



# Opportunities and challenges in application of artificial intelligence in pharmacology

Mandeep Kumar<sup>1</sup> · T. P. Nhung Nguyen<sup>1,2</sup> · Jasleen Kaur<sup>3</sup> · Thakur Gurjeet Singh<sup>4</sup> · Divya Soni<sup>5</sup> · Randhir Singh<sup>5</sup> · Puneet Kumar<sup>5</sup>

Received: 20 October 2022 / Revised: 23 December 2022 / Accepted: 25 December 2022 / Published online: 9 January 2023  
© The Author(s) under exclusive licence to Maj Institute of Pharmacology Polish Academy of Sciences 2023

## Abstract

Artificial intelligence (AI) is a machine science that can mimic human behaviour like intelligent analysis of data. AI functions with specialized algorithms and integrates with deep and machine learning. Living in the digital world can generate a huge amount of medical data every day. Therefore, we need an automated and reliable evaluation tool that can make decisions more accurately and faster. Machine learning has the potential to learn, understand and analyse the data used in healthcare systems. In the last few years, AI is known to be employed in various fields in pharmaceutical science especially in pharmacological research. It helps in the analysis of preclinical (laboratory animals) and clinical (in human) trial data. AI also plays important role in various processes such as drug discovery/manufacturing, diagnosis of big data for disease identification, personalized treatment, clinical trial research, radiotherapy, surgical robotics, smart electronic health records, and epidemic outbreak prediction. Moreover, AI has been used in the evaluation of biomarkers and diseases. In this review, we explain various models and general processes of machine learning and their role in pharmacological science. Therefore, AI with deep learning and machine learning could be relevant in pharmacological research.

**Keywords** Artificial intelligence · Big data · Machine learning · Bioinformatics · Algorithm · Data mining

## Abbreviations

AI	Artificial intelligence	LSSVM	Least square support vector machine
ML	Machine learning	PSOWNN	Particle swarm optimized wavelet neural network
DL	Deep learning	EPNN	Enhanced probabilistic neural network
MGI	McKinsey global institute	CNN	Convolutional neural networks
R&D	Research and development	RDR	Referable diabetic retinopathy
MRI	Magnetic resonance imaging	CLI	Critical limb ischaemia
ECG	Electrocardiogram	GLCM	Grey level co-occurrence matrix
NLP	Natural language processing	SVM	Support vector machines
		RF	Random forest
		ANN	Artificial neural network
		GAN	Generative adversarial network
		VAE	Variational autoencoder
		AAE	Adversarial autoencoder
		QSAR	Quantitative structure–activity relationship
		VS	Virtual screening
		SBVS	Structure-based virtual screening
		LBVS	Ligand-based virtual screening
		NMR	Nuclear magnetic resonance
		DNN	Deep neural network
		Knn	K-nearest neighbouring
		BP-NN	Backpropagation network
		DNDD	De novo drug design

✉ Puneet Kumar  
puneet.bansal@cup.edu.in; punnubansal79@gmail.com

<sup>1</sup> Department of Pharmacy, Unit of Pharmacology and Toxicology, University of Genoa, Genoa, Italy

<sup>2</sup> Department of Pharmacy, Da Nang University of Medical Technology and Pharmacy, Da Nang, Vietnam

<sup>3</sup> Department of Pharmacology and Toxicology, National Institute of Pharmaceutical Education and Research (NIPER), Lucknow, Uttar Pradesh 226002, India

<sup>4</sup> Chitkara College of Pharmacy, Chitkara University, Rajpura, Punjab, India

<sup>5</sup> Department of Pharmacology, Central University of Punjab, Ghudda, Bathinda, Punjab 151401, India

ADMET	Absorption, distribution, metabolism, excretion, and toxicity
PPB	Plasma protein binding
MAH	Marketing authorization holder
ICSR	Individual case safety report
GWAS	Genome-wide association studies

molecular basis, epidemiology, and related disciplines at the population level. In this review, we briefly discuss the significant contribution of AI/ML in target identification and drug discovery (Fig. 1).

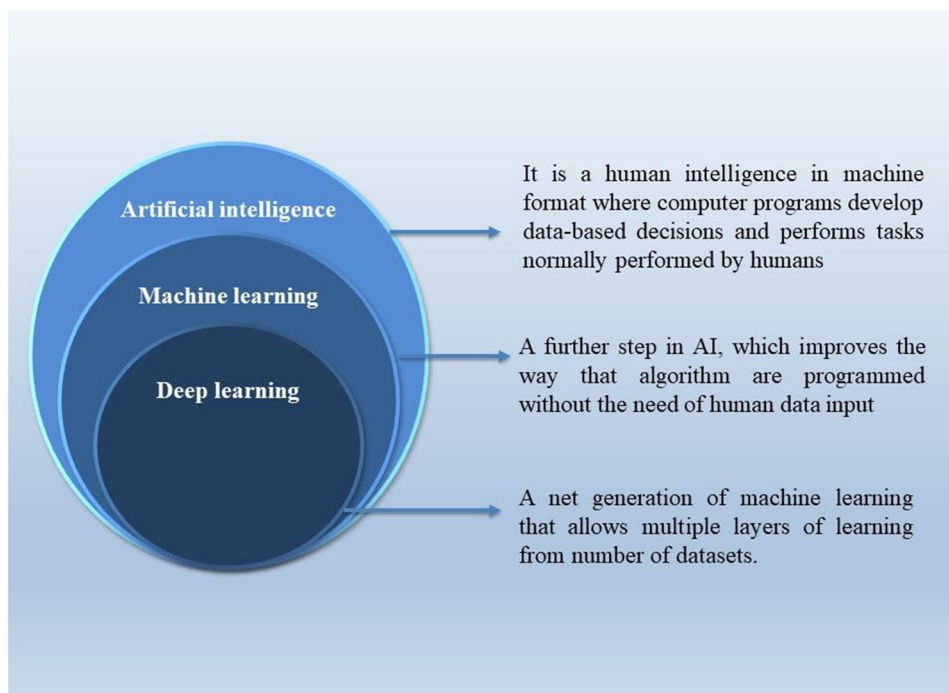
## Introduction

Artificial intelligence (AI) is a multi-disciplinary approach that has achieved great success in recent years especially in the area of machine learning (ML) and deep learning (DL) [1]. We are living in a technological world that generates huge amount of data every year in almost every field. AI provides assistance in handling the vast amount of data for humans. Thus, scientific societies generating big databases are considered as part of ML. Nowadays, machines are empowering humans in physical work and can provide rapid development at every stage. Machines have advantages over humans in various skills such as analysing, learning and understanding communication. ML can assess big data without any human intervention by developing and utilizing complex computer programmes and can also assist in various stages of drug discovery, including pharmacological research such as identification of lead compounds [2]. AI uses advanced mathematical processes and investigative procedures to process all types of data and overpower human intelligence. It plays a significant role in clinical pharmacology, including,

## AI/ML in pharmaceutical sciences

Pharmaceutical sciences consist of a wide range of scientific procedures related to drug discovery and development. It requires many efforts for the expansion of health care services. AI provides the best approach for a better health care system [3]. To get better results, AI and ML generally need a big amount of data and most of the pharmaceutical and healthcare sectors have extensive data [4, 5]. For example, in 2019, The McKinsey global institute (MGI) generated up to \$100 billion value worth of data in the US healthcare system [6, 7]. AI and ML are helping to manage the best health care options for physicians, consumers, insurers, and regulators. Data is generated from several sources such as academic domain, research and development (R&D) organisations, industrial units, clinical and community pharmacy [8]. AI and ML provides a better option to synchronize the large healthcare information and improve the healthcare system and treatments [9]. The use of AI and ML is already in process in pharmaceutical industries for different tasks such as drug discovery and evaluation of active compound, disease diagnosis, clinical trial, radiotherapy, and smart electronic health record [10, 11].

**Fig. 1** Demonstration of AI integration with machine learning and deep learning. *AI* artificial intelligence



## AI/ML in pharmacology

AI/ML has a significant influence on the pharmacological sciences. The use of various analytical techniques such as magnetic resonance imaging (MRI), X-ray, electrocardiogram (ECG) and histopathological imaging give more refined results with the help of variable sensor and data acquisition systems when compared with the conventional technologies [12–14]. It also improves the healthcare-based data via detecting the retina and dermal response by analysing input and stored data [15, 16]. Furthermore, AI provides health care services associated with animal research, such as animal behaviour, movement, physiological and pathological changes, as well as selection of suitable drugs in specific conditions [17].

## ML processes and their models

The ML process is quite similar to the learning behaviour of the human brain. The human brain works with billions of neurons by interpreting various perceptions such as images, sound, smell, structures, movements, and recognition of dimensional patterns. Similarly, machines work through the electronic nose to analyse and calculate the data [18]. ML process mainly requires two components. The first component is Input (all the data presented for analysis) and the second component is Output (The result of the calculation done by the ML algorithm) [19, 20]. The ML process starts from the preparation of high-quality data like ML which randomises the data to exclude any anomalies or duplicates [20]. It divides the data into three sections: training data, validation data, and test data [21]. Before testing the different datasets, it undergoes some training and validation, and after the data process, the next step is to select the algorithm and the learning model [22]. The selection of algorithm and learning model depends on the type of data and what task needs to be automated. In this procedure, usually, supervised models are used, but it requires labelled data points. However, in case of insufficient data: unsupervised, semi-supervised, or self-supervised learning models can be used [23]. In the training phase, the algorithm matches the outcome with the previous output, and if there is any kind of error the model is amended and another iteration is tested [24].

## Model of ML

As discussed above, based on the type of data, ML is divided into three models: supervised learning, unsupervised learning, and reinforcement learning. The supervised model

requires well-labelled input data for learning and after training, the algorithm can analyse non-labelled data [25]. In this model, the elements are sorted into groups with predefined features, and the value is predicted based on training data calculations [26]. Whereas in unsupervised learning, the machine tries to find patterns and correlations between non-labelled data points [27]. Consequently, an algorithm must assemble data by characteristics which differentiates them from other groups of objects. Data types such as MRI scans, digital photographs, and audio signals are characterized by high dimensions that indicate the number of features for each observation [28]. ML decreases the amount of data by selecting important attributes or combining similar traits.

## AI in medical diagnosis

A vast number of articles published on diverse applications of AI demonstrate that AI is successfully deployed in medical diagnoses like skin cancers, neurological diseases, strokes, Alzheimer's disease, acute ischaemic stroke, etc. Natural language processing (NLP), the first reading device to provides excellent flexibility for doctors to study the descriptors of X-rays on the chest are also used to treat infectious diseases [29] and least square support vector machine (LSSVM) used in cancer diagnosis were introduced based on AI [30]. After two years, SVM was used to identify neurological as well as psychiatric disease imaging biomarkers [31]. Breast defects can be diagnosed by particle swarm optimized wavelet neural network (PSOWNN) [32]. Various combinations of SVM are used to detect the initial stage of Alzheimer's disease [33]. Similarly, Parkinson's disease can be diagnosed with the help of an enhanced probabilistic neural network (EPNN) [34]. Convolutional Neural Networks (CNN) was proposed [35] in the study of referable diabetic retinopathy (RDR) in diabetic patients. The Naïve Bayes classification was proposed that observed and rectified heavy stroke lesions of people T1-weighted MRI scans [36]. Recently it has been demonstrated that an 11-layer deep, multi-scale, 3D CNN used for the lesion segmentation in multimodal brain MRI [37].

CNN is also used to diagnose, stratify, and treat congenital cataracts [38]. A study improved the laboratory AE reporting using NLP to study EMR-based AE ascertainment and grading substantially [39]. NLP is also used to implement the EMR's power to determine a group of patients with intracranial aneurysms [40]. A study used EMR and NLP to evaluate the suicidal behaviour of pregnant women [41]. Moreover, various studies used NLP-based algorithms to diagnose different diseases such as critical limb ischaemia (CLI) [42], acute ischaemic stroke (AIS) [43], and ischaemic Stroke Thrombolysis [44].

Apart from the human-based databases, AI also is employed in veterinary and agricultural sciences such as hybrid prediction, cattle behaviours, disease identification, chewing patterns, etc. Dairy cattle are used to develop ANN models for rumen fermentation patterns [45]. Previous studies introduced grey level co-occurrence matrix (GLCM) to determine patterns with several variations, robustness to geometrical distortions, and simple transformation of the cattle race, in contrast, energy, and homogeneity [46]. AI is also involved in the development of agricultural techniques [47]. A study evaluated an automated and accurate detection method in rice seedlings of the pathogen, which took less time than a naked eye examination [48]. A tool based on ANN/XY-Fusion identifies and differentiates the infected and healthy plants during the growth of vegetation with 95.16% accuracy [49]. Similarly, a technique based on the SVM image differentiated the parasites and thrips in strawberries [50]. Another CNN-based method differentiates the fresh and diseased leaves [51].

## AI in drug discovery and development

Generally, a molecule takes approximately 13.5 years to reach the point of approval with total research and development (R&D) estimated costs around \$2.6 billion [52]. However, AI development has immensely impacted drug discovery and development with numerous advantages such as speeding up the time, reducing cost-consuming lengthy protocols due to the better utilization of available resources [53]. The creative collaboration between mind and machine has been anticipated for better decisions in the process of drug designing, chemical synthesis, and analysis of biological tests. The general workflow of building an AI model in drug discovery includes four stages: first, the definition of the problem that is specific or general needs to be clarified; second, choosing suitable AI algorithm and setting initial values for hyperparameters that provides appropriate AI architecture; third, the input data need to be prepared that are satisfactory in quality and quantity, representation characteristic, fitting proportion; lastly, the model training and evaluation such as training algorithms, optimization strategies, evaluation mechanisms and metrics, and the hyperparameter tuning algorithms are chosen [54].

In the first and second stages, various frequent algorithm models like support vector machines (SVM), random forest (RF), artificial neural network (ANN), deep Boltzmann machine, deep belief network, generative adversarial network (GAN), variational autoencoder (VAE), adversarial autoencoder (AAE), symbolic learning, and meta-learning can be applied in the construction model [54, 55]. Notably, ANN is one of the most powerful nonlinear data models-it has gained broad popularity in the past 2 decades as it

implements into a quantitative structure–activity relationship (QSAR) and virtual screening [56, 57]. Meanwhile, the GAN technique has been contributed to medicinal chemistry as molecular de novo design and biochemical science as de novo peptide and protein design, dimension reduction of single-cell data in preclinical development [58].

There are two main types of data including input X and output Y for building AI models; the input X data can be a fixed-length input vector (molecular descriptors, fingerprints), a sequence (SMILES strings, biomacromolecule structures), or molecular structure graph; the output data can be real-valued numbers, binary values, integer values, fixed-size vector, sequential data, single data column or multiple data columns [54]. Due to the large volumes of input data, several database libraries have been summarized to provide comprehensive information: for example, in biomolecule target identification, the platforms DisGeNET, CTD, LinkedOmics, Open-Target platform, DepMap portal, HMDD, STRING, Therapeutic Target Database-TTD have supported in managing the heterogeneous omics data [43].

To the last stage, for model training, a one-shot learning algorithm has been introduced, set up the iterative refinement of the long short-term memory architecture, and received strong results in predictive power meaningful for low-data drugs discovery. A continuous, data-driven auto encoder model has been applied in chemical design as optimization methods in the reconstruction of SMILES strings and capturing characteristic features of a molecular training set.

## AI in peptide-based drug discovery

Peptide as a therapeutic option in drug discovery process has received more attention over the last few decades, particularly for anticancer peptides, antimicrobial, and anti-inflammatory peptides. In comparison to small molecules, peptide-based therapy has high specificity and low toxicity. Therefore, peptides are widely used in discovery and design of newer drugs [59]. Moreover, biopeptide discovery and development are labour-intensive, time taking process and mostly dependent on a variety of factors. However, the ML approach can predict the use of therapeutic peptides quickly and effectively. It can improve decision making and discovery for well-defined queries. Some of the machine-based methods such as support random forest, extremely randomised tree, and DL methods are effective in peptide-based drug development. These approaches make to predict functional peptides with greater accuracy [60]. To predict therapeutic peptides, three tools are proposed: blood–brain barrier penetrating peptides (BBPs), antihypertensive peptides, and antiparasitic peptides. Prediction of BBPs using a random forest method accelerating the discovery of new

drugs to treat several brain diseases [61]. Additionally, anticancer peptides (ACPs) are therapeutic peptide drugs that have been shown to target and kill cancer cells. ACPs based on ML and DL have several benefits, including high specificity, and low toxicity under normal physiological circumstances [62]. During the COVID-19 pandemic, a peptide library was created to combat the SARS-Cov-2 virus of which four peptides were found effective, having high binding affinity for protease enzyme using AI [63]. Furthermore, researchers studied the role of dietary peptide in immunomodulating activity with the help of AI algorithm and found that this bioactive peptide has a high affinity for inflammatory receptors and suppresses the expression of pro-inflammatory cytokines such as tumour necrosis factor (TNF- $\alpha$ ), and nitrogen oxides [64].

### Structure-based virtual screening (VS)

The identification of drug-target interactions is crucial in drug development and hence, sophisticated ML techniques in VS via verifying the physicochemical properties related to compound structures and/or target receptors have been demonstrated to generate predictive models [65, 66]. VS can be divided into structure-based virtual screening (SBVS) and ligand-based virtual screening (LBVS). Structure-based virtual methods utilize 3D structures of the targets and compounds that have been confirmed by X-ray crystallography or nuclear magnetic resonance (NMR), respectively. Molecular docking a major technique in VS contains two steps: first, a ligand from a database platform is virtually docked into the binding site of the receptor based on the steric and physical, chemical properties; then, a mathematical scoring function will calculate the energetic binding affinity [66, 67]. Some of the most popular docking tools are AutoDock, Glide, DOCK [68–70].

Recently, AI algorithms have been applied in nonparametric scoring functions to estimate the binding affinity and correct the disadvantages of classical methods as well as improving accuracy [43, 54, 71]. The basic techniques that are applied to improve scoring function in AI include naïve Bayes, SVM, RF, feed-forward ANNs, and deep neural network (DNN) approaches [54, 72–74]. Currently, the novel method as Similarity of the Interaction Energy Vector Score inspired from finger-print methods is also proposed with accuracy improvements [75].

In the context of ML approaches, RF and SVM have been applied to improve docking scoring functions [71]. ML-based RF-Score has been formerly introduced that extremely improves training set size and obtained protein–ligand binding affinity predictions in the diverse test set [76]. It has been presented recently that ALADDIN- an integrated ML and docking approach and the RF classifier implicated the

accuracy in establishing all-against-all ensemble docking in VEGFR2, p38 $\alpha$  MAPK, and GCR and solved the challenges in docking and scoring functions such as protein flexibility, solvation [77]. Several algorithms have been conducted such as k-nearest neighboring (kNN), Neural Network, RF, and SVM based on leave-one-out random sampling model to establish novel P-glycoprotein inhibitors (the input compounds were retrieved from the ChEMBL database) then the RF algorithm performed better in learning and validation. Recently, traditional problem was fixed in the naïve Bayesian model (the classifying performance of compounds could decrease when more receptor structures were added to the ensemble) via several ML models such as kNN, logistic regression, SVM, and RF for 20 protein kinases [73].

Also, deep learning (DL) a subset of ML used to improve docking results has been extensively implemented in drug design and development [55, 71]. For example, a study developed a DL neural network architecture that the input data were protein voxels and ligand fingerprints, and the output linear data were RMSD<sub>min</sub>, RMSD<sub>ave</sub>, and nRMSD by DockBench [78]. Previously, introduced, DeepVS, based on CNN has been introduced which also achieved good results (the best AUC ROC has ever reported) without human-defined parameters [79]. In general, the AI application in structure-based virtual screening is a promising tool; however, the output depends on multiple features such as the dataset, AI models and the definition of precise parameters [74, 80].

### Ligand-based virtual screening

LBVS is the first choice when the 3D structure of the target compound is not available. It is based on the hypothesis that if the structures are similar, then the biological effects are the same [54, 81]. So far, the AI approach has been applied in the field of QSAR based LBVS successfully [81]. Similar AI algorithms to the SBVS methods (as mentioned above) implemented in QSAR based LBVS include ANN, RF, SVM, Bayesian algorithm, DNN, kNN [81, 82].

ANN the most popular paradigm for nonlinear modeling in QSAR aims to imitate the human nervous system workflow and contains several neuron layers. ANN is integrated with adaptive neuro-fuzzy inference systems and multiple linear regression (MLR) to the dataset consisting of 90 pyridinylimidazole-based compounds (inhibitors of p38Rmitogen-activated protein kinases); the performance of ANN was a better predictor model (ANN vs MLR, R<sup>2</sup> training: 0.8520, 0.4049, respectively) to establish physicochemical properties and output descriptors relationship [83]. A study optimized ANN architectures and interpreted six different methods (partial derivative-PaD, pairwise partial derivative, weights, perturbation, profile methods, and

sum of ranking differences analysis) to figure out the relationship between quantum mechanical molecular descriptors and output (Trolox-equivalent antioxidant capacity of 33 flavonoids). The authors concluded that the PaD and profile methods were most stable [84]. There are several subtypes of ANNs such as feed-forward backpropagation network (BP-NN), radial basis function networks and probabilistic neural networks, and linear regression was combined with nonlinear BP-NN-QSAR model to investigate inhibitory activities of pyridinone derivatives with HIV-1 reverse transcriptase; the results showed that the model was robust and cost-effective for pIC<sub>50</sub> estimation, capable of prediction of complex relationships [85]. Previously, it has been presented as three molecular fingerprints (namely FP2, MACCS, and ECFP6) combined ANN-QSAR (FANN-QSAR) to predict biological activities of cannabinoid ligands; after validation, ECFP6-ANN-QSAR required no alignment in the training process, and its performance consistently across diverse data sets was better than others [82, 86]. In summary, the ANN algorithm achieved good generalization, high prediction accuracy, and pattern recognition ability for unseen data [54, 86].

Among the other methods, DNN is a powerful tool that can deal with large data without manual engineering. It is built on six datasets collected from the ChEMBL database (EGFR inhibitors); the DNN model also obtained high performance in comparison with the RF method by cross-validation or screening a large compound library (PubChem, ChemDiv database) [87]. Besides, the least-squares- SVM (LS-SVM) and genetic algorithm-MLR algorithms has been conducted to predict IC<sub>50</sub> of poly ADP-ribose polymerase-1 inhibitors for breast cancer, the outcomes ( $R^2$ , F, RMSE,  $Q^2_{cv}$ ) proved that LS-SVM had good potential mathematical optimization base in comparison with MLR [88]. Moreover, RF and Multiple Partial Least Squares Regression, RF model was more stable, reliable, and precise in toxicity prediction of nano-TiO<sub>2</sub> on Hk-2 cells [89].

In other studies, the authors developed multiple algorithms to investigate QSAR [90–92]. For example, A study combined kNN, SVM, MLR, neural nets to predict the brain uptake ability of 25 known drugs obtained from PubChem [90]. Previously, seven methods (naïve Bayes classifier, Sequential Minimal Optimization-SMO, Instance-Based Learning, Decorate, Hyper Pipes, PART, and RF) have been conducted on the thirteen data sets (HIV-1 integrase inhibitors from ChEMBL database), the SMO was the highest efficiency to classify compounds [91]. Screened for histone deacetylase-3 inhibitors has been done by five ML classifiers (kNN, SVM, RF, DNN, eXtreme Gradient Boosting-XGBoost), and the best-performing one was the XGBoost. Altogether, developing numerous AI algorithms and larger data sizes have been devoted to the LBVS growth and acted as a useful tool in drug R&D [92].

## De novo drug design

The concept of de novo drug design (DNDD) based on AI with diverse techniques (e.g.: the autoencoders—AE, graph neural networks—GNNs, GAN, CNN, and the recurrent neural networks—RNNs) aims at generating novel compounds (previously unknown) with desired properties [93–95]. Normally, the algorithms contained two steps: firstly, from the worthy databases (ChEMBL, ZINC, PubChem), the model automatically generated new molecules based on rules (SMILES, molecular graph); secondly, reinforcement learning methods speed up to explore the novel regions to design structures with promising activities [94]. The benefits of DNDD are tremendous such as broaden exploration of chemical space, lessen costs, time-efficient manner, designed structures with intellectual properties; however, the challenges remain as synthetic procedure of the formulas, the regulatory acceptance and standard of the models, analysis, sharing platforms, or training datasets [25].

RNN-based generative models are suitable for sequential data (SMILES) and can be applied for multi-objective evolutionary DNDD [95]. RNN model composed of three layers with 512 gated recurrent units per layer in a multi-objective approach targeting neuraminidase, acetylcholinesterase, novel SARS-CoV-2 main protease. This framework was suited for lead generation and optimization phases; the compounds were generated with relevant physicochemical properties (MW, logP, HBA, HBD) [94]. RNN trained by reinforcement learning with a special exploration strategy (Drugex) to design inhibitor ligands of adenosine A<sub>2A</sub> receptor; for the training process, 10,000 SMILES sequences were constructed; this strategy generated molecules with diverse chemical activities while maintaining the similarity to the known ligands [96]. Furthermore, the recently established bidirectional generative RNNs for SMILES-based molecule design (BIMODAL), 30,000 unique and novel SMILES samples were used in the training set, this method suited for scaffold diversity and chemical-biological relevance (evaluated by FCD values- with 1024 hidden units,  $FCD = 1.59 \pm 0.03$  when the starting point was fixed,  $FCD = 1.62 \pm 0.04$  when the starting point was random- were lower than other models) [97].

In other studies, the CNN model has been successfully employed in image processing in both training and test phases as a feature detector [95], classify images, score protein–ligand, and pose prediction [98]. Lately, the GNN model can be operated on the graph structure data and can be applied in molecule scoring, generation, and optimization [99]. Meanwhile, GAN architecture provides better results in image generation processing [58] for example, implementation of Mol-CycleGAN—a CycleGAN

framework to optimize compounds from ZINC, ChEMBL databases with high structural similarity and 99.75% success [100]. Lastly, autoencoder models can be categorized into three subtypes such as VAE (the algorithm that converts discrete representations to multidimensional continuous values of molecules), sequence-to-sequence autoencoder (an algorithm that transfers an input sequence to a fixed-sized vector), AAE (a model utilized to generate molecules with the desired properties in the form of fingerprints) [95]. Hence, the diversity of these methods allows the development of AI approaches in drug discovery and tremendous advantages, as mentioned above.

## AI applications in ADMET predictions

Currently, *in silico* absorption, distribution, metabolism, excretion, and tolerable toxicity (ADMET) assessments have progressed tremendously alongside the rise of AI tools, the more accurate models have been developed [101, 102]. Methods such as DNNs, ANNs, RFs, and SVMs, k-NN have opened broad boulevard in this field with good performance, however, several factors should be considered, such as encoding functions, the quantity, and quality of input data [95]. Intestinal drug absorption is one of the most influential in bioavailability, several studies have established for accuracy with acceptable range [95, 103]. An innovative hierarchical SVM scheme was developed to forecast colon carcinoma cell layer (Caco-2) on a dataset containing 104 and 26 molecules (training and test set, respectively), the model showed excellent qualitative performance, unveiled great accuracy [103]. Earlier, DNN architecture was proposed using 209 molecular descriptors and obtained good discriminant power for cellular permeability prediction in Caco-2 cell lines of compounds [104]. In a larger dataset (1272 compounds) four methods MLR, partial least-squares PLS, SVM, and boosting to predict permeability, and the boosting model was combined together which was the most suitable with the highest  $Q^2$ , RMSE, CV,  $R^2$  values compared to others [105].

Plasma protein binding (PPB) is a significant pharmacokinetic property that influences the drug volume distribution. Due to the time-consuming and high cost of common experiments, *in silico* construction of predictive models using heterogeneous data has been developed [106]. Prediction of the PPB values of cyclic peptides by two algorithms (enumerating lasso solutions -ELS and forward beam search—FBS), the ELS was more robust in predicting diverse molecules with high generalization ability [107]. In a study, the authors presented six models (SVM, ANN, k-NN, PLS, Probabilistic neural network, Linear discriminant analysis) in a 736 compounds dataset; all methods had high efficiency in binary classification and PPB prediction,

SVM demonstrated the best accuracy, sensitivity, specificity, precision, and  $F1$  score [108]. In addition, prediction models have also been established to deepen knowledge on other natural barriers (e.g., the blood–brain barrier–BBB) that affect drug distribution [109]. DL-based RNN has been proposed to predict compounds penetration to the CNS, the accuracy and specificity scores were 96.53% and 98.08%, respectively; the dataset comprised compounds encoded in SMILES (1803 BBB<sup>+</sup> and 547 BBB<sup>-</sup> compounds) [110]. For small molecules, several binary classifiers and logistic regression were proposed (18 models) to predict BBB ability of marine-derived kinase inhibitors, the accuracies of RF, gradient boosting, and logistic regression was the top performed model. Briefly, the extensive applicable scope of AI approaches can reduce the workload of many clinical trials on drug distribution research [111].

Several studies of AI in the field of metabolism predictions have been carried out to investigate the location sites of metabolism, the isoforms that were responsible for the procedure, and metabolic pharmacokinetics, pharmacodynamics (drug-drug interactions) [95, 112]. DeepLoc (DNN-based model) has been presented to predict protein subcellular localization from the dataset extracted from the UniProt database, the accuracy obtained was highest (compared to other methods: LocTree2, MultiLoc2, MultiLoc2, YLoc, CELLO, iLoc-Euk, WoLF PSORT) in the independent test set [113]. Application of the Laplacian-modified naïve Bayesian method to categorise the inhibition potency of 4500 compounds on five CYP isoforms [114]. Furthermore, a multilabel kNN, twin SVM, and five network-based label space division (NLSD)-based methods to study CYP450—substrate selectivity on the dataset of 484 compounds and 1299 compound/isoform pairs; NLSD-XGB achieved the best performance in both CV and HO methods [115]. A study presented a Super CYPsPred web server based on the RF algorithm focussed on five CYPs isoenzymes, containing 17 143 substances to investigate the CYP inhibition ability and CYP inhibitors interaction [116]. A new sequence method has been proposed based on distance-weighted k-NN to identify G-protein coupled receptor (GPCR) drug interaction [117]. A study predicted drug exposure (AUC,  $C_{max}$ , and  $T_{max}$ ) by the BIOiSIM platform integrated coarse-tuning and fine-tuning algorithms [118].

The development of AI-based excretion predictors to investigate the clearance pathways have grown recently. Recent development of CPathPred—the SVM-based predictor on the dataset containing 141 approved drugs for major clearance pathways and increased easily molecular descriptors [119]. It has been accessed that total plasma clearance ( $Cl_{tot}$ ) of 1114 compounds by StarDrop used eight different techniques (PLS, radial basis function fitting—RBF, RF, Gaussian process models—GP with the two-dimensional search for parameters (GP2DS), fixed hyperparameters

(GPFixed) and hyperparameters were obtained by forwarding variable selection (GPFVS) and rescaled procedure (GPRFVS), and by conjugate gradient optimization (GPOPT)); these *in silico* models showed better predictivity compared to *in-vitro* assay [120]. Therefore, the utility of ML approaches improves the screening paradigm in the early phase of drug R&D with multiple benefits.

To predict drug toxicity, advanced AI algorithms have been applied to construct several web tools and packages: BlueDesc, ChemoPY, Mole dB, PaDEL-Descriptor, DRAGON, AdmetSAR 2, Lazar, ProTox II [95, 121]. Variety ML methods such as SVM, RF, naive Bayesian, back propagation neural network, k-NN, C4.5 decision tree (C4.5 DT) have been developed for toxicity evaluation, such as mitochondrial toxicity, protein toxicity, reproductive toxicity, haemolytic toxicity [122–124]. For example, A study introduced seven ML methods on the dataset of 284 food and drug administration (FDA) approved drugs, the naïve Bayes classifier was the best predictive performance and stability [125]. A study tested 246 compounds that were implemented in four models (RF, Gradient, Boosting, DL); the combination of ML and structural alerts was a powerful tool to forecast mitochondrial toxicity [122]. In a large dataset (2487 compounds) six ML methods (SVM, ANN, C4.5 DT, RF, kNN, and Naive Bayes) were used and, the SVM classifier performed the most accurate in predicting reproductive toxicity [126]. For the first time in literature, evaluation of haemolytic toxicity of 452 saponins by four ML methods (SVM, k-NN, RF, gradient boosting machine) and developed “e-Hemolytic-Saponin” programme to predict toxicity automatically [124]. Briefly, the breakthrough of ML has the utility of the early design stage and is becoming an attractive, reliable tool; nevertheless, many obstacles are remaining as the chosen model architecture, model overfitting, data source [121].

Overall, ADMET evaluation via ML approaches has provided high reliability and robust advancement. Currently, the AI-based ADMET predictor tools remain not yet replacing the *in vitro*, *in-vivo* measurements. The strategy has grown and reduced time and cost consuming in drug hit-to-lead and lead-optimization processes.

## Role of AI in adverse drug reactions

The 3 aspects of monitoring, detection and prevention of the adverse reactions produced by the newly developed or developing drugs come under the scope of Pharmacovigilance. The unwanted effects that can be seen in the population from a normal tested dose of the drug are the major concern that extends in the post-marketing surveillance of the drug. A large number of databases are generated for the report of the adverse event. To address the issues such as under-reporting

of certain events which are rare, certain statistical and computational tools are used. Ensuring drug safety is divided into two phases. The first phase involves the evaluation of drugs for their adverse events and toxicities before launching it in the market when the drug is in developmental stages, and the other phase involves reporting adverse events in the population when the drug is marketed.

The databases generated to report and analyse the adverse drug reactions are huge, and developing countries usually adopt systematic collection methods associated with the marketing authorization holder (MAH) industries and Adverse drug reaction monitoring centres. The extended processes incorporate the detection of adverse drug events and their severity, generation of a technical database, reporting of drug-drug interactions, and generating the safety reports. These protocols are extensively time-consuming and need human intervention, which creates more chances of errors. Data collected worldwide is so vast and requires new technological progress to keep track of such data [127].

AI plays a crucial role in Pharmacovigilance. First, it uniquely identifies the incoming data and kind of adverse drug events and also helps in reducing time and burden for processing the data. It also perks up the quality of information and also evaluates the case studies without any human interference. However, with this system economic aspect remains questionable [128]. AI deploys novel methods such as ML and DL techniques from the data generated before and after the marketing of the drug candidate.

The electronic reporting system collects mixed elements from various healthcare facilities. To generate meaningful data with utmost accuracy, DL has an insightful effect as it incorporates features such as image and speech recognition and processing of language in the natural form. Various reports also suggest that neural networks and their massive webwork have improved the analytical application of DL. The current models of DL significantly mould the raw data and recognize the clinical outcomes with accuracy [129]. ML, on the other hand, is an algorithmic technique that defines boundaries between variables and generates a model based on the given data to make accurate predictions [130].

Individual case safety report (ICSR) is a body that operates in accordance to the FDA regulations and provides information for adverse events, product defects, and consumer complaints. A quite number of ML techniques are adopted to increase work efficiency and reduce the labour. First, all the raw facts and figures are inserted in a structured or unstructured manner. Then natural learning processes and ML processes are used to dig out ICSR content, which is usually not refined. At this point, AI plays a significant role in listing out the events, classifying the drugs on the required basis, and carrying out the necessary correlations [131].

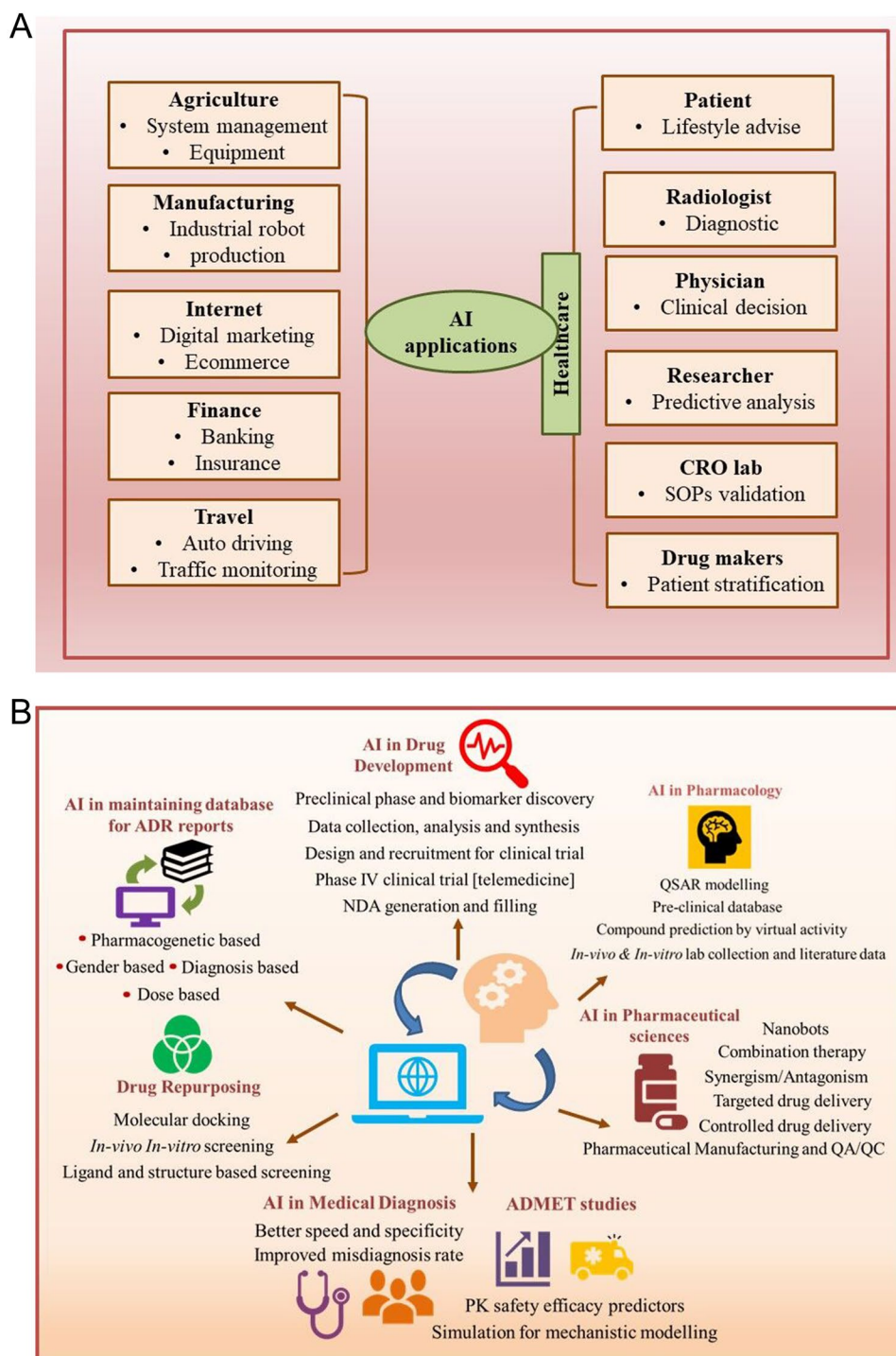
There are specific tools of AI which are used to assist these functions. For instance, to carry out the analysis of



the recorded structured data, VigiBase is used. About 20 million adverse drug report databases are generated using this tool. To access the VigiBase, another tool named VigiAccess is used [132]. VigiFlow is another web-based platform that monitors the online sharing and collection of data to execute a functional analysis. And to obtain the clinically relevant information score of an individual case

report, VigiGrade is employed. VigiRank is another interface for the detection of statistical signals [133]. Another important domain is the clinical evaluation, which according to the Bayesian Confidence Propagation Neural Network is carried out by the WHO-UMC [134]. Although AI applications are abundant, the only shortcoming is the economic impact of such systems (Fig. 2A and B).

**Fig. 2 A, B** Applications of AI in different fields. *AI* artificial intelligence; *NDA* new drug application; *QA/QC* quality assurance/quality control; *QSAR* quantitative structure–activity relationship



## Role of AI in drug repurposing

Drug development is a process that takes a tremendous amount of effort, time, and cost. Drug repurposing provides a fundamental opportunity to use the existing candidates for different therapeutic purposes, exploiting the fact that a known candidate can have more than one target site [135]. This has been made easier with the help of various computational approaches such as molecular docking that has developed an extensive database for the evaluation of the drug effects on different targets. One such example is the connectivity map (CMap) used for mRNA and gene expression, GWAS (genome-wide association studies). Excellent platforms of AI technologies are being approached for drug repurposing in the pharmaceutical industry for mixed data sources such as PREDICT [136], Netlap RLS [137], DTINet [138]. Most of the studies till date reported have employed learning algorithms to develop accurate predictions on the basis of creating a significant correlation between drugs, the targets and the disease. The heterogeneous data sources from the small molecules under the investigation, disease-dependent phenotypes and the biological pathways. However, the relative importance of each factor remains unclear [138].

To execute this approach, three distinct kernels are employed for different level of information. Starting from the kernel-based on a structure which gathers information regarding the analogy between the chemical configurations. The other kernel is based on the transcriptional information that gives the information of the gene expression based on the similarities that exists among the drugs and the kernel gives the information regarding the targets like the distance between the targets and the interaction between the proteins of the given targets. This data is further integrated and output predictions are generated and come under as supervised learning. However, in certain cases where the data for high-quality samples is not available certain unsupervised and semi-supervised learning algorithms are employed.

The unsupervised computational approach uses the already known predictors about the drug labels for training purposes. These are based on clustering algorithms [139]. This unsupervised algorithm and topographical pharmacophore descriptor (CATS) are used for clustering algorithms. The prediction accuracy of unsupervised learning is however moderate. The semi-supervised learning paradigm is a substantial model for a small amount of labelling training sets and tremendous amounts of unlabelled data information. One such example is LapRLS that generates algorithms for drug-target interactions knowing that FDA approval was based only on target predictors. However, the feature of simultaneous prediction sets a good score for this method. Other method includes BLM-NII, Net CBP, LPMIHN.

However, still, AI-based drug repurposing is in its initial stages of applicability. To ensure its diverse applicability in the field, the system must first overcome the prediction accuracy attained manually by the experts [54].

## Role of AI in clinical pharmacology

The most integral part of the drug development process is the study of its clinical aspects. Any failure at this step is a great deal of impairment for the economy and time. So, any failure at the level of patient recruitment and inefficient monitoring system may lead to a skyrocketing loss. Therefore, to improve the scores of these trials, technologies based on AI and ML has been emerged with a vision of better and more accurate predictions at the level of designing the studies and up to the execution of trials [140].

IBM Watson is a system that utilises the electronically generated medical record of volunteers to generate a database for Clinical trial matching based on the eligibility criteria of the patients. This system enriches the enrolment criteria without complex protocols for the manual sorting and analysis of profiles for clinical findings. Some DL-based models measure the outcomes of the clinical trials at different phases. Probabilities of the side-effects and points of the pathway activation are used for training purposes to generate models that can accomplish prediction for the clinical trial outcomes [141]. Various ongoing innovations aim to create the virtual framework to mimic the data points regarding the physiological and pathological build-up of the human system that can help design drug regimens, prognosis, diagnosis, and treatment criteria [142].

Clinical pharmacology can benefit in multiple ways from these advancing technologies. For instance, many web-based platforms make a point by generating interfaces for rational use of drugs and medical-based tools to help patients and all the web platform users for drug-related information. The major challenge in clinical pharmacology is the connectivity of different layers of the drug development process, which is operated at any single level but at different levels and with different people, making it difficult to gather structured information. Moreover, very few studies are known to incorporate large databases with multiple numbers of parameters. This makes the scope of AI and ML in clinical pharmacology more disrupt. However, the next-generation technologies seem to create a better navigation path for this field with the involvement of intelligent personal assistants [142] (Table 1).

## Authors' opinion and critical view

Mimicking of human intelligence using machine is known as AI and in current scenario use of AI in healthcare system has gradually increased, including a broad range of

**Table 1** Various methods of artificial Intelligence and their applications

S.No	Methods	Applications
1	MRI, X-ray, ECG	Pharmaceutical sciences
2	NLP	Medical diagnosis
3	LSSVM	Medical diagnosis
4	PSOWNN	Medical diagnosis
5	EPNN	Medical diagnosis
6	CNN	Medical diagnosis
7	GLCM	Medical diagnosis
8	NN/XY-Fusion	Medical diagnosis
9	QSAR and	Drug discovery and development
10	GAN, DNDD	Drug discovery and development
11	BBPs, ACPs	Peptide-based drug discovery
12	X-ray, NMR	Drug discovery and development
13	DNNs, ANNs, RFs, and SVMs, k-NN	ADMET
14	VigiGrade, VigiRank	Adverse effect
15	CMap, GWAS	Drug repurposing
16	LapRLS	Drug repurposing

*MRI* magnetic resonance imaging; *ECG* electrocardiogram; *NLP* natural language processing; *LSSVM* least square support vector machine; *PSOWNN* particle swarm optimized wavelet neural network; *EPNN* enhanced probabilistic neural network; *CNN* convolutional neural networks; *RDR* referable diabetic retinopathy; *GLCM* grey level co-occurrence matrix; *SVM* support vector machines; *RF* random forest; *ANN* artificial neural network; *GAN* generative adversarial network; *QSAR* quantitative structure–activity relationship; *NMR* nuclear magnetic resonance; *DNN* deep neural network; *kNN* K-nearest neighbouring; *BP-NN* backpropagation network; *DNDD* de novo drug design; *ADMET* absorption, distribution, metabolism, excretion, and toxicity; *BBPs* blood–brain barrier penetrating peptides; *ACPs* anticancer peptides; *VS* virtual screen; *GWAS* genome-wide association studies; *CMap* connectivity mat

applicability in various fields of pharmacology. Now AI technologies are used at every stage of the drug development process, which reduces the health risks associated with preclinical and clinical trials while also significantly lowering the cost. It has the potential to improve patient care, helps in diagnosis of various diseases, nurture profitable development, and improve outcomes. It can also be used to discover treatments for various neurodegenerative disorders like Parkinson's and Alzheimer's disease. AI may track patient data more proficiently than traditional methods of care. Thus, offering doctors more time to concentrate on treatments. However, it has some limitations such as high cost and security breaches with regard to data privacy. Another major disadvantage of AI is that it can not compensate for the *in-vivo* study in preclinical drug discovery process. *In-vivo* experiments are still required in drug development process to confirm the safety and efficacy of drugs.

## Summary

AI has significant impact on pharmacology and drug discovery process, providing numerous benefits such as accelerating the process, limiting time-consuming and costly protocols, and making judicious use of available resources. The role of AI in pharmaceutical sciences encompass a wide range of scientific procedures associated with drug discovery and development such as use of analytical techniques such as X-ray, ECG, and histopathological imaging in diagnosis of various disorders. In drug development process AI is used to various stages of preclinical study and data collection. AI is also used for prediction of safety, efficacy and pharmacokinetics of drug molecule. Additionally, in pharmaceutical sciences AI is used for drug targeting study, combination study and manufacturing processes. The ML process mimic the behaviour learning of human and randomising the data to eliminate any anomalies or duplicates. Moreover, AI has grown in popularity as it is used in QSAR and virtual screening for prediction of virtual pharmacological activity of compound. The GAN technique has been also contributed to medicinal chemistry as molecular de novo design. So far, the AI approach has been successfully applied in LBVS for molecular docking and *in-vivo*, *in-vitro* screening of compound. It is the first choice when the target compound's 3D structure is unavailable. It is predicated on the idea that if the structures are similar, the biological effects will be the same. Furthermore, AI and ML are being used to improve drug trial results. Its goal is to make long-term decisions about drug design, chemical synthesis, and biological test.

## Conclusion

This review has demonstrated the various functions of AI in pharmacological research and drug development. It will influence pharmacological researchers to promote newer innovative outcomes for the improvement of health care services. AI can reduce the risk of drug failure in clinical study by predicting the target and potency. It also reduces the economic burden by reducing the time and expenses of drug discovery method.

**Acknowledgements** The authors would like to thank the Department of Science and Technology for providing INSPIRE fellowship to Divya Soni [IF200156].

**Author contributions** Conceptualization Conceived, designed the experiments and final approval: PK; Analysed the data and drafting manuscript: MK, TPNN. Wrote the manuscript: MK, TPNN, JK, DS. Editing of the Manuscript: PK, TGS, RS.

**Funding** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Data availability statement** All data generated or analysed during study are included in this published article.

## Declarations

**Conflict of interest** The authors declare that they have no competing interests. All authors read and approved the final manuscript.

## References

- Chaturvedula A, Calad-Thomson S, Liu C, Sale M, Gattu N, Goyal N. Artificial intelligence and pharmacometrics: time to embrace, capitalize, and advance? *CPT Pharmacomet Syst Pharmacol*. 2019;8:440. <https://doi.org/10.1002/psp4.12418>.
- Murali N, Sivakumaran N. Artificial intelligence in healthcare—a review. *J Mod Comput Inf Commun Technol*. 2018;1:103–10. <https://doi.org/10.13140/RG.2.2.27265.92003>.
- Jabeen A, Ranganathan S. Applications of machine learning in GPCR bioactive ligand discovery. *Curr Opin Struct Biol*. 2019;55:66–76. <https://doi.org/10.1016/j.sbi.2019.03.022D>.
- Durairaj M, Ranjani V. Data mining applications in healthcare sector: a study. *Int J Sci Technol Res*. 2013;2:29–35.
- Palanisamy V, Thirunavukarasu R. Implications of big data analytics in developing healthcare frameworks—a review. *J King Saud Univ*. 2019;31:415–25. <https://doi.org/10.1016/j.jksuci.2017.12.007>.
- Seyhan AA, Carini C. Are innovation and new technologies in precision medicine paving a new era in patients centric care? *J Transl Med*. 2019;17:1–28. <https://doi.org/10.1186/s12967-019-1864-9>.
- Bertucci F, Anne-Gaëlle LC-S, Monneur A, Fluzin S, Viens P, Maraninchi D, et al. E-health and Cancer outside the hospital walls. Big Data and artificial intelligence. *Bull Cancer*. 2019;107:102–12. <https://doi.org/10.1016/j.bulcan.2019.07.006>.
- Hemingway H, Asselbergs FW, Danesh J, Dobson R, Maniadakis N, Maggioni A, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J*. 2018;39:1481–95. <https://doi.org/10.1093/eurheartj/ehx487>.
- Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, et al. Deep learning intervention for health care challenges: some biomedical domain considerations. *JMIR mHealth uHealth*. 2019;7:e11966. <https://doi.org/10.2196/11966>.
- Jain R, Sontisirikit S, Iamsirithaworn S, Prendinger H. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. *BMC Infect Dis*. 2019;19:1–16. <https://doi.org/10.1186/s12879-019-3874-x>.
- Stephenson N, Shane E, Chase J, Rowland J, Ries D, Justice N, et al. Survey of machine learning techniques in drug discovery. *Curr Drug Metab*. 2019;20:185–93. <https://doi.org/10.2174/1389200219666180820112457>.
- Park YW, Oh J, You SC, Han K, Ahn SS, Choi YS, et al. Radiomics and machine learning may accurately predict the grade and histological subtype in meningiomas using conventional and diffusion tensor imaging. *Eur Radiol*. 2019;29:4068–76. <https://doi.org/10.1007/s00330-018-5830-3>.
- de Oliveira L, Portugal LC, Pereira M, Chase HW, Bertocci M, Stiffler R, et al. Predicting bipolar disorder risk factors in distressed young adults from patterns of brain activation to reward: a machine learning approach. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2019;4:726–33. <https://doi.org/10.1016/j.bpsc.2019.04.005>.
- Alhussein M, Muhammad G, Hossain MS. EEG pathology detection based on deep learning. *IEEE Access*. 2019;7:27781–8. <https://doi.org/10.1109/ACCESS.2019.2901672>.
- Teikari P, Najjar RP, Schmetterer L, Milea D. Embedded deep learning in ophthalmology: making ophthalmic imaging smarter. *Ther adv ophthalmol*. 2019;11:2515841419827172. <https://doi.org/10.1177/2515841419827172>.
- Jetswaart R, Arat S, Chen AX, Farahmand S, Kim B, DuMouchel W, et al. Machine learning guided association of adverse drug reactions with in vitro target-based pharmacology. *EBioMedicine*. 2020;57:102837. <https://doi.org/10.1016/j.ebiom.2020.102837>.
- Mirzaei G, Adeli A, Adeli H. Imaging and machine learning techniques for diagnosis of Alzheimer's disease. *Rev Neurosci*. 2016;27:857–70. <https://doi.org/10.1515/revneuro-2016-0029>.
- Karakaya D, Ulucan O, Turkan M. Electronic nose and its applications: A survey. *Int J Autom Comput*. 2020;17:179–209. <https://doi.org/10.1007/s11633-019-1212-9>.
- Stulp F, Sigaud O. Many regression algorithms, one unified model: a review. *Neural Netw*. 2015;69:60–79. <https://doi.org/10.1016/j.neunet.2015.05.005>.
- Zhang S, Zhang C, Yang Q. Data preparation for data mining. *Appl Artif Intell*. 2003;17:375–81. <https://doi.org/10.1080/713827180>.
- Ten CD. quick tips for machine learning in computational biology. *BioData Min*. 2017;10:1–17. <https://doi.org/10.1186/s13040-017-0155-3>.
- Boulesteix A-L. Ten simple rules for reducing overoptimistic reporting in methodological computational research. *PLOS San Francisco*. 2015;11:e1004191.
- Koteluk O, Wartecki A, Mazurek S, Kołodziejczak I, Mackiewicz A. How do machines learn? artificial intelligence as a new era in medicine. *J Pers Med*. 2021;11:32. <https://doi.org/10.3390/jpm11010032>.
- Altae-Tran H, Ramsundar B, Pappu AS, Pande V. Low data drug discovery with one-shot learning. *ACS Cent Sci*. 2017;3:283–93. <https://doi.org/10.1021/acscentsci.6b00367>.
- Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet*. 2015;16:321–32. <https://doi.org/10.1038/nrg3920>.
- Jyoti B, Sharma AK. AntMiner: bridging the gap between data mining classification rule discovery and bio-inspired algorithms. In: *Proceedings of international conference on IoT inclusive life (ICIIL 2019)*, NITTTR Chandigarh, India. Cham: Springer; 2020.
- Lloyd S, Mohseni M, Rebentrost P. Quantum algorithms for supervised and unsupervised machine learning. *Arxiv Preprint*. 2013. <https://doi.org/10.48550/arXiv.1307.0411>.
- Ayesha S, Hanif MK, Talib R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Inf Fusion*. 2020;59:44–58. <https://doi.org/10.1016/j.inf-fus.2020.01.005>.
- Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J Am Med Inf Assoc*. 2000;7:593–604. <https://doi.org/10.1136/jamia.2000.0070593>.
- Sweilam NH, Tharwat A, Moniem NA. Support vector machine for diagnosis cancer disease: a comparative study. *Egypt Inform J*. 2010;11:81–92. <https://doi.org/10.1016/j.eij.2010.10.005>.
- Orru G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev*. 2012;36:1140–52. <https://doi.org/10.1016/j.neubiorev.2012.01.004>.
- Dheeba J, Singh NA, Selvi ST. Computer-aided detection of breast cancer on mammograms: a swarm intelligence optimized

- wavelet neural network approach. *J Biomed Inf.* 2014;49:45–52. <https://doi.org/10.1016/j.jbi.2014.01.010>.
33. Khedher L, Ramirez J, Górriz JM, Brahim A, Segovia F, Initiative AsDN. Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing.* 2015;151:139–50. <https://doi.org/10.1016/j.neucom.2014.09.072>.
  34. Hirschauer TJ, Adeli H, Buford JA. Computer-aided diagnosis of Parkinson's disease using enhanced probabilistic neural network. *J Med Syst.* 2015;39:1–12. <https://doi.org/10.1007/s10916-015-0353-9>.
  35. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316:2402–10. <https://doi.org/10.1001/jama.2016.17216>.
  36. Griffis JC, Allendorfer JB, Szaflarski JP. Voxel-based Gaussian naïve Bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans. *J Neurosci Methods.* 2016;257:97–108. <https://doi.org/10.1016/j.jneumeth.2015.09.019>.
  37. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal.* 2017;36:61–78. <https://doi.org/10.1016/j.media.2016.10.004>.
  38. Long E, Lin H, Liu Z, Wu X, Wang L, Jiang J, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nat Biomed Eng.* 2017;1:1–8. <https://doi.org/10.1038/s41551-016-0024>.
  39. Miller TP, Li Y, Getz KD, Dudley J, Burrows E, Pennington J, et al. Using electronic medical record data to report laboratory adverse events. *Br J Haematol.* 2017;177:283–6. <https://doi.org/10.1111/bjh.14538>.
  40. Castro VM, Dligach D, Finan S, Yu S, Can A, Abd-El-Barr M, et al. Large-scale identification of patients with cerebral aneurysms using natural language processing. *Neurology.* 2017;88:164–8. <https://doi.org/10.1212/wnl.0000000000003490>.
  41. Zhong Q-Y, Karlson EW, Gelaye B, Finan S, Avillach P, Smoller JW, et al. Screening pregnant women for suicidal behavior in electronic medical records: diagnostic codes vs. clinical notes processed by natural language processing. *BMC Med Inf Decis Mak.* 2018;18:1–11. <https://doi.org/10.1186/s12911-018-0617-7>.
  42. Afzal N, Mallipeddi VP, Sohn S, Liu H, Chaudhry R, Scott CG, et al. Natural language processing of clinical notes for identification of critical limb ischemia. *Int J Med Inf.* 2018;111:83–9. <https://doi.org/10.1016/j.ijmedinf.2017.12.024>.
  43. Kim C, Zhu V, Obeid J, Lenert L. Natural language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke. *PLoS ONE.* 2019;14:e0212778. <https://doi.org/10.1371/journal.pone.0212778>.
  44. Bacchi S, Zerner T, Oakden-Rayner L, Kleinig T, Patel S, Jannes J. Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: a pilot study. *Acad Radiol.* 2020;27:e19–23. <https://doi.org/10.1016/j.acra.2019.03.015>.
  45. Craninx M, Fievez V, Vlaeminck B, De Baets B. Artificial neural network models of the rumen fermentation pattern in dairy cattle. *Comput Electron Agric.* 2008;60:226–38. <https://doi.org/10.1016/j.compag.2007.08.005>.
  46. Santoni MM, Sensuse DI, Arymurthy AM, Fanany MI. Cattle race classification using gray level co-occurrence matrix convolutional neural networks. *Proced Comput Sci.* 2015;59:493–502. <https://doi.org/10.1016/j.procs.2015.07.525>.
  47. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: a survey. *Comput Electron Agric.* 2018;147:70–90. <https://doi.org/10.1016/j.compag.2018.02.016>.
  48. Chung C-L, Huang K-J, Chen S-Y, Lai M-H, Chen Y-C, Kuo Y-F. Detecting Bakanae disease in rice seedlings by machine vision. *Comput Electron Agric.* 2016;121:404–11. <https://doi.org/10.1016/j.compag.2016.01.008>.
  49. Pantazi XE, Moshou D, Oberti R, West J, Mouazen AM, Bochtis D. Detection of biotic and abiotic stresses in crops by using hierarchical self organizing classifiers. *Precis Agric.* 2017;18:383–93. <https://doi.org/10.1007/s11119-017-9507-8>.
  50. Ebrahimi M, Khoshtaghaza MH, Minaei S, Jamshidi B. Vision-based pest detection based on SVM classification method. *Comput Electron Agric.* 2017;137:52–8. <https://doi.org/10.1016/j.compag.2017.03.016>.
  51. Ferentinos KP. Deep learning models for plant disease detection and diagnosis. *Comput Electron Agric.* 2018;145:311–8. <https://doi.org/10.1016/j.compag.2018.01.009>.
  52. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ.* 2016;47:20–33. <https://doi.org/10.1016/j.jhealeco.2016.01.012>.
  53. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. *Drug Discov Today.* 2021;26:80. <https://doi.org/10.1016/j.drudis.2020.10.010>.
  54. Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of artificial intelligence for computer-assisted drug discovery. *Chem Rev.* 2019;119:10520–94. <https://doi.org/10.1021/acs.chemrev.8b00728>.
  55. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers.* 2021;25:1315–60. <https://doi.org/10.1007/s11030-021-10217-3>.
  56. Dobchev D, Karelson M. Have artificial neural networks met expectations in drug discovery as implemented in QSAR framework? *Expert Opin Drug Deliv.* 2016;11:627–39. <https://doi.org/10.1080/17460441.2016.1186876>.
  57. Korotcov A, Tkachenko V, Russo DP, Ekins S. Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. *Mol Pharm.* 2017;14:4462–75. <https://doi.org/10.1021/acs.molpharmaceut.7b00578>.
  58. Lin E, Lin C-H, Lane H-Y. Relevant applications of generative adversarial networks in drug design and discovery: molecular de novo design, dimensionality reduction, and de novo peptide and protein design. *Molecules.* 2020;25:3250. <https://doi.org/10.3390/molecules25143250>.
  59. Wu Q, Ke H, Li D, Wang Q, Fang J, Zhou J. Recent progress in machine learning-based prediction of peptide activity for drug discovery. *Curr Top Med Chem.* 2019;19:4–16. <https://doi.org/10.2174/1568026619666190122151634>.
  60. Basith S, Manavalan B, Hwan Shin T, Lee G. Machine intelligence in peptide therapeutics: a next-generation tool for rapid disease screening. *Med Res Rev.* 2020;40:1276–314. <https://doi.org/10.1002/med.21658>.
  61. Ge R, Dong C, Wang J, Wei Y. Machine learning for peptide structure, function, and design: frontiers in Genetics. *Frontier Media SA;* 2022.
  62. Hwang JS, Kim SG, Shin TH, Jang YE, Kwon DH, Lee G. Development of anticancer peptides using artificial intelligence and combinational therapy for cancer therapeutics. *Pharmaceutics.* 2022;14:997. <https://doi.org/10.3390/pharmaceutics14050997>.
  63. Kabra R, Singh S. Evolutionary artificial intelligence based peptide discoveries for effective COVID-19 therapeutics. *Biochim Biophys Acta Mol Basis Dis.* 2021;1867:165978. <https://doi.org/10.1016/j.bbadis.2020.165978>.
  64. Rein D, Ternes P, Demin R, Gierke J, Helgason T, Schön C. Artificial intelligence identified peptides modulate

- inflammation in healthy adults. *Food Funct.* 2019;10:6030–41. <https://doi.org/10.1039/C9FO01398A>.
65. Kim H, Kim E, Lee I, Bae B, Park M, Nam H. Artificial intelligence in drug discovery: a comprehensive review of data-driven and machine learning approaches. *Biotechnol Bioprocess Eng.* 2020;25:895–930. <https://doi.org/10.1007/s12257-020-0049-y>.
66. Rifaioglu AS, Atas H, Martin MJ, Cetin-Atalay R, Atalay V, Doğan T. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. *Brief Bioinform.* 2019;20:1878–912. <https://doi.org/10.1093/bib/bby061>.
67. Morris GM, Lim-Wilby M. Molecular docking. In: *Molecular modeling*. Humana Press; 2008. p. 365–82.
68. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31:455–61. <https://doi.org/10.1002/jcc.21334>.
69. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem.* 2004;47:1739–49. <https://doi.org/10.1021/jm0306430>.
70. Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des.* 2001;15:411–28. <https://doi.org/10.1023/a:1011115820450>.
71. Pinzi L, Rastelli G. Molecular docking: shifting paradigms in drug discovery. *Int J Mol Sci.* 2019;20:4331. <https://doi.org/10.3390/ijms20184331>.
72. Kadioglu O, Efferth T. A machine learning-based prediction platform for P-glycoprotein modulators and its validation by molecular docking. *Cells.* 2019;8:1286. <https://doi.org/10.3390/cells8101286>.
73. Chandak T, Mayginnes JP, Mayes H, Wong CF. Using machine learning to improve ensemble docking for drug discovery. *Proteins Struct Funct Genet.* 2020;88:1263–70. <https://doi.org/10.1002/prot.25899>.
74. Maia EHB, Assis LC, De Oliveira TA, Da Silva AM, Taranto AG. Structure-based virtual screening: from classical to artificial intelligence. *Front Chem Front Chem.* 2020;8:343. <https://doi.org/10.3389/fchem.2020.00343>.
75. Yasuo N, Sekijima M. Improved method of structure-based virtual screening via interaction-energy-based learning. *J Chem Inf Model.* 2019;59:1050–61. <https://doi.org/10.1021/acs.jcim.8b00673>.
76. Ballester PJ, Mitchell JB. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics.* 2010;26:1169–75. <https://doi.org/10.1093/bioinformatics/btq112>.
77. Fan N, Bauer CA, Stork C, de Bruyn KC, Kirchmair J. ALAD-DIN: Docking approach augmented by machine learning for protein structure selection yields superior virtual screening performance. *Mol Inform.* 2020;39:1900103. <https://doi.org/10.1002/minf.201900103>.
78. Jiménez-Luna J, Cuzzolin A, Bolcato G, Sturlese M, Moro S. A deep-learning approach toward rational molecular docking protocol selection. *Molecules.* 2020;25:2487. <https://doi.org/10.3390/molecules25112487>.
79. Pereira JC, Caffarena ER, Dos Santos CN. Boosting docking-based virtual screening with deep learning. *J Chem Inf Model.* 2016;56:2495–506. <https://doi.org/10.1021/acs.jcim.6b00355>.
80. Batool M, Ahmad B, Choi S. A structure-based drug discovery paradigm. *Int J Mol Sci.* 2019;20:2783. <https://doi.org/10.3390/ijms20112783>.
81. Dai W, Guo D. A ligand-based virtual screening method using direct quantification of generalization ability. *Molecules.* 2019;24:2414. <https://doi.org/10.3390/molecules24132414>.
82. Lima AN, Philot EA, Trossini GHG, Scott LPB, Maltarollo VG, Honorio KM. Use of machine learning approaches for novel drug discovery. *Expert Opin Drug Deliv.* 2016;11:225–39. <https://doi.org/10.1517/17460441.2016.1146250>.
83. Abdolmaleki A, Ghasemi JB. Inhibition activity prediction for a dataset of candidates' drug by combining fuzzy logic with MLR/ANN QSAR models. *Chem Biol Drug Des.* 2019;93:1139–57. <https://doi.org/10.1111/cbdd.13511>.
84. Žuvela P, David J, Wong MW. Interpretation of ANN-based QSAR models for prediction of antioxidant activity of flavonoids. *J Comput Chem.* 2018;39:953–63. <https://doi.org/10.1002/jcc.25168>.
85. Barzegar A, Zamani-Gharehchamani E, Kadkhodaie-Ilkhchi A. ANN QSAR workflow for predicting the inhibition of HIV-1 reverse transcriptase by pyridinone non-nucleoside derivatives. *Future Med Chem.* 2017;9:1175–91. <https://doi.org/10.4155/fmc-2017-0040>.
86. Myint KZ, Xie X-Q. Ligand biological activity predictions using fingerprint-based artificial neural networks (FANN-QSAR). In: *Methods in molecular biology*. Springer; 2015. p. 149–64.
87. Xiao T, Qi X, Chen Y, Jiang Y. Development of ligand-based big data deep neural network models for virtual screening of large compound libraries. *Mol Inform.* 2018;37:1800031. <https://doi.org/10.1002/minf.201800031>.
88. Abbasi-Radmoghaddam Z, Riahi S, Gharaghani S, Mohammadi-Khanaposhtani M. Design of potential anti-tumor PARP-1 inhibitors by QSAR and molecular modeling studies. *Mol Divers.* 2021;25:263–77. <https://doi.org/10.1007/s11030-020-10063-9>.
89. Yuan B, Wang P, Sang L, Gong J, Pan Y, Hu Y. QNAR modeling of cytotoxicity of mixing nano-TiO<sub>2</sub> and heavy metals. *Ecotoxicol Environ Saf.* 2021;208:111634. <https://doi.org/10.1016/j.ecoenv.2020.111634>.
90. Geldenhuys WJ, Bloomquist JR. Development of an a priori computational approach for brain uptake of compounds in an insect model system. *Bioorganic Med Chem Lett.* 2021;40:127930. <https://doi.org/10.1016/j.bmcl.2021.127930>.
91. Kureczyk A, Warszycki D, Musiol R, Kafel R, Bojarski AJ, Polanski J. Ligand-based virtual screening in a search for novel anti-HIV-1 chemotypes. *J Chem Inf Model.* 2015;55:2168–77. <https://doi.org/10.1021/acs.jcim.5b00295>.
92. Li S, Ding Y, Chen M, Chen Y, Kirchmair J, Zhu Z, et al. HDAC3i-finder: a machine learning-based computational tool to screen for HDAC3 inhibitors. *Mol Inform.* 2021;40:2000105. <https://doi.org/10.1002/minf.202000105>.
93. Merk D, Friedrich L, Grisoni F, Schneider G. De novo design of bioactive small molecules by artificial intelligence. *Mol Inform.* 2018;37:1700153. <https://doi.org/10.1002/minf.201700153>.
94. Domenico A, Nicola G, Daniela T, Fulvio C, Nicola A, Orazio N. De novo drug design of targeted chemical libraries based on artificial intelligence and pair-based multiobjective optimization. *J Chem Inf Model.* 2020;60:4582–93. <https://doi.org/10.1021/acs.jcim.0c00517>.
95. Mouchlis VD, Afantitis A, Serra A, Fratello M, Papadiamantis AG, Aidinis V, et al. Advances in de novo drug design: from conventional to machine learning methods. *Int J Mol Sci.* 2021;22:1676. <https://doi.org/10.3390/ijms22041676>.
96. Liu X, Ye K, van Vlijmen HW, van Ijzerman AP, Westen GJ. An exploration strategy improves the diversity of de novo ligands using deep reinforcement learning: a case for the adenosine A2A receptor. *J Cheminform.* 2019;11:1–16. <https://doi.org/10.1186/s13321-019-0355-6>.
97. Grisoni F, Moret M, Lingwood R, Schneider G. Bidirectional molecule generation with recurrent neural networks. *J Chem*

- Inf Model. 2020;60:1175–83. <https://doi.org/10.1021/acs.jcim.9b00943>.
98. Sunseri J, King JE, Francoeur PG, Koes DR. Convolutional neural network scoring and minimization in the D3R 2017 community challenge. *J Comput Aided Mol Des*. 2019;33:19–34. <https://doi.org/10.1007/s10822-018-0133-y>.
99. Xiong J, Xiong Z, Chen K, Jiang H, Zheng M. Graph neural networks for automated de novo drug design. *Drug Discov Today*. 2021;26:1382–93. <https://doi.org/10.1016/j.drudis.2021.02.011>.
100. Maziarka Ł, Pocha A, Kaczmarczyk J, Rataj K, Danel T, Warchoń M. Mol-CycleGAN: a generative model for molecular optimization. *J Cheminform*. 2020;12:1–18. <https://doi.org/10.1186/s13321-019-0404-1>.
101. Kumar A, Kini SG, Rathi E. A recent appraisal of artificial intelligence and in silico ADMET prediction in the early stages of drug discovery. *Mini Rev Med Chem*. 2021;21:2788–800. <https://doi.org/10.2174/1389557521666210401091147>.
102. Ferreira LL, Andricopulo AD. ADMET modeling approaches in drug discovery. *Drug Discov Today*. 2019;24:1157–65. <https://doi.org/10.1016/j.drudis.2019.03.015>.
103. Ta GH, Jhang C-S, Weng C-F, Leong MK. Development of a hierarchical support vector regression-based in silico model for CACO-2 permeability. *Pharmaceutics*. 2021;13:174. <https://doi.org/10.3390/pharmaceutics13020174>.
104. Shin M, Jang D, Nam H, Lee KH, Lee D. Predicting the absorption potential of chemical compounds through a deep learning approach. *IEEE/ACM Trans Comput Biol Bioinform*. 2016;15:432–40. <https://doi.org/10.1109/TCBB.2016.2535233>.
105. Wang N-N, Dong J, Deng Y-H, Zhu M-F, Wen M, Yao Z-J, et al. 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;56:763–73. <https://doi.org/10.1021/acs.jcim.5b00642>.
106. Sun L, Yang H, Li J, Wang T, Li W, Liu G, et al. In silico prediction of compounds binding to human plasma proteins by QSAR models. *ChemMedChem*. 2018;13:572–81. <https://doi.org/10.1002/cmdc.201700582>.
107. Tajimi T, Wakui N, Yanagisawa K, Yoshikawa Y, Ohue M, Akiyama Y. Computational prediction of plasma protein binding of cyclic peptides from small molecule experimental data using sparse modeling techniques. *BMC Bioinform*. 2018;19:157–70. <https://doi.org/10.1186/s12859-018-2529-z>.
108. Kumar R, Sharma A, Siddiqui MH, Tiwari RK. Prediction of drug-plasma protein binding using artificial intelligence based algorithms. *Comb Chem High Throughput Screen*. 2018;21:57–64. <https://doi.org/10.2174/1386207321666171218121557>.
109. Shaker B, Yu M-S, Song JS, Ahn S, Ryu JY, Oh K-S, et al. LightBBB: computational prediction model of blood–brain-barrier penetration based on light GBM. *Bioinformatics*. 2021;37:1135–9. <https://doi.org/10.1093/bioinformatics/btaa918>.
110. Alsenan S, Al-Turaiqi I, Hafez A. A recurrent neural network model to predict blood–brain barrier permeability. *Comput Biol Chem*. 2020;89:107377. <https://doi.org/10.1016/j.compbiolchem.2020.107377>.
111. Plisson F, Piggott AM. Predicting blood–brain barrier permeability of marine-derived kinase inhibitors using ensemble classifiers reveals potential hits for neurodegenerative disorders. *Mar Drugs*. 2019;17:81. <https://doi.org/10.3390/md17020081>.
112. Xiong Y, Qiao Y, Kihara D, Zhang H-Y, Zhu X, Wei D-Q. Survey of machine learning techniques for prediction of the isoform specificity of cytochrome P450 substrates. *Curr Drug Metab*. 2019;20:229–35. <https://doi.org/10.2174/1389200219666181019094526>.
113. Wang D, Zhang Z, Jiang Y, Mao Z, Wang D, Lin H, et al. DM3Loc: multi-label mRNA subcellular localization prediction and analysis based on multi-head self-attention mechanism. *Nucleic Acids Res*. 2021;49:e46. <https://doi.org/10.1093/nar/gkab016>.
114. Lee J, Basith S, Cui M, Kim B, Choi S. In silico prediction of multiple-category classification model for cytochrome P450 inhibitors and non-inhibitors using machine-learning method. *SAR QSAR Environ Res*. 2017;28:863–74. <https://doi.org/10.1080/1062936X.2017.1399925>.
115. Shan X, Wang X, Li C-D, Chu Y, Zhang Y, Xiong Y, et al. Prediction of CYP450 enzyme–substrate selectivity based on the network-based label space division method. *J Chem Inf Model*. 2019;59:4577–86. <https://doi.org/10.1021/acs.jcim.9b00749>.
116. Banerjee P, Dunkel M, Kemmler E, Preissner R. SuperCYPsPred—a web server for the prediction of cytochrome activity. *Nucleic Acids Res*. 2020;48:W580–5. <https://doi.org/10.1093/nar/gkaa166>.
117. Wu W, Song L, Yang Y, Wang J, Liu H, Zhang L. Exploring the dynamics and interplay of human papillomavirus and cervical tumorigenesis by integrating biological data into a mathematical model. *BMC Bioinform*. 2020;21:1–8. <https://doi.org/10.1186/s12859-020-3488-8>.
118. Maharao N, Antontsev V, Hou H, Walsh J, Varshney J. Scalable in silico simulation of transdermal drug permeability: application of BIOISIM platform. *Drug Des Devel Ther*. 2020;14:2307. <https://doi.org/10.2147/DDDT.S253064>.
119. Toshimoto K, Wakayama N, Kusama M, Maeda K, Sugiyama Y, Akiyama Y. In silico prediction of major drug clearance pathways by support vector machines with feature-selected descriptors. *Drug Metab Dispos*. 2014;42:1811–9. <https://doi.org/10.1124/dmd.114.057893>.
120. Kosugi Y, Hosea N. Direct comparison of total clearance prediction: computational machine learning model versus bottom-up approach using in vitro assay. *Mol Pharm*. 2020;17:2299–309. <https://doi.org/10.1021/acs.molpharmaceut.9b01294>.
121. Vo AH, Van Vleet TR, Gupta RR, Liguori MJ, Rao MS. An overview of machine learning and big data for drug toxicity evaluation. *Chem Res Toxicol*. 2019;33:20–37. <https://doi.org/10.1021/acs.chemrestox.9b00227>.
122. Hemmerich J, Troger F, Füzi BF, Ecker G. Using machine learning methods and structural alerts for prediction of mitochondrial toxicity. *Mol Inform*. 2020;39:205. <https://doi.org/10.1002/minf.202000005>.
123. Vishnoi S, Matre H, Garg P, Pandey SK. Artificial intelligence and machine learning for protein toxicity prediction using proteomics data. *Chem Biol Drug Des*. 2020;96:902–20. <https://doi.org/10.1002/jat.3772>.
124. Zheng S, Wang Y, Liu H, Chang W, Xu Y, Lin F. Prediction of hemolytic toxicity for saponins by machine-learning methods. *Chem Res Toxicol*. 2019;32:1014–26. <https://doi.org/10.1021/acs.chemrestox.8b00347>.
125. Zhang H, Mao J, Qi H-Z, Ding L. In silico prediction of drug-induced developmental toxicity by using machine learning approaches. *Mol Divers*. 2020;24:1281–90. <https://doi.org/10.1007/s11030-019-09991-y>.
126. Jiang C, Yang H, Di P, Li W, Tang Y, Liu G. In silico prediction of chemical reproductive toxicity using machine learning. *J Appl Toxicol*. 2019;39:844–54. <https://doi.org/10.1002/jat.3772>.
127. Basile AO, Yahi A, Tatonetti NP. Artificial intelligence for drug toxicity and safety. *Trends Pharmacol Sci*. 2019;40:624–35. <https://doi.org/10.1016/j.tips.2019.07.005>.
128. van Dessel M, Patti C (2011) Just give me the facts: literalism vs. symbolism in B2B advertising. Advice from the top: the expert guide to B2B marketing, pp 111–9
129. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *J Am Med Inf Assoc*. 2018;25:1419–28. <https://doi.org/10.1093/jamia/ocy068>.

130. Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol*. 2019;19:1–18. <https://doi.org/10.1186/s12874-019-0681-4>.
131. White RW, Wang S, Pant A, Harpaz R, Shukla P, Sun W, et al. Early identification of adverse drug reactions from search log data. *J Biomed Inf*. 2016;59:42–8. <https://doi.org/10.1016/j.jbi.2015.11.005>.
132. Shankar PR. VigiAccess: promoting public access to vigibase. *Indian J Pharmacol*. 2016;48:606. <https://doi.org/10.4103/0253-7613.190766>.
133. Caster O, Sandberg L, Bergvall T, Watson S, Norén GN. vigi-Rank for statistical signal detection in pharmacovigilance: first results from prospective real-world use. *Pharmacoepidemiol Drug Saf*. 2017;26:1006–10. <https://doi.org/10.1002/pds.4247>.
134. Ventola CL. Big data and pharmacovigilance: data mining for adverse drug events and interactions. *Pharmacol Ther*. 2018;43:340.
135. Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL. Global mapping of pharmacological space. *Nat Biotechnol*. 2006;24:805–15. <https://doi.org/10.1038/nbt1228>.
136. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol*. 2011;7:496. <https://doi.org/10.1038/msb.2011.26>.
137. Xia Z, Wu L-Y, Zhou X, Wong ST. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst Biol*. 2010. <https://doi.org/10.1186/1752-0509-4-S2-S6>.
138. Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun*. 2017;8:1–13. <https://doi.org/10.1038/s41467-017-00680-8>.
139. Giuliani S, Silva AC, Borba JV, Ramos PI, Paveley RA, Muratov EN, et al. Computationally-guided drug repurposing enables the discovery of kinase targets and inhibitors as new schistosomocidal agents. *PLoS Comp Biol*. 2018;14:e1006515. <https://doi.org/10.1371/journal.pcbi.1006515>.
140. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci*. 2019;40:577–91. <https://doi.org/10.1016/j.tips.2019.05.005>.
141. Gayvert KM, Madhukar NS, Elemento O. A data-driven approach to predicting successes and failures of clinical trials. *Cell Chem Biol*. 2016;23:1294–301. <https://doi.org/10.1016/j.chembiol.2016.07.023>.
142. Zhavoronkov A, Vanhaelen Q, Oprea TI. Will artificial intelligence for drug discovery impact clinical pharmacology? *Clin Pharm Therap*. 2020;107:780–5. <https://doi.org/10.1002/cpt.1795>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.