Artificial Intelligence in Toxicology and Pharmacology

Abstract:

Methods that utilize machine learning and artificial intelligence have transformed a wide variety of fields, including the field of toxicology. Physiologically based pharmacokinetic (PBPK) modeling, quantitative structure-activity relationship modeling for toxicity prediction, adverse outcome pathway analysis, high-throughput screening, toxicogenomics, big data, and toxicological databases are just some of the areas that are covered in this review. By leveraging machine learning and artificial intelligence approaches, it is now possible to develop PBPK models for hundreds of chemicals in an efficient manner, to create in silico models to predict toxicity for a large number of chemicals with similar accuracies compared with in vivo animal experiments, and to analyze a large amount of data of various types (toxicogenomics, high-content image data, etc.) to generate new insights into toxicity mechanisms rapidly, which was previously impossible. This is an improvement over the previous situation The field of toxicological sciences faces a number of challenges that must be overcome before it can make further progress. These challenges include the following: (1) not all machine learning models are equally useful for a particular type of toxicology data; therefore, it is important to test different methods to determine the optimal approach; (2) the current toxicity prediction is primarily based on bioactivity classification (yes/no); therefore, additional studies are required to predict the intensity of effect or dose-response relationship.

1. Introduction:

Toxicology is a branch of science that focuses on the study of the harmful effects and the underlying mechanisms of toxicity caused by chemicals, substances, or situations on humans, animals, and the environments [1, 2]. It also focuses on the prevention and amelioration of such harmful effects, in addition to the application of toxicology knowledge to the safety evaluation and risk assessment of xenobiotics [3]. Toxicology encompasses a wide range of topics that can be broken down into a number of subfields [4], such as chemical toxicology (the study of the toxicity of various chemical classes, such as pesticides, metals, etc.), organ systems toxicology (the study of the effects of toxicity on various target organs), nonorgan-

directed toxicity (the study of carcinogenesis, genetic toxicology, and developmental toxicology), toxicokinetics (such as physiologically based pharmacokinetic [5-7].

Figure 1 Subfields of toxicology

According to one research. [8], artificial intelligence is a subfield of computer science that is undergoing significant expansion, and its primary objective is to design and build computers or computational models that are capable of carrying out a wide range of cognitive tasks at a level that is on par with or even superior to that of human intellect [9]. The phrase "artificial intelligence" (AI) can refer to a number of various things depending on the context [10, 11]. The term "machine learning" is used throughout this current study to refer to the applications of various machine learning approaches in the prediction and evaluation of chemical toxicokinetic (also known as absorption, distribution, metabolism, and excretion [ADME]) and toxicity properties [12-14]. Machine learning is a subfield of artificial intelligence that refers to mathematical or computer algorithms meant to teach or train a computational model to solve a problem or perform difficult tasks based on some input parameters [15, 16].

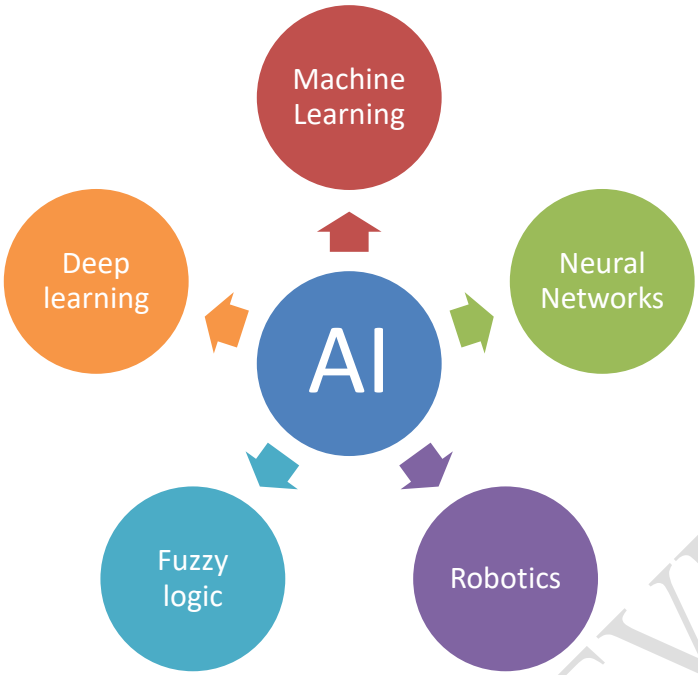

Figure 2 Few types of AI

Machine learning is a subfield of artificial intelligence that refers to mathematical or computer algorithms designed to teach or train a computational model to perform complex tasks [17]. Learning on a machine can typically be broken down into one of three categories: supervised learning, unsupervised learning, or reinforcement learning [18-20].

There are many distinct kinds of artificial neural networks, and the vast majority of these networks have at least some applications in the fields of medicine and the biological sciences [21-23]. Information moves in only one direction through feed forward ANNs, whereas feedback ANNs are responsible for feeding information back. Single-layer perceptrons, often known as SLPs, are a type of basic feed forward artificial neural network that is typically utilized for the linear binary categorization of data [24]. In addition to an input layer and an output layer, more complicated multi-layer perceptrons comprise one or more hidden layers of fully linked neurons. Furthermore, in contrast to SLPs, these perceptrons use a nonlinear activation function [25, 26]. Techniques that fall under the umbrella of deep learning (DL) include supervised machine learning with the use of MLPs [27]. A convolutional neural network, often known as a CNN, is another type of artificial neural network (ANN) that is used for deep learning. A CNN is essentially a regularized form of a multilayer perceptron (MLP), and it consists of a layer called the convolutional layer that typically applies a rectified linear unit activation function [28, 29]. In CNNs (Convolutional deep neural networks), the input data are convolved with individual neurons in the convolutional layer receiving data exclusively for a particular receptive field. This is in contrast to MLPs, which

include neurons that are fully connected to one another. Because of this, the likelihood of data over fitting, which is the most significant drawback of MLPs, is reduced. CNNs are frequently utilized in the fields of medicine and biology for the purpose of analysis and categorization of signals in two dimensions [30, 31]. The usage of network models that make use of Bayesian inference, often known as Bayesian neural networks (BNN), is one of the more modern alternatives to these traditional neural network methodologies. By undertaking weight marginalization rather than optimization, as is the case with many other machine learning approaches, Bayesian neural networks (BNNs) have the potential to, under certain conditions, lessen the likelihood of data over fitting and increase the accuracy of their predictions [32-35].

In recent years, artificial neural networks have been utilized on a number of instances for the purpose of classifying and predicting the hazardous effects of a wide variety of biologically active chemicals. It is safe to expect that deep learning will in the not too distant future become a very essential component of numerous diagnostic and research protocols in the field of toxicology [36-38]. The application of deep learning in toxicology is a novel method that is undergoing rapid development. This article is a condensed assessment of recent discoveries and research methodologies regarding the application of artificial neural networks (ANNs) to the processing of complex toxicological data[39, 40].

1.1. Machine Learning Techniques:

ML is the fundamental paradigm that is comprised of a number of method-based domains and a number of algorithms in order to recognize patterns within the data. DL and ML are used in every automation-based technique, but the difference between the two is held constant[41]. ML may be further subdivided into numerous types, and one of these types is called deep learning. This imitates the transmission of electrical impulses and is analogous to the biological neurons found in humans. This model is a well-established mathematical model that relates to the process of learning strategies for forecasting future data [42]. It reveals underlying patterns that are present in the data and information that are accessible. The success of ML can be attributed to the successful resolution of difficult mathematical problems, which is employed in many areas of contemporary biology. In terms of generalized machine learning approaches, it has been used to accurately forecast the data set that has not yet been observed in order to select the method that will perform the best in difficult circumstances. In order to prevent brute force sensitivity and optimize specifically by better

understanding the viewpoint in various model architectures, training a single dataset with many models at the same time is common practice. Reinforcement learning is another significant category of ML methods that processes mostly towards the application in dealing with complicated environments. This type of learning teaches the ideal series of actions in response to opposed environmental information to the output in a manner that is comparable to supervised learning [43]. Instead of displaying scoring functions that incorporate the Drug score [44], the ML techniques that are used in AutoDock are well-prepared and trained for the particular types of protein/DNA, active site residues, small molecules, and medications that are being considered [45]. The ML addresses a number of applications, each of which offers an excessive amount of success in one or more phases of the HTVS that are characteristic of drug development in terms of comprehending novel medications and lead components[46].

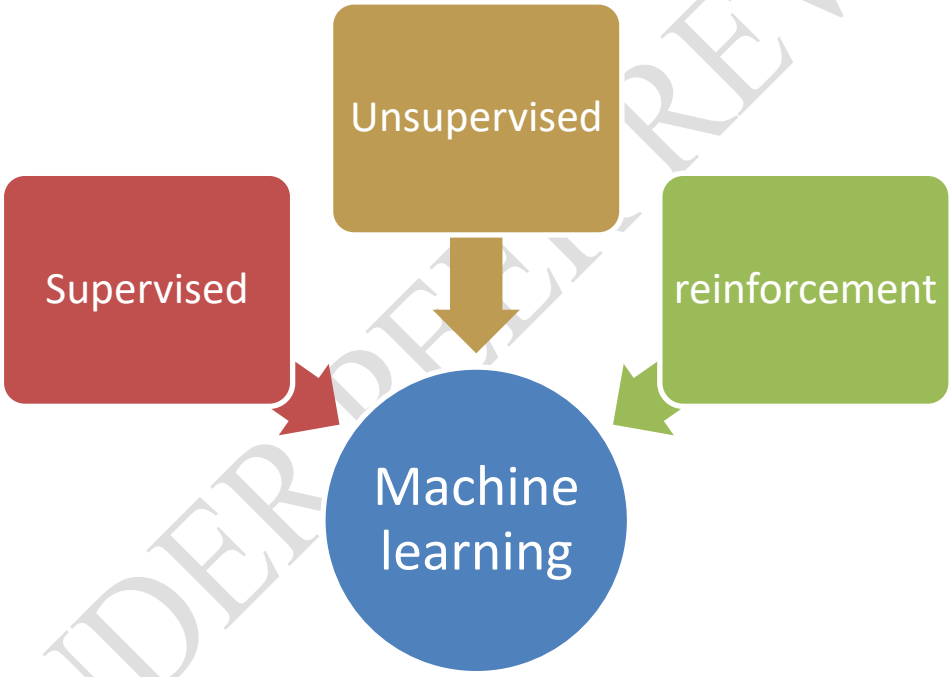

Figure 3 Machine learning Techniques

1.2. Deep learning Techniques:

The DL approach reflects the neural network by having several hidden layers, and it is widely utilized because of its adaptability to learn arbitrary complex functions [41, 47]. It is able to learn as much as it possibly can with sufficient data and the investment of sufficient computational effort, resulting in extremely dependable outputs. The numerous layers that are buried behind DL patterns provide a flexible access point for learning arbitrarily complicated modules, which directly provide relevant neurons and trained sets. End-to-end differentiation

is achieved with the application of a backpropagation algorithm and a gradient-based optimization method by the DL [48]. These methods make it possible for neural networks to be trained. In addition to this, the feed-forward networks are connected to layers, conventional architecture, and graph convolutional architecture, all of which contributed to the growth of a variety of data kinds and domains. The DL approach is being propelled forward by the modern trend of data reading as well as by technological advances in algorithmic and computational hardware. Because of the difficulty they have in training and comprehending small sets, neural networks have garnered a lot of attention within the realm of deep networks [49]. From an algorithmic perspective, deep learning networks with additional layers have frequently been plagued by disappearance gradients, which inhibit the ability of models to train in an effective manner. Significant improvements in effectively trained deep networks can be attributed to new method initialization strategies, neural activation functions, and gradient-based optimization methods[50]. Recurrent neural networks, also known as RNNs, are recurrent units that are primarily employed for the purpose of capturing the temporal dependency in sequence-level data input. Conventional neural networks, often known as CNNs, are becoming increasingly popular for image processing. CNNs do this by utilizing learning filters to capture both local and spatial correlations [50]. Graph neural networks, also known as GNNs, are typically used to function with unordered data, such as when doing an analysis of social networks; nonetheless, this model is ideally suited for the representation of tiny molecules [51].

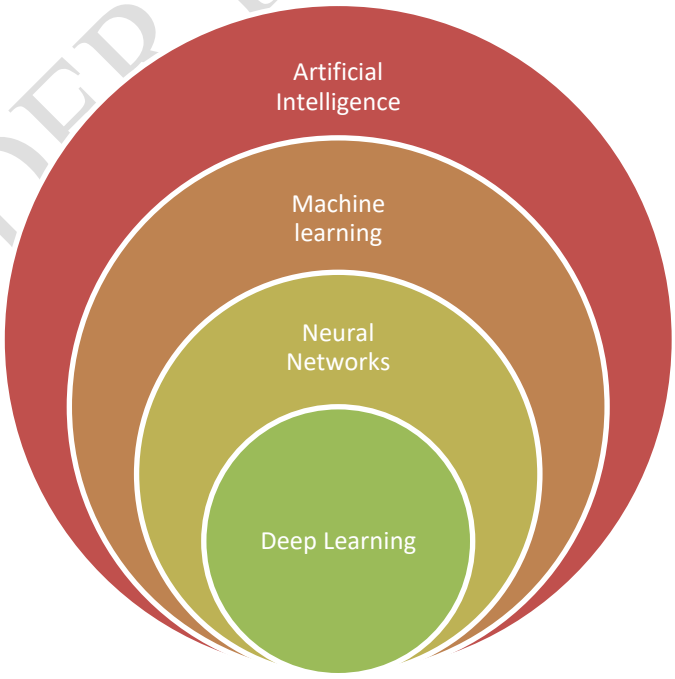

Figure 4 AI

2. TOXICOGENOMICS

Toxicogenomics is a sub-discipline of toxicology that applies genomic technologies (such as gene expression profiling, proteomics, metabolomics, and other related methods) to the study of the harmful effects of chemicals or xenobiotic substances at the gene and/or protein levels within specific cells or tissue(s) of an organism [52]. Toxicogenomics is a relatively new field of study. Toxicogenomics has emerged as an important tool in the identification of potential molecular mechanisms of toxicity in response to exposure to environmental chemicals at the gene, protein, or metabolite level in the cells or tissues of organisms. It also serves as biomarkers for predictive toxicology [53]. Recent advances in computational technologies have made it possible to integrate toxicogenomics with computational models (for example, machine learning and PBPK models), which allows researchers to correlate molecular endpoints derived from toxicogenomics data with in vivo regulatory-relevant phenotypic toxicity or toxicokinetic endpoints [54]. Data for toxicogenomics could be obtained either from in vitro or from in vivo experiments. In vivo toxicogenomics data are desirable; however, it is both impracticable and unethical to gather toxicogenomics data for thousands of chemicals using animal tests on varying dose groups and treatment periods. This is because it would violate both the animal's and the researcher's rights [55]. A recent study developed an artificial intelligence-based Tox-GAN framework by applying a deep generative adversarial network (GAN) approach [56].

2.1. The Application of Artificial Intelligence to Assess the Toxicology and Safety of Drugs

One of the most powerful tools that modern medicine may employ in the fight against disease is called interventional pharmacology. These medications, on the other hand, can have serious adverse effects, so their use needs to be carefully watched. The scientific discipline known as pharmacovigilance focuses on the monitoring, detection, and prevention of adverse drug reactions (ADRs) [57]. Efforts to ensure patient safety begin during product development with in vivo and in vitro testing, continue through clinical trials, and extend through postmarketing surveillance of adverse drug reactions in actual patients and populations. The boundaries of these traditional techniques will be tested in the future by difficulties related to toxicity and patient safety, including rising polypharmacy and patient

diversity. Improving drug safety science through the use of artificial intelligence (AI) and machine learning presents a unique opportunity made possible by the recent availability of massive volumes of newly compiled data [58].

3. AI In The Treatment Of Drug Addiction:

In the fight against complex diseases, in which many patients require completely novel treatment protocols, the application of AI as an aid in the selection and administration of personalized drug regimens is vital [59]. Modern medicine has achieved its massive success through the use of technologies such as vaccinations and antibiotics, which generally work without much adjustment from patient to patient [60]. Artificial intelligence (AI) is an aid that can be used in the selection and administration of personalized drug regimens [61]. Precision and individualized medical approaches are being developed in order to treat the diseases that have been left behind, diseases for which there is no one cure that will work for everyone [62, 63]. The massive amounts of biological, chemical, and medical data that need to be integrated in order to completely comprehend and combat complex diseases such as malignancies can only be truly leveraged by the deployment of computational tools that are dependent on AI engines[64, 65] . There is currently no artificial intelligence that can serve as a skeleton key and offer therapies for all illnesses. Instead, tools are being developed to perform certain jobs in the process of selecting drugs and administering them to patients, and these tools are being tailored to each condition and patient individually [66-69]. The functions that are determined by the various approaches of AI modeling are at their most accurate when they are not extrapolated over broad output areas, much like other mathematical approximation methods [70-72]. Increasing the accuracy of models can be accomplished by restricting the space in which their predictions can be made. This reduces the number of possible outputs or the range of outputs, depending on whether the problem is one of classification or regression [73, 74]. In artificial intelligence technologies, the complexity of the data and modeling methodologies that are depended upon often results in the formation of a decision-making process that is utterly incomprehensible to human operators. This is unavoidable since it is physically impossible for humans to take into account and incorporate every applicable piece of biological, chemical, and medical information that is important to each patient in the process of designing a treatment protocol that is tailored specifically to meet the needs of that patient [75-77]. If we were capable of deciphering the information, we would not require the assistance of the AI. In the field of drug therapy, the goal of artificial intelligence (AI) is to condense the vast amounts of data

into something that doctors can comprehend, so providing them with access to information that they were unable to take into account in the past [78-80]. In order for a transparent reduction process to work, its steps would need to be straightforward enough for the operator to comprehend them. This kind of simplification frequently results in a loss in the strength of prediction. Black box decision support solutions that are powered by AI contribute a higher level of accuracy but have less openness. Because their decision-making process is not necessarily rationalizable to humans, there has been a lot of anxiety around the utilization of so-called black box models; nevertheless, this stressor ought to be relieved if they prove to be successful and patient survival rates increase [81, 82]. Because of the consequences of this kind of software and the fact that it frequently comes with a lack of interpretability, many trials are required in order to justify any medical black box model's predictive powers. Before the models can be used in clinics, it is necessary to provide evidence that they will unquestionably lead to better outcomes for patients. It will be significantly more challenging for models to attain transparency as a result of the introduction and wider-scale usage of DL for the selection and optimization of medicinal therapies [83-85].

4. Modeling Using Artificial Intelligence and Big Data For Drug Discovery:

The term "big data" refers to a collection of data sets that are both so huge and complicated that it would be impossible to process them using the conventional methods and tools for doing data analysis [86-89]. The use of big data is becoming increasingly prevalent in clinical investigations as well as other types of research that are driven by biological data [90-92]. The era of big data has arrived in the field of modern drug discovery, which is one of the fields responsible for producing large amounts of data. The research community faces both new obstacles and new opportunities as a result of the requirement for fresh computational tools. These techniques may include data mining and production, curation, storage, and management. In the last ten years, in addition to the development of HTS techniques in a variety of screening centers, other data-sharing efforts have also been started. For instance, PubChem is a public archive for chemical structures as well as the biological features of the molecules that make them up[93]. The number of PubChem compounds has expanded from 25 million in 2008 [94] to 96 million in 2022 [95], representing a tenfold increase in that time period. A publicly accessible big data resource for compounds, including the majority of medications and drug prospects, with a variety of target response information is constituted by the enormous amount of bioassay data that is stored in PubChem and is updated on a daily

basis. A database that is quite similar to PubChem and that contains binding, functional, ADME, and toxicity data for a large number of chemicals is called ChEMBL [96]. In comparison to PubChem, ChEMBL has a significant amount of data that has been personally curated from the relevant literature. (https:// www.ebi.ac.uk/chembl/) The ChEMBL database currently contains more than 2.2 million compounds that have been evaluated against more than 12,000 targets [95]. As a result, the database contains activity data for 15 million compound-target pair combinations. There are a number of additional data sources that have been developed expressly for medications and drug prospects. For instance, there is a database known as DrugBank (https://www.drugbank.ca) that is open to the public and contains information on all approved medications, including their mechanisms, interactions, and key targets [97]. The present version of the DrugMatrix database stores extensive gene expression data obtained from the tissues of rats that were given over 600 different medicines, the majority of which targeted many important organs such as the liver. The Binding Database, often known as BindingDB, is a resource on the web that is open to the public and contains information about the drug-target binding that has been measured [98]. The proteins and enzymes that are deemed to be drug targets are the targets that are included in BindingDB. Currently, there are 1,587,753 binding data stored in BindingDB. These data are associated with 7,235 protein targets and 710,301 small molecules (https://www.bindingdb.org/bind/index.jsp). One such way to classify the public sources of big data is according to the dimensions of the associated digital files for the data sets. For instance, the present version of the PubChem bioassay database contains around 240 million different bioactivities. These bioactivities are organized into 30 gigabytes worth of XML files. To handle and analyze these available massive data sets, rather than utilizing personal computers with central processing units, it is required to leverage novel hardware techniques such as cloud computation (41, 51) and graphics processing units (GPUs) [98].

5. The Role Of Artificial Intelligence And Machine Learning In The Decision-Making Process From Lead To Candidate And Beyond:

The application of artificial intelligence (AI) and machine learning (ML) in pharmaceutical research and development has, up until this point, been primarily focused on research [99, 100]. This research includes target identification; docking-, fragment-, and motif-based generation of compound libraries; modeling of synthesis feasibility; rank-ordering likely hits according to structural and chemometric similarity to compounds having known activity and

affinity to the target(s); optimizing a smaller library for synthesis and high-throughput. When it comes to predicting absorption, distribution, metabolism, excretion, and toxicological features, progress has been sluggish when using AI and ML algorithms to lead optimization and lead-to-candidate (L2C) decision-making [101].

6. Molecular Design Using AI:

The fields of machine learning and artificial intelligence have transitioned from theoretical study to applications in the actual world. The subject of cheminformatics, and particularly QSAR, has traditionally been one of the earliest adopters of statistical approaches and machine learning; however, over the course of the past few years, the number of unique algorithms that have been developed in this sector has significantly expanded. Novel techniques, such as deep neural nets (DNNs), convolutional neural nets (CNNs), or recurrent neural nets (RNNs), have been increasingly recognized as valuable additions to the toolbox of chemoinformaticians [9, 89, 102-104]. This is in addition to more conventional models, such as Random Forest, Gradient Boosted Trees, or Gaussian Processes, which have been applied very successfully in the past[105]. CNNs are particularly appealing in this context due to the fact that they provide an alternative, data-driven approach to the extraction of molecular characteristics [106]. The promise of these innovative methodologies stems not only from somewhat superior performance metrics in retrospective evaluations, but also, and perhaps more importantly, in an inherent ability to interpret unstructured material as well as navigate and manipulate the "latent" environment. This has resulted in the development of a number of specialized artificial intelligence tools that are able to carry out tasks that are not feasible using "traditional" machine learning algorithms [107-109]. This is especially helpful for noisy and smaller data sets, for which data collection experiments are time consuming and expensive, for example in ADMET predictions [42, 110]. Another series of publications has shown that deep neural nets have the ability to use matrices of experimental observations (multitasking) rather than vectors to improve predictive accuracy [43, 111, 112].

Making all of these innovative machine-learning models and technologies practical in an industrial environment is a key issue to consider. This involves deployment, access, repeatability, monitoring, and maintenance. In addition, these brand new machine learning systems introduce brand new technical obstacles into industrial contexts, many of which are not immediately visible [44].

7. The Application of Artificial Intelligence to the Fields of Forensic Medicine and Toxicology:

The rise of artificial intelligence (AI) as the dominant technology will usher in the subsequent Industrial Revolution[113]. Artificial intelligence will transform all aspects of the business world. When it comes to the investigation of crimes, forensic medicine and toxicology is an important branch, and with the help of AI, this branch has a significant amount of room for growth and progress [114]. Any field of study or method cannot exist in isolation from developing technologies for very long. The conventional approach to conducting an autopsy and formulating an opinion has a number of drawbacks, all of which are potentially surmountable with the assistance of AI [115].In the field of forensic medicine, various procedures, such as the analysis of toxins, the collection of the various samples of medicolegal importance from body cavities, the detection of pathological changes in various organs of the body, the detection of various stains on the body, the detection of a weapon used in a crime, time since death calculations, and so on, are the areas in which AI will play a key role in framing the various opinions of medicolegal importance. In addition, AI may be included into pre-existing procedures for testing and analysis, which will make the entire process more efficient and accurate. It is possible that AI will play a significant role in the practice of forensic medicine and toxicology in the future [116]. Forensic medicine professionals, namely those who work in the fields of autopsy and toxicological analysis, stand to profit in a variety of ways from artificial intelligence. Finding data of sufficient quality for use in training the AI system is the primary obstacle faced when attempting to deploy AI in forensic medicine and toxicology. Therefore, initially, forensic medical professionals all around the world will need to exert a significant amount of work in order to offer extremely precise data to machines [117, 118]. The essential data for the machines may comprise a variety of postmortem findings, together with high-quality supporting photos and precise opinions regarding the pattern of the injury; it will also include a variety of statistical inputs pertaining to biomarkers, in addition to the analytical methodology. Providing data will take a significant amount of time, and it will also be necessary to periodically update the machine's data. However, in the event that specialists from the forensic sector are successful in overcoming these first obstacles, the AI tool will have the ability to have an advantage in the field of forensic medicine to frame diverse viewpoints of medicolegal significance. Computer technology and AI algorithms will improve forensic investigation methods with

greater accuracy and promptness. This is the classic morphological perspective. It is possible that AI will become an essential component of forensic medicine and toxicology procedures once it is included into the many testing and analysis processes that are already in place [119, 120].

Another aspect of artificial intelligence that raises concerns is how it would fare in a legal setting. Although the court might not see an opinion that was generated by AI as definitive proof, this view can nonetheless serve as corroborative evidence because the output of any machine is dependent on the facts that it is supplied. Before passing judgment on anything, even the judicial system needs to have a solid understanding of how the machine in question actually functions.In this kind of situation, the judges can seek out expert opinion regarding the reliability of the equipment. This transformation, on the other hand, will take place with the passage of time and the maturation of the method of using AI [121-123].

It will be a difficult task for policymakers in developing nations like India to establish high-tech infrastructure in the field of forensic medicine because the country's healthcare system is still in the early stages of development and is concentrated only in urban regions. The provision of healthcare services to each and every member of a developing nation's population in every region should currently be the top priority for developing nations.

8. Future Perspectives:

In the pharmaceutical industry, ML and DL methods of artificial intelligence can be utilized extensively to comprehend the chemical structure and activity relationship of lead molecules by analyzing a large amount of pharmaceutical data. Because of their superior data mining capabilities, AI-based technologies have become increasingly prevalent in recent years in the field of computer-assisted drug development. The performance of both deep learning and machine learning methods is directly influenced by the amount of data that is stored in the data mining technology. This method does have certain drawbacks, though. In spite of the fact that the successful construction of deep learning methods has a potentially effective approach to overcoming these issues, one of the most significant drawbacks is that the mechanism behind deep learning models is still not fully understood. In addition, brain models of formation are involved in changing various parameters; nevertheless, the optimization of these models has only been achieved through a limited number of practical guidelines. In recent years, there has been a primary increase in the number of applications of AI-based methodologies. The drug discovery process, which includes activities like de novo design and the identification of lead compounds, is formed by the enormous data sets. It is

reasonable to anticipate that, as a result of developments in a variety of sectors, there will be a tendency toward the more automated drug discovery process with the assistance of computers. This will produce results that are more accurate than those produced by other approaches. Therefore, additional study in vital and missing areas, innovative concepts in biological research sectors, and a drug discovery pipeline could potentially yield various findings that can be incorporated into the process of designing drugs.

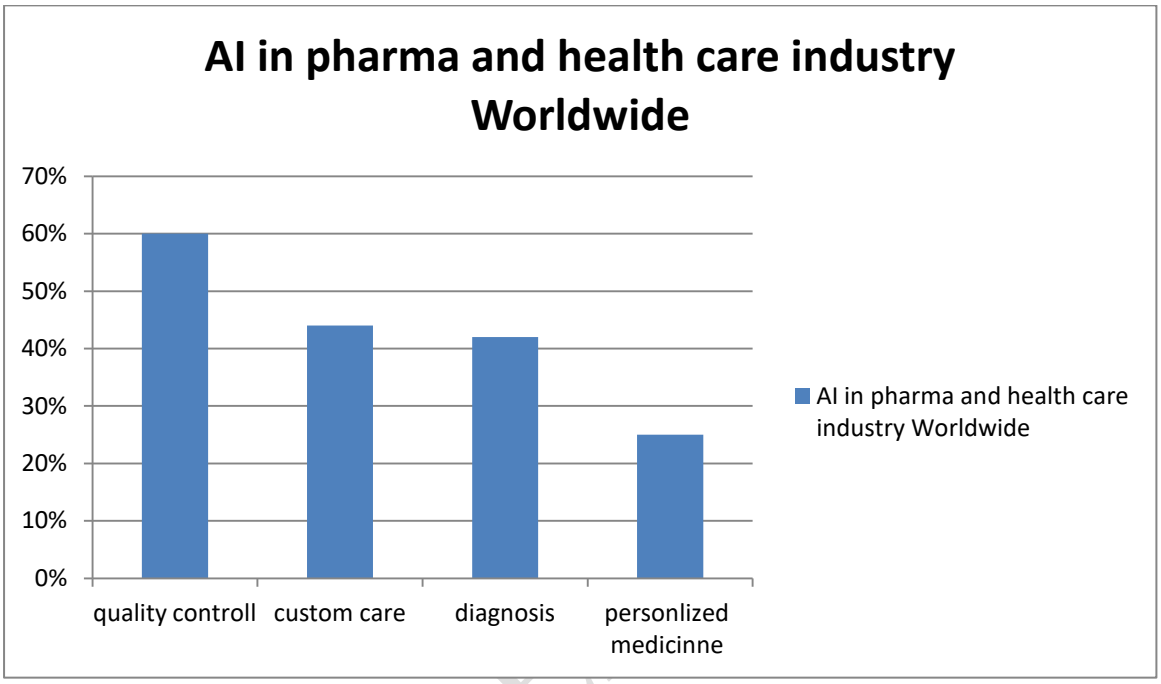

Figure 5 AI in pharma and health care

Conclusions:

The proliferation of computing power and other technologies has permeated every branch of scientific inquiry. The field of computer science that deals with artificial intelligence (AI) is one of the most important subfields, and its influence can be seen in many areas of science and technology, from fundamental engineering to medical treatments. Therefore, AI is currently being utilized in the fields of pharmaceutical chemistry and health care. In recent years, traditional approaches to drug design have been phased out in favor of computer-aided designs of various pharmaceuticals. There has been a significant increase in the usage of AI in recent years, which has led to significant improvements in drug design methods and production times. Additionally, the target proteins may be easily discovered by applying AI, which increases the likelihood that the proposed medicine will be successful. The application of AI technology in each stage of the drug design process reduces the risks to patients' health that are associated with preclinical testing and also brings down the overall cost of the project significantly. Based on the vast amounts of pharmaceutical data and the process of machine

learning, artificial intelligence is a useful tool for data mining. As a result, AI has been applied to de novo drug design, activity scoring, virtual screening, and in silico evaluation of the features of a pharmacological molecule, including absorption, distribution, metabolism, excretion, and toxicity.

In the past few years, artificial intelligence has rapidly become a standard analytical tool in the discovery of new drugs. This results in a significant number of new developments and improvements in our overall level of understanding regarding pharmacology. When putting AI models into practice, however, caution is necessary, and one must be aware of the potential hazards. To be more specific, the performance of any artificial intelligence model can only be as good as the data that is used to train it. If there are inherent flaws in the data that are used to train a model (for example, biases related to ethnicity, gender, or disease, or measurement errors), then such inaccuracies will also be present in the model, which will make the model less universal and more difficult to apply. Therefore, whenever selecting the data that will be used to train an AI model, extreme caution should be exercised at all times. The "accuracy-interpretability trade-off" presents yet another obvious barrier to the widespread implementation of AI in the healthcare sector. When it comes to AI models, the rule of thumb is that the more precise they are, the more difficult they are to interpret. In the realm of preclinical pharmacology, this is less of an issue because patients are not yet included in the research. Transparency and interpretability, on the other hand, take on a far greater significance in the clinical sector. This trade-off will compel healthcare practitioners to pick between two models: one that is simpler, more closely matches conventional statistics, and is easier to read, but provides less accurate results. The first model is highly accurate, but it is difficult to understand what exactly it does. In spite of this, there is an infinite number of ways in which these kinds of models can be utilized once one is aware of the benefits and drawbacks associated with various AI methodologies. In the not-too-distant future, we may anticipate that AI-based strategies will gradually replace the more traditional models that are being used at the moment. In addition, AI will begin to make its way into clinical pharmacology in the form of in silico clinical trials and AI-based decision support tools that can be classified as medical devices. These applications will begin to emerge in the next few years. In preparation for the latter application, the Food and Drug Administration (FDA) in the United States is now working on a guideline for the credibility of computational models used in medical device regulatory applications and other regulatory applications. This is being done in expectation of more widespread applications of AI in healthcare in general. This will, in the long run, result in more efficient paths for the development of new drugs,

and it will also allow us to better optimize the pharmacological therapy of each particular patient.

References:

1. *NextMove Software|SmallWorld. Available at https://www.nextmovesoftware.com/smallworld.html. Accessed 24 May 2019.*
2. *Ertl P, Lewis R, Martin E, Polyakov V (2017) In silico generation of novel, drug-like chemical matter using the LSTM neural network. arXiv preprint arXiv:171207449.*
3. *Jin W, Barzilay R, Jaakkola T (2018) Junction tree variational autoencoder for molecular graph generation. arXiv preprint arXiv:180204364.*
4. *Reaxys. In: Reaxys. Available at www.reaxys.com. Accessed 1 Jan 2020.*
5. *Cui J, Zhang H, Han H et al (2018) Improving 2D Face Recognition via Discriminative Face Depth Estimation. 2018 International Conference on Biometrics (ICB).*
6. *Carlini N, Liu C, Kos J, et al (2018) The secret sharer: measuring unintended neural network memorization & extracting secrets. arXiv preprint arXiv:180208232.*
7. *Raschka S (2018) Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:181112808.*
8. *Eykholt K, Evtimov I, Fernandes E, et al (2018) Robust Physical-World Attacks on Deep Learning Visual Classification. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.*
9. Alber, M., S. Lapuschkin, and P. Seegerer, *iNNvestigate neural networks.* J Mach Learn Res, 2019. **20**.
10. Almeida, A.F., et al., *Synthetic organic chemistry driven by artificial intelligence.* Nat Rev Chem, 2019. **3**.
11. Altae-Tran, H., et al., *Low data drug discovery with one-shot learning.* ACS Cent Sci, 2017. **3**.
12. Borrel, A., N.C. Kleinstreuer, and D. Fourches, *Exploring drug space with ChemMaps.com.* Bioinformatics, 2018. **34**.
13. Brown, N., et al., *GuacaMol: benchmarking models for de novo molecular design.* J Chem Inf Model, 2019. **59**.
14. Cha, K.H., N. Petrick, and A. Pezeshk, *Evaluation of data augmentation via synthetic images for improved breast mass detection on mammograms using deep learning.* J Med Imag (Bellingham), 2020. **7**.

15. Chen, H., O. Engkvist, and Y. Wang, *The rise of deep learning in drug discovery.* Drug Discov Today, 2018. **23**.
16. Chuang, K.V. and M.J. Keiser, *adversarial controls for scientific machine learning.* ACS Chem Biol, 2018. **13**.
17. Coley, C.W., W.H. Green, and K.F. Jensen, *Machine learning in computer-aided synthesis planning.* Acc Chem Res, 2018. **51**.
18. Coley, C.W., W. Jin, and L. Rogers, *A graph-convolutional neural network model for the prediction of chemical reactivity.* Chem Sci, 2019. **10**.
19. Dearden, J.C., M.T.D. Cronin, and K.L.E. Kaiser, *How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR).* SAR QSAR Environ Res, 2009. **20**.
20. Elton, D.C., et al., *Deep learning for molecular design—a review of the state of the art.* Mol Syst Design Eng, 2019. **4**.
21. Engkvist, O., P.O. Norrby, and N. Selmi, *Computational prediction of chemical reactions: current status and outlook.* Drug Discov Today, 2018. **23**.
22. Fechner, N., et al., *Estimation of the applicability domain of kernel-based machine learning models for virtual screening.* J Cheminform, 2010. **2**.
23. Fei-Fei, L., J. Deng, and K. Li, *ImageNet: constructing a large-scale image database.* J Vision, 2010. **9**.
24. Free, S.M. and J.W. Wilson, *A mathematical contribution to structure-activity studies.* J Med Chem, 1964. **7**.
25. Ghiandoni, G.M., M.J. Bodkin, and B. Chen, *Enhancing reaction-based de novo design using a multi-label reaction class recommender.* J Comput Aided Mol Des, 2020.
26. Gómez-Bombarelli, R., J.N. Wei, and D. Duvenaud, *automatic chemical design using a data-driven continuous representation of molecules.* ACS Cent Sci, 2018. **4**.
27. Goodnow, R.A., C.E. Dumelin, and A.D. Keefe, *DNA-encoded chemistry: enabling the deeper sampling of chemical space.* Nat Rev Drug Discov, 2017. **16**.
28. Green, D.V.S., S. Pickett, and C. Luscombe, *BRADSHAW: a system for automated molecular design.* J Comput Aided Mol Des, 2019.
29. Hansch, C. and T. Fujita, *p-σ-π Analysis. A method for the correlation of biological activity and chemical structure.* J Am Chem Soc, 1964. **86**.
30. Heaven, D., *Why deep-learning AIs are so easy to fool.* Nature, 2019. **574**.

31. Hoffmann, T. and M. Gastreich, *The next level in chemical space navigation: going far beyond enumerable compound libraries.* Drug Discov Today, 2019. **24**.
32. Holzgrabe, U., *QSAR: Hansch analysis and related approaches, H. Kubiny, VCH, Weinheim 1993. 232 Seiten, 60 Abb. und 32 Tab. 158,– DM. ISBN 3-527-30035-X.* Pharm Unserer Zeit, 1994. **23**.
33. Hristozov, D., M. Bodkin, and B. Chen, *ChemInform abstract: validation of reaction vectors for de novo design.* ChemInform, 2012. **43**.
34. Ivakhnenko, A.G. and V.G. Lapa, *Cybernetics and forecasting techniques*. 1967, New York: American Elsevier Pub. Co.
35. Kearnes, S., K. McCloskey, and M. Berndl, *Molecular graph convolutions: moving beyond fingerprints.* J Comput Aided Mol Des, 2016. **30**.
36. Kim, S., J. Chen, and T. Cheng, *PubChem 2019 update: improved access to chemical data.* Nucleic Acids Res, 2019. **47**.
37. Kola, I. and J. Landis, *Can the pharmaceutical industry reduce attrition rates?* Nat Rev Drug Discov, 2004. **3**.
38. Lapuschkin, S., S. Wäldchen, and A. Binder, *Unmasking Clever Hans predictors and assessing what machines really learn.* Nat Commun, 2019. **10**.
39. Lewis, R.A., *A general method for exploiting QSAR models in lead optimization.* J Med Chem, 2005. **48**.
40. Lin, A., B. Beck, and D. Horvath, *Diversifying chemical libraries with generative topographic mapping.* J Comput Aided Mol Des, 2019.
41. Mohan, A., Z. Sun, and S. Ghosh, *A machine-learning derived Huntington's disease progression model: insights for clinical trial design.* Mov Disord., 2022. **37**.
42. Liu, Q., et al., *Application of machine learning in drug development and regulation: current status and future potential.* Clin Pharm Ther., 2020. **107**.
43. Maharao, N., et al., *Entering the era of computationally driven drug development.* Drug Metab Rev., 2020. **52**.
44. Manolis, E., et al., *EMA Modelling and simulation working group.* CPT Pharmacometrics Syst Pharmacol., 2017. **6**.
45. Marshall, S., et al., *Model-informed drug discovery and development: current industry good practice and regulatory expectations and future perspectives.* CPT Pharmacometrics Syst Pharmacol., 2019. **8**.

46. Massard, C., S. Michiels, and C. Ferté, *High-throughput genomics and clinical outcome in hard-to-treat advanced cancers: results of the MOSCATO 01 trial.* Cancer Discov., 2017. **7**.
47. Moehler, M., M. Dvorkin, and N. Boku, *Phase III trial of avelumab maintenance after first-line induction chemotherapy versus continuation of chemotherapy in patients with gastric cancers: results from JAVELIN Gastric 100.* J Clin Oncol., 2021. **39**.
48. Morrissey, K.M., M. Marchand, and H. Patel, *Alternative dosing regimens for atezolizumab: an example of model-informed drug development in the postmarketing setting.* Cancer Chemother Pharmacol., 2019. **84**.
49. Parekh, V. and M.A. Jacobs, *Radiomics: a new application from established techniques.* Expert Rev Precis Med Drug Dev., 2016. **1**.
50. Paul, D., et al., *Artificial intelligence in drug discovery and development.* Drug Discov Today., 2021. **26**.
51. Sun, R., E.J. Limkin, and M. Vakalopoulou, *A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study.* Lancet Oncol., 2018. **19**.
52. Terranova, N., J. French, and H. Dai, *Pharmacometric modeling and machine learning analyses of prognostic and predictive factors in the JAVELIN Gastric 100 phase III trial of avelumab.* CPT Pharmacomet Syst Pharmacol., 2022. **11**.
53. Walden, D.M., et al., *Molecular simulation and statistical learning methods toward predicting drug-polymer amorphous solid dispersion miscibility, stability, and formulation design.* Molecules., 2021. **26**.
54. Wang, Y., et al., *Model-informed drug development: current US regulatory practice and future considerations.* Clin Pharmacol Ther., 2019. **105**.
55. Zhao, X., J. Shen, and V. Ivaturi, *Model-based evaluation of the efficacy and safety of nivolumab once every 4 weeks across multiple tumor types.* Ann Oncol., 2020. **31**.
56. *Leinfuss E. Changing the drug development playbook – model-informed drug development has arrived. PharmaVOICE, Nov-Dec 2016, https://www.certara.com/app/uploads/Resources/Articles/AR_ChangingDrugDevPlaybook.pdf. Accessed 17 Jan 2022.*
57. Zhavoronkov, A., Q. Vanhaelen, and T.I. Oprea, *Will artificial intelligence for drug discovery impact clinical pharmacology?* Clin Pharmacol Ther, 2020. **107**.

58. Zhang, M., Q. Su, and Y. Lu, *Application of machine learning approaches for protein-protein interactions prediction.* Med Chem, 2017. **13**.
59. Lowe, D.M., *Extraction of chemical structures and reactions from the literature*. 2012, Cambridge: PhD University of Cambridge.
60. Ma, J., R.P. Sheridan, and A. Liaw, *Deep neural nets as a method for quantitative structure-activity relationships.* J Chem Inf Model, 2015. **55**.
61. Mayr, A., G. Klambauer, and T. Unterthiner, *Large-scale comparison of machine learning methods for drug target prediction on ChEMBL.* Chem Sci, 2018. **9**.
62. Mendez, D., A. Gaulton, and A.P. Bento, *ChEMBL: towards direct deposition of bioassay data.* Nucleic Acids Res, 2019. **47**.
63. Méndez-Lucio, O., B. Baillif, and D.A. Clevert, *De novo generation of hit-like molecules from gene expression signatures using artificial intelligence.* Nat Commun, 2020. **11**.
64. Reker, D. and G. Schneider, *Active-learning strategies in computer-assisted drug discovery.* Drug Discov Today, 2015. **20**.
65. Reker, D., P. Schneider, and G. Schneider, *Multi-objective active machine learning rapidly improves structure-activity models and reveals new protein-protein interaction inhibitors.* Chem Sci, 2016. **7**.
66. Montanari, F., et al., *Modeling physico-chemical ADMET endpoints with multitask graph convolutional networks.* Molecules, 2020. **25**.
67. Morrison, C., *AI developers tout revolution, drugmakers talk evolution.* Nat Biotechnol, 2019.
68. Mullard, A., *New drugs cost US$2.6 billion to develop.* Nat Rev Drug Discov, 2014. **13**.
69. Olivecrona, M., et al., *Molecular de-novo design through deep reinforcement learning.* J Cheminform, 2017. **9**.
70. Patel, H., et al., *Knowledge-based approach to de novo design using reaction vectors.* J Chem Inf Model, 2009. **49**.
71. Prykhodko, O., S. Johansson, and P.C. Kotsias, *A de novo molecular generation method using latent vector based generative adversarial network.* J Cheminform, 2019. **11**.
72. Ramsundar, B., B. Liu, and Z. Wu, *Is multitask deep learning practical for pharma?* J Chem Inf Model, 2017. **57**.
73. Reymond, J.L., *The chemical space project.* Acc Chem Res, 2015. **48**.

74. Robinson, M.C., R.C. Glen, and A.A. Lee, *Validating the validation: reanalyzing a large-scale comparison of deep learning and machine learning models for bioactivity prediction.* J Comput Aided Mol Des, 2020.
75. Ruddigkeit, L., et al., *Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17.* J Chem Inf Model, 2012. **52**.
76. Samek, W. and K.R. Müller, *Towards explainable artificial intelligence explainable. AI: interpreting, Explaining and Visualizing Deep Learning*. 2019, Cham: Springer.
77. Sanchez-Lengeling, B. and A. Aspuru-Guzik, *Inverse molecular design using machine learning: generative models for matter engineering.* Science, 2018. **361**.
78. Sanchez-Lengeling, B., et al., *Optimizing distributions over molecular space An objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC).* ChemRxiv., 2017.
79. Schneider, G., *Automating drug discovery.* Nat Rev Drug Discov, 2018. **17**.
80. Schneider, P., W.P. Walters, and A.T. Plowright, *Rethinking drug design in the artificial intelligence era.* Nat Rev Drug Discov, 2019.
81. Sculley, D., G. Holt, and D. Golovin, *Hidden technical debt in machine learning systems.* Adv Neural Inf Process Syst, 2015. **2**.
82. Searls, D.B., *Data integration: challenges for drug discovery.* Nat Rev Drug Discov, 2005. **4**.
83. Segler, M.H.S., et al., *Generating focused molecule libraries for drug discovery with recurrent neural networks.* ACS Cent Sci, 2018. **4**.
84. Segler, M.H.S., M. Preuss, and M.P. Waller, *Planning chemical syntheses with deep neural networks and symbolic AI.* Nature, 2018. **555**.
85. Sheridan, R.P., *Interpretation of QSAR models by coloring atoms according to changes in predicted activity: how robust is it?* J Chem Inf Model, 2019. **59**.
86. Sheridan, R.P., et al., *Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR.* J Chem Inf Comput Sci, 2004. **44**.
87. Barile, F.A., *Clinical toxicology: principles and mechanisms*. 2010: CRC Press.
88. Basile, A.O., A. Yahi, and N.P. Tatonetti, *Artificial intelligence for drug toxicity and safety.* Trends Pharmacol Sci, 2019. **40**.
89. Baud, F.J. and P. Houzé, *Introduction to clinical toxicology*, in *An introduction to interdisciplinary toxicology*. 2020, Elsevier.

90. Bijlsma, N. and M.M. Cohen, *Environmental chemical assessment in clinical practice: Unveiling the elephant in the room.* Int J Environ Res Public Health, 2016. **13**.
91. Buch, V.H., I. Ahmed, and M. Maruthappu, *Artificial intelligence in medicine: current trends and future possibilities.* Br J Gen Pract, 2018. **68**.
92. Chary, M.A., et al., *The role and promise of artificial intelligence in medical toxicology.* J Med Toxicol, 2020. **16**.
93. Ciallella, H.L. and H. Zhu, *Advancing computational toxicology in the big data era by artificial intelligence: data-driven and mechanism-driven modeling for chemical toxicity.* Chem Res Toxicol, 2019. **32**.
94. Panch, T., P. Szolovits, and R. Atun, *Artificial intelligence, machine learning and health systems.* J Glob Health, 2018. **8**.
95. Sullivan, D.W. and S. Gad, *Clinical toxicology and clinical analytical toxicology*, in *Information resources in toxicology*. 2020, Elsevier.
96. Vatansever, S., et al., *Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: state-of-the-arts and future directions.* Med Res Rev, 2020. **41**.
97. El-Khateeb, E., et al., *Physiological-based pharmacokinetic modeling trends in pharmaceutical drug development over the last 20-years; in-depth analysis of applications, organizations, and platforms.* Biopharm Drug Dispos, 2020. **42**.
98. Whitehead, T.M., B.W.J. Irwin, and P. Hunt, *Imputation of assay bioactivity data using deep learning.* J Chem Inf Model, 2019. **59**.
99. Vamathevan, J., D. Clark, and P. Czodrowski, *Applications of machine learning in drug discovery and development.* Nat Rev Drug Discov, 2019. **18**.
100. Walters, W.P. and M. Murcko, *Assessing the impact of generative AI on medicinal chemistry.* Nat Biotechnol, 2020. **38**.
101. Stokes, J.M., K. Yang, and K. Swanson, *A deep learning approach to antibiotic discovery.* Cell, 2020. **180**.
102. Barrett, J.S., et al., *Role of disease progression models in drug development.* Pharm Res., 2022. **39**.
103. Antontsev, V., et al., *A hybrid modeling approach for assessing mechanistic models of small molecule partitioning in vivo using a machine learning-integrated modeling platform.* Sci Rep., 2021. **11**.

104. Chakravarty, K., et al., *Accelerated repurposing and drug development of pulmonary hypertension therapies for COVID-19 treatment using an AI-integrated biosimulation platform.* Molecules., 2021. **26**.
105. Chen, E.P., R.W. Bondi, and P.J. Michalski, *Model-based target pharmacology assessment (mTPA): an approach using PBPK/PD modeling and machine learning to design medicinal chemistry and DMPK strategies in early drug discovery.* J Med Chem., 2021. **64**.
106. Combes, F.P., et al., *Model-informed drug development for everolimus dosing selection in pediatric infant patients.* CPT Pharmacometrics Syst Pharmacol., 2020. **9**.
107. DiMasi, J.A., H.G. Grabowski, and R.W. Hansen, *Innovation in the pharmaceutical industry: new estimates of R&D costs.* J Health Econ., 2016. **47**.
108. Khotimchenko, M., et al., *In silico simulation of the systemic drug exposure following the topical application of opioid analgesics in patients with cutaneous lesions.* Pharmaceutics., 2021. **13**.
109. Le-Rademacher, J.G., et al., *Application of multi-state models in cancer clinical trials.* Clin Trials., 2018. **15**.
110. Lesko, L.J., *Perspective on model-informed drug development.* CPT Pharmacometrics Syst Pharmacol., 2021. **10**.
111. Madabushi, R., et al., *The US Food and Drug Administration's model-informed drug development paired meeting pilot program: early experience and impact.* Clin Pharm Ther., 2019. **106**.
112. Maharao, N., et al., *Scalable in silico simulation of transdermal drug permeability: application of BIOiSIM platform.* Drug Des Devel Ther., 2020. **11**.
113. Zador, A.M., *A critique of pure learning and what artificial neural networks can learn from animal brains.* Nat Commun, 2019. **10**.
114. You, J., R.D. McLeod, and P. Hu, *Predicting drug-target interaction network using deep learning model.* Comput Biol Chem, 2019. **80**.
115. Yang, X., Y. Wang, and R. Byrne, *Concepts of artificial intelligence for computer-assisted drug discovery.* Chem Rev, 2019. **119**.
116. Yamashita, R., M. Nishio, and R.K.G. Do, *Convolutional neural networks: an overview and application in radiology.* Insights Imag, 2018. **9**.
117. Yamashita, F., S. Wanchana, and M. Hashida, *Quantitative structure/property relationship analysis of Caco-2 permeability using a genetic algorithm-based partial least squares method.* J Pharm Sci, 2002. **91**.

118. Xu, Y., J. Pei, and L. Lai, *Deep learning based regression and multiclass models for acute oral toxicity prediction with automatic chemical feature extraction.* J Chem Inf Model, 2017. **57**.
119. Wilton, D.J., R.F. Harrison, and P. Willett, *Virtual screening using binary kernel discrimination: analysis of pesticide data.* J Chem Inf Model, 2006. **46**.
120. Wang, T., X.S. Yuan, and M.B. Wu, *The advancement of multidimensional QSAR for novel drug discovery—where are we headed?* Expert Opin Drug Discov, 2017. **12**.
121. Wang, S., S. Sun, and Z. Li, *Accurate de novo prediction of protein contact map by ultra-deep learning model.* PLoS Comput Biol, 2017. **13**.
122. Wang, N.N., J. Dong, and Y.H. Deng, *ADME properties evaluation in drug discovery: prediction of Caco-2 cell permeability using a combination of NSGA-II and boosting.* J Chem Inf Model, 2016. **56**.
123. Volk, M.J., I. Lourentzou, and S. Mishra, *Biosystems design by machine learning.* ACS Synth Biol, 2020. **9**.