC6225477.
34. Zhu, H. (2020) Big data and artificial intelligence modeling for drug discovery. Annual Review of Pharmacology and Toxicology 60, 573–589
35. Ciallella, H.L. and Zhu, H. (2019) Advancing computational toxicology in the big data era by artificial intelligence: data-driven and mechanism-driven modeling for chemical toxicity.Chemical Research in Toxicology 32, 536–547.
36. Chan, H.S. et al. (2019) Advancing drug discovery via artificial intelligence. Trends in Pharmacological Sciences 40(8), 592–604.
37. Brown, N. (2015) In Silico Medicinal Chemistry: Computational Methods to Support DrugDesign, Royal Society of Chemistry.
38. Pereira, J.C. et al. (2016) Boosting docking-based virtual screening with deep learning. Journal of Chemical Information and Modeling 56, 2495–2506
39. http://hitdexter2.zbh.uni-hamburg.de
40. https://github.com/chengwang88/deltavina
41. https://github.com/HIPS/neural-fingerprint
42. https://deepmind.com/blog/alphafold
43. https://zenodo.org/record/1481731
44. Lauv Patel, Tripti Shukla, Xiuzhen Huang, David W. Ussery, Shanzhi Wang. "Machine Learning Methods in Drug Discovery", Molecules, 2020
45. Yang, Y.; Adelstein, S.J.; Kassis, A.I. Target discovery from data mining approaches. DrugDiscov. Today 2009,14, 147–154.
46. Mahout Logistic Regression,https://mahout.apache.org/users/classification/logistic-regression.html.
47. Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*, 359-366.
48. Bengio Y (2009) Learning deep architectures for AI. Now Publishers Inc, Norwell.
49. Ramsundar B, Liu B, Zhenqin W, Verras A, Tudor M, Sheridan RP, Pande V (2017) Is

multitask deep learning practical for pharma? J Chem Inf Model 57(8):2068–2076.
50. Multilayer Perceptron, https://en.wikipedia.org/wiki/Multilayer_perceptron
51. Ashish Gupta, “Learning Apache Mahout Classification”, page 144, 2015.
52. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. Commun. ACM 63, 11 (November 2020), 139–144. DOI:https://doi.org/10.1145/3422622
53. Guimaraes, Gabriel & Sanchez, Benjamin & Farias, Pedro & Aspuru-Guzik, Alán. (2017). Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models.
54. Kadurin A, Nikolenko S, Khrabrov K, Aliper A, Zhavoronkov A. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. Mol Pharm. 2017 Sep 5;14(9):3098-3104. doi: 10.1021/acs.molpharmaceut.7b00346. Epub 2017 Aug 4. PMID: 28703000.
55. Chen, R.; Liu, X.; Jin, S.; Lin, J.; Liu, J. Machine Learning for Drug-Target Interaction Prediction. Molecules, 2018, 23, 2208.
56. Rifaioglu, A.S.; Atas, H.; Martin, M.J.; Cetin-Atalay, R.; Atalay, V.; Doğ an, T. Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools and databases. Brief Bioinform. **2019**, 20, 1878–1912.
57. Zhang, F., Wang, M., Xi, J., Yang, J. & Li, A. A novel heterogeneous network-based method for drug response prediction in cancer cell lines. Sci. Rep. 8, 3355 (2018).
58. He, X., Folkman, L. & Borgwardt, K. Kernelized rank learning for personalized drug recommendation. Bioinformatics 34, 2808–2816 (2018).
59. Chang, Y. et al. Cancer drug response profifile scan (CDRscan): a deep learning model thatpredicts drug effectiveness from cancer genomic signature. Sci. Rep. 8, 8857 (2018).
60. Rampášek, L. et al Improving drug response prediction via modeling of drug perturbation effects. Bioinformatics. https://doi.org/10.1093/bioinformatics/ btz158 (2019).
61. Gupta, S. et al. Prioritization of anticancer drugs against a cancer using genomic features of cancer cells: A step towards personalized medicine. Sci. Rep. 6, 23857 (2016).
62. Costello, J. C. et al. A community effort to assess and improve drug sensitivity prediction algorithms. Nat. Biotechnol. 32, 1202–1212 (2014).
63. M. A. S. D. T. T. A. Sofi, “Attention-based Conditional VAE for Lung Cancer Drug Generation,” in 10th International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 2023, pp. 924–928.
64. O. S. B. A. T. T. A. Zargar, “Feature Selection, Importance and Missing Value Imputation in Diabetes Mellitus Prediction,” in 10th International Conference on Computing for Sustainable Global Development (INDIACom), IEEE, 2023, pp. 914–919.
65. T. A. Teli and R. Yousuf, “Deep Learning for Bioinformatics,” in Applications of Machine Learning and Deep Learning on Biological Data, Boca Raton: Auerbach Publications, 2023, pp. 181–196. doi: 10.1201/9781003328780-11.
66. T. A. Teli, F. S. Masoodi, and Z. Masoodi, “Application of ML and DL on Biological Data,” in Applications of Machine Learning and Deep Learning on Biological Data, Boca Raton: Auerbach Publications, 2023, pp. 159–180. doi: 10.1201/9781003328780-10.

67. O. Shafi Zargar, A. Baghat, and T. A. Teli, “A DNN Model for Diabetes Mellitus Prediction on PIMA Dataset,” 2022. Accessed: Dec. 23, 2022. [Online]. Available: https://infocomp.dcc.ufla.br/index.php/infocomp/article/view/2476
68. Sidiq S Jahangeer, Shafi Ovass, Zaman Majid, and Teli Tawseef Ahmed, “Big Data and Deep Learning in Healthcare,” in Applications of Artificial Intelligence, Big Data and Internet of Things in Sustainable Development, Ist.Taylor & Francis CRC Press, 2022, pp. 145–160.
69. M. A. Qureshi, I. A. Mir, T. Ahmad, and M. Iqbal, “Prediction of Heart Diseases Using Decision Tree and Neural Network Data Mining Techniques-A Review,” IJERT, vol. 5, pp. 2278–0181, 2018.
70. M. A. Qureshi, I. A. Mir, and T. A. Teli, “Hybrid Heart Diseases Prediction Model using Data Mining Techniques,” 2019. [Online]. Available: www.ijsrcsams.com
71. O. Shafi, S. J. Sidiq, T. A. Teli, and M. Zaman, “A Comparative Study on Various Data Mining Techniques for Early Prediction of Diabetes Mellitus,” in Global Emerging Innovation Summit (GEIS-2021), O. Shafi, S. J. Sidiq, T. A. Teli, and M. Zaman, Eds., BENTHAM SCIENCE PUBLISHERS, 2021, pp. 51–61. doi: 10.2174/9781681089010121010009.
72. S. Mushtaq, A. Roy, and T. A. Teli, “A Comparative Study on Various Machine Learning Techniques for Brain Tumor Detection Using MRI,” in Global Emerging Innovation Summit (GEIS-2021), S. Mushtaq, A. Roy, and T. A. Teli, Eds., BENTHAM SCIENCE PUBLISHERS, 2021, pp. 125–137. doi: 10.2174/9781681089010121010016.
73. T. Ahmed Teli and F. Masoodi, “Blockchain in Healthcare: Challenges and Opportunities,” SSRN Electronic Journal, 2021, doi: 10.2139/ssrn.3882744.
74. O. Shafi, J. S. Sidiq, T. Ahmed Teli, and K. -, “EFFECT OF PRE-PROCESSING TECHNIQUES IN PREDICTING DIABETES MELLITUS WITH FOCUS ON ARTIFICIAL NEURAL NETWORK,” 2022.