ISSN 2582 – 6050

Available Online at www.ijpscr.info
International Journal of Pharmaceutical Sciences and Clinical Research 2022; 2(1):6-15

REVIEW ARTICLE

Drug Discovery Involving Artificial Intelligence and Big Data Modeling: Review

Aditya Dixit[1], Pradeep Golani, Devendra Singh Lodhi, Megha Verma[1], Sanjay Nagdev[1], Aakash Singh Panwar[2]

[1]*Department of Pharmacy, Gyan Ganga Institute of Technology and Sciences, Jabalpur, Madhya Pradesh, India,*
[2]*Department of Pharmaceutics, Institute of Pharmaceutical Sciences, SAGE University, Indore, Madhya Pradesh, India*

Received: 03-11-2021; Revised: 15-12-2021; Accepted: 10-01-2022

ABSTRACT

Modern drug discovery has progressed to the big data age as a result of the huge data sets accessible for medication candidates. The development of artificial intelligence (AI) techniques to apply creative modeling based on the dynamic, diverse, and enormous character of pharmacological data sets is at the heart of this transition. As a result, recently established AI methodologies such as deep learning and relevant modeling studies provide new ways to drug candidate efficacy and safety evaluations based on large data modeling and analysis. The models that resulted gave researchers a lot of information about the whole process, from chemical structure through *in vitro*, *in vivo*, and clinical results. Recent modeling research has benefited greatly from the use of innovative data mining, duration, and management strategies. In conclusion, advances in AI in the big data age have prepared the way for future rational medication development and optimization, which will have a substantial influence on drug discovery methods and, eventually, public health.

Keywords: Artificial intelligence, Big data, Computer-aided drug discovery, Deep learning, Machine learning, Rational drug design

INTRODUCTION

Drug discovery is a time-consuming, costly, and difficult procedure with a high failure rate. Medication turnover costs a lot of money in clinical trials, and right now, 9 out of 10 new compounds fail between phase I clinical studies and approval processes. When compared to traditional animal models, *in vitro* and *in silico* approaches get the ability to significantly decrease the price of new drugs. The use of *in vitro* and *in silico* tools initially in the drug development process can help in the reduction of false positives.[1] Drug attrition can be minimized by identifying and rejecting candidate molecules that have sufficient pharmacological efficacy inappropriate chemicals with negative adverse effects, On the other hand, the results of *in vitro* and *in vivo* in most cases, in-silico testing has low correlations with drug actions in vivo, notably in terms of efficacy and safety side effects that are complicated.[2]

The ability of computers to learn from previously recorded data is known as artificial intelligence (AI), or deep learning.[3] The use of AI-based machine learning to assess drug possible bioactivity and toxicity is a promising tool. Current computer systems, like those focus on quantitative structure-activity relationship (QSAR) methods, can be used to quickly predict a huge number of novel drugs for a wide range of biological endpoints. [3] Current models (for example, that employed in professional clinical research tools) could only

***Corresponding Author:**
Aditya Dixit,
E-mail: adityadixit076@gmail.com

predict basic physical, chemical factors such as logP or liquidity.[4] Designs for complex human physiology (for example, medicinal effectiveness and side effects) are inefficient but are accurate in predicting the pharmacological activity of novel medications based on basic processes[5] [Figure 1]. Previous QSAR modeling studies have serious shortcomings such as light training sets, testing examples, experimental data errors, and a shortage of experimental verifications.[6] The QSAR model's calculations of newer drugs were challenged due to its restricted chemical space coverage. Activity cliffs and overfitting is the initial assumption of QSAR modeling (that identical compounds will have identical functions) has been shown for being incorrect on a few occasions implying that data frames containing simply biochemical shape information and objective activities are insufficient to address all mentioned issues.[7]

Big industrial collections were a vital source of novel production of compounds as medicinal chemistry has progressed rapidly during the 1990s. This initiative has also helped in the development of elevated screening high-throughput screening (HTS) techniques in the last 10 years.[8] HTS is a method of screening hundreds to millions of chemicals in a short amount of time and a well-defined procedure. To test a chemical library, current HTS approaches are frequently integrated with robotic methods and need moderate resources.[9] HTS data processing and assay in parallel microelectronics are becoming increasingly common in the pharmaceutical and regulatory industries as a result of the fact that they considerably diminish the cost of conducting experiments.[10] HTS chemical-response data continue to develop on a regular basis, helping to create a massive data infrastructure. Modern screening procedures provide vast, huge quantity of biological information, specifically about reaction of a medication to a particular target, thanks to the mixed initiatives of HTS or production of compounds in conjunction.[10,11]

The "four Vs" are the issues faced by big data: volume (data scale), velocity (data increase), variety (source diversity), and veracity (data uncertainty).[12] The sets of data easily obtainable for new drug, particularly in the Pharma industry, might include a large number of variables. Compounds (ranging between 100000 and many millions) have been tested compared to a variety of goals. Typical QSAR modeling and AI methodologies are not always appropriate while dealing with complex problems. [13]

Moreover, one of the most significant barriers to accepting big data is the inconsistency of available data (also known as data sparsity) regrettably. When mixed with more complicated biological systems like medication reactions, the shortage and classes of data can be overwhelming. From *in vitro* to *in vivo* research, the data produced increased sharply [Figure 1]. To expect therapeutic effectiveness and adverse effects in living organisms, this concept of huge data necessitates the growth of novel artificial algorithms facing with high-volume, multidimensional, and high-sparsity data sources.

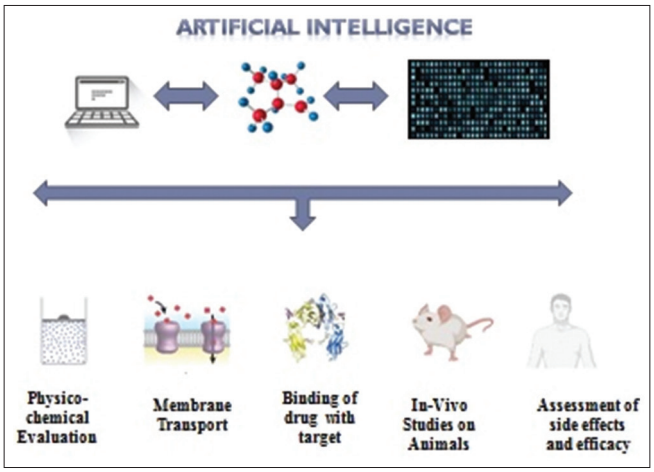

Figure 1: Challenges of data-driven artificial intelligence modeling in modern, computer-aided drug discovery

The above mentioned limitations, as well as the usage of different forms of data (for example, photos), have necessitated the creation of fresh AI methodologies to forecast ahead modeling in the latest compound discovery. In today's era of big data world, the most common AI approaches are deep learning. The 1st objectives of deep learning were in the era of medicine. The 2012 QSAR machine learning study looked at the process of drug development in the Pharma industry.[14] In this challenge, deep learning models showed significantly better predictivity than traditional machine learning approaches for 15 absorption, distribution, metabolism, and excretion (ADME) and toxicity data sets for drug candidates developed at Merck Since then, and with the development of neural network approaches [e.g., convolutional neural networks (CNNs)], deep learning has been widely applied to drug discovery approaches. Deep learning was frequently used in compound development methodologies, particularly using CNNs.[15] Due to the fact that deep learning is still considered a black-box algorithm, the present advancement of AI helped by deep learning has shown significant potential in creation of sensible drug in the area of big data. The challenges of big data modeling for pharmaceuticals and drug prospects, particularly for those in research and important Artificial Intelligence. The primary focus of this review is on deep learning and other innovative techniques.

BIG DATA IN DRUG DISCOVERY

The word "big data" refers to system that is big and complicated and the normal data analysis techniques are not able to handle them. Big data is growing rapidly on clinical trials and more eras of research that are information derived from biology.[16] As one of the fields generating a massive amount of data, modern drug discovery has moved into the big data era. The demand for new artificial techniques, such as data gathering, selection, retention, and monitoring, presents the scientific community with new difficulties and opportunities. In the last 10 years, several data-sharing projects have been launched in equal with the discovery of HTS methods in many screening infrastructure.[17] PubChem, for example, is a local database (DB) of compound shape and biological features. The number of PubChem substances has expanded tenfold in 10 years, from 25 million in 2008 to 96 million in 2018.[18] Over the same time span, the number of stored bioassays in PubChem grew from 1197 in 2008 over a million in 2018. According to PubChem's most recent figures, there are 97.3 million chemicals and 1.1 million bioassays in the DB (https://pubchem.ncbi.nlm.nih.gov). The huge volume of daily updated PubChem's bioassay data creates a locally available big data source for chemicals, having mostly medicines and drug applicant, applicants, having a wide range of goal respond data. ChEMBL, like PubChem, is a DB that contains binding, functional, ADME, and adversity data to a variety of chemicals. In comparison to PubChem, ChEMBL has a lot more data that have been carefully managed presently; the ChEMBL. DB has over 2.2 million chemicals that have been tested against bacteria from that more than 12,000 targets yielding in 15 million compound-target combinations with activity data (https://www.ebi.ac.uk/chembl/).

A number of other data sets are dedicated to pharmaceuticals and drug possibilities. DrugBank (https://www.drugbank.ca) is a locally viewable data set that contains various permitted medications, as well as their mode of action, compatibility, and possible goals.[19] DrugBank (version 5.1.2, December 20, 2018) now has 12,110 drug listings, having 2553 authorized microparticle pharmaceuticals, 1280 permitted biotech (protein/peptide) drugs, 130 nutraceuticals, and over 5842 drugs for practical (https://ntp.niehs.nih.gov/results/). DrugMatrix (https://ntp.niehs.nih.gov/results/drugmatrix/index.html) focused on drug toxicogenomic data to shorten the time it takes to create a foreign substance toxicity prospect.

The latest Drug Matrix DB contains substantial gene expression profiles using tissue of rats that were given over 600 medications, the majority of which were sedatives.[20] Various key organs are being targeted (e.g., liver). The Binding DB is a publically available, web-based DB of drugs – target binding information, given as binding

energies that were assessed. Proteins and enzymes that are thought to be pharmacological receptors are included in Binding DB. Binding DB contains 1,587,753 binding data for 7235 target proteins and 710,301 small molecules[21] (https://www.bindingdb.org/bind/index.jsp).

The size of electronic files for various data sets can also be used to categorize public big data sources. The current PubChem's bioassay data sets, for example, contain around 240 million biological actions in 30 GB of XML documents. Rather than employing private desk stop, a central computer is used. Instead of using personal computers with central processing units, the use of new hardware techniques such as cloud computation and graphics processing units is necessary to process and analyze these available big data.[22]

BIG DATA MODELING CHALLENGES: MISSING DATA AND BIASED DATA

Figure 2 shows the response data of 2,118 approved drugs tested against 531 PubChem assays (each assay having at least 25 active responses among these drug molecules). The results were created with the use of an in-house computerized data designing equipment (http://ciipro.rutgers.edu/). This response profile has over a million data points. Despite this, there were multiple responses. There were missing data in this profile [Figure 2]. Furthermore, the active-to-passive ratio responses that are not active are likewise influenced (approximately 1:6 in this data). For example, two well-known drugs were included in this profile: acetaminophen (CAS 103-90-2), which has 16 active and 213.

Each assay against all drug molecules (one column) has at least 25 active responses (red spots) this Data is obtained from Drug Bank (https://www.drugbank.ca) and Pub Chem (https://pubchem.ncbi.nlm.nih.gov). HTS data often contain far some active reactions than inactive reactions, particularly for pharmaceuticals, due to the nature of the HTS methodologies. An early assessment of 275,000 unique compounds out of 4.8 million came up in the pharmaceutical area.[21] When tested against 1,036 targets (or more), active

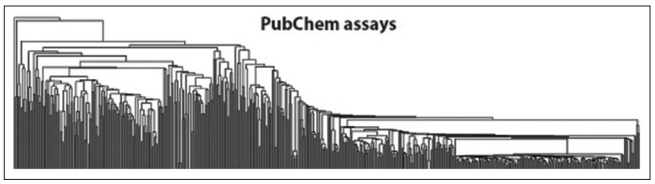

Figure 2: Bioprofile of 2,118 approved drugs from DrugBank (x-axis) represented by the response data obtained from 531 PubChem assays (y-axis)

responses were seen, suggesting that the majority of the testing was completed. Outcome was not favorable. Notably, the medications with the most active responses in local big data profiles are those used for cancer therapy. Those are known to have serious adverse effects other allergic reactions. Bortezomib (CAS 179324-69-7) is an example of chemotherapy medicine used to treat cancer. Multiple myeloma and mantle cell lymphoma are two types of cancer. It has the highest number of active responses (258 actives and 49 inactives). In Figure 2's response profile missing data is a major issue in big data modeling. A frequent solution in past studies was creating QSAR models for individual assays and then uses the results. Compounds that have not been evaluated in these assays. Only when the expected data utilized for model construction were basic, could this strategy of Biological processes like logPs or structural difficult target bindings can be used. Due to the expectations mistakes through QSAR models, however, this technique still brought uncertainty into the modeling process. Progressed statistical procedures like multiple assumptions are required when working with segregated and complex data (e.g., clinical data). Active rather than inactive outcomes should be preferred during modeling to reflect the unbalanced nature of HTS data. Procedure pharmacophore modeling was frequently used in early-stage computational investigations to identify chemical characteristics important for significant bioactivities. Later, modeling studies that used machine learning methodologies required preprocessing of unbalanced training to balance active and inactive outcomes utilizing various strategies such as resembling.

ADVANCING AI FROM MACHINE LEARNING TO DEEP LEARNING

The historical progress of AI coupled with the increase of the data size used for model development and hardware improvement in drug discovery is summarized in Figure 3. The concept of AI was emerged in the 1950s and was used in drug discovery after the first study of QSAR in the 1960s. The most common computational methodologies utilized for model generation in the prestages of pharmaceutical research (that is, before the 1990s) were linear regressions.[23] The chemical descriptors used for analysis in this early research were shorted to chemical structural information, such as atomic type and fragmental descriptors. The production of innovative chemical descriptors like topological descriptors aided discoveries in the beginning. Descriptors and molecular modeling both of are of large size/classes of descriptors derived through internship data sets. Descriptor selection was included in the modeling technique, such as the evolutionary algorithm, and simulated instead of using all possible descriptors. Newer data mining techniques due to the nonlinear forecasting models such as k-nearest neighbors and support vector machines are being developed, and strange forest was commonly utilized in modeling research from the 1990s instead of linear regression. From the 1990s to the 2000s, concept evaluation was stressed and recognized as a must-have element of modeling throughout this time period. Rather than just displaying identity, the built ideas applying these novel machine learning methodologies were forever tested by applying cross-evaluation external evaluations, and/or practical evaluation. In addition, using the applicability domain in model creation has become common practice. QSAR was created in the pre-2000s. Modeling, in combination to appropriate investigations (e.g., docking), evolved into a well-developed methodology.[24] On the above mentioned advancements in AI [Figure 3]. These AI compound discovery achievements are other evaluations that have stressed this point.

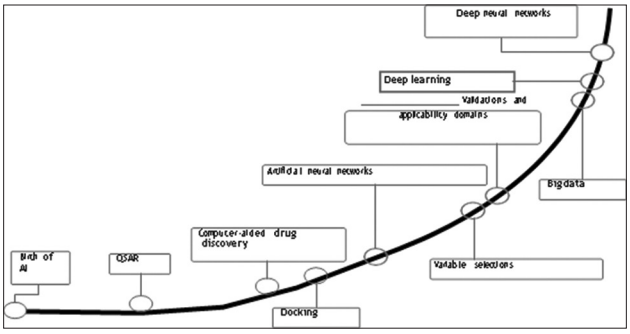

Figure 3: The historical progress of artificial intelligence in drug discovery coupled with increasing data size and computer power (shown as processor improvement)

Aside from AI advancements, the computational capability of hardware and the amount of data available are all factors to consider. To help with this, the data for modeling were greatly improved [Figure 3]. Simple algorithms (e.g., linear) are used to model tiny training sets in the early stages of computational modeling. Regressions did not necessitate a lot of processing resources. With the improvement of computing capability and affordability of biologic information for pharmaceuticals, novel modeling tools like vast scale connections were able to be used to clarify drug discovery challenges. The synapse connection that was created even as technique for computation in the 1980s was used for the first time in this application. One of the early efforts of applying deep learning in the drug discovery process in pharmaceutical industries was the 2012 QSAR machine learning challenge supported by Merck. In this challenge, deep learning models showed significantly better predictivity than traditional machine learning approaches for 15 absorption, distribution, metabolism, and excretion (ADME) and toxicity data sets for drug candidates developed at Merck. Since then, and with the development of neural network approaches.[25] This method is based on biological neural networks like those found in humans' brain. ANN techniques form with a variety of input variables (for example, chemical descriptors). Hundreds of artificial neurons are linked by weighted interactions in the brain network's shape. Although a single neuron may be capable of forecasting output but the total number of neurons in the system must be considered from the network which consists of 100's or 1000's of synapse that makes the actual predictions.[26] ANNs are a good machine learning tool for building nonlinear relations between variables and target

biological processes since they learn from their input data. Artificial systems with innovative features based on different machine learning techniques, like ANNs, are being developed. Technology advances in the 1990s like Powerful computers, and benefited directly. Deep learning was first introduced in the 1980s in conjunction with ANNs.

When the data utilized for model construction are limited, yet synapse connections do not exist compared to other machine teaching methods, it does not show significant advantages. Desktop hardware was quite insufficient for training synapse networks with numerous invisible sections throughout the 1990s and 2000s. The information sets for model construction were not enough when the sets of data for model construction are vast. In the 2010s, hardware construction hit a critical point with the use of GPUs and cloud computing, which helped neural network–modeling research directly [Figure 3]. Deep neural networks (DNNs), also known as deep neural nets, are complex neural networks with numerous hidden layers that were built to answer difficult queries such as recognition of speech. An AI algorithm based on a neural network was developed as part of the Google Deep Mind project in 2015. DNN, which has 13 invisible levels, 1st conquered the game of Go, which has long been regarded as the most difficult game in the world. It was the most difficult from the classic games for AI. The seminal publication on deep learning and the big data concept were both published practically simultaneously. Deep learning was used in the science of life almost quickly, demonstrating its potential to find classes in life systems that are complex. Deep learning was used for the very first time in this research A Merck-sponsored QSAR machine learning challenge indicated that approaches performed much better compared to other data study options for drug development. The National Center for Advancing Translational Sciences of the National Institutes of Health (NIH) conducted a similar project to model approximately 12000 molecules, adding several medications, for 12 different diseases. DeepTox, a computational toxicity model based on machine learning, won this competition. DNNs beat other machine learning-based models in this study.

In the last 3 years, there have been a number of specific deep learning experiments for pharmaceutical research, in addition to modeling issues outlined above. For example, depending on 15,524 drug-target combinations gathered from the DrugBank DB, Ching T *et al*. described a deep learning model created to assume compatibility between medications and their targets.[27] Another case that is similar to the deep learning analysis employed transcriptase data from the Library of Integrated Network-Based Cellular Signatures program. In addition, multifunctional education based on DNNs is a modeling opportunity that enables for the modeling of numerous related activities at the same time. Multitask learning was used to simulate numerous physiologically linked finished points (that is, biological actions with comparable processes) for drug development objectives, and it outperformed typical QSAR models by reducing overfitting, resolving data bias concerns, and discovering factors from the data.Responsibilities that are connected with these DNN models' outstanding performance highlights the benefits of utilizing deep learning algorithms to design big data profiles and choose important characteristics. Yet, there have been some latest findings which reveal inconsistent outcomes when comparing deep learning and AI modeling. There is no common eligibility for picking appropriate modeling dimensions and/or creating modeling routine because deep learning is a novel concept that is being used in desktop drug discovery.

OTHER AREAS OF COMPUTATIONAL MODELING UTILIZING AI FOR DRUG DISCOVERY

Rational nanomaterial design

Current nanoscience has a significant impact on drug development by providing friendly nanoparticles (for example, nanomedicines with desirable pharmacological actions and minimum adverse effects) to drug discovery process, particularly as flexible yet stable contains for drug delivery to patients to Cure malignancies and other sickness that affects the body. Molecular

dynamic (MD) simulations were used in the early stages of applying AI in nanoprofiling for drug development. Several investigations employing MD simulations, for example, discovered the injection of nano-materials into the plasma membranes of receiver cells as well as an alteration in the cell membrane anatomy overall. After the equal strategy was utilized, the efficacy of carbon nanotubes for chemical compounds was calculated using this method. Another study found that potential membrane disturbance of a collection of nanomaterials was investigated outside the body in 4 cell types. Nanoparticles' effects were investigated. The results of the size, density, position, distribution, length, and type of drug molecule on the physiological properties of nonmaterials also are evaluated using MDs under specific circumstances (e.g., interactions with and even getting passed via membranes), and also the influence of the size, density, position, distribution, length, and type of surface legends on the biocompatibility of the nanoparticles. MD simulations have the benefits of being able to exactly simulate molecular structures, but they also have the disadvantage of being computationally expensive and unable to offer speedy results. Due to existing computational resource scarcity, forecasts for large DB s are not possible. Traditional QSAR modeling methods can also be used for nonmaterials as a computational approach. The QSAR technique, for example, has been utilized to develop expected structures for nanomaterials with the same or distinct metal cores.[28] Membrane-nanoparticle compatibility has recently been studied in detail. Recently, membrane-nanoparticle interactions were modeled based on the atomization energy of the metal oxide, the period of the nanoparticle metal, and the primary size of the nanoparticle.

Because of the deficiency of acceptable chemical descriptors, the modern parameters of AI techniques in nanoprofiling are confined to creating novel nanoparticles. Although descriptors wise choices from surface drug molecule are helpful in assuming particular biological actions/characteristics of nanoparticles, as previously stated, the effects of the nanoparticles size/shape, density, position, and distribution, as well as the effects of the nanoparticles size/shape, density, position, and distribution, as well as the effects of the, In these research, the length and class of surface drug molecule were not taken into account. Descriptors generated from practical parameters (for example, nanoparticle size) or biological profile (for example, proteomics profile) have been used in certain other nanomodeling research. Puzyn *et al*. suggested that no global nano-QSAR exists because of the multiplicity and difficulty of nanoparticle modeling. The biological characteristics of changeable nanomaterials can be correctly predicted using this approach. Figure 4 illustrates throughout the modeling procedure, a new approach for nanostructure simulation. To summarize, the nano-materials' characteristics and bioactivities are substantially governed by their surface chemistry. Surface ligand orientations and functional group accessibility had to be taken into account in the calculations to accurately replicate nanosurface chemistry [Figure 4]. Heavy atoms and functional groups, for example, had a role in early modeling of nanohydrophobicity. The accessibility of nano logP values to solvent molecules was associated. An enhanced approach of incorporating the solvent-accessible surface into computations was recently demonstrated in a recent study to be used as a nano logP calculator that is universal. A modeling method that is comparable to that used previously has been used to simulate nanocellular uptake capabilities, as well as a variety of other nanobioactivities. The models developed were used to design and synthesize a number of novel nanomaterials with the required Nano biological actions.

CNN and image modeling

CNN is a neuroscience-inspired connectivity design technique for replicating visuals in the cerebral system, where single neurons are connected. Only their receptive fields react to changes. Different neurons can partially overlap each other in order to cover the complete number of parameters. The CNN system is designed in such a manner that secret levels are very good at

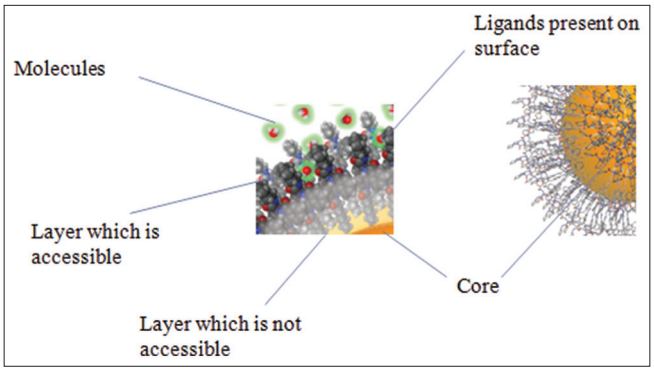

Figure 4: Nanomaterial surface simulations for computational modeling: surface ligand orientations and accessibility assessments

filtering. multitasking data such as red, green, and blue saturation numbers gathered from 1000's of people the number of pixels in an image.[29] The CNN method employs kernels and grids in the training process nerve cells react to previously defined aspects to assess the photo and begin to spot specific crucial aspects such as lines and contours for a facial image, and the CNN is just a specialized network analysis method motivated by brain science to simulate photographs inside this visual cortex, in which individual neurons react to stipredefined dimensions to inspect the photograph and start understanding specific significant points such as lines and contours for a human face. CNNs were firstly given in the 1980s for photograph identification, but they did not gain popularity until the 2010s. Since 2012, this approach has taken all image analysis challenges, and it is currently the foundation for image/speech recognition, video review, communication translation, and other technologies.

CNNs have already utilized for picture modeling in medical diagnosis like in cancer, Alzheimer's disease, and cardiac disorders and it is the most popular deep learning methodologies. CNNs were also utilized in traditional drug discovery to analyze picture profile produced by practical drug analysis, like HTS results.[30] CNNs were also utilized to recognize 3-D practical and virtual pictures to conclude drug-target binding due to their particular advantages in image recognition of ligand-protein interactions. In other investigations, CNNs were used in conjunction with other computational methods to achieve specified objectives. CNNs, for example, were employed as a novel method for recognizing molecular characteristics in drug molecular graphs. In this study, drug particles were interpreted like 2D. Statistics that contain atom characteristics and for purpose of training, the CNN was used to produce new molecular graphs based on the features of the input molecular graphs. Another study used an updated CNN model-based lifespan deep convolution network to predict patients' cancer outcomes based on histology images and genetic diagnostic data. Using a text-mining technique, CNNs also are able to extract adverse drug events data from clinical publications.

Personalized medicine

A drug's affinity and adversity are strongly influenced through interactions with many goals, involving both on- and off-targets. Many genetic, epigenetic, and atmospheric factors influence how a medication molecule affects a single biosystem (like a diseased person). Personalized medicine was developed to detect this hidden hierarchical information.[31] It is Developed to respond each patient's problems. Personalized medicine is based on a scientific knowledge of how a patient's unique traits, like molecular or genetic data, make him or her prone to disease and susceptible to treatment. Treatment for a disease since the late 1990s, for the treatment of diseases 100's of genes and the credit goes to biomarker research that plays a vital role in human sickness and patient genetic heterogeneity has also been used to differentiate between individual reactions to a variety of therapies.

Computational modeling has become one of the most essential tools for customized treatment, in addition to the massive amounts of data collected by research like the Human Genome Project.

Many latest advances in this era rely on artificial methods, including binding site expectations, metabolic connection modeling, and local genomic pattern analysis modeling. Many data generation and sharing projects are part of the NIH Precision Medicine initiative. To facilitate the expansion of precision, initiatives and computational modeling efforts have emerged medicine. For example, the National Cancer Institute's Genomic Data Commons program intends to establish a data source that allows data communication among cancer genomic investigations in favor of précised medicine.[32] https://gdc.cancer.gov/ site has received and shared 33,549 case studies so far and it is a website dedicated to cancer research. Although it is not the subject of this article, genome sequencing is an important topic. Analysis has been a commonly used AI technique, and there have been numerous assessments. This popular bioinformatics topic is now available.

CONCLUSIONS

By offering early-stage evaluations of therapeutic compounds, AI has the potential to significantly cut the cost and time associated with drug research and pharmaceutical data rapidly to rise at a high rate, requiring the growth of new AI process to facing with large data sets. For this problem, current deep learning modeling research has revealed benefits over classic AI. However, in order to deep learning models to be useful, common criteria and modeling methods are still required. Beyond traditional drug discovery, AI's uses have been greatly expanded. AI and contemporary deep learning data have prepared the path, when combined with data set duration, web-linked progression as data source network, and improvements in desktop technology in the development of modern drugs.

REFERENCES

1. Maia EH, Assis LC, de Oliveira TA, da Silva AM, Taranto AG. Structure-based virtual screening: From classical to artificial intelligence. Front Chem 2020;8:343.
2. Merlot C. Computational toxicology – A tool for early safety evaluation. Drug Discov Today 2010;15:16-22.
3. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 2019;28:73-81.
4. Fischhoff B, Davis AL. Communicating scientific uncertainty. Proc Natl Acad Sci U S A 2014;111 Suppl 4:13664-71.
5. Colombo S. Applications of artificial intelligence in drug delivery and pharmaceutical development. In: Artificial Intelligence in Healthcare. England: Elsevier; 2020. p. 85-116.
6. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature 2018;559:547-55.
7. Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S, *et al.* New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol 2014;34:1-18.
8. Phatak SS, Stephan CC, Cavasotto CN. High-throughput and *in silico* screenings in drug discovery. Expert Opin Drug Discov 2009;4:947-59.
9. Clayson IG, Hewitt D, Hutereau M, Pope T, Slater B. High throughput methods in the synthesis, characterization, and optimization of porous materials. Adv Mater 2020;32:e2002780.
10. Szymański P, Markowicz M, Mikiciuk-Olasik E. Adaptation of high-throughput screening in drug discovery-toxicological screening tests. Int J Mol Sci 2012;13:427-52.
11. Pandey UB, Nichols CD. Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery. Pharmacol Rev 2011;63:411-36.
12. Wüest RO, Zimmermann NE, Zurell D, Alexander JM, Fritz SA, Hof C, *et al.* Macroecology in the age of big data-where to go from here? J Biogeogr 2020;47:1-12.
13. Rodrigues JF Jr., Florea L, De Oliveira MC, Diamond D, Oliveira ON Jr. A survey on Big Data and Machine Learning for Chemistry. arXiv Prepr. arXiv1904.10370; 2019.
14. Neves BJ, Braga RC, Melo-Filho CC, Moreira-Filho JT, Muratov EN, Andrade CH. QSAR-based virtual screening: Advances and applications in drug discovery. Front Pharmacol 2018;9:1275.
15. Karimi M, Wu D, Wang Z, Shen Y. DeepAffinity: Interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks. Bioinformatics 2019;35:3329-38.
16. Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. Genome Med 2016;8:134.
17. Wang Y, Cheng T, Bryant SH. PubChem BioAssay: A Decade's development toward open high-throughput screening data sharing. SLAS Discov 2017;22:655-66.
18. Huang G, Yan F, Tan D. A review of computational

methods for predicting drug targets. Curr Protein Pept Sci 2018;19:562-72.
19. Duran-Frigola M, Pauls E, Guitart-Pla O, Bertoni M, Alcalde V, Amat D, *et al.* Extending the small-molecule similarity principle to all levels of biology with the Chemical Checker. Nat Biotechnol 2020;38:1087-96.
20. Alexander-Dann B. Understanding Compound-Induced Histopathology in Rat Liver Using Gene Expression Network Methods. England: University of Cambridge; 2020.
21. Zhu H. Big data and artificial intelligence modeling for drug discovery. Annu Rev Pharmacol Toxicol 2020;60:573-89.
22. Tang W, Feng W. Parallel map projection of vector-based big spatial data: Coupling cloud computing with graphics processing units. Comput Environ Urban Syst 2017;61:187-97.
23. Ma DL, Chan DS, Leung CH. Drug repositioning by structure-based virtual screening. Chem Soc Rev 2013;42:2130-41.
24. Rasulev B. Ecotoxicological QSAR modeling of nanomaterials: Methods in 3D-QSARs and combined docking studies for carbon nanostructures. In: Ecotoxicological QSARs. Methods in Pharmacology and Toxicology, Springer nature, Switzerland; 2020. p. 215-33.
25. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. Heliyon 2018;4:e00938.
26. Cui Y, Ahmad S, Hawkins J. Continuous online sequence learning with an unsupervised neural network model. Neural Comput 2016;28:2474-504.
27. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, *et al.* Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface 2018;15:20170387.
28. Muratov EN, Bajorath J, Sheridan RP, Tetko IV, Filimonov D, Poroikov V, *et al.* QSAR without borders. Chem Soc Rev 2020;49:3525-64.
29. Eraslan G, Avsec Ž, Gagneur J, Theis FJ. Deep learning: New computational modelling techniques for genomics. Nat Rev Genet 2019;20:389-403.
30. Whitebread S, Hamon J, Bojanic D, Urban L. Keynote review: *In vitro* safety pharmacology profiling: An essential tool for successful drug development. Drug Discov Today 2005;10:1421-33.
31. Melnykova N, Shakhovska N, Melnykov V, Melnykova K, Lishchuk-Yakymovych K. Personalized data analysis approach for assessing necessary hospital bed-days built on condition space and hierarchical predictor. Big Data Cogn Comput 2021;5:37.
32. Jensen MA, Ferretti V, Grossman RL, Staudt LM. The NCI Genomic Data Commons as an engine for precision medicine. Blood 2017;130:453-9.