Archives of Computational Methods in Engineering (2022) 29:2513–2529
https://doi.org/10.1007/s11831-021-09661-z

REVIEW ARTICLE

Exploring Artificial Intelligence in Drug Discovery: A Comprehensive Review

Rajneet Kaur Bijral[1] · Inderpal Singh[2] · Jatinder Manhas[3] · Vinod Sharma[1]

Received: 16 June 2021 / Accepted: 1 October 2021 / Published online: 12 October 2021

Abstract
Drug discovery and development process is very lengthy, highly expensive and extremely complex in nature. Traditional methods involve expensive techniques and take many years to bring a new drug to the market. With the advent of new tools and technologies in this field, the major challenge is to reduce the time and cost required for the development of a new drug. These complex problems involve extremely high computations and can be addressed with the help of Artificial Intelligence based techniques. In this paper, we have broadly discussed different emerging applications of artificial intelligence in the field of drug discovery and development including identification of gene targets for diseases, repurposing of existing drugs through pathway networks, improvements in structure modelling, virtual screenings and hit identification, ADMET prediction, lead identification, clinical trials etc. using various artificial intelligence methods and their inter comparisons. This review presents the literature survey of different research articles published in reputed journals of international publishers such as Springer, Science Direct, IEEE Xplore, Elsevier etc. This is a systematic review of 143 publications to provide an organized summary. In addition to the in-depth analysis the foreseen challenges and existing limitations associated with drug discovery and development process are also pointed out in bold and humble suggestions have been made for necessary improvements. Readers, who are new to the field, will find it useful for enhancing their view about the field.

1 Introduction

Human genome has approximately 568 protein kinases and 156 protein phosphatases that play an important role in indispensable biological processes such as differentiation, proliferation and apoptosis. The activation or deactivation of these protein kinases is achieved in different ways; such as binding with activator or inhibitor proteins; to kinase itself, through autophosphorylation or dimerization induced cis-phosphorylation etc. Under physiological conditions, their expression and activation is precisely regulated inside cells. However, just like pulling the strings of an intricately woven net, deregulation of kinase activity changes the spatio-temporal landscape of gene expressions which further leads to various disease conditions including development of tumors. Now, with the advent of science and instrumentation, the mechanistic defects that cause disease are deeply understood which have created an immense scope for the development of new drugs for their remedy. Enormous sums of money is being spent every year for the realization of these remedies, but endeavour goes down the drain when nine out of ten drugs fail in between the phase 1 trials or regulatory approval thus creating a huge gap in demand and delivery. To fill the gap, computational technology could be explored to rescue the problem. In the past few years Artificial Intelligence (AI) has become a pertinent topic in the field of drug discovery with an aim to reduce time, research expenses, and failure rates in clinical trials. The availability of large datasets for life sciences and rapid evolution of machine learning (ML) algorithms have led many AI based companies to focus on drug discovery [1]. AI has a wide range

✉ Rajneet Kaur Bijral
rajbijral@gmail.com

Inderpal Singh
ipsinghbijral@gmail.com

Jatinder Manhas
manhas.jatinder@gmail.com

Vinod Sharma
vnodshrma@gmail.com

[1] Department of Computer Science and IT, University of Jammu, Jammu, J&K, India

[2] Bioinfores, Jammu, J&K, India

[3] Department of Computer Science and IT, Bhaderwah Campus, University of Jammu, Jammu, J&K, India

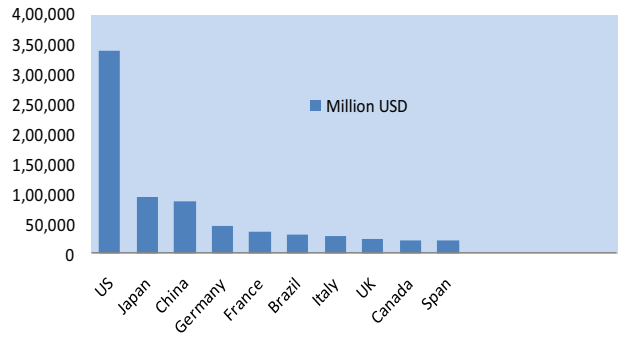

Fig. 1 Countries having biggest pharmaceutical market share in the world

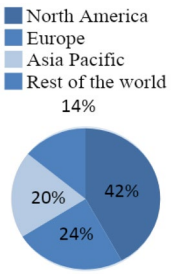

Fig. 2 Percentage wise contributions of four regions for AI based drug discovery

of applications in medical sectors from hospital to clinical research to cut down cost and to improve the outcome of the patients. Many pharmaceutical companies have been using AI in the development of drugs. Figure 1 shows market share of the top ten companies across the world [141]. The global market of drug discovery is segmented into four regions; North America, Europe, Asia-Pacific countries (APAC), and Rest of the World [142]. North America comprising the US, Canada and Mexico is the fastest growing AI based drug discovery market and the US is the major contributor. Figure 2 shows the percentage wise contribution of four regions in AI aided drug discovery. Application of AI in the field of drug discovery along with molecular dynamics simulations could automate and fleet the drug discovery process.

1.1 Motivation of the Study

Protein Kinases play a vital role in the cellular activation processes. Phosphorylation of protein kinase is the critical process that regulates different cellular activities including cell cycle, growth, motility, proliferation, differentiation, apoptosis, etc. With the recent advancement in our understanding regarding the fundamental mechanisms related to the cell signalling have shown that deregulation of the kinases activity leads to oncogenesis. Identification and characterization of new diseases and their causative defects have created a huge scope for development of new drugs for therapeutic intervention. Traditional medicinal pipelines are time consuming, costly and alone cannot fill the demand and delivery gap, thus AI methods have come to rescue this problem. Advent of AI in the field of drug discovery and development has exponentially lowered the time and hence cost required to bring a new drug to the market.

1.2 Contribution of the Study

In this review, we have covered the systematic review of recent research trends in the field of drug discovery using AI, that includes the application of AI in the different phases of drug design and development: i) Identification of target, ii) Drug screening for hit identification, iii) Lead identification, iv) Clinical trial, v) Drug repurposing. Article discusses different AI techniques used to extract the patterns for the identification of drug-targets that are difficult for humans to identify through traditional methods alone. Second, its application to virtually screen the targets against millions of compounds significantly reduces the cost and time required in wet labs by reducing the experimentally screenable compounds. Third, generation and assessment of optimized structures by AI models for lead optimization. Fourth, AI application in the clinical trials and drug repurposing that have shown recommendable results is discussed. Different AI based tools available online for predicting 3D structure of protein and ligand binding site prediction are also covered in the review. Lastly, we have also discussed the critical issues and limitations associated with each stage of the drug discovery process along with the future directions associated with them.

2 Drug Discovery Process

The overall process of drug discovery is time consuming, complex and depends on numerous factors. It starts with identification of the biological target i.e. cause of the disease, then the identification of the first chemical compound that shows activity against the given target, this first compound is called a 'hit'. Hits are found by screening the chemical libraries or isolated naturally from bacteria, plants and fungi [2]. The next step is to isolate the lead compound; it is the compound that shows propitious potential for the development of the drug against the given target. The selected lead is further modified for its enhanced specificity and potency even at lower concentration; this process is known as lead optimization. Then the clinical trial of the drug is done to know its effect. Overall, phases of drug discovery and development are shown in Figure 3.

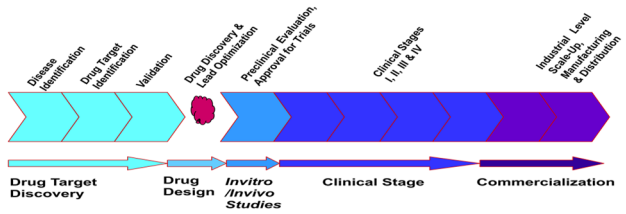

Fig. 3 Overall drug discovery process

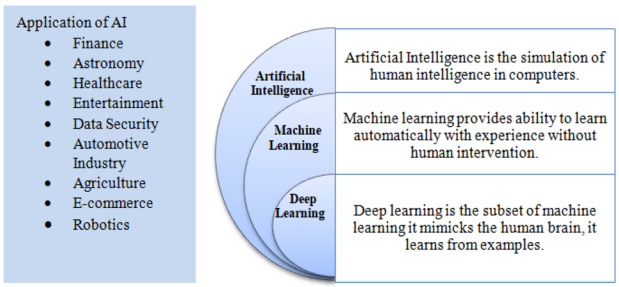

Fig. 4 Application of AI in different sectors

3 AI, machine learning and deep learning

AI enables the system to perform tasks that require human intelligence. In the process, information is acquired, rules are developed for using information, approximate or definite conclusions are drawn and then self-correction is done [3]. The main advantage of an AI approach is that they learn from examples and even develop a model if our understanding of the underlying process is limited [4]. AI has various applications in the field of health care which include diagnosis and treatment of many diseases. ML is the subfield of AI that has the ability to automatically improve with experience. ML makes predictions using computational statistics, which can be classified into unsupervised, supervised and reinforcement learning. In unsupervised learning techniques, hidden patterns in the data are extracted and this information is used to form clusters in meaningful ways. Disease target identification by clustering through feature methods can be done by unsupervised ML. In supervised learning techniques the model is trained on input data that have output associated with it, then the model thus developed is used to predict the output for unseen input data. Classification and regression methods are used to develop the predictive model based on labelled data. Supervised Learning algorithms can be used for disease diagnosis, clinical and medical research [5]. The reinforcement learning system learns by interacting with the environment and by using feedback from its experiences and actions. DL is the subset of ML in which the system learns without human intervention from both unstructured and unlabeled data. AI has its applications in various fields; speech recognition [129], health care [124], gaming [125], automobiles [126], social media [128], agriculture [127] etc. Figure 4 shows the application of AI in different sectors. K. Das *et al.* [120] proposed a model for the treatment of Epilepsy based on Electroencephalogram (ECG) signals. The framework thus proposed consists of feature extraction (current maxima, lower threshold and target point selection), second module consists of pattern matching (segment and domain matching) and in the third module they were able to detect epilepsy seizure from ECG signals. The accuracy and F1 score of the proposed model was reported as 92.66% and 94.86 % respectively. D. D. Chakladar *et al.* [123] proposed a framework for classifying the cognitive state of a user by using the Filter Bank Cognitive State Pattern (FBCSP) method and Long Short-Term Memory (LSTM) based Deep ensemble. For the experimentation purpose the ECG signals were divided into equal size multiple frequency bands and extraction of features was done by the Common Spatial Pattern (CSP) algorithm. Deep ensemble models thus proposed consisted of multiple LSTM networks connected in parallel. Accuracy of 87% was obtained and was able to estimate the cognitive state in a low computing environment. Das. Chaklandar *et al.* [122] presented a hybrid model based on bidirectional long short-term memory (BLSTM) and LSTM for the classification of workload during multitasking mental activities of humans. "STEW" Data set consists of two tasks: "no task" and "simultaneous capacity (SIMKAP) based multitasking activity" was used for experimentation. The presented model was able to attain the classification accuracy of 86.33% and 82.57% for "no task" and "simultaneous capacity (SIMKAP)- based multitasking activity" respectively. S. Mukherjee *et al.* [121] a DL based model for the automatic detection of four classes of diseases in plant leaves. For classification they have adopted GoogleNet to identify disease types and an accuracy of 85.04% has been reported. From the literature it has been reported that AI has a wide range of applications. AI can be employed in the drug discovery and development process to improve the decision making process involving abundant, high-quality data, thus promoting data-driven decision making and reducing the failure rates in drug discovery [4].

4 AI in Drug Discovery Process

4.1 Application of AI in Target Disease Gene Prediction

AI has been appreciated in the past few years and has been successfully applied in various stages of drug discovery. The Human Genome Project was completed in 2013 since then numerous updates of draft assemblies have been made available to the academicians and industry for understanding the

Human Genome and the quest to understand the genomic factors associated with the disease. Thousand Genome Project and HapMap project have also provided wealth of information regarding the alternate loci and alternate genotypes (i.e. Insertions, Deletions and Substitutions etc.). Screening of the clinical cases and comparisons with the healthy controls led to the identifications of risk alleles (i.e. an alternate genotype at a certain locus with high correlation with the disease phenotype), massive genome wide association studies (GWAS) further identified the causative genotypes from the correlated milieu as a function of tradeoff between stringency and sensitivity addressed in the statistical methods. Interestingly, after the development and clinical validation of the Drugs, variable response amongst patients prompted genotyping which led to identification of certain germline variants and somatic mutations. Structural and biophysical investigations on the target protein and drug molecules further elucidated the atomistic explanation for the variable responses which included identification of nonsynonymous variations causing sensitivity to the therapy or resistance perse. ML methods have been used in the identification of various disease targets. Its classifiers can be trained on vast genetic data usually in excess of gigabytes and Gene Ontology for predicting disease gene association [6]. Decision Tree based classifiers trained on the morbid and druggable dataset, metabolic and transcriptional interactions, protein–protein interaction, tissue expression and sub-cellular localization are network attributes to predict the genes that are associated with morbidity and may be druggable. Based on these applications, researchers have concluded that plasma membrane localization and transcription factors are vital factors for druggability and morbidity. In a recent study, Random Forest (RF) based model outperformed other classification methods of ML such as Naive Bayes (NB), linear and radial SVM (SVM) with an accuracy of 80% for prediction of Autism Spectrum Disorder (ASD) genes [7]. In the 2019 [8] manifold learning-based method was proposed by assuming that the distance between genes and its associated disease is shorter as compared to non-associated disease-gene pairs. The model thus developed is capable of identifying new disease-gene associations when studied for Lung Cancer and Bladder Cancer. A Deep Neural Network-based technique has also been developed for the identification of association between infectious disease and host genes by considering sequence and protein interaction as network features [9]. They found that out of 100 highly infectious disease-genes associated with them, 73 were verified experimentally. A model was developed to track the changes that occur in human muscles due to age [10]. Several ML supervised methods were compared, out of which linear-Kernel SVM and Deep feature selection models were the best model suited for the identification of ageing biomarkers. They concluded that ageing biomarkers could be used for anti-aging therapy. E. Ferrero et al. [11] trained different ML based classifiers such as SVM, RF, gradient boosting machine and neural network to explore disease-gene association. It was reported that neural network based classifiers achieved an accuracy of $>71\%$. They used the developed model for the prediction of 1431 novel targets. Joen et al. [12] proposed an SVM based model to classify cancer drug targets and non-cancer drug targets for pancreatic cancers (PaCa), ovarian cancers (OvCa) and breast cancers (BrCa). Thirteen biological and network features were identified, out of which relevant were selected using SVM-recursive feature elimination method the key features thus obtained were mRNA expression, gene essentiality, somatic mutation pattern, DNA copy number and protein–protein interaction. 257 antibody targets were identified, out of which 30 affected all the three types, whereas 88, 28 and 53 were specific for PaCa, BrCa and OvCa, resp. They were also able to identify 345 peptide targets. DGLinker [135] is a web server developed for the prediction of candidate genes for human diseases. It is a user-friendly interface that uses the biomedical information from various biological and phenotypic databases and uses ML based techniques for the prediction of new disease-associated genes. From the past few decades ML based algorithms have been applied in bioinformatics for the prediction of disease-gene association [136, 137].

4.1.1 Critical Issues and Future Directions

Different databases from different sources are available for target identification. The major issue is to manage the heterogeneity among these databases. Data recorded in these databases are collected under different experimental conditions, and the format of recording is not similar. To address this problem integrated or curated databases have been created. The curated databases are DisGeNet [109], Therapeutic target database (TTD) [110], STRING [111], LinkedOmics [112], Open-Target platform [113], DepMap portal [114], HMDD [115] and Comparative Toxicogenomics Databases (CTD) etc.[116]. Curated databases have major limitations, such as lack of validation for the target-disease association. Some of the curated databases have a number of publications as supporting evidence but they lack direct correlation with potency of target modification. Another limitation of these curated databases is that they lack target druggability information. Target druggability softwares have been developed including TractaViewer [117] and Drug Tragetor [118] to find out the molecular Ligand abilities and potential safety risks. Usage of curated databases should be increased. Moreover, curated databases lack programmatic accessibility. There is a need to increase program accessibility to accelerate usage of curated databases.

4.2 Application of AI in Drug Screening

To reduce the R&D expenditure in the process of drug discovery, AI techniques can be explored for the identification of target-specific molecules that involve screening of large compound libraries. Selecting drug candidates for particular targets with desirable properties is an important step in the drug discovery process. Physical properties and chemical properties of the compound must be considered as they substantially affect the bioavailability, toxicity and bioactivity of drug molecules. Virtual-ligand based or structure-based design approach is applied on available data for profile selection which substantially reduces the final size of the compound library for in-vitro validation.

4.2.1 AI in Predicting Physical Properties

Physical properties greatly affect the biological properties of the drug by modifying its solubility, stability, protein binding and absorption. Drug's Physical properties i.e. hydrophobicity, pKa and solubility affects the bioavailability of the drug. The concept of inverse design is introduced that starts with the desired properties of the compound and then the probable molecules are searched. Generative modeling of ML has been employed [13] for the joint probability distribution of both molecular representation and physical properties to retrieve inverse design. Molecular fingerprint, coulomb matrix, potential energy functions, bag of molecular bonds and fragments, density of electrons, atom and bond weighted graph, atomic charge association in 3D and SMILES strings are used by AI based tools for molecular representation.

4.2.2 AI in Predicting Bioactivity

The traditional ML based techniques namely gradient boosting machines (GBMs) [14], deep neural network (DNNs) [15] and RF [16] have been applied to interpolate the transformation done in drug-compounds by retrosynthesis. In recent years matched molecular pair (MMP) analysis, i.e. the impact of bioactivity and molecular properties by introducing a single chemical transformation in a drug compound [17] has been widely used in de novo design. DNN coupled with MMP performs better than GBMs and RF in predicting the bioactivity of the compound [18]. With the availability of large dataset for public domain, ML along with the MMP has been used to predict the bioactivity properties namely absorption, distribution, metabolism, and excretion (ADME) [19], intrinsic clearance [20], oral exposure [21] and mode of action.

4.2.3 AI in Predicting Toxicity

Drug toxicities and its side effects are an important issue in the regulatory clearance of drugs. Traditional in vitro and in vivo tests are performed to scrutinize drug safety. "organ on a chip" an in vitro model [22] has been developed in recent years to reduce the cost, but still this approach is time-consuming and costly. Computational methods have shown considerable dominance in comparison to experimental methods as they are fast, inexpensive, more accurate and can be applied prior to the synthesis of compounds [23]. In recent years various ML based methods include probabilistic neural network [24], SVM [25] and NB [26] have been used to predict chemical carcinogenesis of the compound. DeepTox [27] is a DL based tool for predicting toxicity, the tool first normalizes the chemical representation of the compound, chemical descriptors are computed and they are used as the input to ML methods. The descriptors are classified into two categories: static or dynamic. Static Descriptors include surface areas, atom counts and the absence or presence of a predefined substructure in a compound; different infinite numbers of dynamic features are calculated. The DeepTox Algorithm predicts the toxicity of compounds with good accuracy. S. Jain et al.[138] developed a model using RF, deep neural network, conventional and graph convolutional neural network approach for prediction of toxicity of small molecules using ChemIDplus dataset consisting of > 80,000 compounds having measurements against 59 acute toxicity endpoints. They were able to predict 36 out of 59 points. Some of the tools used for predicting toxicity, chemical synthesis and molecular properties of the compounds are listed in Table 1.

4.2.4 Critical Issues and Future Directions

The major issue is the quality and quantity of data for screening the drugs. Small amount of data is available and data is dispersed across many literatures and is ambiguous. Curated databases (MoleculeNet [133]) can solve this problem. The transfer learning concept can be explored in this area to have effective results. ML based models are based on intrinsic feature and have low interpretability [134] can be solved by building a data driven feature generation model. Another challenge prediction of ligand based property is the activity cliff. Activity Cliff means chemicals having similar structure but exhibiting different properties. To solve this problem, one needs to have the information beyond the structure of the compound which is quite challenging.

4.3 Application of AI in Lead Optimization

Drug-like molecules for specific targets involve extensive virtual screening of compound libraries. Once

Table 1 Tools for prediction toxicity, chemical synthesis and molecular properties

Tools	Description	Websites	References
Chemputer	Tool for reporting a chemical synthesis procedure	https://zenodo.org/record/1481731	[92]
ORGANIC	Molecular generation tool for creation of molecules with desired properties	https://github.com/aspuru-guzik-group/ORGANIC	[93]
DeepNeuralNet- QSAR	Tool for prediction of Molecular activity	https://github.com/Merck/DeepNeuralNet-QSAR	[94]
Hit Dexter	ML based models for the predicting molecules that might respond to biochemical assays	http://hitdexter2.zbh.uni-hamburg.de	[95]
ODDT	Chemoinformatics and molecular modeling toolkit	https://github.com/oddt/oddt	[96]
REINVENT	RNN (recurrent neural network) and RL (reinforcement learning) based Molecular de novo design	https://github.com/MarcusOlivecrona/REINVENT	[97]
SCScore	A model to evaluate scoring function of the complexity of a molecule synthesis	https://github.com/connorcoley/scscore	[107]
NeuralGraph Fingerprints	Novel molecules Property prediction tool	https://github.com/HIPS/neural-fingerprint	[108]
DeepTox	Tool for Toxicity predictions	www.bioinf.jku.at/research/DeepTox	[27]

drug-candidate is identified then it is further refined or modified to make it more target specific and effective that involves two step processes, first step is retrosynthesis that is to recursively transform the drug molecule into smaller fragment and second step is to find out the organic reaction that will transform the fragment into target compound. Finding the suitable organic reaction is quite cumbersome as it requires scanning of a large number of reactions. AI techniques can be explored to pick the most feasible reaction. Previously the Expert based system was used to solve the problem of prediction of reaction and retrosynthesis but these techniques were not widely used by chemists because these algorithms required human intervention, as the dataset thus used do not have molecular context knowledge. A deep neural network based model [28] has been reported to solve reaction conflict that occurs in the early rule-based system; the model was trained on 3.5 million reactions. They attained an accuracy of 95% (for top10) in retrosynthesis and 97% of accuracy for reaction prediction. AI and Monte Carlo tree search [29] have been combined for the synthesis of organic molecules by retrosynthesis. The performance of MCTS, neural Best First Search (BFS) and heuristic Best First Search (BFS) for 497 different molecules has been studied; 92% of the test set was solved by MTCS, whereas neural BFS and Heuristic BFS solved 71% and 4% respectively. Search problem in retrosynthesis is complex and deep reinforcement learning [30] is used for identification of reactions in each step of the retrosynthesis. Neural networks have been trained to estimate the cost of the molecule on the basis of its molecular structure. AI techniques can be trained on available dataset to predict the probability of the selection of the reaction for the transformation of molecules from one stage to another, linking each transmission with its predecessor and also considering the yield and cost of the transformation. Auto in-silico ligand directing evolution (AILDE) [139] has been developed for lead optimisation. In the developed Framework compound library were constructed, molecular dynamics simulation was used for conformational sampling and fragment growing for ligand modifications. However, an assumption was made that there were no changes in the binding mode. AILDE was not able to perform well in case of an activity cliff. AI can be employed for the prediction of optimal and feasible retrosynthesis routes for drug molecules.

4.3.1 Critical Issues and Future Directions

Although the emergence of AI based methods in retrosynthesis looks promising. There are critical issues associated with these methods. The first issue is that a mostly similarity based approach is used for the prediction of the next step in the reaction based on existing reaction knowledge. The result is based on an empirical approach for the automation of retrosynthesis. As the model is restricted to operate and give suggestions on the basis of the data provided to it. Significant amount of uncertainty is suggested by the model extrapolating outside its training data. The second issue associated with AI based retrosynthesis is lack of high quality data. The major issue associated with the data are; kinetic associated with the reaction and order of catalysts and reagents. Need is to accelerate standardized matrix and quality shirt data set with common benchmark to have favourable results from AI based models.

4.4 Application of AI in Clinical Trial

It has been reported that to bring a single drug to market it takes about 1.5–2.0 billion USD [31]. Clinical trials of drugs take about half of the time of the whole drug development process. Failure of it is not only a waste of time but also money spent in preclinical phases of drug development. The success of the clinical trial depends on various factors including recognition of the disease, identification of target

and finding out the effect of the drug molecule in the patient. AI technique's' ability to automatically identify the pattern from large datasets could be explored to reduce the time required in clinical trials. AI platform (AiCure) has been used on mobile devices to measure medication adherence phase II for patients suffering from schizophrenia [32]. The comparison of AI platform and modified directly observed therapy (mDOT) shows that mean cumulative adherence reported for AI platform and mDOT was 89.7% and 71.9% respectively for patients receiving ABT-126.

4.4.1 Critical Issues and Future Directions

The major issue in application of AI in clinical trials is the necessity of a high volume of labelled data sets for the training of models. Another issue is the need for regulating relevant ethical issues (patient privacy, securing data, retaining confidentiality) for using healthcare data for AI. To effectively explore the numerous steps of clinical trials, data scientists and medical scientists need to work together to have promising results. Data should be collected in such a way that it carries information about correlation of trial design features and trial performance.

4.5 Application of AI in Drug Repurposing

Drug repurposing is the process in which reuse of existing drugs is explored and implemented for new medical therapy. The advantage of drug repurposing is that already approved drugs can omit the phase I of the clinical trial and toxicity testing thus reducing the development time and risk in drug development. A deep neural network (DNN) model [33] has been proposed by using transcriptional data to classify therapeutic categories of different drugs. The data set consisted of 433, 454 and 308 drugs for PC3, MCF7 and A549 cell lines respectively. The proposed model was able to classify drugs based on their toxicity and therapeutic use. Study performed by Li et al. [34] suggested a DL based drug repurposing approach based on chemical structure and transcriptome expression data. They were able to report the repurposing of Pimozide used in the treatment of Tourette's Disorder for the treatment of non-small cell lung cancer. Zeng et al. [35] proposed deepDR, a DL based approach for drug repositioning by predicting a new drug-disease association. Proposed model consisted of a heterogeneous network: seven networks of drug-drug and one network for drug-side-effect, drug-disease and drug-target. deepDR first constructed the PPMI (positive point wise mutual information) matrices for each network then the features were extracted by Multimodal Deep Autoencoder and finally features were used by collective Variational Autoencoder (cVAE) for the prediction of drug-disease association. The accuracy thus obtained by the deepDR for predicting association between drug and disease was 82.6%. A deep generative adversarial autoencoder (AAE) and variational autoencoder (VAE) were implemented and compared by Kadurin et al. [36] for the identification of molecular properties that had known anticancer properties. AAE and VAE performance was compared by conducting three experiments, in the first experiment they compared models with reconstruction error, and it was found that AAE is better than VAE with reconstruction error of 9.52 as compared to VAE with reconstruction error of 14.60. In the second experiment, VAE and AAE were compared by their ability to generate molecular vectors. VAE performed better than AAE in terms of coverage. In the third experiment two models were compared in terms of feature extraction where both the models performed well. The association between drug-disease can be used for drug repurposing Zhang et al. [37] represent the association between drug-disease as bipartite networks. They proposed a similarity-based inference method (NTSIM) for prediction of unknown association between drug-disease and similarity-based classification method (NTSIM-C) for classification of therapeutic association. Moghadam et al. [38] presented scored mean kernel fusion (SMKF) method to predict drug candidate by considering six features that are drug chemical structure, drug side effect, drug's receptor phenotype, protein–protein interactions, drug sequence alignment with receptor protein and disease phenotype. The model was developed to know the effect of disease and drug features in predicting drug-disease association.

4.5.1 Critical Issues and Future Directions

Major issue in drug repurposing is intellectual property consideration. Legal issues related to patenting the new medical use for already existing drugs impede the drug repurposing. Electronic Health Recorder (EHR) has been used to overcome the limitation related to data availability for drug repurposing. The need is to advance technology for integration and extraction of heterogeneous large scale data. Options including patent pools, open licensing should be explored for rare and neglected diseases from intellectual property prospective to enhance drug repurposing.

5 AI in Predicting the 3D Structures of Protein and Protein Ligand Binding Site Prediction

5.1 AI in Predicting the 3D Structure of Protein

Proteins are complex macromolecules in our cells which regulate physiology. Knowing the 3D structure from the sequence of the proteins is vital for drug discovery as it helps to determine its function, topology and druggable pockets.

It also helps to find the drug molecule that will bind with it. Prediction of protein structure by experimentation is a very complex, time consuming and tedious task. AI-based techniques are implemented in this area to increase the accuracy and efficiency of structure prediction. AlphaFold, an AI based system was developed [39] by combination of three neural networks for protein structure prediction based on distance prediction between the residue pairs, quantification of the candidate structure was done by Global Distance Test (GDT_TS) and then the protein structure was generated. The system was able to predict the structure with high accuracy because it was based on the distance between residues and angle between peptide bonds and the system was made to learn probability distribution to generate the structure of protein. FoldRec [140] is the model for recognition of protein folding to incorporate the interaction among proteins. In the proposed model recognition is done by combination of cluster-to-cluster model and protein similarity network. Some of the tools for 3D structure prediction of proteins are listed in Table 2.

5.2 AI in Protein–Ligand Binding Site Prediction

Onco or disease marker proteins with aberrant activity require binding with other bio-molecules or ions to form specific interactions to attain specific functions. These bio-molecules or ions are called ligands, specific positions or key amino residues in proteins where the ligand binds are called ligand binding sites (LBSs). The identification of these LBSs helps us to effectively explore the mechanism behind the pathogenesis of diseases, thus helping in the process of drug design and development. With the development of computer technology in recent years, AI algorithms have been used not only in ligand binding site prediction but also for binding affinity prediction. Identification of protein–ligand interactions has an extensive impact in the field of drug discovery as it not only helps to identify the lead hits but also in the process of drug repositioning. With the emergence of large collections of protein–ligand complexes complemented by binding data, as found in PDBbind or BindingMOAD, new opportunities for parameterization and evaluating scoring functions have arisen. With huge data collections available, it becomes feasible to fit scoring functions in a QSAR style, *i.e.*, by defining protein–ligand interaction descriptors and analyzing them with modern ML methods [40]. Some of the ML and DL based LSB prediction and protein–ligand prediction methods are listed in Table 3.

Protein ligand binding sites are a class imbalance and dichotomous problem. Many ML algorithms have been implemented to predict the protein ligand binding sites including; Linear regression, Support Vector Machine, Naïve Bayes classifier, RF algorithm and KNN algorithm some of them are demonstrated in Fig. 5.

Linear regression is simple to implement but its accuracy is poor because of under-fitting. Naïve Bayes classifier is a simple, fast and effective algorithm for classification problems but it requires prior probability and not effective for data that have correlation between samples. Although the KNN algorithm is quick, simple and has less training cost, it performs poorly for class imbalance problems. RF algorithm works on a decision tree that performs poorly for class imbalance. SVM (SVM) has excellent classification accuracy, high generalization ability and exceptional ability to classify high-dimensional small sample data; it has been used recently in the field of LSB prediction and protein–ligand interaction. Some of the published research using SVM is discussed below.

In 2009 Chauhan J S et al. [45] developed the ATPint web server to identify ATP binding residue in the protein. Two SVM based models are developed; the first model is developed using the primary sequence of the proteins and the second model is developed by using PSI-BLAST generated position specific scoring matrix (PSSM). The first model attains the maximum Matthews's Correlation Coefficient (MCC) of 0.33 with accuracy of 66.25% and second model performance is recorded as MCC 0.5, which is better than the first. In 2011, MetaDBSite server based on SVM developed by Jingna Si et al. [51] predicted the protein-DNA binding residues by considering sequence information. MetaDBSite integrates the results of six predictive tools: BindN-rf [82], DNABindR [83], BindN [84], DISIS [85], DBS-PRED [86] and DP-Bind [87]. Input parameters of the SVM model are attained by the result of DNABindR, BindN, DISIS, and BindN-rf, while DBS-PRED and DP-Bind provide auxiliary parameters. The output obtained by MetaDBSite is better than any single prediction model. On

Table 2 Tools for predicting 3D structure of proteins

Tools	Description	Websites	References
RaptorX	Tool for protein function and structure prediction	http://raptorx.uchicago.edu/	[88]
AlphaFold	Tool for the prediction of Protein 3D structure	https://deepmind.com/blog/alphafold	[39]
ESyPred3D	Tool for homology modeling	http://www.fundp.ac.be/urbm/bioinfo/esypred/	[89]
Phyre and Phyre2	Tool for for protein structure prediction	http://www.sbg.bio.ic.ac.uk/phyre2	[90]
HHpred	Tool for template detection, alignment, 3D modeling	http://protevo.eb.tuebingen.mpg.de/hhpred	[91]

Table 3 Machine learning and deep learning based LSB prediction and protein–ligand prediction methods

Authors/year	Method	Machine learning and deep learning technique
W. Deng et al. (2004)	Knowledge-based scoring functions, structure–activity relationship (QSAR) approach [41]	Kernel-Partial Least Squares
T. Guo et al. (2005)	Novel statistical descriptor (the Oriented Shell Model) [42]	support vector machine
K. Ye et al. (2007)	Multi-RELIEF [43]	RELIEF algorithm
C. Sotriffer et al. (2008)	SFCscore [44]	Partial Least Squares Analysis Multiple Linear Regression
J. S. Chauhan et al. (2009)	ATPint [45]	Support Vector Machine
B. Huang et al.(2009)	MetaPocket [46]	Hierarchical Clustering Algorithm
J. A. Capra et al.(2009)	ConCavity [47]	K-Means algorithm
P. J. Ballester et al. (2010)	RF-Score [48]	Random Forest algorithm
J. D. Durrant et al. (2010)	NNScore [49]	Artificial Neural Network
J. D. Durrant et al. (2011)	NNScore 2.0 [50]	Artificial Neural Network
J. Si et al. (2011)	MetaDBSite [51]	Support Vector Machine
K. Chen et al. (2012)	NsitePred [52]	Support Vector Machine
Y. Dou et al. (2012)	L1pred [53]	L1-Logreg Regression classifier
M. Brylinski et al.(2013)	eFindSite [54]	Support Vector Machine
J. Yang et al. (2013)	COACH [55]	Support Vector Machine
B. Panwar et al. (2013)	VitaPred [56]	Support Vector Machine
D. J. Yu et al. (2013)	TargetS [143]	Support Vector Machine
P. Chen et al. (2014)	LigandRFs [57]	Random Forest algorithm
M. Suresh et al. (2015)	Naïve Bayes method [58]	Naïve Bayes classifier
Y. Komiyama et al. (2015)	Utprot [59]	Support Vector Machine Artificial Neural Network The Random Forest algorithm Genetic Algorithm
D. J. Yu et al. (2015)	OSML[60]	Support Vector Machine
R. Krivák et al. (2015)	PRANK [61]	Random Forest algorithm
B. Alipanahi et al. (2015)	DeepBind [62]	Convolutional Neural Networks (CNNs)
P. Chen et al. (2016)	LigandDSES [64]	Random Forest algorithm
J. W. Jian et al. (2016)	ISMBLab-LIG [63]	Artificial Neural Network
M. M. et al. (2016)	SAnDReS [65]	Regression Analysis
S. Zhang et al. (2016)	Multimodal Deep Belief Network [66]	Deep Belief Network (DBN)
J. Jiménez et al. (2017)	DeepSite [67]	Deep Convolutional Neural Networks (DCNNs)
M. Wen et al. (2017)	DeepDTIs [68]	Deep Belief Network
Q. Wu et al. (2018)	COACH-D [69]	Support Vector Machine
R. Krivák et al. (2018)	P2Rank [70]	Random Forest algorithm
H. Öztürk et al. (2018)	DeepDTA [71]	Deep Convolutional Neural Networks (DCNNs)
J. Jiménez et al.(2018)	KDEEP [72]	3D-Convolutional Neural Networks (DCNNs)
I. Lee et al. (2019)	DeepConv-DTI [73]	Deep Convolutional Neural Networks (DCNNs)
L. Zheng et al. (2019)	OnionNet [74]	Deep Convolutional NeuralNetworks (DCNNs)
Z. Zhao et al. (2019)	SXGBsite [75]	Synthetic Minority Over-Sampling Technique (SMOTE) and Extreme Gradient Boosting (XGBoost)
Y. Cui et al. (2019)	DeepCSeqSite [76]	Deep Convolutional Neural Networks (DCNN)
L. Pu et al. (2019)	DeepDrug3D [77]	Deep Convolutional Neural Networks (DCNN)
da Silva. et al. (2020)	Taba [78]	Linear Regression Least Absolute Shrinkage, and Selection Operator (Lasso) Lasso with cross-validation Ridge with cross-validation
H. Zhang et al. (2020)	DeepBindPoc [79]	Deep Convolutional Neural Networks (DCNN)
I. Kozlovskii et al. (2020)	BiteNet [80]	3D-Convolutional Neural Networks (DCNNs)

Table 3 (continued)

Authors/year	Method	Machine learning and deep learning technique
M. M. Stepniewska-Dziubinska et al. (2020)	Kalasanty [81]	U-Net: Convolutional networks for biomedical image segmentation
N. Verma et al.(2021)	SSnet [130]	Deep Convolutional Neural Networks (DCNN)
F. Hu et al.(2021)	Multi-PLI [132]	Deep Convolutional Neural Networks (DCNN)

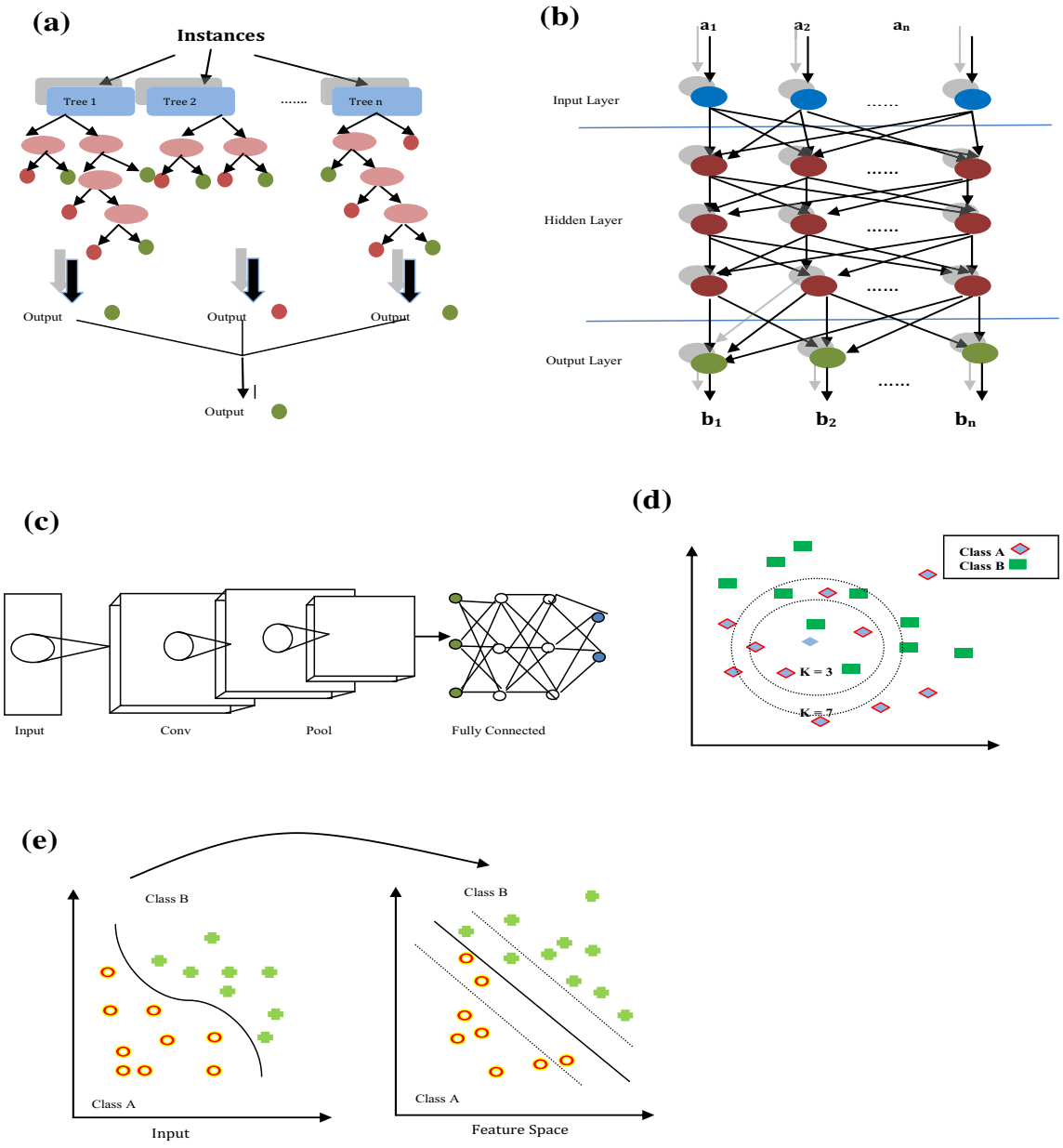

Fig. 5 Diagrammatically representation of various Artificial Intelligence techniques used in the field of Drug discovery and development: **a** Random Forest **b** neural network **c** Convolution neural network **d** KNN and **e** Support Vector Machine

the test set, MetaDBSite reported an ACC, Spe, Sen of 0.77 and MCC of 0.32. Ke Chen et al. [52] in 2012 proposed the NsitePred algorithm SVM based model for predicting binding residues for ATP, ADP, AMP, GTP and GDP from protein sequences. First, secondary structure, dihedral angles and relative solvent accessibility are extracted, then the PSSM profile of the given protein sequence is tested, and an eigenvector for the given protein is generated for describing the residue. NsitePred performs better than ATPint [45] and GTPbinder [98]. Yank Zhang et al. [55] in 2013 presented a SVM-based method, combining sequence information-based prediction and template-based methods TM_SITE and S-SITE along with the prediction result of the three methods FINDSITE [99], ConCavity[100], and COFACTOR[101] for training SVM. COACH model performs better than classical prediction algorithms with MCC = 0.54 and Pre = 0.59.

In the early 2000s, DL has surpassed ML in various fields such as speech recognition [102], image recognition [103], text classification [106], image segmentation [104] and semantic modeling [105]. DL solve the complex problem even if the dataset is very large, inter-connected and unstructured that make it well suited to solve the complex problem in the field of drug discovery and medical domain such as image-base diagnosis of diseases, predicting the chemical activity of compound, designing the chemical structure of compound and protein LBS prediction. In the past few years, DL has been used by researchers for protein ligand-binding-sites prediction. Some of the DL based LBS prediction methods are discussed below.

Durrant et al. [49] proposed a scoring function based on a neural network trained on 4141 protein–ligand complexes from the Protein Data Bank. Neural network having one hidden layer.

and five neurodes was trained to find out the influence of the training set size and architecture of the network on the accuracy and robustness of the network output DL based deep belief network (DBN) named as DeepDTIs was proposed by Wen et al. [68] for effective prediction of drug-target interaction. They tested their model with an independent test set and compared them with other algorithms: RF, Bernoulli Naïve Bayesian (BNB) and Decision Tree (DT). The dataset being used was taken from the DrugBank database containing 1412 drugs, 1520 targets and 2,146,240 DTPs. The features being used included—Extended Connectivity Fingerprints (ECFP) and protein sequence composition descriptors (PSC) for drugs and target representation. DBN performed better than BNB and DT. Testing was done on external EDTIs data extracted from the DrugBank database containing1412 drugs, 1520 targets and 2,146,240 DTPs. The performance of the model was measured on parameters including Area under the ROC Curve (AUC), accuracy, sensitivity and specificity 91.58, 85.88, 82.27 and 89.53, respectively. 3D-Convolutional Neural Networks KDEEP [71] predicted binding affinity and compared it with another ML approach. Dataset PDBbind (v.2016) has been used to contain 13,308 protein – ligand complexes and their corresponding experimentally determined binding affinities. The 3D convolution neural network was compared with RF-Score, X-Score, and cyScore. They reported the comparison of Pearson's correlation coefficient (R) of 0.82 and a RMSE of 1.27 in pK units between experimental and predicted affinity. A study done by Ztu et al. [70] considered only sequential information of both target and drug for binding affinity prediction using a DL based model. CNN based model DeepDTA was developed on two Kinase dataset Davis and KIBA dataset. The Concordance Index (CI) was used for the performance measure of the model and it was compared with the Kronecker Regularized Least Squares (KronRLS) based approach and SimBoost based approach. Nested cross validation was used to decide the best parameters for each test set. Lee et al. [72] proposed a Drug Target Interaction (DTI) prediction model based on convolution neural networks by performing convolution on various lengths of amino acids sub sequences. The model was trained which contained data from three databases: DrugBank, KEGG, and IUPHAR; duplicates from the dataset were removed. Dataset contained 11,950 compounds, 3,675 proteins and 32,568 DTIs. Negative DTI dataset was inevitably generated randomly. Biasness from randomly generated negative DTIs dataset was reduced by building ten sets from a positive dataset. Validation dataset was created from the MATADOR dataset, and all DTIs observed in the training dataset were excluded. Evaluation of the model was done on two independent test datasets from the PubChem BioAssay database and ChEMBL KinaseSARfari. Hyperparameters such as learning rate and window sizes are tuned during cross validation to increase the performance of the model. They first determined the learning rate of the model, then the selection of the activation function and regularization parameters were set. Grid-search method was employed for optimization of other hyperparameters for neural networks. DeepCSeqSite proposed by Cui et al. [75], in which several convolutional layers were stacked on each other to extract hierarchical features. Convolutional kernels combined extracted features and for prediction softmax was used. In 2020, Zhang, H. et al. proposed DeepBindPoc based on DL; the model was developed by incorporating the information of the binding pocket and associated ligand. The model contains densely connected 16 layers outputting 100 units. The ReLU activation function is used for the hidden layer and output layer employed by the sigmoid activation function. One of the advantages of DNN is that it learns more high-level and abstract features of very complex data. BiteNet [80] DL based model identifies binding sites by spatiotemporal features identification. It represents the structure of the protein as a 3D image with a channel corresponding to atomic

densities. It has been observed that it takes about 0.1 s for the analysis of single conformation and about 1.5 min for analyzing MD trajectory of 1000 frames, each frame containing about 2000 atoms. SSnet [130] a deep neural network based model was developed for prediction of protein ligand interaction by utilising secondary structure information and torsion of the protein backbone. It was observed that SSnet is not biased towards any specific conformation and was able to extract information for protein ligand interaction prediction. DEELIG [131] a DL based model uses the spatial relationship among data for prediction of affinity and binding of proteins. Multi-PLI [132] a DL based model was developed to overcome generalizability and heterogeneous data issues that occur in structure based models. Classification was used to find out whether it was binding or not binding. Regression was used for finding binding affinity. It was found out that the model was able to predict amino acids that were essential for the binding of ligands.

5.3 Critical Issues and Future Directions

Tools have been developed for protein structure prediction but still there are many issues. Energy functions needed for prediction are approximated for computational efficiency as a result of which accurately balancing non-polar and polar interaction at the interface is a challenging task. However, only modelling subset with ordered water molecules can be done but it is a computationally costly process. The need is to develop both robust techniques for prediction of protein structure and conformation. Analysis of the energy landscape along with molecular dynamics trajectory can be exploded to capture a flexible dynamics system. The binding site prediction depends on structure information of protein. As the protein structure database will grow in future it will open up the opportunities for the improvement of binding site prediction and functions.

6 Open Research Challenges and Opportunities in Drug Discovery

The major challenge faced by pharmaceutical companies to develop a new drug is its cost and time required. AI technologies have been successfully applied in various fields: Natural Language Processing, Signal Processing, computer vision, agriculture Sector etc. and has the potential to reduce the time and cost required developing a new drug. Many researchers have shown that the future of drug discovery is very promising as covered in our review. Still, application of AI in the field of drug discovery is very challenging. Drug discovery is a very complex process and it requires knowledge of various fields (chemistry, biology and medicine). Second, reliability and safety are the major issues in the decision-making process of the discovery as it directly affects public health. High quality data is the main concern. Data labelling in drug discovery is very complicated. Moreover, data available is very less as compared to the large amount of information available in records, as open data sharing is not common in pharmaceutical companies. The data that is available is not in uniform format. The solution of this problem is to start an initiative to share data for the betterment of Public Health. To deal with heterogeneity of the data a "one-short learning" algorithm has been developed by Stanford University [119]. Fourth, Lead optimization is a challenging phase to develop effective drugs with desired properties and sometimes these parameters are incompatible and independent. This makes the process very complicated. Optimizing each parameter individually and improving our model this problem can be solved. Another major challenge faced by companies using AI in drug discovery is that they have to undergo a rigorous process to have copyright for their work as most of the countries do not give patents to these inventions.

7 Discussion and Conclusion

Drug discovery and development is a complex process and typically costs billions of USD and takes about 10–12 years to bring a drug to the market. To address this problem different AI based techniques have been explored. An example of this is drug screening, there are millions of drug-like compounds at online databases and laboratory screening of each of these compounds traditionally costs 60–100 USD and takes several months to screen a significant batch, yet, it remains unfeasible to screen all available compounds even through high throughput robotics. With the advent of AI in the drug screening process, billions of compounds can be screened in a few days. Many AI based tools and methods have been developed to facilitate the different phases of drug development process from 3D structure prediction, target disease gene prediction, protein–ligand binding site prediction, drug screening, predicting physical property, toxicity and bioactivity, lead optimization, clinical trial to drug repurposing. In the past few years, AI computational methods have shown a great impact in the field of drug discovery that lead many pharmaceutical companies to invest in AI-based R&D programs and to have collaboration with AI start-ups and academic institutions. Takeda Pharmaceuticals Company and MIT's School of Engineering have collaborated to work together to start a drive to explore the application of AI in the field of healthcare and drug discovery. However, there are still some challenges to overcome. Just like organic chemists have over the years adopted a universal nomenclature of chemical compounds, the universality of health care record format and curation of metadata

of patients is still to be achieved. Enormous expanses of experimental screening information available at the research journal archives do not follow a common format and thus often requires manual reformatting before funnelling into an AI algorithm that in itself limits the efficient use of AI in this field. Even after successful identification of the genes implicated in disease development and identification of structural details of its protein product, its druggable pocket and key target amino acid residues identification through a bunch of AI algorithms, yet there is no integrated workflow which address this process end to end and this necessitates development of such a platform. Proteins are drug targets and are dynamic in nature, the excessive reliance on their experimentally available structures alone for AI based drug discovery can bias the results, thus conformational ensembles generated through the molecular dynamics simulations could also be included in such a procedure to boost identification of novel compound scaffolds for intellectual property. By and large, we feel safe to say that a strong foothold of AI is already into the drug discovery and development and due to realization of its strength; the associated issues will be addressed.

Author contributions R.K, I.S, J.M and V.S conceived the area of review. R.K and I.S did the literature review and wrote the article. J.M and V.S did edition and corrections.

Declarations

Conflicts of interest Authors declare no conflict of interest.

Human and animal rights No human and animal participants were used.

References

1. Agatonovic-Kustrin S, Beresford R (2000) Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal 22(5):717–727. https://doi.org/10.1016/S0731-7085(99)00272-1
2. Zhu T et al (2013) Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis (in eng). J Med Chem 56(17):6560–6572. https://doi.org/10.1021/jm301916b
3. Mak K-K, Pichika MR (2019) Artificial intelligence in drug development: present status and future prospects. Drug Discovery Today 24(3):773–780. https://doi.org/10.1016/j.drudis.2018.11.014
4. Wlodzislaw D, Swaminathan K, Meller J (2007) Artificial intelligence approaches for rational drug design and discovery. Curr Pharm Des 13(14):1497–1508. https://doi.org/10.2174/138161207780765954
5. Sidey-Gibbons JAM, Sidey-Gibbons CJ (2019) Machine learning in medicine: a practical introduction. BMC Med Res Methodol 19(1):64. https://doi.org/10.1186/s12874-019-0681-4
6. Costa PR, Acencio ML, Lemke N (2010) "A machine learning approach for genome-wide prediction of morbid and druggable human genes based on systems-level data," (in eng). BMC Genomics 11(Suppl 5):S9. https://doi.org/10.1186/1471-2164-11-s5-s9
7. Asif M, Martiniano H, Couto F (2018) "Identifying disease genes using machine learning and gene functional similarities, assessed through Gene Ontology. PLoS ONE 13:e0208626. https://doi.org/10.1371/journal.pone.0208626
8. Luo P, Xiao Q, Wei P-J, Liao B, Wu F-X (2019) "Identifying disease-gene associations with graph-regularized manifold learning. Front Genet. https://doi.org/10.3389/fgene.2019.00270
9. Barman RK, Mukhopadhyay A, Maulik U, Das S (2019) Identification of infectious disease-associated host genes using machine learning techniques. BMC Bioinform. https://doi.org/10.1186/s12859-019-3317-0
10. Mamoshina P et al (2018) "Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification," (in english). Front Genet. https://doi.org/10.3389/fgene.2018.00242
11. Ferrero E, Dunham I, Sanseau P (2017) In silico prediction of novel therapeutic targets using gene–disease association data. J Transl Med 15(1):182. https://doi.org/10.1186/s12967-017-1285-6
12. Jeon J et al (2014) "A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. Genome Med 6(7):57. https://doi.org/10.1186/s13073-014-0057-7
13. Sanchez-Lengeling B, Aspuru-Guzik A (2018) "Inverse molecular design using machine learning: Generative models for matter engineering. Science 361(6400):360–365. https://doi.org/10.1126/science.aat2663
14. Sheridan RP, Wang WM, Liaw A, Ma J, Gifford EM (2016) "Extreme gradient boosting as a method for quantitative structure-activity relationships," (in eng). J Chem Inf Model 56(12):2353–2360. https://doi.org/10.1021/acs.jcim.6b00591
15. Wallach I, Dzamba M, Heifets A (2015) AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery, *ArXiv,* abs/1510.02855
16. Pereira JC, Caffarena ER, dos Santos CN (2016) Boosting docking-based virtual screening with deep learning. J Chem Inf Model 56(12):2495–2506. https://doi.org/10.1021/acs.jcim.6b00355
17. Tyrchan C, Evertsson E (2017) Matched molecular pair analysis in short: algorithms, applications and limitations. Comput Struct Biotechnol J 15:86–90. https://doi.org/10.1016/j.csbj.2016.12.003
18. Turk S, Merget B, Rippmann F, Fulle S (2017) Coupling matched molecular pairs with machine learning for virtual compound optimization. J Chem Inform Model 57(12):3079–3085. https://doi.org/10.1021/acs.jcim.7b00298
19. Keefer CE, Chang G, Kauffman GW (2011) Extraction of tacit knowledge from large ADME data sets via pairwise analysis. Bioorg Med Chem 19(12):3739–3749. https://doi.org/10.1016/j.bmc.2011.05.003
20. Paixão P, Gouveia LF, Morais JA (2010) Prediction of the in vitro intrinsic clearance determined in suspensions of human hepatocytes by using artificial neural networks. Eur J Pharm Sci 39(5):310–321. https://doi.org/10.1016/j.ejps.2009.12.007
21. Leach AG et al (2006) Matched molecular pairs as a guide in the optimization of pharmaceutical properties; a study of aqueous solubility, plasma protein binding and oral exposure. J Med Chem 49(23):6672–6682. https://doi.org/10.1021/jm0605233
22. Huh D, Hamilton GA, Ingber DE (2011) From 3D cell culture to organs-on-chips. Trends Cell Biol 21(12):745–754. https://doi.org/10.1016/j.tcb.2011.09.005

23. Segall MD, Barber C (2014) Addressing toxicity risk when designing and selecting compounds in early drug discovery. Drug Discov Today 19(5):688–693. https://doi.org/10.1016/j.drudis.2014.01.006
24. Singh KP, Gupta S, Rai P (2013) Predicting carcinogenicity of diverse chemicals using probabilistic neural network modeling approache. Toxicol Appl Pharmacol. https://doi.org/10.1016/j.taap.2013.06.029
25. Tanabe K, Kurita T, Nishida K, Lučić B, Amić D, Suzuki T (2013) Improvement of carcinogenicity prediction performances based on sensitivity analysis in variable selection of SVM models. SAR QSAR Environ Res 24(7):565–580. https://doi.org/10.1080/1062936X.2012.762425
26. Zhang H, Cao ZX, Li M, Li YZ, Peng C (2016) Novel naïve Bayes classification models for predicting the carcinogenicity of chemicals. Food Chem Toxicol 97:141–149. https://doi.org/10.1016/j.fct.2016.09.005
27. Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) "Deeptox: toxicity prediction using deep learning," (in english). Front Environ Sci. https://doi.org/10.3389/fenvs.2015.00080
28. Segler MHS, Waller MP (2017) "Neural-symbolic machine learning for retrosynthesis and reaction prediction," (in eng). Chemistry 23(25):5966–5971. https://doi.org/10.1002/chem.201605499
29. Segler MHS, Preuss M, Waller MP (2018) "Planning chemical syntheses with deep neural networks and symbolic AI," (in eng). Nature 555(7698):604–610. https://doi.org/10.1038/nature25978
30. Schreck JS, Coley CW, Bishop KJM (2019) Learning retrosynthetic planning through simulated experience. ACS Cent Sci 5(6):970–981. https://doi.org/10.1021/acscentsci.9b00055
31. Harrer S, Shah P, Antony B, Hu J (2019) Artificial intelligence for clinical trial design. Trends Pharmacol Sci 40(8):577–591. https://doi.org/10.1016/j.tips.2019.05.005
32. Bain EE et al (2017) Use of a novel artificial intelligence platform on mobile devices to assess dosing compliance in a phase 2 clinical trial in subjects with schizophrenia (in eng). JMIR Mhealth Uhealth 5(2):e18. https://doi.org/10.2196/mhealth.7030
33. Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. Mol Pharm 13(7):2524–2530. https://doi.org/10.1021/acs.molpharmaceut.6b00248
34. Li B et al (2020) A novel drug repurposing approach for non-small cell lung cancer using deep learning. PLoS ONE 15(6):e0233112–e0233112. https://doi.org/10.1371/journal.pone.0233112
35. Zeng X, Zhu S, Liu X, Zhou Y, Nussinov R, Cheng F (2019) deepDR: a network-based deep learning approach to in silico drug repositioning. Bioinformatics 35(24):5191–5198. https://doi.org/10.1093/bioinformatics/btz418
36. Kadurin A, Nikolenko S, Khrabrov K, Aliper A, Zhavoronkov A (2017) druGAN: An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in Silico. Mol Pharm 14(9):3098–3104. https://doi.org/10.1021/acs.molpharmaceut.7b00346
37. Zhang W, Yue X, Huang F, Liu R, Chen Y, Ruan C (2018) Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network. Methods 145:51–59. https://doi.org/10.1016/j.ymeth.2018.06.001
38. Moghadam H, Rahgozar M, Gharaghani S (2016) Scoring multiple features to predict drug disease associations using information fusion and aggregation. SAR QSAR Environ Res 27(8):609–628. https://doi.org/10.1080/1062936x.2016.1209241
39. Senior AW et al (2019) Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). Proteins 87(12):1141–1148. https://doi.org/10.1002/prot.25834
40. Ballester PJ, Mitchell JBO (2010) A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics (Oxford, England) 26(9):1169–1175. https://doi.org/10.1093/bioinformatics/btq112
41. Deng W, Breneman C, Embrechts MJ (2004) Predicting protein-ligand binding affinities using novel geometrical descriptors and machine-learning methods. J Chem Inf Comput Sci 44(2):699–703. https://doi.org/10.1021/ci034246+
42. Guo T, Shi Y, Sun Z (2005) "A novel statistical ligand-binding site predictor: application to ATP-binding sites," (in eng). Protein Eng Des Sel 18(2):65–70. https://doi.org/10.1093/protein/gzi006
43. Ye K, Feenstra KA, Heringa J, Ijzerman AP, Marchiori E (2008) Multi-RELIEF: a method to recognize specificity determining residues from multiple sequence alignments using a Machine-Learning approach for feature weighting (in eng). Bioinformatics 24(1):18–25. https://doi.org/10.1093/bioinformatics/btm537
44. Sotriffer CA, Sanschagrin P, Matter H, Klebe G (2008) SFC-score: scoring functions for affinity prediction of protein-ligand complexes. Proteins 73(2):395–419. https://doi.org/10.1002/prot.22058
45. Chauhan JS, Mishra NK, Raghava GPS (2009) Identification of ATP binding residues of a protein from its primary sequence. BMC Bioinformatics 10(1):434. https://doi.org/10.1186/1471-2105-10-434
46. Huang B (2009) MetaPocket: a meta approach to improve protein ligand binding site prediction. OMICS 13(4):325–330. https://doi.org/10.1089/omi.2009.0045
47. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA (2009) Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput Biol 5(12):e1000585. https://doi.org/10.1371/journal.pcbi.1000585
48. Ballester PJ, Mitchell JB (2010) A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics 26(9):1169–1175. https://doi.org/10.1093/bioinformatics/btq112
49. Durrant JD, McCammon JA (2010) NNScore: a neural-network-based scoring function for the characterization of protein-ligand complexes. J Chem Inf Model 50(10):1865–1871. https://doi.org/10.1021/ci100244v
50. Durrant JD, McCammon J (2011) NNScore 2.0: a neural-network receptor ligand scoring function. J Chem Inf Model 51:2897–2903
51. Si J, Zhang Z, Lin B, Schroeder M, Huang B (2011) "MetaDBSite: a meta approach to improve protein DNA-binding sites prediction," (in eng). BMC Syst Biol 5(Suppl 1):S7. https://doi.org/10.1186/1752-0509-5-s1-s7
52. Chen K, Mizianty MJ, Kurgan L (2012) Prediction and analysis of nucleotide-binding residues using sequence and sequence-derived structural descriptors. Bioinformatics 28(3):331–341. https://doi.org/10.1093/bioinformatics/btr657
53. Dou Y, Wang J, Yang J, Zhang C (2012) L1pred: a sequence-based prediction tool for catalytic residues in enzymes with the L1-logreg classifier. PLoS ONE 7(4):e35666. https://doi.org/10.1371/journal.pone.0035666
54. Brylinski M, Feinstein WP (2013) eFindSite: improved prediction of ligand binding sites in protein models using meta-threading, machine learning and auxiliary ligands. J Comput Aided Mol Des 27(6):551–567. https://doi.org/10.1007/s10822-013-9663-5
55. Yang J, Roy A, Zhang Y (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment (in eng). Bioinformatics 29(20):2588–2595. https://doi.org/10.1093/bioinformatics/btt447

56. Panwar B, Gupta S, Raghava GPS (2013) Prediction of vitamin interacting residues in a vitamin binding protein using evolutionary information. BMC Bioinformatics 14(1):44. https://doi.org/10.1186/1471-2105-14-44
57. Chen P, Huang J, Gao X (2014) LigandRFs: Random forest ensemble to identify ligand-binding residues from sequence information alone. BMC Bioinformatics 15(Suppl 15):S4. https://doi.org/10.1186/1471-2105-15-S15-S4
58. Suresh MX, Gromiha MM, Suwa M (2015) Development of a machine learning method to predict membrane protein-ligand binding residues using basic sequence information. Adv Bioinformatics. https://doi.org/10.1155/2015/843030
59. Komiyama Y, Banno M, Ueki K, Saad G, Shimizu K (2016) Automatic generation of bioinformatics tools for predicting protein-ligand binding sites. Bioinformatics 32(6):901–907. https://doi.org/10.1093/bioinformatics/btv593
60. Yu DJ, Hu J, Li QM, Tang ZM, Yang JY, Shen HB (2015) Constructing query-driven dynamic machine learning model with application to protein-ligand binding sites prediction. IEEE Trans Nanobiosci 14(1):45–58. https://doi.org/10.1109/tnb.2015.2394328
61. Krivák R, Hoksza D (2015) Improving protein-ligand binding site prediction accuracy by classification of inner pocket points using local features. J Cheminform 7(1):12. https://doi.org/10.1186/s13321-015-0059-5
62. Alipanahi B, Delong A, Weirauch MT, Frey BJ (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol 33(8):831–838. https://doi.org/10.1038/nbt.3300
63. Jian JW et al (2016) Predicting ligand binding sites on protein surfaces by 3-dimensional probability density distributions of interacting atoms (in eng). PLoS ONE 11(8):e0160315–e0160315. https://doi.org/10.1371/journal.pone.0160315
64. Chen P et al (2016) "A sequence-based dynamic ensemble learning system for protein ligand-binding site prediction," (in eng). IEEE/ACM Trans Comput Biol Bioinform 13(5):901–912. https://doi.org/10.1109/tcbb.2015.2505286
65. Xavier MM et al (2016) SAnDReS a computational tool for statistical analysis of docking results and development of scoring functions (in eng). Comb Chem High Throughput Screen 19(10):801–812. https://doi.org/10.2174/1386207319666160927111347
66. Zhang S et al (2016) A deep learning framework for modeling structural features of RNA-binding protein targets (in eng). Nucleic Acids Res 44(4):e32. https://doi.org/10.1093/nar/gkv1025
67. Jiménez J, Doerr S, Martínez-Rosell G, Rose AS, De Fabritiis G (2017) DeepSite: protein-binding site predictor using 3D-convolutional neural networks (in eng). Bioinformatics 33(19):3036–3042. https://doi.org/10.1093/bioinformatics/btx350
68. Wen M et al (2017) Deep-learning-based drug-target interaction prediction. J Proteome Res 16(4):1401–1409. https://doi.org/10.1021/acs.jproteome.6b00618
69. Wu Q, Peng Z, Zhang Y, Yang J (2018) COACH-D: improved protein-ligand binding sites prediction with refined ligand-binding poses through molecular docking (in eng). Nucleic Acids Res 46(W1):W438-w442. https://doi.org/10.1093/nar/gky439
70. Krivák R, Hoksza D (2018) P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. Journal of Cheminformatics 10(1):39. https://doi.org/10.1186/s13321-018-0285-8
71. Öztürk H, Özgür A, Ozkirimli E (2018) DeepDTA: deep drug-target binding affinity prediction (in eng). Bioinformatics 34(17):i821–i829. https://doi.org/10.1093/bioinformatics/bty593
72. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G (2018) KDEEP: protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. J Chem Inf Model 58(2):287–296. https://doi.org/10.1021/acs.jcim.7b00650
73. Lee I, Keum J, Nam H (2019) DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences (in eng). PLoS Comput Biol 15(6):e1007129. https://doi.org/10.1371/journal.pcbi.1007129
74. Zheng L, Fan J, Mu Y (2019) OnionNet: a multiple-layer intermolecular-contact-based convolutional neural network for protein-ligand binding affinity prediction. ACS Omega 4(14):15956–15965. https://doi.org/10.1021/acsomega.9b01997
75. Zhao Z, Xu Y, Zhao Y (2019) SXGBsite: prediction of protein-ligand binding sites using sequence information and extreme gradient boosting (in eng). Genes (Basel) 10(12):965. https://doi.org/10.3390/genes10120965
76. Cui Y, Dong Q, Hong D, Wang X (2019) Predicting protein-ligand binding residues with deep convolutional neural networks. BMC Bioinformatics 20(1):93. https://doi.org/10.1186/s12859-019-2672-1
77. Pu L, Govindaraj RG, Lemoine JM, Wu H-C, Brylinski M (2019) DeepDrug3D: Classification of ligand-binding pockets in proteins with a convolutional neural network (in eng). PLoS Comput Biol 15(2):e1006718–e1006718. https://doi.org/10.1371/journal.pcbi.1006718
78. da Silva AD, Bitencourt-Ferreira G, de Azevedo WF, Jr. (2020) Taba: a tool to analyze the binding affinity (in eng). J Comput Chem 41(1):69–73. https://doi.org/10.1002/jcc.26048
79. Zhang H et al (2020) DeepBindPoc: a deep learning method to rank ligand binding pockets using molecular vector representation (in eng). Peer J 8:e8864–e8864. https://doi.org/10.7717/peerj.8864
80. Kozlovskii I, Popov P (2020) Spatiotemporal identification of druggable binding sites using deep learning (in eng). Commun Biol 3(1):618–618. https://doi.org/10.1038/s42003-020-01350-0
81. Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P (2020) Improving detection of protein-ligand binding sites with 3D segmentation (in eng). Sci Rep 10(1):5035–5035. https://doi.org/10.1038/s41598-020-61860-z
82. Wang L, Yang MQ, Yang JY (2009) Prediction of DNA-binding residues from protein sequence information using random forests. BMC Genomics 10(1):S1. https://doi.org/10.1186/1471-2164-10-S1-S1
83. Yan C, Terribilini M, Wu F, Jernigan RL, Dobbs D, Honavar V (2006) Predicting DNA-binding sites of proteins from amino acid sequence. BMC Bioinformatics 7(1):262. https://doi.org/10.1186/1471-2105-7-262
84. Wang L, Brown SJ (2006) BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences (in eng). Nucleic Acids Res 34:W243–W248. https://doi.org/10.1093/nar/gkl298
85. Ofran Y, Mysore V, Rost B (2007) Prediction of DNA-binding residues from sequence (in eng). Bioinformatics 23(13):i347–i353. https://doi.org/10.1093/bioinformatics/btm174
86. Ahmad S, Gromiha MM, Sarai A (2004) Analysis and prediction of DNA-binding proteins and their binding residues based on composition, sequence and structural information (in eng). Bioinformatics 20(4):477–486. https://doi.org/10.1093/bioinformatics/btg432
87. Hwang S, Gou Z, Kuznetsov IB (2007) DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins (in eng). Bioinformatics 23(5):634–636. https://doi.org/10.1093/bioinformatics/btl672
88. Källberg M et al (2012) Template-based protein structure modeling using the RaptorX web server (in eng). Nat Protoc 7(8):1511–1522. https://doi.org/10.1038/nprot.2012.085
89. Lambert C, Léonard N, De Bolle X, Depiereux E (2002) ESyPred3D: Prediction of proteins 3D structures (in eng).

Bioinformatics 18(9):1250–1256. https://doi.org/10.1093/bioinformatics/18.9.1250
90. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis (in eng). Nat Protoc 10(6):845–858. https://doi.org/10.1038/nprot.2015.053
91. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction (in eng). Nucleic Acids Res 33:W244–W248. https://doi.org/10.1093/nar/gki408
92. Gromski P, Granda J, Cronin L (2019) Universal chemical synthesis and discovery with 'the chemputer.' Trends Chem. https://doi.org/10.1016/j.trechm.2019.07.004
93. Sanchez B, Outeiral C, Guimaraes G, Aspuru-Guzik A (2017) *Optimizing distributions over molecular space. An Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry (ORGANIC)*.
94. Xu Y, Ma J, Liaw A, Sheridan RP, Svetnik V (2017) Demystifying multitask deep neural networks for quantitative structure-activity relationships. J Chem Inf Model 57(10):2490–2504. https://doi.org/10.1021/acs.jcim.7b00087
95. Stork C, Wagner J, Friedrich NO, de Bruyn Kops C, Šícho M, Kirchmair J (2018) Hit dexter: a machine-learning model for the prediction of frequent hitters (in eng). ChemMedChem 13(6):564–571. https://doi.org/10.1002/cmdc.201700673
96. Wójcikowski M, Zielenkiewicz P, Siedlecki P (2015) Open drug discovery toolkit (ODDT): a new open-source player in the drug discovery field. Journal of Cheminformatics. https://doi.org/10.1186/s13321-015-0078-2
97. Blaschke T et al (2020) REINVENT 2.0: an ai tool for de novo drug design (in eng). J Chem Inf Model 60(12):5918–5922. https://doi.org/10.1021/acs.jcim.0c00915
98. Chauhan JS, Mishra NK, Raghava GPS (2010) Prediction of GTP interacting residues, dipeptides and tripeptides in a protein from its evolutionary information. BMC Bioinformatics 11(1):301. https://doi.org/10.1186/1471-2105-11-301
99. Brylinski M, Skolnick J (2008) A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. Proc Natl Acad Sci 105(1):129. https://doi.org/10.1073/pnas.0707684105
100. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA (2009) Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure (in eng). PLoS Comput Biol 5(12):e1000585. https://doi.org/10.1371/journal.pcbi.1000585
101. Roy A, Yang J, Zhang Y (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation (in eng). Nucleic Acids Res 40:W471–W477. https://doi.org/10.1093/nar/gks372
102. Nassif A, Shahin I, Attili I, Azzeh M, Shaalan K (2019) Speech recognition using deep neural networks: a systematic review. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2896880
103. Xin M, Wang Y (2019) Research on image classification model based on deep convolution neural network. EURASIP J Image Video Process. https://doi.org/10.1186/s13640-019-0417-8
104. Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D (2020) *Image Segmentation Using Deep Learning: A Survey*.
105. Huang X, Zanni-Merk C, Crémilleux B (2019) Enhancing Deep Learning with Semantics: an application to manufacturing time series analysis. Procedia Comput Sci 159:437–446. https://doi.org/10.1016/j.procs.2019.09.198
106. Hassan A, Mahmood A (2017) Efficient Deep Learning Model for Text Classification Based on Recurrent and Convolutional Layers. *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA))*. 1108–1113
107. Coley CW, Rogers L, Green WH, Jensen KF (2018) SCScore: synthetic complexity learned from a reaction corpus. J Chem Inf Model 58(2):252–261. https://doi.org/10.1021/acs.jcim.7b00622
108. Duvenaud D, Maclaurin D, Aguilera-Iparraguirre J, Gómez-Bombarelli R, Hirzel T, Aspuru-Guzik A, Adams R (2015) Convolutional networks on graphs for learning molecular fingerprints. *Advances in Neural Information Processing Systems (NIPS)*.
109. Piñero J et al (2020) The DisGeNET knowledge platform for disease genomics: 2019 update (in eng). Nucleic Acids Res 48(D1):D845-d855. https://doi.org/10.1093/nar/gkz1021
110. Wang Y et al (2020) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics (in eng). Nucleic Acids Res 48(D1):D1031–D1041. https://doi.org/10.1093/nar/gkz981
111. Szklarczyk D et al (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets (in eng). Nucleic Acids Res 47(D1):D607-d613. https://doi.org/10.1093/nar/gky1131
112. Vasaikar SV, Straub P, Wang J, Zhang B (2018) LinkedOmics: analyzing multi-omics data within and across 32 cancer types (in eng). Nucleic Acids Res 46(D1):D956-d963. https://doi.org/10.1093/nar/gkx1090
113. Carvalho-Silva D et al (2019) Open Targets Platform: new developments and updates two years on (in eng). Nucleic Acids Res 47(D1):D1056–D1065. https://doi.org/10.1093/nar/gky1133
114. DepMap portal. https://depmap.org/portal/.
115. Huang Z et al (2019) HMDD v3.0: a database for experimentally supported human microRNA-disease associations (in eng). Nucleic Acids Res 47(D1):D1013-d1017. https://doi.org/10.1093/nar/gky1010
116. Davis AP et al (2018) The comparative Toxicogenomics database: update 2019. Nucleic Acids Res 47(D1):D948–D954. https://doi.org/10.1093/nar/gky868
117. Pearson N et al (2019) TractaViewer: a genome-wide tool for preliminary assessment of therapeutic target druggability (in eng). Bioinformatics 35(21):4509–4510. https://doi.org/10.1093/bioinformatics/btz270
118. Gaspar HA, Hübel C, Breen G (2019) Drug Targetor: a web interface to investigate the human druggome for over 500 phenotypes (in eng). Bioinformatics 35(14):2515–2517. https://doi.org/10.1093/bioinformatics/bty982
119. Altae-Tran H, Ramsundar B, Pappu AS, Pande V (2017) Low data drug discovery with one-shot learning. ACS Cent Sci 3(4):283–293. https://doi.org/10.1021/acscentsci.6b00367
120. Das K, Daschakladar D, Roy PP, Chatterjee A, Saha SP (2020) Epileptic seizure prediction by the detection of seizure waveform from the pre-ictal phase of EEG signal. Biomed Signal Process Control 57:101720. https://doi.org/10.1016/j.bspc.2019.101720
121. Mukherjee S, Kumar P, Saini R, Roy PP, Dogra DP, Kim B-G (2017) Plant disease identification using deep neural networks. J Multimed Inform Sys 4(4):233–238
122. Das Chakladar D, Dey S, Roy PP, Dogra DP (2020) EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm. Biomed Signal Process Control 60:101989. https://doi.org/10.1016/j.bspc.2020.101989
123. Chakladar DD, Dey S, Roy PP, Iwamura M (2021) EEG-based cognitive state assessment using deep ensemble model and filter bank common spatial pattern. In *2020 25th International Conference on Pattern Recognition (ICPR)*. https://doi.org/10.1109/ICPR48806.2021.9412869.
124. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P (2021) The role of artificial intelligence in healthcare: a structured literature review. BMC Med Inform Decis Mak 21(1):125. https://doi.org/10.1186/s12911-021-01488-9

125. Aiolli F, Palazzi C (2008) Enhancing artificial intelligence in games by learning the opponent's playing style. *international federation for information processing digital library; First IFIP Entertainment Computing Symposium on "New Frontiers for Entertainment Computing (ECS-2008),* 279. https://doi.org/10.1007/978-0-387-09701-5_1.
126. Demlehner Q, Schoemer D, Laumer S (2021) How can artificial intelligence enhance car manufacturing? A Delphi study-based identification and assessment of general use cases. Int J Inf Manage 58:102317. https://doi.org/10.1016/j.ijinfomgt.2021.102317
127. Ben Ayed R, Hanana M (2021) Artificial intelligence to improve the food and agriculture sector. J Food Qual. https://doi.org/10.1155/2021/5584754
128. Capatina A, Kachour M, Lichy J, Micu A, Micu A-E, Codignola F (2019) Matching the future capabilities of an artificial intelligence-based platform for social media marketing with potential users' expectations. Technol Forecast Soc Chang. https://doi.org/10.1016/j.techfore.2019.119794
129. Abreu Araujo F et al (2020) Role of non-linear data processing on speech recognition task in the framework of reservoir computing. Sci Rep 10(1):328. https://doi.org/10.1038/s41598-019-56991-x
130. Verma N et al (2021) SSnet: A Deep Learning Approach for Protein-Ligand Interaction Prediction (in eng). Int J Mol Sci. https://doi.org/10.3390/ijms22031392
131. Ahmed A, Mam B, Sowdhamini R (2021) DEELIG: a deep learning approach to predict protein-ligand binding affinity. Bioinform Biol Insights 15:11779322211030364. https://doi.org/10.1177/11779322211030364
132. Hu F, Jiang J, Wang D, Zhu M, Yin P (2021) Multi-PLI: interpretable multi-task deep learning model for unifying protein–ligand interaction datasets. J Cheminform 13(1):30. https://doi.org/10.1186/s13321-021-00510-6
133. Wu Z et al (2018) MoleculeNet: a benchmark for molecular machine learning. Chem Sci 9(2):513–530. https://doi.org/10.1039/C7SC02664A
134. Chakravarti SK, Alla SRM (2019) Descriptor free QSAR modeling using deep learning with long short-term memory neural networks (in english). Front Artif Intell. https://doi.org/10.3389/frai.2019.00017
135. Hu J, Lepore R, Dobson RJB, Al-Chalabi A, Bean DM, Iacoangeli A (2021) DGLinker: flexible knowledge-graph prediction of disease–gene associations. Nucleic Acids Res 49(W1):W153–W161. https://doi.org/10.1093/nar/gkab449
136. Shu J, Li Y, Wang S, Xi B, Ma J (2021) Disease gene prediction with privileged information and heteroscedastic dropout. Bioinformatics 37(Supplement1):i410–i417. https://doi.org/10.1093/bioinformatics/btab310
137. Kolosov N, Daly MJ, Artomov M (2021) Prioritization of disease genes from GWAS using ensemble-based positive-unlabeled learning. Eur J Hum Genet. https://doi.org/10.1038/s41431-021-00930-w
138. Jain S et al (2021) Large-scale modeling of multispecies acute toxicity end points using consensus of multitask deep learning methods (in eng). J Chem Inf Model 61(2):653–663. https://doi.org/10.1021/acs.jcim.0c01164
139. Wu F, Zhuo L, Wang F, Huang W, Hao G, Yang G (2020) Auto in silico ligand directing evolution to facilitate the rapid and efficient discovery of drug lead (in eng). iScience 23(6):101179. https://doi.org/10.1016/j.isci.2020.101179
140. Shao J, Yan K, Liu B (2020) FoldRec-C2C: protein fold recognition by combining cluster-to-cluster model and protein similarity network. Brief Bioinform. https://doi.org/10.1093/bib/bbaa144
141. https://www.worldatlas.com/articles/countries-with-the-biggest-global-pharmaceutical-markets-in-the-world.html
142. https://www.marketsandmarkets.com/Market-Reports/ai-in-drug-discovery-market-151193446.html
143. Yu DJ, Hu J, Yang J, Shen HB, Tang J, Yang JY (2013) Designing template-free predictor for targeting protein-ligand binding sites with classifier ensemble and spatial clustering. IEEE/ACM Trans Comput Biol Bioinform 10(4):994–1008. https://doi.org/10.1109/TCBB.2013.104