Check for
updates

From understanding diseases to drug design: can artificial intelligence bridge the gap?

Anju Choorakottayil Pushkaran[1] · Alya A. Arabi[1]

Abstract

Artificial intelligence (AI) has emerged as a transformative technology with significant potential to revolutionize disease understanding and drug design in healthcare. AI serves as a remarkable accelerating tool that bridges the gap between understanding diseases and discovering drugs. Given its capacity in the analysis and interpretation of massive amounts of data, AI is tremendously boosting the power of predictions with impressive accuracies. This allowed AI to pave the way for advancing all key stages of drug development, with the advantage of expediting the drug discovery process and curbing its costs. This is a comprehensive review of the recent advances in AI and its applications in drug discovery and development, starting with disease identification and spanning through the various stages involved in the drug discovery pipeline, including target identification, screening, lead discovery, and clinical trials. In addition, this review discusses the challenges that arise during the implementation of AI at each stage of the discovery process and provides insights into the future prospects of this field.

Keywords Artificial intelligence · Machine learning · Disease identification · Drug discovery

1 Introduction

The classical drug discovery process is long and expensive. It takes around 10 to 15 years for a drug to be in the market, at an approximate cost of around $161 million to $4.54 billion (Schlander et al. 2021). Despite the investment of money, efforts, and resources, nearly 90% of the potential drug candidates fail in clinical trials (Sun et al. 2022). This is because of their reduced clinical efficacy, poor pharmacokinetic properties, or adverse side effects (Waring et al. 2015). More efforts are being put forward to develop alternative methods that can accelerate the drug discovery process while reducing the cost associated with it and increasing the success rate of lead compounds in clinical trials. Over the last decades, many methods, with AI and machine learning (ML) being at the forefront,

✉ Alya A. Arabi
alya.arabi@uaeu.ac.ae; alya.arabi@dal.ca

1   College of Medicine and Health Sciences, Department of Biochemistry and Molecular Biology, United Arab Emirates University, P. O. Box: 15551, Al Ain, United Arab Emirates

have been developed and successfully implemented at several stages of drug design, starting from disease identification to clinical trials. This extensive focus on AI research is exponentially growing, as evident from the number of scholarly outputs published over the years (Fig. 1).

With the rapid advances in computer-based technology, computational methods have quickly become indispensable for medical research. For instance, in the past decades, many efforts have been put into developing computational chemistry tools that can predict drug properties and their interactions in silico. Such tools have helped reduce the heavy dependence on wet-lab measurements, which tend to be expensive and time-consuming. These tools include molecular docking and molecular dynamics methods, both of which can be applicable to bulky biochemical systems; as well as quantum mechanics (QM) methods, which offer notable improvements in accuracy, yet are too computationally expensive to be tractable for the relatively large systems studied in drug design (Bolcer and Hermann 2007). Recently, more attention has been given to computer science, ML, and statistical methods that can predict the properties of large macrosystems with the accuracy of quantum methods, but at low computational costs. Such ML models are used as building blocks for developing AI tools. AI involves the development of machines with the ability to perform tasks that require human intelligence and predictive power. AI models are demonstrated to potentially have high accuracies in predictions and, thus, have the tendency to be reliable in decision support (Manallack and Livingstone 1999).

There are different classes of ML methods, among which the most commonly used methods in the drug discovery process are supervised learning, unsupervised learning, semi-supervised learning, ensemble learning, and deep learning (Patel et al. 2020). Table 1 provides a list of some important summary tables and figures reported in the literature about AI in drug discovery and below is a brief description of the most used classes and subclasses of AI algorithms in this review.

Before describing how AI connects disease diagnosis with drug development, a concise overview of the classes and subclasses of AI algorithms recurrently mentioned in this review is provided. Supervised learning is central to drug discovery. The key requirement to develop a supervised learning model is having a labeled dataset. For
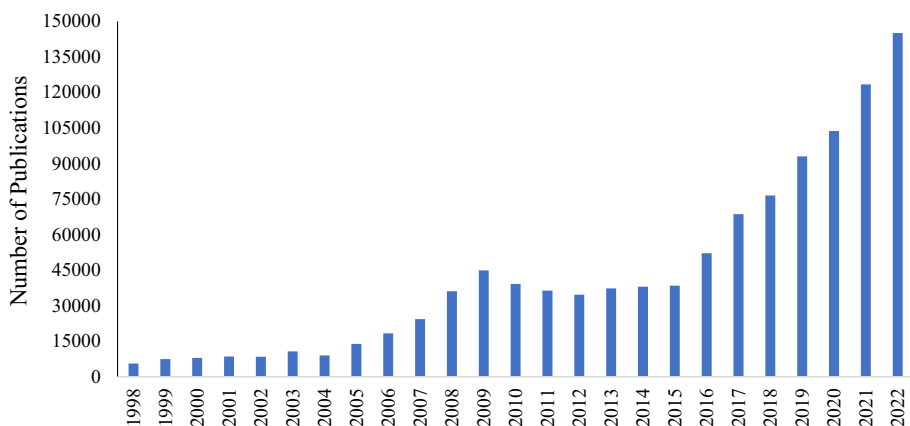
Fig. 1 Number of research articles related to AI that were published between 1998 and 2022. The data was obtained from the Social/Scopus database as of January 6, 2024. The total number of scholarly output was extracted for the 'World' from the 'Artificial Intelligence' subject under the 'Computer Science' subject area

Table 1 A list of important summary tables and figures reported in the literature about AI in drug discovery, along with a brief description of each item and its references

Item	Description	References
Figure 2	Schemes used by random forest and support vector machine in drug discovery.	Patel et al. (2020)
Figure 2	An overview of supervised and unsupervised AI algorithms and their utilization in drug discovery.	Dara et al. (2021)
Figure 1	History of AI in healthcare.	Gupta et al. (2021)
Figure 2	A flowchart of conventional ML and deep-learning models for drug-drug interaction predictions.	Vo et al. (2022)
Table 1	ML algorithms used specifically in personalized and genomic medicine.	Quazi (2022)
Figure 1	A timeline of the key events and developments that have shaped the use of network-based and ML-based approaches in target identification for cancer.	You et al. (2022)
Figure 4	Advantages and challenges of AI implementations in clinical trials for infectious diseases, endocrinology, cardiovascular, and oncology.	Askin et al. (2023)
Table 1	Applications of Explainable AI (XAI)-based algorithms in drug discovery.	Jiménez-Luna et al. (2020)

example, assessing the activity of chemical compounds against a specific target involves using a dataset containing information about compounds, along with their corresponding biological assay results (i.e., active or inactive). This labelling enables the model to learn the relationship between the chemical features of the compounds and their biological activity. The model can then be used to predict the biological activity of novel compounds. Examples of supervised learning algorithms are Support Vector Machine (SVM), Random Forest (RF), and naïve Bayes (Yang et al. 2019; Dara et al. 2021).

SVM is a binary classifier method that can be extended for multi-class classification tasks. SVM can perform both classification and regression tasks. It begins with labeled training data. Normalization or scaling of data is essential to ensure optimal results. During training, an SVM algorithm finds the optimal hyperplane or decision boundary to separate the data into different classes, calculated by finding support vectors and maximizing the margin (i.e., the distance between support vectors and the boundary). Mathematical optimization techniques are employed to optimize the margin while minimizing classification errors. The trained SVM model can then classify new, unseen data points (Yang et al. 2019). SVM is versatile and effective in handling complex data separation tasks. However, it is sensitive to the choice of hyperparameters, such as the regularization parameter and kernel parameter, where a kernel is a function that computes the similarity between data points in a higher-dimensional space. SVM can also be affected by overfitting when the input data is noisy or when the kernel function is not well-suited to the problem (Vamathevan et al. 2019). It has a wide range of applications in drug discovery such as virtual screening, predicting pharmacokinetic properties, and predicting toxicity (Heikamp and Bajorath 2013).

RF is a supervised and ensemble learning method that combines multiple decision trees to make predictions. It begins with bootstrapped sampling, where multiple subsets of the training data are created through random sampling with replacement. This introduces diversity in the dataset for each tree in the forest. Additionally, at each split point in the decision tree construction, a random subset of features is chosen. The next step involves growing multiple decision trees, each independently trained on these bootstrapped datasets and random feature subsets. This process generates a collection of diverse decision trees. When making predictions, the results from all the individual trees are combined. In classification tasks, this is achieved through majority voting, while in regression problems, predictions are averaged. Its advantages include high accuracy, resistance to overfitting, and suitability for various types of data (Patel et al. 2020). RF can handle large datasets with many features and is capable of ranking feature importance. However, its disadvantages involve reduced interpretability compared to individual decision trees. The model may not perform well on very noisy data. While RF is robust, it might not be the best choice for tasks requiring precise probability estimates.

Naïve Bayes is a probabilistic ML algorithm used for classification tasks. It is created by modeling the relationships between features and their associated classes using Bayes' theorem. The 'naïve' assumption is that the features are conditionally independent, simplifying the modeling process and making it computationally efficient. The key advantages of this model involve the speed, and suitability for high-dimensional data. However, its naïve independence assumption may not hold in complex real-world datasets, potentially affecting accuracy. It also struggles with rare events and might require extensive data preprocessing (Yang et al. 2019).

Unsupervised learning deals with unlabeled data. Its primary objective is to uncover underlying structures and features within the data to facilitate the grouping of input samples into clusters or reduce dimensionality. These algorithms are useful in applications where

predefined target outcomes are not available. Unsupervised learning is distinguished by the absence of feedback signals for assessing solution quality. Notable techniques include clustering methods (such as k-means and hierarchical clustering) and dimensionality reduction approaches (for example, principal component analysis and self-organizing maps) (Dara et al. 2021).

Semi-supervised learning is a hybrid between supervised and unsupervised learning, proving particularly valuable in scenarios where an abundance of input data is available with limited labeled samples. It has predictive accuracy with minimal additional real-world experimental costs. Semi-supervised models are trained to use available labeled data to predict labels for unlabeled data, and their performance heavily relies on the amount and quality of labeled data available (Yang et al. 2019).

Reinforcement Learning (RL) is an ML algorithm where the model learns to make sequences of decisions through interaction with an environment. It is built on the idea of an agent taking actions to maximize cumulative rewards over time. The agent learns by exploring different actions to alter that environment and receiving feedback in the form of rewards or punishments. RL typically involves defining a reward function, selecting an appropriate RL algorithm (e.g., Q-learning, Deep Q Networks), and fine-tuning the agent's policies through repeated interactions. While RL has been successful in various applications, it comes with challenges like the exploration-exploitation trade-off and can require significant computational resources and time. The key advantage of the RL model is that the training can be done even in sparse environments where few or no examples are available. It is especially suitable for sequential decision-making (Lutz et al. 2023).

Deep learning is focused on artificial neural networks (ANN) with multiple layers (deep neural networks). They consist of input and output layers, along with several hidden layers that learn progressively more abstract features. These networks are designed to automatically learn and extract features from raw data through a hierarchy of layers. Provided that they have sufficient data and resources, deep learning models can scale to handle complex problems such as natural language processing and speech recognition (Nag et al. 2022). Deep learning models are prone to overfitting on small datasets. Deep learning models are generally used to analyze and process large amounts of data for e.g., clinical imaging (Rajpurkar et al. 2018; Ding et al. 2019; Narin et al. 2021), virtual screening (Carpenter et al. 2018; Gentile et al. 2022), and bioactivity predictions (Bule et al. 2021). Examples of deep learning algorithms are Deep Neural Networks (DNN) and Convolutional Neural Networks (CNN).

DNNs are basic feedforward neural networks with multiple hidden layers. They are constructed by stacking multiple layers within which artificial neurons are interconnected and employ activation functions to introduce non-linearity. DNNs are trained using labeled data to minimize prediction errors through backpropagation and gradient descent. DNNs can be applicable in both supervised and unsupervised learning scenarios. A major limitation of DNNs is their complex nature which makes them challenging to interpret and they may involve meticulous hyperparameter tuning (Vamathevan et al. 2019; Nag et al. 2022).

CNNs are suited for image and grid-structured data. They employ convolutional layers to detect local patterns. CNNs are constructed with several key components: convolutional layers, pooling layers, and densely connected layers. The network begins with an input layer, followed by convolutional layers that extract features from the input data. The characteristics of the convolutional layer are specified in terms of its three dimensions (width, height, and depth). It operates by scanning and capturing information from a small receptive field, typically a square of pixels, and the depth

corresponds to different channels of information sources in images. Activation functions introduce non-linearity, while pooling layers reduce spatial dimensions. Densely connected layers connect neurons across layers, leading to the final output layer, which produces predictions. CNNs use a loss function to quantify the prediction error, and optimization algorithms to adjust model parameters (Yang et al. 2019). They are trained on labeled data, evaluated for performance, and then used for making predictions. CNN architecture can be customized for specific tasks and datasets. One of the major disadvantages of this method is that it might not be the best choice for all types of data.

Ensemble learning combines multiple individual models or algorithms to create a more robust and accurate predictive model with reduced overfitting issues, and enhanced generalizations. Among the common ensemble methods are RF (as mentioned above) and Gradient Boosting (GB). GB is an ensemble of multiple combined models used for regression and classification tasks of complex datasets. The algorithm works iteratively, it focuses on the errors made by the previous models and optimizes the following model to correct those errors. Gradient Boosting requires careful parameter selection, and potentially longer training time (K and Mohan 2022).

As demonstrated in this review (see Fig. 2), the use of AI resulted in substantial advancements that helped bridging the gap between disease diagnosis and drug development, ultimately increasing the chances of drug approval. Figure 2 depicts the key stages of drug discovery, along with their corresponding timelines. These stages include disease diagnosis, target identification, lead identification, lead optimization, preclinical trials, clinical trials, and drug approval. For each of these stages, a set of tasks that can benefit from AI is listed. For an extensive list of such AI tools, please refer to Table 2. For example, as depicted in Fig. 2 and Table 2, target identification can be enhanced by using ML methods for 3D structure prediction, image reconstruction, and druggability prediction. Similarly, lead identification can benefit from AI in speeding up virtual screening, pharmacophore modeling, designing synthetic routes, and predicting bioactivity and toxicity. For instance, using AI, the development of a drug called DSP-1181
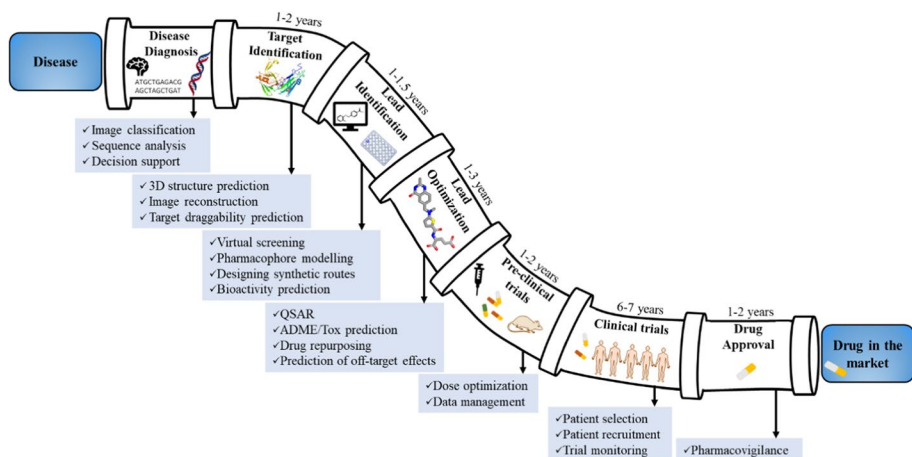
Fig. 2 A schematic diagram illustrating the drug discovery pipeline, including all stages from disease identification to drug marketing. The timeline provided for each stage is the average time needed without AI support. The points in the blue boxes are the steps where AI can be beneficial in speeding up task completion

Table 2 A list of recent and/or dominant AI platforms and tools used in drug discovery, along with their application and limitations. The reference for the tool and the link to access it (whenever available) are also provided

AI Tool /Platform	Application	Limitations	Reference	URL to the application or source code
Clinical Diagnosis				
NoduleX	A deep convolutional neural network to predict lung nodule malignancy from CT images.	• Trained on smaller set of data (only in the range of thousands of images).	Causey et al. (2018)	https://github.com/jcausey-astate/NoduleX_code
IDx-DR	A device based on deep learning to detect retinopathy in diabetic patients.	• Not able to detect mild diabetic retinopathy. • Requires superior quality retinal images captured through Topcon TRC-NW400 camera.	Heijden et al. (2017)	It is proprietary software. https://www.healthvisors.com/en/idx-dr/
Fabric GEM	An advanced AI algorithm to identify genetic causes of rare diseases by analyzing sequence data.	• Requires high-performance computing.	Vega et al. (2021)	It is proprietary software. https://fabricgenomics.com/fabric-gem/
Personalized Medicine				
IBM Watson for oncology	A logistic regression classifier model to analyze patient data and to provide treatment options for breast, lung, colon, rectal, gastric, and cervical cancer.	• Expensive. • Predictions are evidence-based, may not consider patients' unique characteristics like novel mutation. • Not trained with diverse patient dataset.	Fu et al. (2015)	It is proprietary software. https://www.ibm.com/docs/en/announcements/watson-oncology
CURATE.AI	A machine learning tool to advance personalized medicine through dose optimization and to improve the efficacy of drug therapies.	• Currently optimized only for a certain category of diseases such as cancer.	Blasiak et al. (2020)	Under development.
Gene Sequence Analysis				
DeepVariant	A deep neural network model to call genetic variants from next generation sequencing data.	• May not work well for somatic cell variant calling. • Needs additional tools or expertise for result interpretation. • Complex to set up and configure. • Less accurate for *de novo* variants. • Prediction is slower compared to DeepTrio.	DePristo and Poplin (2017)	https://github.com/google/deepvariant

Table 2 (continued)

| AI Tool/Platform	Application	Limitations	Reference	URL to the application or source code
DeepTrio	A deep learning model to call genetic variants of duos or trios.	• May not work well for somatic cell variant calling. • Needs additional tools or expertise for result interpretation. • Complex to set up and configure. • Not suitable for long deletions. • More suitable for SNPs and short INDELS. • Suitable only for trio or duo samples.	Kolesnikov et al. (2021)	https://github.com/google/deepvariant
DeepSV	A deep learning tool to predict long deletions from sequence reads.	• Requires labeled data for training.	Cai et al. (2019)	https://github.com/CSupe rlei/DeepSV/
Target Identification and Validation				
DeepDigest	A deep learning tool to predict the cleavage sites of protease enzymes.	• Limited to only 8 protease enzymes. • Limited to predicting the cleavage sites of proteins in their native state.	Yang et al. (2021a)	https://bio.tools/deepdigest
BANDIT	A Bayesian approach for putative target identification from small molecule input.	• Does not predict the off-target effects.	Madhukar et al. (2019)	Not available publicly.
deepDTnet	A deep learning model for target prediction.	• Literature-dependent and could introduce errors in prediction because of the incompleteness of biomedical networks.	Zeng et al. (2020)	https://github.com/ChengF-Lab/deepDTnet
CryoDRGN	A deep neural network to reconstruct heterogeneous cryo-EM structures.	• Users must decide on the dimensionality of the latent space, which can influence the quality of the results. • Computationally intensive.	Zhong et al. (2021)	https://github.com/zhonge/cryodrgn
AlphaFold	A deep learning algorithm to predict 3D structures of proteins.	• Fails to predict disordered regions. • Predicts only one conformer. • Limited ability to detect protein folding defects caused by point mutations.	Jumper et al. (2021)	https://www.deepmind.com/open-source/alphafold

Table 2 (continued)

AI Tool /Platform	Application	Limitations	Reference	URL to the application or source code
PandaOmics	A deep learning platform to predict druggable targets.	• Not publicly available.	Pun et al. (2022)	It is proprietary software. https://insilico.com/pandaomics
Lead Identification and Optimization				
DeepDocking	A deep learning tool to speed up molecular-docking-based virtual screening.	• Time consuming.	Gentile et al. (2020)	https://github.com/vibudh2209/D2
DeepTox	A deep learning algorithm to predict the toxicity of chemical compounds.	• Computationally intensive for large data sets.	Mayr et al. (2016)	Not available publicly.
DeepAffinity	A deep learning tool to predict the protein-ligand binding affinities.	• Needs improvements to include 2-D graph representation for compounds rather than SMILES.	Karimi et al. (2019)	https://github.com/Shen-Lab/DeepAffinity
Chemputer	A chemical computer system based on ML for the automated chemical synthesis of small molecules.	• Requires expertise in chemical description language to program the tool.	Steiner et al. (2019)	Not available publicly.
Chemistry42	A deep learning software platform for small molecule design and optimization.	• Not publicly available.	Ivanenkov et al. (2023)	It is proprietary software. https://insilico.com/chemistry42
Hit Dexter	A binary classification tool to predict the likelihood of small molecules to provide a positive response in bioassays.	• Time consuming when the input size is larger (~ 1000 molecules take 4 h).	Stork et al. (2019)	http://hitdexter2.zbh.uni-hamburg.de
Clinical Trials				
AiCure	A deep learning neural network platform for patient recruitment and management of clinical trials.	• Requires a smartphone. • Data export to external platforms is not supported.	Salcedo et al. (2021)	It is proprietary software. https://aicure.com/

took only 12 months, from start to pre-clinical studies, compared to 4–5 years in the classical drug discovery process. The compound is developed by a British pharmaceutical company, called ExScientia, in collaboration with Sumitomo Dainippon Pharma in Japan (Burki 2020), more details are discussed in Sect. 3.4.

The present review first discusses how AI techniques can assist in disease identification, clinical diagnosis, genome analysis, and precision medicine, with a focus on diseases that have been extensively explored in AI studies such as infectious diseases, lifestyle-disorders, neurodegenerative disorders, and cancer (Sect. 2). Section 3 highlights the application of AI techniques in target and lead identification, followed by examples of AI-enhanced clinical trials in Sect. 4. At the end of each section, we present a critical review on the AI-methods discussed, mainly the advantages and disadvantages. Section 5 discusses the challenges in implementing AI in drug discovery and its future perspectives.

2 AI in disease identification and clinical diagnosis

Laboratory investigations and clinical examinations are the most common methods used in clinical diagnosis, which is a the fundamental step in providing high-quality treatments. The remarkable ability of AI techniques in Clinical Diagnosis Decision Support (CDDS) has acquired a significant interest in medical research in recent years. The incorporation of AI in clinical workflows provides abundant opportunities to reduce clinical errors, improve treatment outcomes, lower treatment costs, detect diseases at
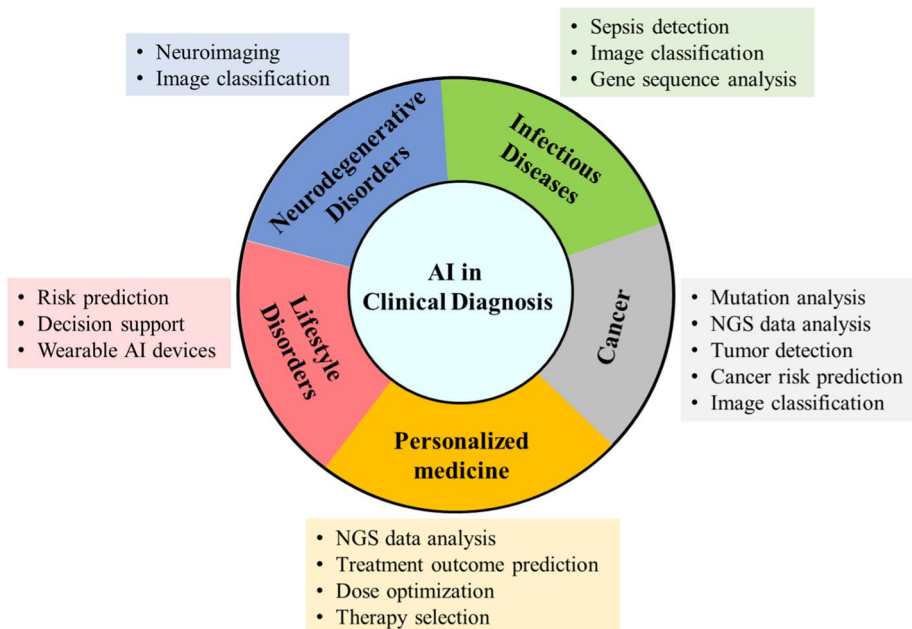
Fig. 3 Applications of AI in clinical diagnosis of various diseases

earlier stages, and track treatment progress over time. In this section, we will elaborate on the recent studies that have reported the use of AI technology for clinical disease diagnosis. Furthermore, we will highlight the applications of AI in genome analysis and personalized medicine (see Fig. 3).

2.1 Diagnosis of diseases using AI

AI is revolutionizing the way healthcare professionals identify, manage, and control diseases. AI algorithms can rapidly analyze large datasets of clinical symptoms and laboratory test results to detect diseases at early stages. This early detection allows for timely interventions and containment measures to prevent further spread. This section focuses on the recent AI-facilitated advancements in the diagnosis of both communicable diseases (e.g., infectious diseases such as sepsis, coronavirus disease, urinary tract infections, and bloodstream infections) and non-communicable diseases (e.g., lifestyle disorders such as diabetes, neurodegenerative disorders like Alzheimer's disease, and cancer).

AI-powered diagnostic tools exhibit high accuracy and sensitivity in identifying infectious agents, thereby reducing the chances of misdiagnosis and unnecessary treatments. This leads to better patient outcomes. During the pandemic period of the Coronavirus disease 2019 (COVID-19), there was a particular focus on the development of AI models for its effective diagnosis. Given its availability and low cost, chest X-ray was one of the efficient indirect methods used for COVID-19 diagnosis (Castro et al. 2021). Many ML models have been developed to predict the presence or absence of particular patterns in X-ray radiographs. Panwar et al. reported a supervised deep learning model called 'cornet' as a diagnostic method for COVID-19 (Panwar et al. 2020). The model accepts chest X-ray images as an input and completes analyses for any visual indications such as the hazy or shadowy patches on the lungs. The cornet model was shown to have an accuracy of ~ 97% in identifying COVID-19 patients. Further, Narin et al. proposed an automated CNN based diagnostic model for detecting pneumonia caused by coronavirus (Narin et al. 2021). They developed pre-trained AI models using the X-ray radiographs of healthy individuals, patients with COVID-19, patients with viral pneumonia, and patients with bacterial pneumonia. The reported accuracy in classification reached up to 96%.

In addition to COVID-19, ML models have been built to assist in the diagnosis of other infectious diseases such as urinary tract infections (UTIs), which are often associated with diagnostic errors. Taylor et al. reported a retrospective cohort analysis of approximately 80,387 adults who visited the emergency department with UTI symptoms. Considering symptoms as well as blood and urine sample analyses, six AI algorithms were developed for the diagnosis of UTI: SVM, ANN, elastic net, adaptive boosting, RF, and Extreme Gradient Boosting (XGBoost). The models were built using a full set of 211 factors and a reduced set of 10 variables, e.g., gender, epithelial cells in the urine, history of UTI, and age. The XGBoost algorithm outperformed others in accuracy, with an area under the receiver operating curve (AUROC) of 0.88 and 0.90 for the full and reduced XGBoost models, respectively. The sensitivity and specificity were 61.70 and 94.90 for the full, and 54.70 and 94.70 for the reduced models, respectively (Taylor et al. 2018).

The diagnosis of bloodstream infections, BSI, is yet another example that has benefited from AI technology. Bloodstream infections cause high morbidity and

mortality rates (15-30%) (Verway et al. 2022). However, predictions of the BSI treatment outcomes help in optimizing treatments and, therefore, reducing further complications of the infection. Zoabi et al. reported ML models that use electronic medical records to predict the treatment outcome in BSI patients. The dataset for the study involved medical reports with information on demographics, laboratory results, diagnoses, and medical history of adult patients hospitalized with positive bacterial blood cultures over a six-year period. Predictions from different gradient-boosting architectures were made with the help of decision-tree. The best model has an AUROC of 0.82. Notably, this model outperformed the standard Charlson Comorbidity Index scoring system (with smaller AUROC values of 0.585–0.648) for mortality prediction. This model outperformed other existing models used for similar applications with AUROC of 0.67 (Zhang et al. 2023) and 0.76 ± 0.04 (McFadden et al. 2023), respectively. A major limitation of this study is that it is based on retrospective electronic medical record data, which inherently carry biases (Zoabi et al. 2021). AI has also significantly assisted in the early detection of a life-threatening condition called sepsis, where the body develops an extreme immune response towards infections. ML algorithms have been developed to analyze vast amounts of patient data, including vital signs, laboratory results, and clinical notes in order to identify subtle patterns and changes indicative of sepsis onset, alerting healthcare providers in real-time, and enabling timely interventions. Sepsis Watch is a deep learningbased CDSS for the diagnosis of sepsis. The platform was trained with 50,000 patient records involving over 32 million data points, and it was proven to improve sepsis patient care (Sendak et al. 2020). However, this study was limited to (i) some false positive predictions where the clinical action is prompted for patients who do not ultimately develop sepsis, and (ii) emergency department cases.

Lifestyle disorders such as diabetes, obesity, and hypertension, are associated with the way people live, i.e., their diet, levels of exercise, etc. Many AI-based algorithms have been developed for the early prediction and management of diabetes. Recently, Spänig et al. developed an interactive AI model with the capability of speech recognition and speech synthesis. This model acts as a virtual doctor, it interacts directly with the patients and is able to predict Type-2 diabetes mellitus. This innovative approach involves a virtual doctor cabin equipped with various patient metric-gathering devices to measure weight, height, body mass index etc. An embedded AI system utilizes these metrics to assess potential health issues. The AI then recommends diagnostic steps to healthcare providers, in the context of diabetes. The model recommends whether the patient should or should not perform an HbA1c blood test, which is a long-term blood glucose level indicator based on the glycation of hemoglobin in the blood. To gather additional information about the patients, the system employs speech recognition, interviews, and questions about lifestyle to assess risk factors for developing diabetes. An automated speech recognition system called 'CMUSphinx' is used. The CMUSphinx system converts the spoken language into text with an AUROC of 0.84 (Spänig et al. 2019). Diabetic patients are susceptible to retinopathy (Al-Maskari and El-Sadig 2007) which is generally diagnosed by visual examination of retinal images. Untreated diabetic retinopathy can lead to severe complications including vision loss. Gulshan et al. developed a deep CNN model that bypasses the human capacity at interpreting, evaluating, and classifying retinal images. The model is trained using 128,175 retinal photographs which are evaluated by a panel of clinicians and ophthalmologists. The model is demonstrated to have a high sensitivity and specificity of 97.50% and 93.40%, respectively (Gulshan et al. 2016). In 2018, the U.S. Food and Drug Administration has approved the marketing of the first AI-based medical device called IDx-DR (Heijden et al. 2017) for detecting diabetic retinopathy. The device

has a retinal camera through which the retinal image of the patient is taken and analyzed. The device is autonomous and decides on one of the following results based on the image quality (i) 'more than mild diabetic retinopathy detected: refer to an eye care professional' or (ii) 'negative for more than mild diabetic retinopathy; rescreen in 12 months' (Heijden et al. 2017).

Pulmonary hypertension is a complex cardiovascular disorder characterized by increased pressure in the pulmonary arteries, leading to impaired blood flow to the lungs. Timely identification of hypertension is crucial for early intervention to prevent adverse outcomes. In 2023, Kıvrak et al. reported a classification model to identify pulmonary hypertension from the chest X-ray images (Kıvrak et al. 2023). The model is trained using chest X-ray images of patients with different types of pulmonary hypertension and healthy people (without pulmonary hypertension). Their model was able to attain an accuracy of 86.14% in identifying different types of pulmonary hypertension. The AUROC is calculated to be 0.945. However, the model needs to be improved as it yielded constrained performance outcomes in some patient groups.

Alzheimer's disease (AD) is a neurodegenerative disorder in the brain. The lack of a widely accessible and low-cost screening method for AD can be attributed, in part, to the complexity of its diagnosis. AD diagnosis often relies on invasive tests typically limited to specialized clinical settings. One of the advances in imaging technology, namely fluorine-18-fluorodeoxyglucose positron emission tomography (PET) of the brain, facilitated the early detection of AD. However, the challenge lies in the interpretation of the PET data. Ding et al. developed a deep learning algorithm that interprets PET of the brain for the early prediction of AD. Their model showed a specificity and sensitivity of 82% and 100%, respectively. This model can predict AD, on average, 75.8 months before its diagnosis, with an AUROC of 0.98 (Ding et al. 2019). Recently, Agbavor and Liang developed an end-to-end AI-powered system for the detection of AD as well as to predict the severity of the disease from raw voice recordings (Agbavor and Liang 2022). The dataset used to build the AI model includes a collection of speech recordings where individuals, both cognitively normal individuals and AD patients, provide descriptions of certain pictures. The AI model uses a pre-trained data2vec model, which is a self-supervised algorithm that can work directly with speech data, without the need for human-designed features or manual interventions. The AI model developed in this study is considered 'end-to-end' because it encompasses the entire process, starting from the analysis of raw voice recordings and concluding with AD predictions. This approach eliminates the requirement for distinct manual steps involving feature extraction or pre-processing, as the AI system manages these tasks within a unified framework. In a nutshell, the model directly takes in raw voice data and autonomously processes it to deliver AD-related predictions. The model is tested using data from 'DementiaBank' and it predicted AD with an AUROC of ~0.84. This model can be used as an alternative low-cost diagnostic method for early detection of AD. Also, integrating this model into AD clinical trials can substantially curb the cost and duration of clinical trials which in turn speeds up the drug development process.

Cancer diagnosis and prognosis have highly benefited from the advancements in AI (Tanoli et al. 2021). The key diagnostic methods for cancer are the clinical imaging techniques (Fass 2008) such as X-ray, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI). AI has the potential to improve the speed of analysis and the accuracy of image interpretations. 'AI Dermatologist' (https://ai-derm.com/) is a web-platform based on deep learning to predict skin cancer from photographs. The tool can identify skin cancer from the image uploaded by the user. It can even classify benign and malignant tumors based on asymmetry, boundary, color, diameter, and change over

time. The AI Dermatologist platform is built using deep learning algorithm by training a neural network on a vast database of dermoscopic images assessed by dermatologists. The AI dermatologist was able to achieve 87% sensitivity in picking up cancerous cells from body scans (Longoni et al. 2019). Esteva and co-workers (Esteva et al. 2017), developed a CNN model trained with images of skin lesions to classify different types of skin cancer. Typically, the initial diagnosis of skin cancer is by microscopic examinations of the tissues. However, skin lesions are highly variable from one skin disease to another, making it challenging to have accurate diagnoses. In their study, they have trained their model with a dataset of *ca.* 129,450 images of 2,032 different skin conditions from Stanford University Medical Center and other open-access public repositories. The model was validated in two different ways using three-class and nine-class disease partition. In the three-class disease partition, the CNN showed an overall accuracy of $72.10 \pm 0.9\%$, while the dermatologists obtained ~66.0% accuracy. In the nine-class disease partition, the model's and the dermatologists' accuracies were comparable, $55.40 \pm 1.7\%$ and $53.3 \pm 5.50\%$, respectively. This model is an example of a low-cost diagnostic tool that can be extended to analogous models for other specialties.

In another study, Causey et al. developed an algorithm called NoduleX (see Table 2) for the prediction of malignant lung nodules from clinical CT data. The algorithm is based on a deep-learning CNN model. The authors used over 1,000 images of lung nodules from the Lung Image Database Consortium (LIDC) and the Image Database Resource Initiative (IDRI) cohort for training the model. NoduleX showed high-accuracy predictions with an AUROC of 0.99 (Causey et al. 2018). The tool is still under development to find the best model architectures for analyzing different patterns and features from radiological images. Another future aim of this study is to construct high-quality datasets for training, testing, and validation. Shiri et al. has evaluated the efficiency of different ML approaches developed for predicting the mutation status in the Epidermal Growth Factor Receptor (EGFR) and Kirsten rat sarcoma viral oncogene (KRAS) in Non-Small Cell Lung Cancer (NSCLC) patients. These approaches are based on radiomics analyses, using features extracted from around 150 images from low-dose CT, contrast-enhanced diagnostic quality CT (CTD), and PET imaging techniques. They highlighted multivariate ML-based AUROCs of 0.82 and 0.83 for the EGFR and KRAS mutations, respectively. The primary constraint of this study lies in the relatively limited size of the dataset employed for training and validation purposes (Shiri et al. 2020).

Histological analysis of tissue samples is another method for cancer diagnosis. Histopathologists visually examine tissue samples under the microscope to check for any irregularities in the shape of the cell, tissue distribution, or necrosis. Deep learning techniques that involve CNN models fueled the histological image analysis (Öztürk and Akdemir 2019; Hameed et al. 2020; Srinidhi et al. 2021). Sharma et al. reported a deep CNN model for the classification of gastric carcinoma from images of histopathological samples stained with hematoxylin and eosin. The model is developed for (i) cancer classification using immunohistochemical responses and (ii) necrosis detection in tissues. This model showed an accuracy of 0.699 for cancer classification, and 0.81 for necrosis detection. One of the disadvantages of this preliminary model is the limited size (454 cases) of the data set used. Further, in the proposed CNN configuration, the training takes approximately two days, even with GPU implementations (Sharma et al. 2017). Recently in 2023, Tolkach et al. introduced an AI algorithm designed for tumor tissue detection and tumor regression grading in surgical specimens. These are obtained from patients diagnosed with oesophageal adenocarcinoma or adenocarcinoma of the oesophagogastric junction (Tolkach et al. 2023). The performance of the model is validated on a set of

histopathological slides. During the validation process, the AI tool demonstrated a 63.6% agreement with the analyses performed by a group of twelve pathologists at the case level. Notably, the AI-based regression grading was able to detect small tumor regions initially missed by pathologists. Moreover, AI helped significantly reduce the time required for the diagnosis per case, from 6 min to 1 min. These findings highlight the potential of the AI algorithm to enhance diagnostic accuracy, optimize the evaluation of tumor regression, and improve the efficiency of pathological analyses.

The key to obtaining accurate predictions lies in selecting appropriate models trained on specific data for a given disease in a particular population. In our opinion, algorithmic bias is one of the critical challenges associated with building ML models because it can result in incorrect or unfair diagnoses. It is also important to ensure that the training datasets used include diversity based on race, age, gender, etc. In addition, an efficient human-AI synergy can lead to more reliable decision-making support from AI. This synergy optimizes the benefits by combining human expertise with AI capabilities, such as the natural language processing capability of ML and deep learning methods.

2.2 AI in genome analysis

Around 80% of rare diseases are related to genetic variations (Liu et al. 2019). Hence, the importance of diagnoses provided by genome sequencing. The advancements in Next Generation Sequencing (NGS) technology have led to the collection of vast amounts of data and provided rich information about individual genomes. The bottleneck in NGS lies in the analysis and interpretation of large-scale genome data and the identification of variants (Lucena-Perez et al. 2021). This can take days to weeks. AI-based models, such as deep learning models, opened a new chapter of research related to transforming this 'big data' into meaningful new information. AI technology has been applied in many areas of genomic analysis such as gene annotation, genotype-phenotype correlations, consanguinity diseases, mutation studies, cancer diagnosis, biomarker identification, gene function prediction, and variant calling.

ML models have surpassed the conventional bioinformatics tools for the sequence analysis and identification of variations such as insertions, deletions, or mutations within genomic sequences. Cai et al. developed an ML tool called Concod to detect deletions in DNA. Concod outperformed four existing deletion detection tools (Pindel, SVseq2, BreakDancer, and DELLY) in sensitivity and precision (Cai et al. 2017). However, this tool was limited to identifying only short structural variations in the sequences. Then a visualization-based ML model, DeepSV, was developed (Cai et al. 2019). DeepSV is based on deep learning and is used for identifying long deletions within the genomic sequence. The tool is optimized to work with noisy training data. However, like many other supervised machine learning techniques, DeepSV requires properly labeled training data for its training process (Cai et al. 2019).

DeepVariant and DeepTrio are two AI tools developed by Google for the prediction of genomic variants from the NGS sequence data (DePristo and Poplin 2017; Kolesnikov et al. 2021). DeepVariant is an open-source tool that was released in 2017. This deep learning model is based on CNN and is trained on using images of the sequence reads that are produced from reference genomes. The raw data from sequencing platforms (e.g., Illumina sequencing or polymerase chain reaction sequencing) consists of many reads of overlapping gene fragments. Raw sequence data is mapped to a reference genome and then analyzed by DeepVariant to identify the locations of variations such

as single nucleotide polymorphisms (SNPs) and short insertions and deletions (indels) (DePristo and Poplin 2017). The newer versions of this tool can accept the raw sequence data from Illumina or PacBio sequencing and Oxford Nanopore (DePristo and Poplin 2017; Carroll 2020). One of the major limitations of DeepVariant is that it is well optimized to germline variants. So, it may not be as suitable for somatic variant calling in cancer genomics. DeepTrio further expanded the functionality of DeepVariant. It can predict the genomic variants in duos or trios, meaning that DeepTrio can be used to analyze child-mother-father (trio) or child-father/mother (duo) sequence data. The tool can provide a better understanding of Mendelian diseases and the transmission of genetic traits. DeepTrio specializes in pinpointing novel mutations, which are genetic changes present in the genome of a child but not the genome in either parent (Kolesnikov et al. 2021). Both these tools are excellent in variant calling; however, they have certain limitations. First is the requirement of intensive computational resources. Second, just like any other ML model, the accuracy of the prediction depends on the quality and diversity of the training data used to train these tools. Third, DeepVariant and DeepTrio can be complex to set up and configure, they require expertise in both genomics and machine learning. Last, these models are not self-contained, additional tools and expertise are required to interpret the complex variant calling results they provide.

Another challenge in the analysis of genome sequences is to distinguish benign from disease-causing gene variants for rare genetic disorders (Benowitz 2014). In a collaborative retrospective study between the company Fabric Genomics and Rady Children's Institute, San Diego, researchers built an automated AI algorithm called Fabric GEM (where GEM stands for genomics). They used 179 diagnosed pediatric cases, mostly from the Neonatal intensive care unit (NICU) at Rady Children's Institute, and five other clinics across the world (Vega et al. 2021). Through analyzing NGS data, GEM can swiftly and accurately identify, in 90% of cases, the structural genes responsible for rare genetic disorders. This outperforms the existing variant-calling tools, which correctly identify the structural variants less than 60% of the time (Vega et al. 2021). Fabric GEM utilizes advanced AI technology and integrates genetic, phenotypic, and clinical data to efficiently identify the most likely genetic causes of a medical condition. Unlike many other interpretation platforms that often require a thorough review of 20 to 50 potential genetic variants to pinpoint the causal one, Fabric GEM excels at prioritizing these variants. As a result, it significantly reduces the number of genetic causes that need to be reviewed for a medical condition to fewer than five. This enhanced efficiency streamlines the clinical review process (https://fabricgenomics.com/fabric-gem/). GEM also has the advantage of accurate predictions at low costs.

The need for high-quality data is one of the challenges in AI-assisted genome analysis. The genomic data is often complex and incomplete. Thus, it is important to properly clean data before using it for training the models. Overfitting is also common within AI models used in genome analysis due to the limited availability of data. This is further restricted by confidentiality issues. Nevertheless, it is recommended to have even more stringent rules and regulations to protect the data of patients. Another disadvantage of using AI in genome analysis is that the model needs to be reoptimized based on the population under study. As genome data grows exponentially, it becomes increasingly challenging for algorithms to scale and perform analyses on large datasets within a reasonable time frame. Also, genomes exhibit extensive structural and functional variability. Developing algorithms that can accommodate this variability and

provide robust results is a challenge. Moreover, bridging the gap between raw genomic data and existing biological knowledge databases is a complex process, as it requires advanced natural language processing and semantic integration techniques.

2.3 AI in personalized medicine

Traditionally, clinical practice has been based on the concept of 'one therapy fits all'. However, drug molecules may undergo different metabolic activities in different patients. For example, a drug that works well for a group of people may not be as effective, or may have adverse side effects, for others. These differences in drug metabolism are mostly attributed to the differences in the genetic profile of individuals. Thus, a more futuristic approach is a personalized treatment, also known as precision medicine, where patients are treated based on their genetic profile. The aim is to maximize treatment outcomes while minimizing adverse effects per individual. Thus, different therapies and doses are customized per individual (or per group of patients that share similar genome profiles). AI has fostered considerable improvements in the development of personalized medicine (Boniolo et al. 2021). For example, the AI-derived platform, CURATE.AI, predicts the optimal dosing along with the treatment outcomes based on the individual data of patients. It generates a profile for each patient, using their own medical records, and it dynamically recalibrates the predicted profile over time based on the progression or recession of the disease. CURATE.AI can optimize not only doses of single drugs, but also combinations of drugs (Pantuck et al. 2018; Blasiak et al. 2020). This is helpful given that, nowadays, therapies are becoming more sophisticated with emphasis on combination (or multimodal) treatments. These involve more than one drug or treatment offered either simultaneously or sequentially. Combination therapy is proven to have more efficacy compared to single-drug regimens, especially in the treatment of complex diseases like cancer (Kumar et al. 2005; MacDonald et al. 2017). The limitation of CURATE.AI is that its current version is not integrated with standard electronic medical record systems, which may limit its seamless adoption in healthcare facilities. This lack of integration could lead to challenges in effectively incorporating CURATE.AI into existing medical workflows and systems, potentially making it less accessible or efficient for healthcare professionals (Mukhopadhyay et al. 2022).

Since the efficiency, efficacy, and potency of a drug may vary among individuals, predicting the response of a patient to medications prior to the treatment can assist doctors in selecting the optimal treatment strategy. AI has remarkable applications in this area. To predict the efficacy of a chosen treatment, Kureshi et al. developed an AI decision tree to establish a link between the characteristics of the patient and the tumor response in NSCLC (Kureshi et al. 2016). They used four classifiers (histology, mutation in epidermal growth factor receptor, targeted drugs, and smoking habits) for predicting the response of NSCLC patients to the EGFR tyrosine kinase inhibitors. The method showed an accuracy of 76.6%, and it can support clinicians in choosing the appropriate treatments. One of the drawbacks of this study is the small training set used ($n = 355$), and therefore, the omission of rare patterns such as duplication, deletions, insertions, and point mutations. Using a larger training set could further improve the predictive accuracy of this decision support model. Huang et al. developed an SVM algorithm to predict the response of cancer patients to chemotherapy based on their gene expression profiles. The accuracy levels of this model exceeded 80% (Huang et al. 2018). The 'IBM Watson for oncology' software was designed with the objective of

making a large impact on personalized treatment plans for cancer patients (Fu et al. 2015). The software was trained on thousands of clinical and health records of cancer patients, from medical journals, textbooks, and literature curated by the Memorial Sloan Kettering (MSK) cancer center. This software was determined to make accurate diagnoses and treatment recommendations by identifying related cases from databases of worldwide clinical trials (http://www.clinicaltrials.gov) (Bach et al. 2013). However, a potential disadvantage of this software is its 'bias' towards cancer treatments adopted at the MSK cancer center, possibly resulting in inappropriate recommendations for patients treated elsewhere. Notwithstanding its language processing capabilities, which allowed it to extract insights from unstructured data like clinical notes and summaries, Watson fell short in terms of interpreting data at a level comparable to human doctors. A critical evaluation conducted in 2017, by the news website STAT, revealed that the platform recommended unsafe cancer treatments.

Recently, Sun and Chen reported an interpretable neural network based on deep learning to predict the survival chances of cancer patients based on drug prescriptions and personal transcriptomes (Sun and Chen 2023). The correlation between the predicted and actual months-to-death values is calculated to be 0.937, and the accuracy in classifying long-lived and short-lived cancer patients was 96%. AI has found its way into precision medicine for a wide range of diseases beyond cancer. For instance, Ferrè et al. implemented ML-based methods to identify a genetic signature in the genome of multiple sclerosis patients. They used clinical data along with demographic characteristics to predict the response of patients to a drug called Fingolimod. Using supervised ML methods such as RF, they identified 123 SNPs responsible for the response of patients to this drug. The drug response prediction improved from an AUROC of 0.65 in a model trained exclusively with genetic data to an AUROC of 0.71 in another model trained with both clinical and genetic data. However, the study used a dataset of only 77 patients, which is too small to represent the complexity of genetic data (Ferrè et al. 2023).

AI and ML-based methods have significant potential to revolutionize personalized medicine. However, it is our belief that the concept of personalized medicine is still far from being fully implemented. This is because the concept of personalized medicine suggests an individualized approach to treatment, yet the current implementation often involves treating people in groups based on similarities in their genetic profiles. It is also worth noting that personalized treatment is expected to severely increase treatment expenses. The non-private healthcare sectors may not be ready to accommodate such costs, as they may already be facing financial limitations. To reduce such costs, we suggest improving the data-sharing systems to avoid redundancy in expensive tests while ensuring the protection of data privacy and confidentiality. Additionally, we recommend using AI technology to automate as many steps as possible in the process of personalized medicine. Lastly, ethical considerations, such as the potential for genetic discrimination or warranting the accessibility to personalized treatments for all patients, irrespective of their socio-economic status, must be addressed.

3 AI in target and lead identification

3.1 Target identification

Target identification is about identifying key molecular druggable targets, proteins, or nucleic acids, associated with a given disease. It allows researchers to benefit from drug repurposing and to develop drugs with more successful treatments and improved efficacy. Since many diseases are associated with the upregulation or downregulation of certain proteins, it is important to correctly identify the protein responsible for causing a disease during drug development

3.1.1 AI in target prediction

In 2020, the AI-driven biotech company, Insilico Medicine, launched its AI-powered target discovery platform called 'PandaOmics' (Pun et al. 2022). PandaOmics is an AI platform that searches for new therapeutic targets while significantly reducing the investment of time and resources. This deep learning-based platform has been trained on an extensive wealth of data from 3.8 million patents, 3 million grants, 30 million scientific publications, 1.3 million molecules, 342 thousand clinical trial information, and 5 million omics data. PandaOmics algorithms complete a comprehensive analysis of this vast data, to predict promising new targets and rank them based on multiple critical factors such as novelty, biological relevance, commercial potential, druggability, and safety. This platform is also trained to predict the likelihood of a potential target entering Phase I of clinical trials for various diseases in the upcoming five years. Additionally, it estimates the probability of a successful transition through subsequent trial phases (Pun et al. 2022; Olsen et al. 2023). PandaOmics uses a method called ComBat to reduce batch effects in data analysis. Batch effects are systematic variations in data caused by technical factors such as different experimental conditions or instruments. The ComBat method is employed to minimize these batch effects to improve data quality and accuracy. However, there are certain limitations associated with this batch correction method. First, ComBat is effective only with specific data types such as transcriptomics data, which includes data generated from technologies like microarrays and RNA sequencing (RNAseq). Secondly, there should be at least one dataset that includes both case and control groups within the same dataset, or else ComBat will not be applicable. By implementing PandaOmics in the drug discovery process, the Insilico Medicine company has already showcased successful instances and achieved advancements up to preclinical studies within a period of ~ 18 months (Pun et al. 2022; Ren et al. 2023). For detailed discussion on examples of successfully developed drugs using the PandaOmics platform, please refer to Sect. 3.4 which discusses the treatment of idiopathic pulmonary fibrosis and hepatocellular carcinoma.

Another model used to aid the target identification process is the deepDTnet. It is a network-based deep learning model developed by Zeng et al. (Zeng et al. 2020). The model was trained with chemical, genomic, and cellular network data for the accurate prediction of molecular targets. However, literature dependence and incompleteness of biomedical networks could introduce errors in prediction. The model is shown to have a high accuracy in predicting novel targets for the existing drugs, with an AUROC of 0.96. In addition, using deepDTnet, a drug has been repurposed for the treatment of multiple sclerosis, and it was later found to be effective in the in vivo MS models.

Using AI in target identification, Madhukar et al. developed a Bayesian ML platform, BANDIT, which is capable of predicting drug-binding targets (Madhukar et al. 2019). BANDIT was tested on more than 2000 small molecules and had a prediction accuracy of *ca.* 90%, although it is not able to identify potential drug targets for diseases or conditions that are not well-studied. BANDIT also made novel predictions that were experimentally confirmed through bioassays. This tool, among many others, opens advances in the field of drug repurposing.

In another study, Mamoshina et al. developed ML techniques to analyze human muscle transcriptomic data to discover biomarkers associated with age-related diseases and to identify tissue-specific drug targets (Mamoshina et al. 2018). They developed an AI-assisted approach to monitor age-dependent changes in the human skeletal muscle. The authors constructed a set of tissue-specific biomarkers for aging and used a combination of unsupervised and supervised ML algorithms to identify differentially expressed genes and gene modules that are associated with muscular dystrophy and sarcopenia. The performance of the model was subsequently assessed using gene expression samples from skeletal muscles. Their best model showed a Pearson correlation of 0.80 when predicting the age bin on the external validation set.

In drug discovery, using information from biomedical literature is crucial. Microsoft recently introduced an AI tool, named 'BioGPT', for biomedical text generation and mining (Luo et al. 2022). It is a generative language model based on deep learning. BioGPT is pre-trained on a vast dataset comprising 15 million PubMed abstracts. This tool was tested for various biomedical natural language processing tasks, such as end-to-end relation extraction, question answering, document classification, and text generation. It demonstrated an accuracy of 81% on the question answering task on PubMedQA, a dataset developed to provide yes/no/maybe answers to research queries entered by the users based on the abstracts from PubMed. This surpasses the performance of a single human annotator (78%). BioGPT was used by Zagirova et al. in an application related to the prediction of molecular targets related to aging and age-related diseases (Zagirova et al. 2023). In addition to the 15 million PubMed abstracts used in the BioGPT tool, the authors further trained this tool with a dataset containing information from descriptions of biomedical grants involving target discovery. They identified two potential dual-purpose molecular targets for anti-aging and 14 age-related diseases.

3.1.2 AI in 3D structure prediction

The development of computational tools, high-performance computers, and ML algorithms enabled the generation of myriad drug discovery tools including, but not limited to, three-dimensional (3D) models of protein targets. This is a significant advancement over the experimental techniques that are fraught with challenges. For example, the X-ray diffraction technique is limited to crystallizable samples, which is a major experimental limitation. An alternative experimental technique for determining the structure of biological macromolecules is Cryogenic electron microscopy (Cryo-EM) (Murata and Wolf 2018). Cryo-EM involves producing thousands of two-dimensional (2D) images of frozen protein samples. Computer algorithms are then used to combine these images into a 3D structure representation, a process called 'reconstruction'. Zhong et al. developed a DNN-based software called CryoDRGN for the reconstruction of cryo-EM images using neural networks (Zhong et al. 2021). The software has the potential to reconstruct all the possible 3D conformation of a protein from its 2D cryo-EM images. It encodes 2D particle

images into a low-dimensional latent space, where heterogeneous structures are assumed to exist. The model is trained using stochastic gradient descent and can generate 3D density maps based on latent variables, allowing for the visualization of particle distribution and reconstruction of representative structures. The software can also visualize the movements of proteins. Its remarkable strength lies in its ability to represent a wide range of complex structures without making any restrictive assumptions about the nature of this complexity. One of the limitation is that the users must decide on the dimensionality of the latent space, which can influence the quality of the results (Kinman et al. 2022).

AI has helped advance accuracy and speed in predicting 3D structures of biomolecules, such as proteins, DNA, and RNA. Reinforcement learning has also been instrumental in refining 3D structure predictions and generating energetically stable and biologically relevant conformations (Lu 2022; Yang et al. 2023). DNN has shown abilities in learning complex patterns and representations from vast datasets. An underlying principle of deep learning-based 3D structure prediction is the data-driven learning (Andronico et al. 2011; Hoffmann et al. 2019). Such methods benefit from vast datasets of experimentally determined structures and sequences, to iteratively build relationships between amino acid sequences and their corresponding 3D structures; and ultimately make accurate and rapid predictions for uncharacterized biomolecules. An extensive project on protein structure prediction is DeepMind's AlphaFold (Jumper et al. 2021). AlphaFold is a deep learning tool that employs a two-step process: the fold recognition stage and the model refinement stage. In the fold recognition phase, the software searches for known protein structures by comparing the amino acid sequences of the target and template proteins. AlphaFold uses various tools to perform fold recognition, e.g., multiple sequence alignment (MSA) against structure databases. In the model refinement process, AlphaFold uses a neural network to refine the protein structure predictions by considering MSA, co-evolution, and geometric constraints. The MSA provides information about the evolutionary relationships between the new protein and the known proteins. Co-evolution provides information about the interactions between amino acids in the new protein. Geometric constraints provide information about the spatial arrangement of amino acids in the new protein. The development of this AI tool is substantial in drug discovery as it helped solve the structure of nearly 200 million proteins, that is ~98.5% of the proteins in the human body (Tunyasuvunakool et al. 2021). Together with the European Bioinformatics Institute (EMBL-EBI), a database called AlphaFold DB (https://alphafold.ebi.ac.uk/) is created to store all the structures solved so far with AlphaFold. However, the effect of mutation on the folding of proteins is beyond the capability of AlphaFold (Buel and Walters 2022). It is also limited to predicting only a single state of a given protein, it does not consider the dynamic nature of protein structures (Perrakis and Sixma 2021). Another limitation is that it does not predict other important aspects related to protein structures including co-factors, metal ions, ligands, etc. AlphaFold predictions does not account for post-translational modifications such as glycosylation or phosphorylation, as well as the presence of DNA, RNA, and their respective complexes (Bagdonas et al. 2021).

At the experimental level, mass information about protein fragments can help figure out the identity of a protein and its structure. Mass information can be obtained from Mass spectrometry (MS), which is an experimental technique used to characterize molecules including proteins (Loo et al. 1999). The digestion of proteins by protease enzymes like trypsin is a basic step in protein identification using MS. A few AI tools were developed to efficiently predict the digestion behavior of the protease enzymes (Yang et al. 2021a; Sun et al. 2021). DeepDigest is the first algorithm developed using a deep learning method to predict the proteolytic cleavage sites of eight different protease enzymes (Yang et al.

2021a). The predictive ability of the tool was evaluated by the AUROC, F1 scores, and the Matthews correlation coefficients (MCCs); the values were 0.956–0.98, 0.66–0.90, and 0.65–0.84, respectively. However, this tool is not suitable for predicting the proteolytic sites in modified proteins or peptides *via* glycosylation or phosphorylation.(Yang et al. 2021a).

3.1.3 AI in binding site prediction

Once the structure of the receptor (protein, DNA, etc.) is known, more can be done in order to better understand the properties of the target. For example, in 2021, Kozlovskii and Popov developed a deep learning approach to predict the binding site for small molecules on nucleic acids, DNA, and RNA, based on their 3D structures (Kozlovskii and Popov 2021). Their approach is called BiteNet$_N$ (https://sites.skoltech.ru/imolecule/tools/bitenet/) and it is the first 3D CNN to learn features directly from the nucleic acid structures. They validated the model using two different protein structures, HIV-1 transactivation response element RNA and ATP-aptamer structures. The model showed an AUROC of *ca.* 0.87.

In 2020, Simonovsky and Meyers proposed a CNN-based model called 'DeeplyTough' for pocket matching (Simonovsky and Meyers 2020). The model can convert the 3-D representation of a protein pocket into descriptor vectors. These vectors are then used for comparing ligand binding pockets on protein by calculating pairwise Euclidean distances. The prediction ability of the tool is evaluated using three benchmark datasets. The model had a reasonable performance with AUROC values above 0.83 for all three datasets. This model can be useful in drug repurposing.

In addition to 'pocket matching', AI algorithms can be used to find potential allosteric modulators that could bind to the protein and alter its structure and possibly its function. Tian et al. developed a webserver called PASSer (Prediction of Allosteric Sites Server) to predict allosteric sites in a given target. The webserver uses three ML models, (i) an ensemble learning model, (ii) an automated ML model, and (iii) a learning-to-rank model. PASSer makes remarkably rapid predictions, typically providing allosteric site results within seconds (Tian et al. 2021). The ensemble learning method involved both an XGBoost model and a graph-based CNN. The physical properties of the protein pockets are fed into the former model, and its atomic representation is fed into the latter. The model showed an accuracy of 0.97, a precision of 0.73, and a specificity of 0.98.

In another example, it is useful to predict cryptic pockets that are often involved in allosteric regulation and modulation of protein functions. These pockets are protein cavities that are not apparent from the surface of proteins but can open upon the binding of specific ligands or protein partners. Recently, Miller et al. developed a graph neural network called PocketMiner to predict the cryptic pockets within protein structures (Meller et al. 2023). The model is trained using the residues that are likely to form cryptic pockets identified from over 2,400 simulations of 35 different proteins. The model showed an AUROC value of 0.87.

The AI-driven models discussed in the target identification exhibit several similarities in their approach to drug design. They share a common foundation of data-driven learning, making extensive use of diverse datasets to draw insights and predictions. Deep learning techniques, such as DNN and CNN, are prevalent in these models, allowing them to discern intricate patterns and relationships within the data. XGBoost is also used in a few AI models used for target identification, as discussed in this section. Another important algorithm used in target identification is based on reinforced learning (Tian et al. 2021). Protein structure prediction involves searching for the lowest energy state, where the

protein is most stable. Reinforced learning can help in navigating this energy landscape efficiently. The model can be trained to explore different conformations and refine them iteratively to approach the global energy minimum. This is particularly useful because the energy landscape for proteins is highly complex, with numerous local minima, and traditional optimization methods may get stuck in suboptimal solutions (Lutz et al. 2023). The relevance of these AI-driven models to the future of drug design is indisputable. They bring enhanced efficiency and speed to target identification, protein structure prediction, and drug repurposing, significantly expediting drug discovery. They can reach high precision and accuracy levels, offering a decent level of predictability. On the downside, they are heavily reliant on data, which may not always be comprehensive or readily available. For example, the protein-protein or protein-drug interaction maps are not completely available. These gaps in the data availability affect the performance of the AI models.

3.2 Lead identification

Lead identification involves the discovery of potential small molecules that can bind to the active site of identified targets. Computational virtual screening has made it possible to swiftly screen millions of compounds and identify a few potential molecules for experimental testing. Both structure-based virtual screening (SBVS), and ligand-based virtual screening (LBVS) can benefit from AI (Labbé et al. 2015; Carpenter et al. 2018). In SBVS, the 3D structure of the receptor (nucleic acid or protein) is utilized to screen molecules that can potentially bind to the active site. As mentioned previously, AI is helpful in predicting the 3D structure of the receptor in case it is unavailable or its experimentally determined structure is of poor quality. In addition, AI techniques are also used to enhance the efficiency of computer-aided drug discovery processes, which typically require intensive high-performance computing resources and significant computing hours. For example, Gentile et al. reported an open-source protocol for AI-enabled virtual screening methods to screen libraries with billions of molecules. They used a screening platform called Deep Docking (https://github.com/jamesgleave/DD_protocol) which can accelerate structure-based virtual screenings by 100 folds. The method performs molecular docking for a small subset of a large library, followed by ligand-based prediction of the docking for the rest of the library. A key advantage of this protocol is that it can be used in conjunction with other docking programs such as Glide, Autodock-GPU, and FRED from OpenEye. Although the deep docking method provides faster screening, it is limited to (i) the availability of graphical processing units (GPU) and (ii) the quality and accuracy of the docking program used (Gentile et al. 2022).

In 2021, Yang et al. reported a protocol for hit identification by implementing active learning in the conventional docking protocol. This efficiently scales up the screening process for ultra-large compound libraries (Yang et al. 2021b). First, a small subset of compounds is docked, then these results are used to train the ML model to predict docking scores that are then validated through molecular docking. This data is further incorporated into the ML model for a continued iterative process until the model converges. The authors have tested this protocol to virtually screen a large molecular library against D4, MT1, and AMPC targets. They achieved a notable retrieval rate of over 80% for experimentally validated hits while significantly reducing computational expenses by 14 fold.

LBVS is based on selecting, from databases, molecules that share similar structural features with an active ligand. Pharmacophore-based virtual screening is one of the LBVS techniques. It involves building 2D fingerprints of one or more active ligands using molecular descriptors such as hydrogen-bond donors, hydrogen-bond acceptors, and aromatic rings. These 2D fingerprints are then used to identify molecules, from large chemical libraries, which have matching pharmacophoric features. ML also helps to study the correlation between molecular descriptors [or even atomic descriptors (Matta and Arabi 2011; Osman and Arabi 2022)] and the biological activity of a ligand. This is a broad category of research known as Quantitative Structure-Activity Relationship (QSAR), where the activity of a ligand depends on its pharmacophoric features. Melge et al. developed hybrid inhibitors using the pharmacophore fingerprint of two well-known anti-cancer drugs, Ponatinib and Vorinostat (Melge et al. 2022). They developed a supervised ML approach for 2D-QSAR and 3D-pharmacophore studies to predict the inhibitory activity of novel hybrid molecules. The model had AUROC values of 0.98 and 0.94 for the two different cancer targets, BCR-ABL and Histone deacetylase (HDAC), respectively. Based on in vitro evaluations, the identified novel hybrid molecules showed the potential to develop into lead compounds. Dhamodharan et al. developed three AI models based on genetic function approximation (GFA), SVM, and ANN, to predict the activity of acetylcholinesterase (AChE) and Beta-Secretase 1 (BACE1) dual inhibitors for AD treatment (Dhamodharan and Mohan 2021). The predictive power of the models was evaluated on a test set of 11 inhibitors of AChE and BACE1. The ANN model had the best predictive power with r^2 values of 0.85 and 0.78 for AChE and BACE1, respectively. However, this study is limited to a smaller number of molecules in the dataset used to train and validate the model.

Chemistry42 is an AI-based software platform for the *de novo* designing and optimization of small molecules (Ivanenkov et al. 2023). Since its launch in 2020, Chemistry42 has been utilized by more than 20 pharmaceutical companies. In the first step of the process, users have to upload their data onto the platform. The input data can be the structure of a small molecule, the structure or name of the molecular target, or their chemical properties. The second step, called the generation phase, involves running the platform with many generative models operating in parallel to create new structures. These new structures pass through various filters. Then, in the third step, the molecular structures are evaluated using multiple sets of reward and scoring modules, where the properties of the generated structures based on predefined criteria are evaluated. These modules serve as the cornerstone of Chemistry42's generation protocol based on multiagent reinforcement learning. In the learning phase, the scores of the generated structures are used as feedback to the generative models, reinforcing and guiding the generative process toward producing high-scoring structures. The final step involves ranking the generated structures based on their predicted properties, such as synthetic accessibility, drug likeliness, shape similarity novelty, diversity, and more. The Chemistry42 provides a user-friendly interface and can be easily integrated into other software or platforms (Ivanenkov et al. 2023). Refer to Sect. 3.4 for more details on the successful examples of drugs developed using Chemistry42.

Generative models present a promising approach to small molecule generation, which is key in lead identification. They address the challenge of determining the set of molecules that satisfy a desired set of properties. Generative models are trained to identify the underlying patterns and structures within the training dataset, in order to generate new instances that share similar characteristics with the molecules in the training data. A type of generative model called diffusion model is used by Hoogeboom et al. for generating 3D structures of small molecules from noisy SMILES or structural data (Hoogeboom et al.

2022). This is the first diffusion model developed for predicting small molecules in 3D. In general, diffusion models work by introducing a chain of progressive noising steps, called a diffusion process, where random Gaussian noise is added to the real data until the original sample is unrecognizable. Then, a model is trained in such a way that it can denoise the data. In this study, the authors trained the model with pairs of noisy and clean molecule representations so that the model learns the relationship between noisy data and its underlying structural features. The Euclidean group in 3 dimensions E(3) Equivariant Diffusion Model developed by Hoogeboom et al. learns to denoise, a diffusion process that works with both atom coordinates and atom types. The model utilizes a specific architecture that considers the Euclidean transformations, meaning the generated molecules maintain their identity even when rotated or translated in 3D space. The stability of the atom and the molecule in the predicted structure was compared with the other two existing E(3) models, G-Schnet, and Equivariant Normalizing Flows (E-NF). The E(3) Equivariant Diffusion Model outperformed the two other methods with 98.7% and 82.4% for the atom and molecule stability compared to 85.0% and 4.9% for E-NF, and 95.7% and 68.1% for G-Schnet, respectively. This implies that the E(3) Equivariant Diffusion Model generated, in half the training time, 16 times more stable molecules than the E-NF model.

Bagal et al. reported a generative pre-training model, called MolGPT, for molecular generation (Bagal et al. 2021). This AI-tool can generate small molecules with desired properties. The tool was pre-trained, on a large set of data of SMILES strings from ChEMBL, to learn the basic grammar and syntax of the SMILES molecular representations, and to develop an understanding of common chemical patterns. Using two databases, GuacaMol and MOSES, the model was then fine-tuned to generate molecules having desired properties. GuacaMol contains information of a subset of 1.6 million molecules from ChEMBL, while MOSES contains information on 1.9 million lead-like molecules derived from the ZINC database. MolGPT demonstrates the capability to generate molecules with property values that exhibit minimal deviation from the user-specified scores, with a deviation of 0.31 for partition coefficient, logP, 4.6 for the topological polar surface area metric, 0.2 for the synthetic accessibility score (a measure of difficulty of synthesizing a compound), and 0.075 for the Drug-likeness score. Furthermore, MolGPT can generate molecules that incorporate user-specified scaffolds with 75% of the predicted molecules exhibiting novelty and uniqueness scores exceeding 0.70.

Olivecrona et al. reported a sequence-based generative model (REINVENT 1.0) for the generation of *de novo* molecules with desirable properties (Olivecrona et al. 2017). The authors demonstrated different approaches for this model to generate structures. For example, in the first task, the model was trained to generate molecules with specific structural constraints, e.g., structures devoid of sulfur atoms. This shows the adaptability of the model to such structural constraints in the prediction. In a second task, the model was trained to generate molecules similar to a query structure, e.g., the Celecoxib drug. This showcases the capacity of the model for scaffold hopping and library expansion, demonstrating its utility in diversifying chemical space starting from a single reference molecule. Furthermore, the model also has the ability to generate active compounds against a user-specified molecular target, as tested on the example of the dopamine receptor type 2. Notably, more than 95% of the generated structures are predicted to be active, including experimentally confirmed active compounds. This shows the efficacy of the model in proposing novel chemical entities with potential pharmacological relevance.

Inspired by this model, Blaschke et al. reported REINVENT 2.0, as a production-ready tool for *de novo* design of small molecules in drug discovery (Blaschke et al. 2020). The key components of this tool are the search space, the search algorithm, and the scoring

function. In REINVENT 2.0, a generative model is used as the search space. REINVENT 2.0 is trained using data obtained from ChEMBL and exhibits the capability to generate compounds in the SMILES format. The tool uses reinforcement learning as the search algorithm, which is responsible for generating candidate molecules. The algorithm receives rewards based on the prediction scores per candidate, where the scores are based on several parameters such as calculated properties, pharmacophore shape, similarity criteria, etc. Gradually, the algorithm learns to prioritize actions that generate high-scoring molecules, effectively guiding the search towards promising drug candidates.

Wang et al. proposed a conditional generative pre-trained transformer model, called cMolGPT, for designing target-specific active and drug-like molecules (Wang et al. 2023). The approach taken in this study involves the initial pre-training of the model on the MOSES dataset without incorporating target information. The model is subsequently fine-tuned on three distinct target-specific datasets: EGFR, HTR1A, and S1PR1. The prediction accuracy of the model in generating novel chemical entities tailored for specific targets of interest is compared with eight different models: the Hidden Markov Model (HMM), N-gram generative model, SMILES variational autoencoder (VAE), combinatorial generator, adversarial autoencoder (AAE), junction tree VAE (JTN-VAE), character-level recurrent neural network (CharRNN), and latent vector-based generative adversarial network (LatentGAN). The cMolGPT showed comparatively better performance metrices in terms of the fraction of valid molecules (0.988), uniqueness (~ 1.0), fragment similarity (~ 1.0), and similarity to the nearest neighbor (~ 0.578).

3.3 Interaction energies and toxicity prediction

The activity of drug molecules greatly depends on their binding affinities to the active site of the receptor. Ligands that share similar structural features are likely to exhibit comparable binding affinities when binding to a specific molecular target. Small molecules that exhibit weak binding affinities should be rejected, as they may bind to macromolecules other than their intended target receptor, resulting in toxicity and unfavorable side effects. AI tools such as DeepAffinity (Karimi et al. 2019) and DeltaVina (Wang and Zhang 2016) are capable of predicting binding affinities based on the chemical features of the small molecule and the active site of the receptor.

AI models can predict potential toxicities, helping researchers identify harmful compounds early in the drug development process. The majority of identified lead compounds tend to fail the pre-clinical trials because of their poor pharmacokinetic properties such as absorption, distribution, metabolism, elimination, and toxicity (ADME/Tox) (Sun et al. 2022). The National Institutes of Health, the Environmental Protection Agency, and the US Food and Drug Administration conducted a toxicity prediction challenge called 'Tox 21 Data Challenge' with the goal of comparing computational methods that predict toxicity. As part of the challenge, Mayr et al. developed the best-performing pipeline for toxicity predictions called 'DeepTox' (Mayr et al. 2016). DeepTox first normalizes the chemical structure into standard representations and then computes chemical descriptors such as atom count, surface area, mean polarizability, charge, etc. These descriptors are used as inputs to train the deep learning model, which can then predict the toxicity of new molecules. 'ADMET Predictor' is another AI-based prediction tool that can efficiently predict more than 175 properties including pKa, mutagenicity, logP, absorption, and solubility. Further, using multiscale weighted colored graph theory and gradient boosting decision tree algorithm, Jiang et al. reported a geometric graph-based

toxicity prediction tool called 'CGL-Tox' (Jiang et al. 2021). It uses the gradient boosting decision tree (GBDT) and multiscale weighted colored graph (MWCG) features, which are a type of graph representation that captures the structural and chemical information of molecules. CGL-Tox uses these features to represent the molecular structures of drugs, and then uses the GBDT algorithm to train the model. The model showed an AUROC of ~0.87 in predicting the toxicity of small molecules.

Because of the high complexity in the pathophysiology of diseases, many drugs have off-target binding and are, therefore, dropped out from the pre-clinical trials (Harrison 2016). Reker et al. developed a method to predict molecular targets, including key-target and off-target proteins, of known drugs and computer-generated de novo small molecules. This method is called self-organizing map-based prediction of drug equivalence relationships (SPiDER). Self-organizing maps are a type of ANN that can be used to visualize and analyze high-dimensional data. The software is trained using a manually curated collection of 12,661 active molecules (Reker et al. 2014). A 10-fold cross-validation was performed to estimate the predictive ability of SPiDER, the ROC was in the range of 0.86 to 0.93. Further, in 2022, Naga et al. reported an open-source ML workflow called 'Off-targetP' to predict the off-target binding of small molecules (Naga et al. 2022). This model is generated to assist the chemists in the drug design process, before synthesis, to reduce the attrition rate.

Investigating drug-drug interactions (DDIs) is important in drug development as certain combinations of drugs can cause dangerous interactions, including increased side effects. As the number of possible combinations of drugs can be massive, it is nearly impossible to experimentally test the safety of all combinations AI can assist in identifying DDI that might not be easily detected by traditional methods (Day et al. 2017; Vo et al. 2022). Shukla et al. reported a deep-learning model to predict DDIs (Shukla et al. 2020). Their model is built by integrating CNNs, recurrent neural networks, and mixture density networks. It has an accuracy of $98.50 \pm 0.6\%$. Schwarz et al. reported an Explainable AI (XAI) model called 'AttentionDDI' for DDI predictions. The model is made explainable by adopting the Attention mechanism. AttentionDDI uses a deep learning architecture to learn features from known drug structures and DDIs, and it then uses the Attention mechanism to focus on the most important features for each prediction task. The model showed promising predictions with an area under the Precision-Recall curve (AUPRC) in the range of 0.77 to 0.92 (Schwarz et al. 2021).

As with the AI models used in target identification and any other AI models, the lead identification is also based on data-driven learning, utilizing diverse datasets to make predictions and draw insights. The key algorithms used in the studies discussed above are deep learning, CNN, GFA, SVM, and ANN. One of the key advantages of utilizing AI models in SBVS and LBVS is to minimize the computational resources and time required for the conventional virtual screening methods. Also, AI models are not influenced by human biases (Turon et al. 2023). They provide objective and data-driven results, reducing potential biases in compound selection.

3.4 Examples of successful AI-assisted drug discovery

There are several examples of AI-assisted lead discoveries that made it to clinical trials. In early 2020, the developers of DSP-1181, the first drug created with the assistance of AI, marked a significant milestone as it entered a Phase I clinical trial targeting the treatment of obsessive-compulsive disorder (OCD) (https://www.exscientia.ai/dsp-1181). DSP-1181

is developed as a potent serotonin 5-HT$_{1A}$ receptor agonist. This achievement was made possible through a unique collaboration between Sumitomo Dainippon Pharma in Japan and the UK-based Exscientia. Exscientia uses an AI platform, known as 'Centaur Chemist', for the generation of new molecules and drug targets with a higher likelihood of success in clinical settings (Mak et al. 2021). The platform allowed them to screen through millions of potential small molecules, ultimately selecting and optimizing 10 to 20 candidates for rigorous laboratory experiments. Remarkably, this entire exploratory phase took just 12 months compared to 4–5 years of lead discovery in the conventional drug discovery process. DSP-1181 emerged as the eventual drug candidate.

Zhavoronkov et al. developed inhibitors for the discoidin domain receptor family, member-1 (DDR1) kinase enzyme using the generative tensorial reinforcement learning (GENTRL) method (Zhavoronkov et al. 2019). They trained the models with compounds from the ZINC database and known DDR1 kinase inhibitors. The authors then used this trained model to screen a large database of small molecules and identified several potential DDR1 kinase inhibitors. They then synthesized six compounds and experimentally validated their bioactivity. They further tested one of the promising compounds in vivo in a rodent model. The authors were able to identify lead compounds, including pre-clinical testing, in less than a month.

In early 2022, the AI-driven drug discovery company Insilico (www.insilico.com) devolped a treatment for idiopathic pulmonary fibrosis. The lead compound is currently undergoing clinical trials. The compound, called ISM001-055, is reported to target a novel protein identified through the AI-based target identification platform, PandaOmics. This compound was identified *via* an AI-based lead discovery platform, Chemistry42. This study took only 18 months to reach the clinical trials with an expenditure of $2.6 million. This could have taken up to 15 years through the traditional drug discovery process.

In 2023, Ren et al. reported a new inhibitor for cyclin-dependent kinase 20 (CDK20), utilizing AlphaFold generated structure (Ren et al. 2023). The development of the new inhibitor was based on using multiple AI-based tools. The novel target for the hepatocellular carcinoma, CDK20 was predicted using the PandaOmics target prediction tool and then the structure was modelled using AlphaFold. Further, the putative small molecule inhibitors were generated using the Chemistry42 platform. A total of seven compounds were synthesized and tested in biological assays. Adopting this method, they have identified the small molecule inhibitor within a time span of 30 days after target selection. The developed small molecule inhibitor showed an experimental IC$_{50}$ of 33.4 ± 22.6 nM.

Researchers from the Massachusetts Institute of Technology (Stokes et al. 2020) used a deep learning approach to identify an anti-bacterial lead compound named 'halicin'. They first trained a DNN model with 2,335 molecules which are known to inhibit the growth of *Escherichia coli*. This model was then used to identify and prioritize potential anti-bacterial compounds from large molecular libraries (> 107 million molecules). The ranking of the compounds was done using three criteria: the prediction score, the structural similarity with the known active compounds, and toxicity. They found experimental bactericidal activity for halicin against three bacteria: *E. coli*, carbapenem-resistant Enterobacteriaceae, and *Mycobacterium tuberculosis*.

This section demonstrated the potential of AI to enhance and accelerate target identification and lead discovery. Provided the complex nature of proteins and their interactions, future studies may focus on building prediction models that consider multiple simultaneous factors such as the activation or deactivation of proteins because of conformational changes, molecular interactions, signaling pathways, and allosteric interactions. As discussed above, the experimental determination of biomolecular

structures is a critical step in target discovery, yet it can be a challenging process. Building the training and testing datasets from structures collected using a diversity of experimental techniques such as X-ray crystallography, NMR, and CryoEM can affect the training of AI models. This is because different experimental techniques have different parameters and varying levels of resolution and noise. For example, X-ray crystallography can produce high-resolution structures but cannot capture the dynamic nature of proteins as they need to be crystallized (Srivastava et al. 2018). On the other hand, NMR can capture the dynamic nature of proteins, but at lower resolutions (Sapienza and Lee 2010). Therefore, to train AI models efficiently, we believe that most care must be taken when selecting structural data. Combining data from multiple experimental techniques can help ensure high quality and consistency. Considering the 80:20 data science dilemma, where researchers spend 80% of their time finding and cleaning data, and the varied content and formats across databases like Protein Data Bank, Cambridge Crystallographic Data Centre (CCDC), and National Centre for Biotechnology Information (NCBI) structure database, we propose establishing a comprehensive repository of protein structures with standardized data content and formats, This would streamline AI-driven research. This resource would be highly valuable in enabling convenient training of AI models on a wide range of protein structures from various databases, which can improve the accuracy and generalizability of the models and make them more effective at predicting protein structures.

4 AI in clinical trials and drug marketing

The implementation of AI in clinical trials can shed light on new dimensions regarding patient stratification, patient selection and recruitment, trial design, real-time monitoring, and data analysis. Clinical trials can take around six to seven years before a candidate drug makes it to the market (Norman 2016). Around 50% of the total drug development expenditure is associated with clinical trials, yet only 10% of drugs pass these trials (Harrer et al. 2019; Sun et al. 2022). Clinical trials can fail due to several reasons, including unexpected side effects, insufficient patient enrollment or inadequate patient selection, and challenges in conducting follow-up studies during the trial (Harrer et al. 2019). These challenges can be addressed by building AI models that utilize digital medical records, which are abundantly available in our digitalized world. AI models can also easily collect data from medical journals and efficiently analyze electronic medical reports and clinical trial reports. This AI capability may be used to identify the best-suited individuals to be recruited for clinical trials. For example, AiCure is a mobile application that utilizes AI to monitor treatments in real time during clinical trials (Salcedo et al. 2021). Using digital biomarkers, this platform can monitor the engagement of patients and their response to the tested drug. AiCure can, thus, reduce some of the burdens on medical practitioners during clinical trials, allowing them to be more focused on their patients. However, the limitation of AiCure is that it does not support data export to external platforms. The combination of AI models with wearable devices and the internet of things in medicine also helps in the real-time monitoring of treatment progress. In addition, in clinical trials, placebo control groups can raise serious ethical concerns regarding the potential breach of the rights of patients to receive treatment. This ethical concern must be urgently addressed without any further delays. A well-trained AI model that can predict disease progression may have the potential to replace the placebo control group in clinical trials (Lee and Lee 2020), which can mitigate ethical concerns and improve the accuracy of clinical trial results. Recently,

Insilico Medicine has introduced an AI platform called inClinico (https://insilico.com/inclinico) to predict the success rate of clinical trials. It can also suggest alternative trial designs to improve the success rate of the clinical trial. Although AI has many potential advantages in clinical trials, there are significant risks to patients and liability concerns if the AI predictions (in data analysis, real-time monitoring, or any other application) are inaccurate. To address these concerns, it is recommended that AI predictions should not be relied upon entirely and that clinicians should remain involved to ensure quality control.

After FDA approval, large-scale drug manufacturing can also benefit tremendously from AI technologies. To improve efficiency and ensure high-quality products, drug manufacturing units are increasingly being automated using AI technologies. The team led by Steiner et al. at the University of Glasgow has developed a chemical-robotic laboratory platform, called 'Chemputer', that has the potential to synthesize chemical compounds from any given recipe (Steiner et al. 2019). Steiner's team validated the platform by synthesizing three well-known drugs (diphenhydramine hydrochloride, rufinamide, and sildenafil) without any human intervention. The purity and yield of these drugs were found to be similar to, or even better than, those synthesized using classical methods. This tool has the potential to automate the entire chemical synthesis process, and therefore speed up the drug production phase. The programming of this tool requires expertise in a chemical mark-up language called extensible markup language-based domain-specific language.

After manufacturing, the target is to advertise the drug. Using digital platforms, pharmaceutical companies can expedite the collection of data directly from consumers (Paul et al. 2021). With the ever-expanding volume of data generated by clinical trials, patient records, and scientific literature, machine learning algorithms provide the means to extract valuable insights. These insights help pharmaceutical companies identify market trends, patient preferences, and competitor landscapes (Davenport et al. 2019). Many of the major companies such as Johnson & Johnson, Pfizer, AstraZeneca, and Bristol Myers Squibb use AI for market analysis, trend predictions, and sales improvement. Machine learning models can predict the success of drug candidates, their potential side effects, and even recommend marketing strategies based on historical data. Moreover, they enable the tracking of emerging therapies and their adoption rates, helping pharmaceutical companies stay competitive and responsive to changing market dynamics.

In the area of drug manufacturing and regulatory approval, AI-driven technologies can be helpful. The automated synthesis capabilities of platforms like 'Chemputer' exemplify the potential to enhance efficiency, reduce production timelines, and maintain drug quality. Additionally, utilizing AI for market analysis, trend prediction, and sales promotions streamlines the drug approval process, enabling better-informed decisions and ensuring that products reach patients in need. However, while AI brings numerous advantages, the critical role of human attention cannot be neglected, especially in ensuring that AI-generated predictions and decisions align with quality and safety standards. The synergy between AI technologies and human expertise in drug manufacturing and approval not only offers efficiency but also upholds the highest levels of patient safety and quality control, shaping a promising future for pharmaceutical innovation.

5 Challenges and future perspectives

In summary, the integration of AI in the drug design pipeline has already made considerable improvements. (Arabi 2021) It has been assisting in accelerating the drug discovery process, curbing costs, saving resources and manpower, and reducing attrition rates in clinical trials. In addition, we believe that AI can help minimizing animal sacrifice by reducing the excessive use of in vivo bioassays (Farnoud et al. 2022). Also, it is worth noting that AI is not limited to assisting in drug discovery. AI has the potential to revolutionize the medical world in many other aspects that are beyond the scope of this review. These applications include healthcare management systems such as triage models to improve patient flow (Ivanov et al. 2021), surgeries (Hashimoto et al. 2018), mRNA vaccination (Sharma et al. 2022), preventive treatments (Harmon et al. 2022), nutrigenomics (Kwon 2020), and many more.

Despite their advantages, AI models are fraught with challenges. In addition to the drawbacks listed per model in this review, we discuss here the overall challenges. AI models can have comparable or even better predictive and decision-making abilities than human researchers, yet they are still far from having human intuition. Therefore, we are convinced that the benefit of this technology remains limited to complementing human intelligence, it cannot replace humans. AI models are not perfect and can have detrimental limitations, such as false positive or false negative predictions, especially when dealing with unfamiliar cases. This can compromise the sensitivity and specificity of the model. In addition, AI is highly dependent on the quality of the training data, the appropriateness of the chosen model, the avoidance of bias and overfitting, and more. This is why we advocate for the idea that big data needs big theory (Coveney et al. 2016).

In addition, the challenge with earlier AI models is the lack of explainability, as they are often seen as 'black boxes' that do not provide explanations of how their predictions were made. This drawback makes it difficult to trust the decisions made by AI. Moreover, there are ethical considerations related to patient consent when participating in studies that employ unexplainable AI algorithms. To overcome these challenges, Explainable AI, XAI, has been developed (Mitchell et al. 1986). XAI can provide explanations for its decisions and actions in a way that can be easily understood by humans. However, we do not think that XAI is the ultimate solution, especially that it may involve privacy breach to offer explanations.

Like any other technology, AI is associated with its own drawbacks, and there are always opportunities for further improvements. We would like to highlight that AI technology heavily depends on super-computing power, which is rather costly financially and environmentally with respect to the carbon footprint. We foresee that the future of AI-assisted drug discovery involves the development of a comprehensive virtual human model that encompasses the intricate complexity of human beings. This will enable the virtual testing and accurate prediction of all possible molecular interactions, with the objective of exploring all therapeutic benefits as well as potential adverse effects.

6 Conclusions

The exponential increase in the number of AI-related publications reflects the impact of this technology on society. Given the predictive ability and accuracy of AI models, they have proven to be significant in empowering decision-making in medicine. Overall, this review highlights the broad spectrum of applications of AI-based technology in all phases of drug design, starting as early as diagnosis, through target and lead identifications and clinical trials, to post-marketing analyses. In conclusion, AI can bridge the gap between understanding diseases and developing drugs. AI substantially contributes to the early prediction of diseases, clinical-decision support, development of personalized medicine, NGS analysis, optimization of drug doses, and the prediction of treatment outcomes. Target and lead identifications can be boosted with the help of ML tools that predict protein structures and biological activities of small molecules. AI also helps in the prediction of drug-like properties and off-target effects of *de novo* compounds before experimental validations are performed. In addition, AI technology can improve patient stratification, recruitment, monitoring, and follow-ups in clinical trials. Pharmaceutical companies are adopting AI-driven approaches to assist in various areas including FDA approvals, complete automation of drug synthesis and manufacturing, pharmacovigilance, and even post-market analyses. As detailed in this review, despite all its valuable advantages, AI can still benefit from numerous improvements at the technical level and in other aspects to overcome the challenges associated with its use in drug design and medicine.

Declarations

References

Agbavor F, Liang H (2022) Artificial intelligence-enabled end-to-end detection and assessment of Alzheimer's disease using voice. Brain Sci 13:28. https://doi.org/10.3390/brainsci13010028

Al-Maskari F, El-Sadig M (2007) Prevalence of diabetic retinopathy in the United Arab Emirates: a cross-sectional survey. BMC Ophthalmol 7:1–8. https://doi.org/10.1186/1471-2415-7-11

Andronico A, Randall A, Benz RW, Baldi P (2011) Data-driven high-throughput prediction of the 3-D structure of small molecules: review and progress. J Chem Inf Model 51:760–776. https://doi.org/10.1021/ci100223t

Arabi AA (2021) Artificial intelligence in drug design: algorithms, applications, challenges and ethics. Futur Drug Discov 3(2):FDD59. https://doi.org/10.4155/fdd-2020-0028

Askin S, Burkhalter D, Calado G, Dakrouni SE (2023) Artificial intelligence applied to clinical trials: opportunities and challenges. Health and Technology 13:203–213. https://doi.org/10.1007/s12553-023-00738-2

Bach P, Zauderer MG, Gucalp A et al (2013) Beyond Jeopardy! Harnessing IBM's Watson to improve oncology decision making. J Clin Oncol 31:6508–6508. https://doi.org/10.1200/jco.2013.31.15_suppl.6508

Bagal V, Aggarwal R, Vinod PK, Priyakumar UD (2021) MolGPT: Molecular Generation using a transformer-decoder model. J Chem Inf Model 62:2064–2076. https://doi.org/10.1021/acs.jcim.1c00600

Bagdonas H, Fogarty CA, Fadda E, Agirre J (2021) The case for post-predictional modifications in the AlphaFold protein structure database. Nat Struct & Mol Biology 28:869–870. https://doi.org/10.1038/s41594-021-00680-9

Benowitz SI (2014) Genomics' daunting challenge. Identifying variants that matter. https://www.genome.gov/news/newsrelease/Genomics-daunting-challenge-Identifying-variants-that-matter.

Blaschke T, Arús-Pous J, Chen H et al (2020) REINVENT 2.0: an AI tool for de novo drug design. J Chem Inf Model 60:5918–5922. https://doi.org/10.1021/acs.jcim.0c00915

Blasiak A, Khong J, Kee T (2020) CURATE.AI: optimizing personalized medicine with artificial intelligence. SLAS Technol 25:95–105. https://doi.org/10.1177/2472630319890316

Bolcer JD, Hermann RB (2007) The development of computational chemistry in the United States. Rev Comput Chem. Wiley 5:1–63. https://doi.org/10.1002/9780470125823.ch1

Boniolo F, Dorigatti E, Ohnmacht AJ et al (2021) Artificial intelligence in early drug discovery enabling precision medicine. Expert Opin Drug Discov 16:991–1007. https://doi.org/10.1080/17460441.2021.1918096

Buel GR, Walters KJ (2022) Can AlphaFold2 predict the impact of missense mutations on structure? Nat Struct & Mol Biology 29:1–2. https://doi.org/10.1038/s41594-021-00714-2

Bule M, Jalalimanesh N, Bayrami Z et al (2021) The rise of deep learning and transformations in bioactivity prediction power of molecular modeling tools. Chem Biol & Drug Des 98:954–967. https://doi.org/10.1111/cbdd.13750

Burki T (2020) A new paradigm for drug development. Lancet Digit Health 2:e226–e227. https://doi.org/10.1016/s2589-7500(20)30088-1

Cai L, Chu C, Zhang X et al (2017) Concod: an effective integration framework of consensus-based calling deletions from next-generation sequencing data. Int J Data Min Bioinform 17:153. https://doi.org/10.1504/ijdmb.2017.084267

Cai L, Wu Y, Gao J (2019) DeepSV: accurate calling of genomic deletions from high-throughput sequencing data using deep convolutional neural network. BMC Bioinform 20(1):1–17. https://doi.org/10.1186/s12859-019-3299-y

Carpenter KA, Cohen DS, Jarrell JT, Huang X (2018) Deep learning and virtual drug screening. Future Med Chem 10:2557–2567. https://doi.org/10.4155/fmc-2018-0314

Carroll A (2020) Improving the accuracy of genomic analysis with DeepVariant 1.0. Google AI Blog

Castro AA, Antonio TD, Martinez EC et al (2021) Usefulness of chest X-rays for evaluating prognosis in patients with COVID-19. Radiologia (English Edition) 63:476–483. https://doi.org/10.1016/j.rxeng.2021.05.001

Causey JL, Zhang J, Ma S et al (2018) Highly accurate model for prediction of lung nodule malignancy with CT scans. Sci Rep. 8(1):9286. https://doi.org/10.1038/s41598-018-27569-w

Coveney PV, Dougherty ER, Highfield RR (2016) Big data need big theory too. Philosophical Trans Royal Soc A: Math Phys Eng Sci 374:20160153. https://doi.org/10.1098/rsta.2016.0153

Dara S, Dhamercherla S, Jadav SS et al (2021) Machine learning in drug discovery: a review. Artif Intell Rev 55:1947–1999. https://doi.org/10.1007/s10462-021-10058-4

Davenport T, Guha A, Grewal D, Bressgott T (2019) How artificial intelligence will change the future of marketing. J Acad Mark Sci 48:24–42. https://doi.org/10.1007/s11747-019-00696-0

Day RO, Snowden L, McLachlan AJ (2017) Life-threatening drug interactions: what the physician needs to know. Intern Med J 47:501–512. https://doi.org/10.1111/imj.13404

DePristo M, Poplin R (2017) DeepVariant: highly accurate genomes with deep neural networks. Google AI Blog

Dhamodharan G, Mohan CG (2021) Machine learning models for predicting the activity of AChE and BACE1 dual inhibitors for the treatment of Alzheimer's disease. Mol Diversity 26:1501–1517. https://doi.org/10.1007/s11030-021-10282-8

Ding Y, Sohn JH, Kawczynski MG et al (2019) A deep learning model to predict a diagnosis of Alzheimer Disease by using by using 18F-FDG PET of the brain. Radiology 290:456–464. https://doi.org/10.1148/radiol.2018180958

Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115–118. https://doi.org/10.1038/nature21056

Farnoud A, Ohnmacht AJ, Meinel M, Menden MP (2022) Can artificial intelligence accelerate preclinical drug discovery and precision medicine? Expert Opin Drug Discov 17:661–665. https://doi.org/10.1080/17460441.2022.2090540

Fass L (2008) Imaging and cancer: a review. Mol Oncol 2:115–152. https://doi.org/10.1016/j.molonc.2008.04.001

Ferrè L, Clarelli F, Pignolet B et al (2023) Combining clinical and genetic data to predict response to fingolimod treatment in relapsing remitting multiple sclerosis patients: a precision medicine approach. J Personalized Med 13:122. https://doi.org/10.3390/jpm13010122

Fu J, Gucalp A, Zauderer MG et al (2015) Steps in developing Watson for Oncology, a decision support system to assist physicians choosing first-line metastatic breast cancer (MBC) therapies: improved performance with machine learning. J Clin Oncol 33:566–566. https://doi.org/10.1200/jco.2015.33.15_suppl.566

Gentile F, Agrawal V, Hsing M et al (2020) Deep docking: a deep learning platform for augmentation of structure based Drug Discovery. ACS Cent Sci 6:939–949. https://doi.org/10.1021/acscentsci.0c00229

Gentile F, Yaacoub JC, Gleave J et al (2022) Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking. Nat Protoc 17:672–697. https://doi.org/10.1038/s41596-021-00659-2

Gulshan V, Peng L, Coram M et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316:2402. https://doi.org/10.1001/jama.2016.17216

Gupta R, Srivastava D, Sahu M et al (2021) Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Mol Diversity 25:1315–1360. https://doi.org/10.1007/s11030-021-10217-3

Hameed Z, Zahia S, Garcia-Zapirain B et al (2020) Breast cancer histopathology image classification using an ensemble of deep learning models. Sensors 20:4373. https://doi.org/10.3390/s20164373

Harmon DM, Lopez-Jimenez F, Friedman PA (2022) Introducing artificial intelligence into the preventive medicine visit. Mayo Clinic Proc 97(8):1575–1577. https://doi.org/10.1016/j.mayocp.2022.06.003

Harrer S, Shah P, Antony B, Hu J (2019) Artificial intelligence for clinical trial design. Trends Pharmacol Sci 40:577–591. https://doi.org/10.1016/j.tips.2019.05.005

Harrison RK (2016) Phase II and phase III failures: 2013–2015. Nat Rev Drug Discovery 15:817–818. https://doi.org/10.1038/nrd.2016.184

Hashimoto DA, Rosman G, Rus D, Meireles OR (2018) Artificial intelligence in surgery: promises and perils. Ann Surg 268:70–76 https://doi.org/10.1097/sla.0000000000002693

Heikamp K, Bajorath J (2013) Support vector machines for drug discovery. Expert Opin Drug Discov 9:93–104. https://doi.org/10.1517/17460441.2014.866943

Hoffmann J, Maestrati L, Sawada Y et al (2019) Data-driven approach to encoding and decoding 3-D crystal structures. arXiv preprint. https://doi.org/10.48550/arXiv.1909.00949

Hoogeboom E, Satorras VG, Vignac C, et al (2022) Equivariant diffusion for molecule generation in 3D. International conference on machine learning. Proceedings of Machine Learning Research. https://doi.org/10.48550/arXiv.2203.17003

Huang C, Clayton EA, Matyunina LV et al (2018) Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy. Sci Rep 8(1):16444. https://doi.org/10.1038/s41598-018-34753-5

Ivanenkov YA, Polykovskiy D, Bezrukov D et al (2023) Chemistry42: an AI-driven platform for molecular design and optimization. J Chem Inf Model 63:695–701. https://doi.org/10.1021/acs.jcim.2c01191

Ivanov O, Wolf L, Brecher D et al (2021) Improving ED emergency severity index acuity assignment using machine learning andclinical natural language processing. J Emerg Nurs 47:265–278e7. https://doi.org/10.1016/j.jen.2020.11.001

Jiang J, Wang R, Wei G-W (2021) GGL-Tox: geometric graph learning for toxicity prediction. J Chem Inf Model 61:1691–1700. https://doi.org/10.1021/acs.jcim.0c01294

Jiménez-Luna J, Grisoni F, Schneider G (2020) Drug discovery with explainable artificial intelligence. Nat Mach Intell 2:573–584. https://doi.org/10.1038/s42256-020-00236-4

Jumper J, Evans R, Pritzel A et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

K FM, Mohan M (2022) Ensemble learning models for drug target interaction prediction. International Conference on Applied Artificial Intelligence and Computing

Karimi M, Wu D, Wang Z, Shen Y (2019) DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. Bioinformatics 35:3329–3338. https://doi.org/10.1093/bioinformatics/btz111

Kinman LF, Powell BM, Zhong ED et al (2022) Uncovering structural ensembles from single-particle cryo-EM data using cryoDRGN. Nat Protoc 18(2):319–339. https://doi.org/10.1038/s41596-022-00763-x

Kolesnikov A, Goel S, Nattestad M et al (2021) DeepTrio: variant calling in families. BioRxiv preprint. https://doi.org/10.1101/2021.04.05.438434

Kozlovskii I, Popov P (2021) Structure-based deep learning for binding site detection in nucleic acid macromolecules. NAR Genomics Bioinform 3(4):lqab111. https://doi.org/10.1093/nargab/lqab111

Kumar P, Benedict R, Urzua F et al (2005) Combination treatment significantly enhances the efficacy of antitumor therapy by preferentially targeting angiogenesis. Lab Invest 85:756–767. https://doi.org/10.1038/labinvest.3700272

Kureshi N, Abidi SSR, Blouin C (2016) A predictive model for personalized therapeutic interventions in non-small cell lung cancer. IEEE J Biomedical Health Inf 20:424–431. https://doi.org/10.1109/jbhi.2014.2377517

Kwon DY (2020) Personalized diet oriented by artificial intelligence and ethnic foods. J Ethnic Foods 7(1):1–16. https://doi.org/10.1186/s42779-019-0040-4

Kıvrak T, Yagmur B, Erken H et al (2023) Pulmonary hypertension classification using artificial intelligence and chest X-Ray: ATA AI STUDY-1. medRxiv. https://doi.org/10.1101/2023.04.14.23288561

Labbé CM, Rey J, Lagorce D et al (2015) MTiOpenScreen: a web server for structure-based virtual screening. Nucleic Acids Res 43:W448–W454. https://doi.org/10.1093/nar/gkv306

Lee CS, Lee AY (2020) How artificial intelligence can transform randomized controlled trials. Translational Vis Sci & Technol 9:9. https://doi.org/10.1167/tvst.9.2.9

Liu H-Y, Zhou L, Zheng M-Y et al (2019) Diagnostic and clinical utility of whole genome sequencing in a cohort of undiagnosed Chinese families with rare diseases. Sci Rep 9(1):19365. https://doi.org/10.1038/s41598-019-55832-1

Longoni C, Bonezzi A, Morewedge CK (2019) Resistance to medical artificial intelligence. J Consum Res 46:629–650. https://doi.org/10.1093/jcr/ucz013

Loo JA, DeJohn DE, Du P et al (1999) Application of mass spectrometry for target identification and characterization. Med Res Rev 19:307–319. https://doi.org/10.1002/(sici)1098-1128(199907)19:43.0.co;2-2

Lu J (2022) Protein folding structure prediction using reinforcement learning with application to both 2D and 3D environments. International Conference on Computer Science and Software Engineering. https://doi.org/10.1145/3569966.3570102

Lucena-Perez M, Kleinman-Ruiz D, Marmesat E et al (2021) Bottleneck-associated changes in the genomic landscape of genetic diversity in wild lynx populations. Evol Appl 14:2664–2679. https://doi.org/10.1111/eva.13302

Luo R, Sun L, Xia Y et al (2022) BioGPT: generative pre-trained transformer for biomedical text generation and mining. Brief Bioinform 23(8):bbac409. https://doi.org/10.1093/bib/bbac409

Lutz ID, Wang S, Norn C et al (2023) Top-down design of protein architectures with reinforcement learning. Science 380:266–273. https://doi.org/10.1126/science.adf6591

MacDonald TM, Williams B, Webb DJ et al (2017) Combination therapy is superior to sequential monotherapy for the initial treatment of hypertension: a double-blind randomized controlled trial. J Am Heart Assoc 6(11):e006986. https://doi.org/10.1161/jaha.117.006986

Madhukar NS, Khade PK, Huang L et al (2019) A Bayesian machine learning approach for drug target identification using diverse data types. Nat Commun 10(1):5221. https://doi.org/10.1038/s41467-019-12928-6

Mak K-K, Balijepalli MK, Pichika MR (2021) Success stories of AI in drug discovery - where do things stand? Expert Opin Drug Discov 17:79–92. https://doi.org/10.1080/17460441.2022.1985108

Mamoshina P, Volosnikova M, Ozerov IV et al (2018) Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification. https://doi.org/10.3389/fgene.2018.00242

Manallack DT, Livingstone DJ (1999) Neural networks in drug discovery: have they lived up to their promise? Eur J Med Chem 34:195–208. https://doi.org/10.1016/s0223-5234(99)80052-x

Matta CF, Arabi AA (2011) Electron-density descriptors as predictors in quantitative structure–activity/property relationships and drug design. Future Med Chem 3:969–994. https://doi.org/10.4155/fmc.11.65

Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) DeepTox: toxicity prediction using deep learning. Front Environ Sci 3:80. https://doi.org/10.3389/fenvs.2015.00080

McFadden BR, Inglis TJJ, Reynolds M (2023) Machine learning pipeline for blood culture outcome prediction using Sysmex XN-2000 blood sample results in Western Australia. BMC Infect Dis 23(1):552. https://doi.org/10.1186/s12879-023-08535-y

Melge AR, Parate S, Pavithran K et al (2022) Discovery of anticancer hybrid molecules by supervised machine learning models and in vitro validation in drug resistant chronic myeloid leukemia cells. J Chem Inf Model 62:1126–1146. https://doi.org/10.1021/acs.jcim.1c01554

Meller A, Ward M, Borowsky J et al (2023) Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. Nat Commun 14(1):1177. https://doi.org/10.1038/s41467-023-36699-3

Mitchell TM, Keller RM, Kedar-Cabelli ST (1986) Explanation-based generalization: A unifying view. Mach Learn 1:47–80. https://doi.org/10.1023/a:1022691120807

Mukhopadhyay A, Sumner J, Ling LH et al (2022) Personalised dosing using the CURATE.AI Algorithm: protocol for a feasibility study in patients with hypertension and type II diabetes mellitus. Int J Environ Res Public Health 19:8979. https://doi.org/10.3390/ijerph19158979

Murata K, Wolf M (2018) Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochimica et Biophysica Acta (BBA). - Gen Subj 1862:324–334. https://doi.org/10.1016/j.bbagen.2017.07.020

Nag S, Baidya ATK, Mandal A et al (2022) Deep learning tools for advancing drug discovery and development. 3 Biotech 12(5):110. https://doi.org/10.1007/s13205-022-03165-8

Naga D, Muster W, Musvasva E, Ecker GF (2022) Off-targetP ML: an open source machine learning framework for off-target panel safety assessment of small molecules. J Cheminform 14(1):27. https://doi.org/10.1186/s13321-022-00603-w

Narin A, Kaya C, Pamuk Z (2021) Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Anal Appl 24:1207–1220. https://doi.org/10.1007/s10044-021-00984-y

Norman GAV (2016) Drugs, devices, and the FDA: part 1: an overview of approval processes for drugs. JACC: Basic Transl Sci 1(3):170–179. https://doi.org/10.1016/j.jacbts.2016.03.002

Olivecrona M, Blaschke T, Engkvist O, Chen H (2017) Molecular de-novo design through deep reinforcement learning. J Cheminform 9(1):48. https://doi.org/10.1186/s13321-017-0235-x

Olsen A, Harpaz Z, Ren C et al (2023) Identification of dual-purpose therapeutic targets implicated in aging and glioblastoma multiforme using PandaOmics - an AI-enabled biological target discovery platform. Aging 15(8):2863–2876 https://doi.org/10.18632/aging.204678

Osman AMA, Arabi AA (2022) Quantum and classical evaluations of carboxylic acid bioisosteres: from capped moieties to a drug molecule. ACS Omega 8:588–598. https://doi.org/10.1021/acsomega.2c05708

Öztürk S, Akdemir B (2019) HIC-net: a deep convolutional neural network model for classification of histopathological breast images. Comput Electr Eng 76:299–310. https://doi.org/10.1016/j.compeleceng.2019.04.012

Pantuck AJ, Lee D-K, Kee T et al (2018) Modulating BET bromodomain inhibitor ZEN-3694 and enzalutamide combination dosing in a metastatic prostate cancer patient using CURATE.AI, an artificial intelligence platform. Adv Ther 1:1800104. https://doi.org/10.1002/adtp.201800104

Panwar H, Gupta PK, Siddiqui MK et al (2020) Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. Chaos Solitons & Fractals 138:109944. https://doi.org/10.1016/j.chaos.2020.109944

Patel L, Shukla T, Huang X et al (2020) Machine learning methods in drug discovery. Molecules 25:5277. https://doi.org/10.3390/molecules25225277

Paul D, Sanap G, Shenoy S et al (2021) Artificial intelligence in drug discovery and development. Drug Discov Today 26:80–93. https://doi.org/10.1016/j.drudis.2020.10.010

Perrakis A, Sixma TK (2021) AI revolutions in biology: The joys and perils of AlphaFold. EMBO Rep 22(11):e54046. https://doi.org/10.15252/embr.202154046

Pun FW, Liu BHM, Long X et al (2022) Identification of therapeutic targets for amyotrophic lateral sclerosis using PandaOmics – An AI-enabled biological target discovery platform. Front Aging Neurosci 14:914017. https://doi.org/10.3389/fnagi.2022.914017

Quazi S (2022) Artificial intelligence and machine learning in precision and genomic medicine. Med Oncol 39(8):120. https://doi.org/10.1007/s12032-022-01711-1

Rajpurkar P, Irvin J, Ball RL et al (2018) Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLOS Med 15: https://doi.org/10.1371/journal.pmed.1002686

Reker D, Rodrigues T, Schneider P, Schneider G (2014) Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. Proc Natl Acad Sci 111:4067–4072. https://doi.org/10.1073/pnas.1320001111

Ren F, Ding X, Zheng M et al (2023) AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. Chem Sci 14:1443–1452. https://doi.org/10.1039/d2sc05709c

Salcedo J, Rosales M, Kim JS et al (2021) Cost-effectiveness of artificial intelligence monitoring for active tuberculosis treatment: a modeling study. PLoS ONE 16:e0254950. https://doi.org/10.1371/journal.pone.0254950

Sapienza PJ, Lee AL (2010) Using NMR to study fast dynamics in proteins: methods and applications. Curr Opin Pharmacol 10:723–730. https://doi.org/10.1016/j.coph.2010.09.006

Schlander M, Hernandez-Villafuerte K, Cheng C-Y et al (2021) How much does it cost to research and develop a new drug? A systematic review and assessment. Pharmacoeconomics 39:1243–1269. https://doi.org/10.1007/s40273-021-01065-y

Schwarz K, Allam A, Gonzalez NAP, Krauthammer M (2021) AttentionDDI: Siamese attention-based deep learning method for drug–drug interaction predictions. BMC Bioinformatics 22(1):412. https://doi.org/10.1186/s12859-021-04325-y

Sendak MP, Ratliff W, Sarro D et al (2020) Real-world integration of a sepsis deep learning technology into routine clinical Care: implementation study. JMIR Med Inf 8:e15182. https://doi.org/10.2196/15182

Sharma H, Zerbe N, Klempert I et al (2017) Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. Comput Med Imaging Graph 61:2–13. https://doi.org/10.1016/j.compmedimag.2017.06.001

Sharma A, Virmani T, Pathak V et al (2022) Artificial intelligence-based data-driven strategy to accelerate research, development, and clinical trials of COVID vaccine. Biomed Res Int 2022:1–16. https://doi.org/10.1155/2022/7205241

Shiri I, Maleki H, Hajianfar G et al (2020) Next-generation radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning algorithms. Mol Imaging Biology 22:1132–1148. https://doi.org/10.1007/s11307-020-01487-8

Shukla PK, Shukla PK, Sharma P et al (2020) Efficient prediction of drug–drug interaction using deep learning models. IET Syst Biol 14:211–216. https://doi.org/10.1049/iet-syb.2019.0116

Simonovsky M, Meyers J (2020) DeeplyTough: learning structural comparison of protein binding sites. J Chem Inf Model 60:2356–2366. https://doi.org/10.1021/acs.jcim.9b00554

Spänig S, Emberger-Klein A, Sowa J-P et al (2019) The virtual doctor: an interactive clinical-decision-support system based on deep learning for non-invasive prediction of diabetes. Artif Intell Med 100:101706. https://doi.org/10.1016/j.artmed.2019.101706

Srinidhi CL, Ciga O, Martel AL (2021) Deep neural network models for computational histopathology: a survey. Med Image Anal 67:101813. https://doi.org/10.1016/j.media.2020.101813

Srivastava A, Nagai T, Srivastava A et al (2018) Role of computational methods in going beyond X-ray crystallography to explore protein structure and dynamics. Int J Mol Sci 19:3401. https://doi.org/10.3390/ijms19113401

Steiner S, Wolf J, Glatzel S et al (2019) Organic synthesis in a modular robotic system driven by a chemical programming language. Science 363(6423):eaav2211. https://doi.org/10.1126/science.aav2211

Stokes JM, Yang K, Swanson K et al (2020) A deep learning approach to antibiotic discovery. Cell 180:688–702. https://doi.org/10.1016/j.cell.2020.01.021

Stork C, Chen Y, Šícho M, Kirchmair J (2019) Hit Dexter 2.0: machine-learning models for the prediction of frequent hitters. J Chem Inf Model 59:1030–1043. https://doi.org/10.1021/acs.jcim.8b00677

Sun B, Chen L (2023) Interpretable deep learning for improving cancer patient survival based on personal transcriptomes. Sci Rep 13(1):11344. https://doi.org/10.1038/s41598-023-38429-7

Sun B, Smialowski P, Straub T, Imhof A (2021) Investigation and highly accurate prediction of missed tryptic cleavages by deep learning. J Proteome Res 20:3749–3757. https://doi.org/10.1021/acs.jproteome.1c00346

Sun D, Gao W, Hu H, Zhou S (2022) Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sinica B 12:3049–3062. https://doi.org/10.1016/j.apsb.2022.02.002

Tanoli Z, Vähä-Koskela M, Aittokallio T (2021) Artificial intelligence, machine learning, and drug repurposing in cancer. Expert Opin Drug Discov 16:977–989. https://doi.org/10.1080/17460441.2021.1883585

Taylor RA, Moore CL, Cheung K-H, Brandt C (2018) Predicting urinary tract infections in the emergency department with machine learning. PLoS ONE 13:e0194085. https://doi.org/10.1371/journal.pone.0194085

Tian H, Jiang X, Tao P (2021) PASSer: prediction of allosteric sites server. Mach Learning: Sci Technol 2:035015. https://doi.org/10.1088/2632-2153/abe6d6

Tolkach Y, Wolgast LM, Damanakis A et al (2023) Artificial intelligence for tumour tissue detection and histological regression grading in oesophageal adenocarcinomas: a retrospective algorithm development and validation study. Lancet Digit Health 5:e265–e275. https://doi.org/10.1016/s2589-7500(23)00027-4

Tunyasuvunakool K, Adler J, Wu Z et al (2021) Highly accurate protein structure prediction for the human proteome. Nature 596:590–596. https://doi.org/10.1038/s41586-021-03828-1

Turon G, Hlozek J, Woodland JG et al (2023) First fully-automated AI/ML virtual screening cascade implemented at a drug discovery centre in Africa. Nat Commun 14(1):5736. https://doi.org/10.1038/s41467-023-41512-2

Vamathevan J, Clark D, Czodrowski P et al (2019) Applications of machine learning in drug discovery and development. Nat Rev Drug Discovery 18:463–477. https://doi.org/10.1038/s41573-019-0024-5

van der Heijden AA, Abramoff MD, Verbraak F et al (2017) Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. Acta Ophthalmol 96:63–68. https://doi.org/10.1111/aos.13613

Vega FMDL, Chowdhury S, Moore B et al (2021) Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. Genome Med 13(1):153. https://doi.org/10.1186/s13073-021-00965-0

Verway M, Brown KA, Marchand-Austin A et al (2022) Prevalence and mortality associated with bloodstream organisms: a population-wide retrospective cohort study. J Clin Microbiol 60(4):e0242921. https://doi.org/10.1128/jcm.02429-21

Vo TH, Nguyen NTK, Kha QH, Le NQK (2022) On the road to explainable AI in drug-drug interactions prediction: a systematic review. Comput Struct Biotechnol J 20:2112–2123. https://doi.org/10.1016/j.csbj.2022.04.021

Wang C, Zhang Y (2016) Improving scoring-docking-screening powers of protein-ligand scoring functions using random forest. J Comput Chem 38:169–177. https://doi.org/10.1002/jcc.24667

Wang Y, Zhao H, Sciabola S, Wang W (2023) cMolGPT: a conditional generative pre-trained transformer for target-specific de novo molecular generation. Molecules 28:4430. https://doi.org/10.3390/molecules28114430

Waring MJ, Arrowsmith J, Leach AR et al (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat Rev Drug Discovery 14:475–486. https://doi.org/10.1038/nrd4609

Yang X, Wang Y, Byrne R et al (2019) Concepts of artificial intelligence for computer-assisted drug discovery. Chem Rev 119:10520–10594. https://doi.org/10.1021/acs.chemrev.8b00728

Yang J, Gao Z, Ren X et al (2021a) DeepDigest: prediction of protein proteolytic digestion with deep learning. Anal Chem 93:6094–6103. https://doi.org/10.1021/acs.analchem.0c04704

Yang Y, Yao K, Repasky MP et al (2021b) Efficient exploration of chemical space with docking and deep learning. J Chem Theory Comput 17:7106–7119. https://doi.org/10.1021/acs.jctc.1c00810

Yang K, Huang H, Vandans O et al (2023) Applying deep reinforcement learning to the HP model for protein structure prediction. Physica A 609:128395. https://doi.org/10.1016/j.physa.2022.128395

You Y, Lai X, Pan Y et al (2022) Artificial intelligence in cancer target identification and drug discovery. Signal Transduct Target Ther 7(1):156. https://doi.org/10.1038/s41392-022-00994-0

Zagirova D, Pushkov S, Leung GHD et al (2023) Biomedical generative pre-trained based transformer language model for age-related disease target discovery. Aging 15:9293–9309. https://doi.org/10.18632/aging.205055

Zeng X, Zhu S, Lu W et al (2020) Target identification among known drugs by deep learning from heterogeneous networks. Chem Sci 11:1775–1797. https://doi.org/10.1039/c9sc04336e

Zhang F, Wang H, Liu L et al (2023) Machine learning model for the prediction of gram-positive and gram-negative bacterial bloodstream infection based on routine laboratory parameters. BMC Infect Dis 23(1):675. https://doi.org/10.1186/s12879-023-08602-4

Zhavoronkov A, Ivanenkov YA, Aliper A et al (2019) Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol 37:1038–1040. https://doi.org/10.1038/s41587-019-0224-x

Zhong ED, Bepler T, Berger B, Davis JH (2021) CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. Nat Methods 18:176–185. https://doi.org/10.1038/s41592-020-01049-4

Zoabi Y, Kehat O, Lahav D et al (2021) Predicting bloodstream infection outcome using machine learning. Sci Rep 11(1):20101. https://doi.org/10.1101/2021.05.18.21257369