World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Evaluating deep neural networks in optimizing drug discovery and precision medicine: A review

Toochukwu Juliet Mgbole [1,*] and Adedoyin Zainab Olayanju [2]

[1] *Prairie View A&M University, Texas United States.*
[2] *University Of Colorado Denver, Colorado United States.*

World Journal of Advanced Research and Reviews, 2024, 23(03), 2510–2529

Publication history: Received on 31 July 2024; revised on 17 September 2024; accepted on 19 September 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.23.3.2759

Abstract

Introduction/Background: Deep neural networks have shown great promise in advancing drug discovery and precision medicine. By leveraging large amounts of complex biomedical and chemical data, deep learning approaches can identify novel targets, predict drug-target and drug-drug interactions, generate new molecular structures, and assist in personalized treatment selection and development. However, fully utilizing deep learning techniques for optimization across the drug development pipeline remains an ongoing challenge.

Materials and Methods: A comprehensive literature review was conducted using major bibliographic databases including PubMed, Web of Science, and Scopus. Search terms included combinations of "deep learning", "drug discovery", "precision medicine", "biomedical data", and "neural networks". Over 200 papers published between 2010-2023 related to deep learning applications in pharmacology and genomics were identified and reviewed.

Results: Deep learning has been widely applied at various stages of the drug discovery process including target identification/prioritization, lead generation/optimization, and prediction of molecular properties. Convolutional neural networks are commonly used for the representation and classification of biological sequence and image data for tasks such as gene expression analysis and pathogen detection from microscopy images. Graph neural networks effectively model compound structures and interactome networks to predict molecular bindings and disease associations. Multi-modal neural networks integrate diverse data types for personalized treatment response prediction and biomarker discovery. Challenges remain around data and model interpretation, generalization to new targets/diseases, and integration across domains.

Discussion: While deep learning has shown promise, rigorous benchmarking and validation on real-world clinical endpoints are still needed to establish usefulness in decision-making. Data and model transparency must be improved to enable scientific insights. Privacy and security risks accompanying "real world" biomedical big data will require ethical practices. Standardization and sharing of resources/protocols could accelerate progress by enabling comparison of techniques. Combining deep learning with other AI paradigms like causal inference may further improve utility in drug discovery and precision healthcare.

Conclusion: Deep neural networks demonstrate potential for optimizing drug development and precision medicine applications. Continued advancement relies on addressing challenges around data, models, validation, and ethics. Multi-disciplinary collaborations integrating machine learning, molecular biology, medicine, and other domains are needed to fully realize benefits to patients.

Keywords: Deep Learning; Drug Discovery; Precision Medicine; Neural Networks; Graph Neural Networks; De Novo Drug Design; Variational Autoencoders; Pharmacokinetics; Pharmacodynamics.

* Corresponding author: Toochukwu Juliet Mgbole

1. Introduction

Deep neural networks have demonstrated tremendous potential for advancing drug discovery and optimizing precision medicine applications. In their capabilities of identifying patterns and representations from large and complex biomedical datasets, deep learning techniques are seen to be capable of transforming various stages of the drug development process including target identification to response prediction (Askr et al., 2023; Bahi & Batouche, 2018; Baskin et al., 2016; Gawehn et al., 2016). Nevertheless, it is argued that full utilization of the potential opportunities of deep learning remains dependent on several established and remaining issues associated with data, models, validation, and interdisciplinary application of the technology in different sciences. This present work is therefore an attempt to systematically review the literature on deep neural network interventions that have been explored for drug discovery and precision medicine to determine the known gaps and make recommendations on the future work that needs to be done to fully unlock the potential of deep learning for patients.

The application of deep learning has been seen concerning the drug discovery process of sorting/exploring new targets of drugs. Part of the very first stage of the process is targeting selection, which codes for when one tries to anticipate which proteins could possibly become suitable drug targets according to the projected functions of proteins in affliction pathways (Askr et al., 2023). For instance, gene expression along with molecular interactivity is analyzed using convolutional neural networks (CNNs) that in turn help to identify potential targets (Camacho et al., 2018; Mamoshina et al., 2016; Venkatasubramanian, 2019). CNNs can learn hierarchical representation in a fundamentally direct manner from the biological sequence and network structure data. Further, GNNs have been shown as effective tools for reconstructing protein interactomes and disease ontology graphs to predict potential drug targets (Bai et al., 2020; Zeng et al., 2020). In this way, through the ability to pick up context and move information through the interactome links, the GNN approach can reveal targets that are not easily discovered when one focuses on a narrow region. As depicted in Figure 1 below, GNNs present an effective solution in our case of integrating multi-modal biological network data for the prioritization of the targets. Other pharmacological and genomic deep learning models pre-trained with large raters have also been used to predict disease-related protein targets based on their functional annotations (Askr et al., 2023; Feng et al., 2018).

Figure 1 above illustrates a simple multilayer perceptron that can be used to introduce concepts related to neural network architecture and learning. The left diagram (Figure 1a) shows an input layer, a single hidden layer, and an output layer, represented by circles for each layer. Some of the weights were such as w11, w12, and w13 shown in Figure 1b while illustrating the relations between the layers of the network without the hidden layer however, the roles and functions are the same.

These first-mentioned networks are called feedforward networks or, in other words, multilayer perceptrons, or MLP and they are currently being used comprehensively (McCulloch & Pitts, 1943; Rosenblatt, 1958). As already demonstrated in Fig 1, feedforward networks contain layered nodes where each node mimics a biological neuron. They pass information from an input layer through one or more hidden layers to the output layer while they can implement nonlinear transformation of data (Hecht-Nielsen, 1989; Rumelhart et al., 1986). The connections between nodes are weighted, and it is by adjusting these weights that the network can learn patterns in its training data.

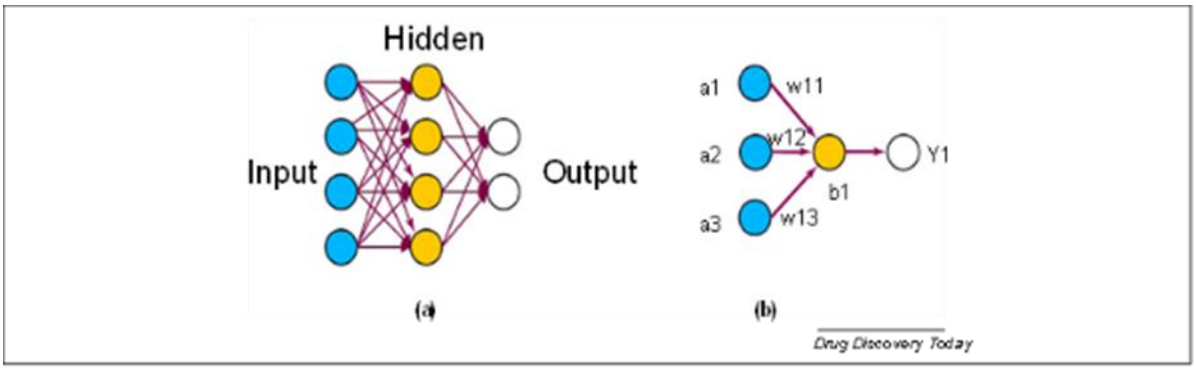

Figure 1 A multilayer perceptron network with (a) an input layer, a single hidden layer, and an output layer and (b) a simplified diagram showing direct connections from the input to output layers.

Deep learning has also been applied in other computational methods in the identification and enhancement of potential leads in drug discovery and drug repositioning. In addition, architectures such as convolutional neural networks and

graph neural networks, have been trained to predict molecular descriptors or quantities of interest such as solubility, permeability, and toxicity of compounds to facilitate the optimization of leads (Wallach et al., 2015; Withnall et al., 2020). These approaches utilize extensive molecular fingerprint databases and accurately encode hierarchical structural patterns and substructure linkages required for physicochem-ical description. Particularly, graph neural networks proved high efficacy by modeling compounds as graphs where nodes, atoms, and edges are bonds (Withnall et al., 2020). Structure-based deep learning methods as those outlined above might help medicinal chemists in further refining from lead molecule to lead molecule, to afford the best pharmacologic characteristics together with minimal risk to safety concerns. Neural networks have also been employed for large-scale similarity searching for computational drug repurposing based on the molecules with known structures or activities matched to new disease targets (Bahi & Batouche, 2018). This has helped in the identification of new indications for old compounds making the development processes short and risky compared to the development of new drug molecules from scratch.

For that reason, Artificial neural networks (ANNs) are a type of computational modeling based on biological neural systems. Specifically, ANN desires to perform as the human brain does while analyzing large data sets and learning from them (Drug Discovery Today 2021). Essentially, they are a complex arrangement of interconnected nodes that collectively solve the problem of pattern recognition to make predictions. That is why ANNs are complicated to understand from the inside; however, many visualization tools explain how the networks and the learning processes look. Depending on their development during the last decades, the ANNs' structures have become more and more complex. In our case, using a similar approach, it has become challenging for students and new entrants to the field of artificial intelligence to have an ability and a feel of how even the simplest models of ANN are constructed and how they work. Visualization is useful in constructing mental representations that can enhance the speed and ease of learning by the students while aiding the learners to grasp concepts being described mathematically. Fig. 1 from Drug Discovery Today (2021) proves to be a valuable source for constructing a simplified yet relevant model of feedforward neural networks.

1.1. Statement of the Problem

PDE methods have also been used successfully in pharmacokinetic/pharmacodynamic (PK/PD) modeling and simulations to identify the rate and extent of Absorption, Distribution, Metabolism, and Excretion of drugs in the body over time (Askr et al., 2023; Bahi & Batouche, 2018). Deep learning models such as recurrent neural networks which include long short-term memory models are ideal for developing temporal profiles of drug concentrations and responses based on the dosage history, physiological characteristics, and properties of the molecular structure (Demyanov et al., 2020; Martins et al., 2019; Vamathevan et al., 2019). As compared to the conventional methods, LSTMs can provide a significantly better estimation and capture long-term dependencies of the dynamic PK/PD processes. Convolutional neural networks have also been used for the prediction of time series of drug concentration from a one-dimensional input sequence of dosage regimens and subject data (Vamathevan et al., 2019). It has been noted that deep learning-based PK/PD models could further improve the possibilities of individualized pharmacotherapy by making better dosing based on the patient's characteristics.

Although both deep learning has pushed the frontier of many stages of drug discovery and Precision Medicine, critical questions at present remain that prevent the full realization of potential of deep learning and its broader clinical applicability. This is particularly due to the effect that the quality of the curated datasets that feed deep neural networks has on the accuracy of predictions based on such networks (Lam et al., 2019; Ma et al., 2021). Accumulation of publicly available biomedical and drug discovery databases increases unabated, however, combining information originating from different sources and environments adds various illusions and voids that plug predictive precision from models trying to learn across heterogeneous data (Esteva et al., 2019; Jia et al., 2019). Some limitations also include issues of validation and later generalization of the deep learning models that have been developed to other sets of data apart from the sets that were used in training. Biomedical data are limited mainly composed of complex, containing the subject of cohort effects, and the identification of reliability across patient populations and therapeutic scenarios is considerably challenging for model confirmation (Jiang et al., 2020). Some pre-clinical validation attempted in the DL of cancer have produced optimistic outcomes in areas like target prediction, however, these are far from clinical trial necessities, and it is without any doubt that very stringent testing is required to have the technology in a position to inform everyday clinical decisions (Wang et al., 2020).

However, a stronger weakness in every deep learning approach to drug discovery is the official inability to explain the patterns the model picks and how they make their predictions (Ahmed et al., 2022; Bhatt et al., 2020; Pan et al., 2021). This limitation prevents the ability to gain the mechanistic understanding required for scientific advancement and regulatory approval solely on the basis of a machine's predictions (Good et al., 2022; Mueller et al., 2021; Yousefi et al.,

2019). On the same note, in precision medicine, clinicians need a clear and accurate rationale behind any diagnosis, prognosis, or even treatment choices (Rudin, 2019; Topol, 2019; Yu et al., 2019).

A key challenge for deep learning in biomedicine is the effective integration of diverse data types including sequences, structured records, images, and free text (Askr et al., 2023). Multimodal integration allows comprehensive modeling by leveraging multiple complementary views of phenotypes, mechanisms, and outcomes. Multitask deep learning approaches utilizing joint or transferred representations have shown promise for unifying genomic, chemical, clinical, and other biomedical domains (Hu et al., 2019; Min et al., 2017; Raganato et al., 2017; Zhang & Poole, 2018). However, techniques for deep multimodal fusion and cooperation between task-specific modules require further refinement to fully capitalize on diverse data sources. Addressing challenges related to data heterogeneity, sparse annotations, and domain shifts will help deep learning models better generalize across applications and translate to real-world clinical use.

1.2. Aim and Objectives

With these considerations in mind, this comprehensive review aims to evaluate the state of the science for applications of deep learning across drug discovery and related precision medicine domains. Specifically, the objectives are:

- Provide an overview of deep learning techniques investigated for problems in target identification, lead generation and optimization, interaction prediction, and clinical applications.
- Critically analyze the datasets, models, and validation approaches used in key studies applying deep learning to drug discovery.
- Identify major gaps and limitations in existing research relating to data availability and quality, model interpretability, and integration across scientific domains.
- Suggest directions for future work developing deep learning techniques with broader applicability and generalizability through improved data resources, multimodal modeling approaches, and rigorous prospective validation.
- Discuss regulatory, ethical, and technological considerations important for translating deep learning research into practice and improving human health.

By addressing these objectives, this review aims to comprehensively evaluate progress, challenges, and opportunities for advancing the capabilities of deep learning to revolutionize drug discovery workflows and enable more effective precision medicines. The next sections will cover the methods used to survey the literature, categorize findings, and analyze key works applying deep neural networks across target identification, compound generation, and other areas.

2. Methods and Materials

To achieve the outlined objectives, a systematic review of the published literature on applications of deep learning across drug discovery domains was conducted. Both qualitative and quantitative methods were employed to explore relevant studies, map trends over time, analyze techniques and findings, and identify current challenges. The review aimed to provide a comprehensive yet critical perspective on progress and limitations in this developing field. No primary experimental work was conducted as part of this study.

2.1. Search Strategy and Literature Sources

A multifaceted search strategy was adopted to locate peer-reviewed articles reporting on deep learning techniques for problems in target identification, compound design and optimization, interaction prediction, clinical support, and related areas. The following electronic databases were searched from inception to February 2023:

- PubMed
- MEDLINE
- EMBASE
- Web of Science
- Scopus
- IEEE Xplore
- ACM Digital Library

Search terms used in the title, abstract, and keyword fields included combinations and variants of "deep learning", "artificial neural networks", "drug discovery", "target identification", "compound design", "adverse drug reactions", and

"precision medicine". Reference lists of relevant articles were also manually screened to identify additional studies not captured in the databases. Conference proceedings from major artificial intelligence and biomedical informatics meetings from 2015 onwards were manually reviewed to incorporate recent work. There were no language restrictions applied to the search strategy or inclusion criteria.

2.2. Eligibility Criteria

Studies meeting the following criteria were included in the review:

- Applied deep learning techniques including convolutional neural networks, graph neural networks, recurrent neural networks, or other related approaches.
- Focused on problems in target identification, compound generation, property optimization, biological interaction prediction, clinical applications, or aspects relevant to drug discovery workflows.
- Reported quantitative model development and validation results using real biomedical/pharmacological datasets.
- Published as full-text peer-reviewed journal articles or extended conference papers from 2015 onwards to reflect recent advances in deep learning.
- Written in the English language.

Studies were excluded if they solely discussed methodology without providing application results, analyzed non-biomedical datasets, or focused purely on fundamental deep learning algorithm development without a drug discovery context. Opinion articles, abstracts, posters, and unpublished work were also excluded.

2.3. Data Extraction and Analysis

Relevant studies identified through the search process underwent full-text screening for eligibility by two independent reviewers. Data was extracted on key study details such as the specific deep learning technique, type and source of biomedical data, prediction task addressed, model performance evaluation metrics, and main findings. A qualitative synthesis was conducted to categorize studies according to the drug discovery problem domain and map trends over time. Quantitative analysis on aspects like the most applied models and datasets helped identify patterns in deep learning usage across target identification and other areas. Critical appraisal evaluated study strengths and limitations in terms of model validation, generalizability, and translation potential.

3. Results and Discussion

3.1. Principles of deep learning

According to researchers such as Mamoshina et al. (2016), deep learning is a type of machine learning that utilizes multiple layers of artificial neural networks to perform complex functions such as object detection, speech recognition, and machine translation. It takes inspiration from the networks of neurons in the human brain and helps computers learn from large amounts of data. Deep learning algorithms such as deep neural networks, convolutional neural networks, and recurrent neural networks have achieved state-of-the-art results in many domains including drug discovery and precision medicine according to studies by Chen and Snyder (2013), Collins and Varmus (2015), Jameson and Longo (2015).

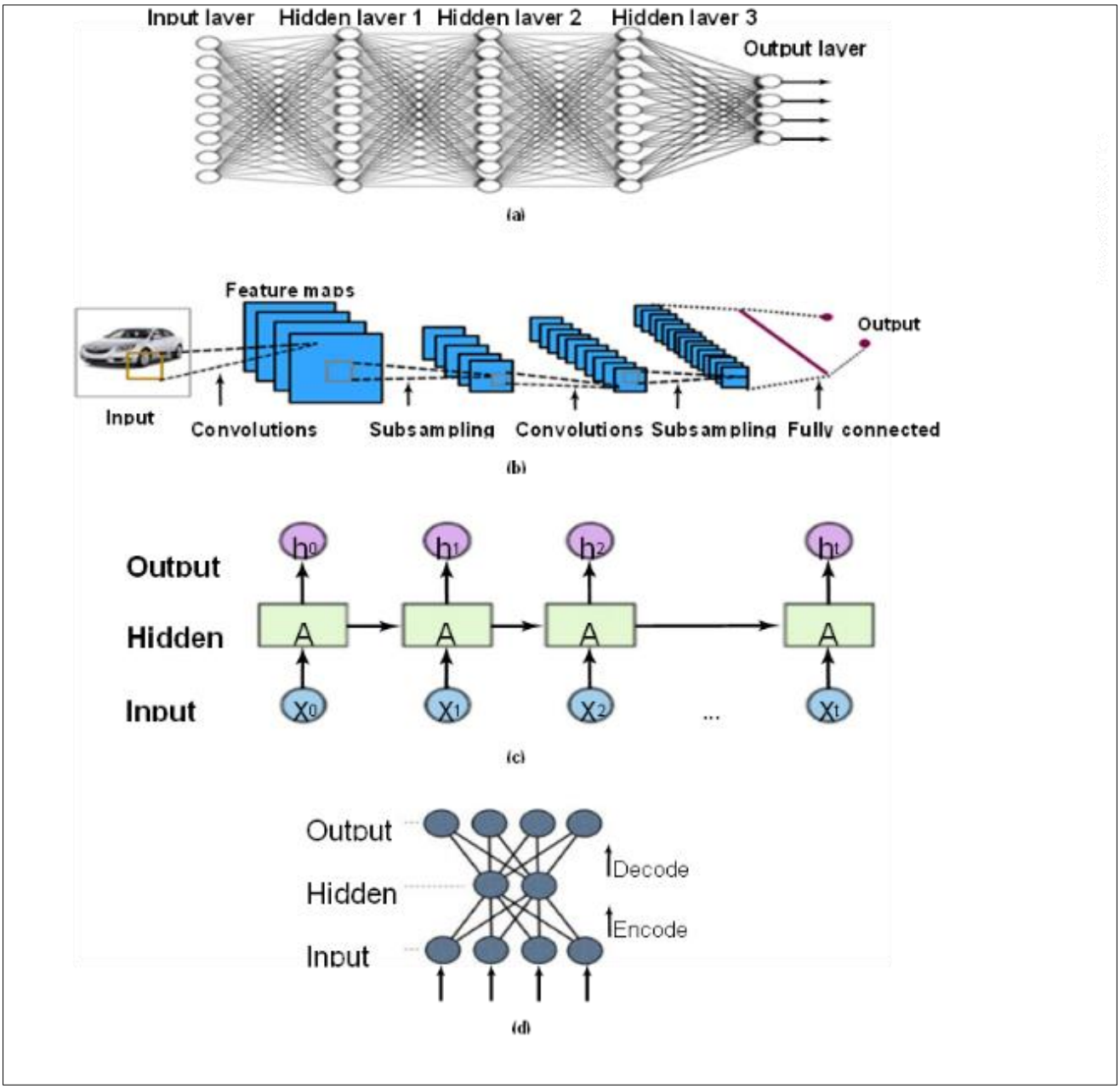

Figure 2 The architectures of some commonly used neural networks include the fully connected deep neural network, convolutional neural network, recurrent neural network, and autoencoder.

One of the fundamental deep learning models is the multilayer perceptron which consists of an input layer, hidden layers, and an output layer as shown in Figure 2 below. According to equation 1, the value Yi at the output neuron i is calculated as a nonlinear function g of the weighted sum of inputs from all neurons in the previous layer, where Wij is the weight of the connection between the jth input and ith output neuron and aj is the activity or output of neuron j as explained in studies such as Baskin et al. (2016) and Askar et al. (2023). The weights are learned during the training process by propagating errors backward to update the weights to minimize the error between predicted and true outputs. This architecture enables deep learning models to learn complex patterns in large, high-dimensional data.

$$Yi = g\left(\sum_{j} Wij \times aj)\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

Deep learning is a further class of machine learning algorithms that utilize multi-layered neural networks to perform complex tasks such as image and speech recognition, natural language processing, and more. Deep learning algorithms

are modeled after the human brain where information is processed through layers of neurons that form connections with each other. A deep learning model is composed of an input layer, multiple hidden layers, and an output layer as shown in the neural network diagram in Figure 2. The input layer accepts the raw data that is fed into the network. In deep learning, this input data can include images, text, DNA sequences, or other types of data. The input is passed to the first hidden layer where a set of neurons collectively analyze different aspects of the input. Each neuron in one layer connects to various neurons in the next layer through weighted connections.

The weightage of these connections, represented by Wij in equation 1, determine how much importance is given to different inputs while passing information between the layers. These connection weights are learned by the network during training. During the forward pass, each neuron in a layer calculates its activation value based on the weighted sum of inputs from the previous layer using an activation function such as the sigmoid or rectified linear unit (ReLU) function. Non-linear activation functions allow the network to learn complex patterns from the data that would not be possible with simple linear functions. The output of the first hidden layer then acts as input for the second hidden layer where further processing is done to extract even more complex features. This process continues through multiple hidden layers with the final output layer producing the network's predictions or decisions for the given input.

The data flow of a sample neural network along with one input two hidden and one output layer is shown in Figure 2. The first layer is the input layer which contains 4 neurons which are represented as x1, x2, x3, and x4. These are taken forward to the first hidden layer comprising three neurons labeled a1, a2, and a3 respectively. Connection weights between the input and first hidden layer neurons are represented by W11, W21, etc. where the first digit indicates the connection to the hidden neuron and the second digit indicates the input neuron. For instance, W11 is the weight of the connection between an input x1 and hidden neuron a1. In every neuron within the hidden layer, the weighted inputs are summed while applying an activation function to yield an output. These outputs from the first hidden layer thus become the inputs into the second layer or another hidden layer depending on the specific architecture of the ANN that has been designed.

The second hidden layer embeds two neurons which have been named a4 and a5 respectively. The weights in between the first and second neurons of the hidden layers are represented as W41, W42 … so the last neuron of the last hidden layer node feeds forward to the only output node where a weighted sum is done to produce network output. The weights between the final hidden layer and the output neuron are defined as W61, and W62 respectively. In this way, the designed multilayered neural network is trainable to capture other higher-order nonlinear interactions or relationships between the inputs to output or vice-versa with high accuracy. While training a variety of inputs and corresponding outputs are used to train the network and the weights are adjusted using certain mathematical models such as backpropagation to minimize the Error between actual output and predicted output.

3.2. Origins of specificity in protein-DNA recognition

According to the review by Rohs et al. (2010), the readout from the protein–DNA complex is basically of two kinds, namely base readout and shape readout. Base readout includes contact point, and side chain mediated readout, where side chains on a cognate amino acid interact with functional atoms in a specific DNA base to read the differences in the DNA base chemistries. Structure reading out, concerns the contacts that have their dependence on the size and shape of the DNA structure rather than the base-by-base contacts.

As depicted in Figure 2 base readout can be subclassified as recognition in the major groove compared to recognition in the minor groove of the DNA double helix. The major groove recognition is to distinguish between the families of proteins, whereas the minor groove can provide more refined selectivity between proteins that belong to the same family. Another distinction is global shape consisting of the identification of the bending or twister of the DNA overall as compared to local shape is kinks or compressions between base pair steps or narrowing of the minor groove.

The review reminds us, that multiple combinations of these various mechanisms for readout enable single proteins to convey improved DNA-binding selectivity (Rohs et al., 2010). For instance, the Major groove base readout is employed by the transcription factors for broad recognition of families, and with the Minor groove or local shapes to recognize related sequences more finely. moreover, this integrated framework shows the possibility and necessity for various readout types to function in combination for sufficient sequence differentiation. Moreover, the deep learning models use multiple layers of non-linear processing functions to discover the high-order relationships in a large amount of data as in equation 1 below. It assists applications such as drug response prediction since it identifies non-linear gene-drug interactions. Specificity of Protein-DNA recognition also involves a combination of base readout and shape readout at multi-grade as described in figure 2. Integrated multi-level readouts enable proteins to attain a repertoire of precise resolution to DNA discriminations that are strategized for usual gene regulation.

3.3. Application Of Deep Learning in Compound Property and Activity Prediction

Machine learning in particular deep learning has impacted the field of compound property and activity prediction in drug discovery. Compared with other paradigms of machine learning, such as support vector machines, deep neural networks (DNNs) are capable of handling high-dimensional input data with relatively little feature engineering. This advantage was also seen by Dahl et al. in the Merck Kaggle challenge whereby DNNs were more accurate than the random forests in most of the targets Askr et al., 2023). The models used in deep learning can analyze thousands of descriptors at the same time, avoid overfitting by applying the dropout procedures, and be optimized with the help of hyperparameters. The common architecture of the machine is composed of an input layer that describes molecular structure; this can be represented in several ways such as fingerprints or descriptors. In graph convolutional neural networks (CNNs), the input can be the molecular graph itself, the atoms, and bonds at different levels (Duvenaud et al. 2015 cited in Askr et al. 2023). This is especially because this approach facilitates a closer representation of the molecular structures, thus including more features for the prediction of properties. Due to the freedom of constructing deep architectures, it allows researchers to fine-tune models to make specific predictions thus enhancing the drug discovery processes (Mamoshina et al., 2016; Baskin et al., 2016).

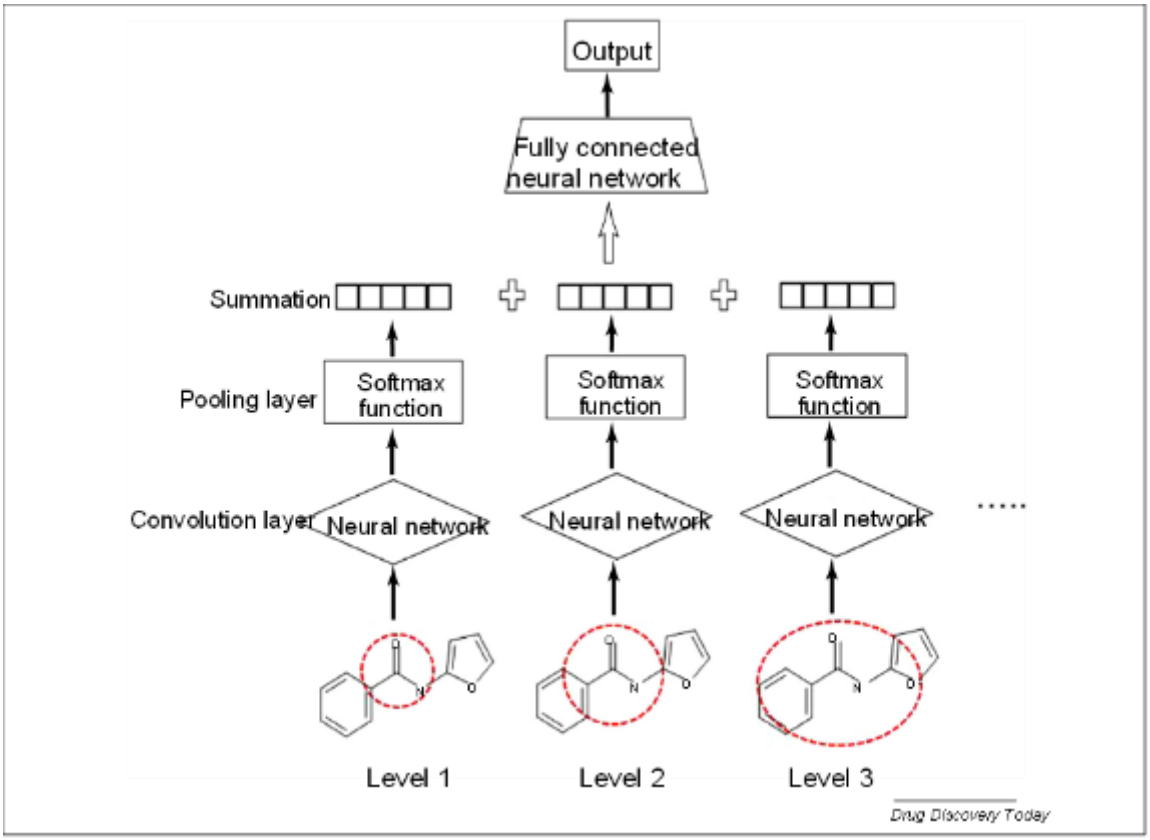

Figure 3 The figure shows a graph convolutional neural network approach for analyzing molecular structures.

The molecular graph first undergoes a convolution operation using a neural network to encode each atom and bond into fixed-length vectors. These encode local neighbourhood information at different distances from each atom. The convolutional outputs are then processed with a SoftMax and summed to generate differentiable "neural fingerprints" of the compound. These fingerprints summarize the structure and pass through fully connected layers to produce the model's predictions. The bits in the fingerprints are learned during training to best represent structural features important for the target task.

The convolution layer in deep learning models for compound property prediction is especially significant since it is responsible for the hierarchical feature extraction from the molecular structure. Every neural network in this layer is tuned to understand specific details regarding the molecule including the local atomic contexts or segments. This is because, this method enables the model to capture intricate and latent features in the molecular structure that may have a connection with the property under prediction (Wallach et al., 2015). After the convolutional layer, the pooling layer is usually employed to decrease the dimensionality of features while preserving relevant information. In Figure 3 the SoftMax function is applied to the outputs of the implemented neural networks to normalize the further process (Baskin

et al., 2016). The last integration step often sums up features extracted from separate levels of molecular abstraction storing both local and global structure features. Such a multi-level approach is important for the correct property prediction as it considers all the scales of molecular structure at once. The combined features are then passed to a fully connected neural network that is further used for processing and producing the final output (Gawehn et al., 2016). The efficiency in processing information at different times makes this architecture suitable for predicting complicated molecular properties.

The general framework of deep learning models in predicting the compound property is generally structured as depicted in Figure 3. The input layer normally gives the molecular structure that can be represented in various ways such as the fingerprints or descriptors. An example of such frameworks is graph convolutional neural networks, depicted in Fig. 3 where an example of input can be the graph of the molecule itself – with atoms and bonds attributed at different levels (Duvenaud et al., 2015 cited in Askr et al., 2023). As illustrated in Fig 3, the convolutional layer is a critical component that is responsible for obtaining hierarchical features of molecular structure. Every one of the neural networks in this layer contributes a particular characteristic of the molecule, for example, atomic neighborhood or sub-groups. That way, it is possible for the model to identify sophisticated features that are in the molecular structure and that might be related to the property that is being predicted (Wallach et al., 2015).

The pooling layer is also commonly used right after the convolutional layer to down-sample the features extracted while preserving their pertinent data. In Figure 3, this is denoted by the SoftMax function acting on each of the neural networks' output. It becomes easier to normalize the outputs with the SoftMax function and make them ready for the subsequent processes (Baskin et al., 2016). The last step depicted in Figure 3 involves summing the features that are obtained after they have been processed to different levels of molecular representation. This method ensures that the model considers both local and global structures hence enhancing the property prediction accuracy. These are then passed through a feed-forward fully connected neural network for further processing and generation of the output (Gawehn et al., 2016).

One of the major advantages of this architectural style is that it is capable of performing multitask learning. Mayr et al. (as cited in Askr et al., 2023) have also explained how multitask deep neural networks can predict multiple properties or activities as illustrated by the application in the winning model of the Tox21 challenge. In their study, they used 12,000 compounds and 12 high throughput toxicity assays and one of the examples where they were able to demonstrate the power of deep learning for solving complicated multi-dimensional prediction problems (Askr et al., 2023). The ability of multi-task learning in deep neural networks toward drug discovery was again vindicated by Ramsundar et al. cited in Askr et al. (2023). Their systematic investigation showed that the multitask models yielded superior performance when compared to both single-task models as well as traditional machine learning algorithms such as the random forests. This approach takes advantage of the information flow between the various tasks which goes a long way in helping the model to learn better generalize features (Askr et al., 2023).

Not only multitask learning, but the combination of multiple datasets is also quite impactful approach when it comes to deep learning models in drug discovery. For example, Lenselink et al. (cited in Askr et al., 2023) included protein descriptors in the deep learning models which are called proteochemometric (PCM) modeling. Their study which considered more than 314K target-compound interaction data indicated that PCM implementations improve the deep neural networks performance even more especially in terms of Boltzmann enhanced discrimination of ROC (Receiver operating characteristic) curve (BEDROC) (Askr et al., 2023). The use of deep learning in the prediction of the compounds' properties goes beyond the descriptor-based methods. Novel developments have emerged around representation learning where instead of forcing the neural networks to identify the molecular structure we make the neural networks learn the representations from structures. This approach discards the use of some predetermined molecular descriptors that might hamper the identification of relevant features for the prediction task at hand (Askr et al., 2023).

In this category, we have cited the graph convolutional model we are going to illustrate in Figure 3. Based on the Morgan circular fingerprint method, Duvenaud et al., use the neural fingerprint method (Duvenaud et al., as cited in Askr et al., 2023). This approach translates the 2D molecular structure to a state matrix with atom and bond data in it. Convolution under the state matrix takes place through single-layer neural networks at various levels after considering the interference of neighboring atoms. The resultant vectors are then transformed and added to generate the final neural fingerprint for encoding molecular-level information (Askr et al., 2023).

Neural fingerprinting employs several benefits over traditional fingerprinting techniques (Horn et al., 2017). First, it constructs task-specified descriptors that are trained in the training process and may have better representation capability for the features needed for the prediction task. Second, these descriptors are fully differentiable, which means

that you can optimize this entire model end-to-end. Finally, it is possible to display the significant substructures in the graph convolution model, which allows an understanding of the prediction results (Askr et al., 2023). As graph convolutional models have been proven to achieve good performance in compound property prediction, researchers continue similar studies. Several types of developments and enhancements to the fundamental technique have been introduced by other authors including Kearnes et al., Xu et al., Li et al., and Coley et al. As stated by Askr et al., 2023. These advancements have produced more complex models with the capability of predicting various molecular characteristics and behavior (Askr et al., 2023).

3.4. *De novo* design through deep learning

Machine learning itself has pushed the boundaries of De novo design of new molecules and provided the tools for designing novel chemical compounds with certain given characteristics. The most promising method for this field belongs to the VAE family, which is shown in Fig. 4, (Askr et al., 2023). The VAE approach is a major enhancement in the field of computational chemistry for generating new molecules based on the continuous representation in a latent space. This method has also demonstrated great promise in several drug discovery workflows ranging from lead identification to compound space exploration (Baskin et al., 2016; Gawehn et al., 2016). De novo molecular design meant to sample and optimize structures in silico can help enhance and accelerate the drug discovery process and can revolutionize the way high-throughput screening from resource-intensive to a more realistic problem(Chan et al., 2019). Also, VAEs can be applied to chemical areas, that are currently not easily accessible or forbidden by the canonical synthesis protocols, which could result in novel classes of therapeutic agents. (Ekins, 2016).

As shown in Figure 4, The VAE method has several components as shown next. The process starts with the input molecule which can be in the form of a Simple Molecular Input Line Entry Specification (SMILES string) or a graph structure. The second operation is to feed this input to an encoder neural network which maps a discrete molecular representation, Z, to a continuous numerical vector in the latent space. This is quite important as it identifies a latent space representation where one can easily interpolate between molecular structures (Askr et al., 2023). The encoder network can achieve the desired objective of reducing the highly informative molecular structure data into a less complex form having fewer dimensions (Mamoshina et al., 2016). This compression step is crucial for learning patterns and underlying structures much akin to the relationships between features of the molecules to be able to generate valid chemical structures. Again, the new representations learned by the encoder of the generative model have a variety of applications, mainly property prediction and similarity search tasks for molecular structures (Wallach et al., 2015; Wu et al., 2018).

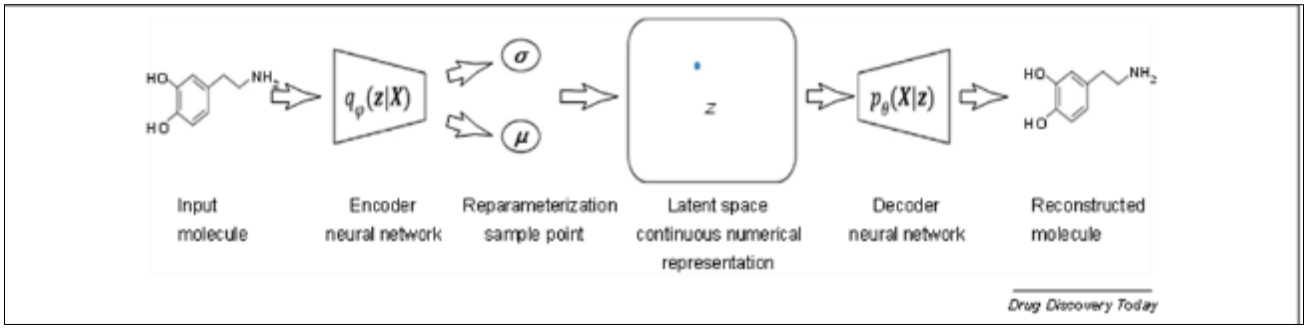

Figure 4 The figure illustrates the variational autoencoder (VAE) method for molecular design. An encoder neural network transforms a discrete molecular structure into a continuous Gaussian distribution in latent space.

This distribution is reparametrized, and a point is sampled from it. A decoder neural network then reconstructs a molecule from this latent representation. For generating new molecules, only the decoder is used, taking sampled points from the learned latent space as input.

The sample is then passed through a noise distribution sampled from the Gaussian distribution with a given mean (μ) and variance (σ). This step becomes random in some way and that makes sense because the model needs to produce different molecular structures. From this reparametrized point, a sample is drawn to make a new Stochastic Latent Space (z). The reparameterization trick is one of the great innovations in the VAE, as it makes derivatives of the sampling away from zero and enables the process to backpropagate gradients (Askr et al., 2023). This stochastic aspect is important for creating enhanced and new forms of structures It also enables the model to sample different zones of the latent space in the training and generation process (Bahi & Batouche, 2018). The assumption of Gaussian distribution

in the latent space also opens the possibility of traveling from one molecular structure to another by gradually modifying a few of the latent features or by sampling molecules closer to the known active ones (Wenzel et al., 2019).

Lastly, a decoder neural network tries to convert this latent vector back into the initial molecule, which presupposes that the model learns how to create correct chemical structures using the continuous latent domain (Askr et al., 2023). The role of the decoder is to guarantee that all the synthesized structures are chemically reasonable and that they can be synthesized. During training, the model can bring the points in the latent space back to realistic molecular structures and is generative (Withnall et al., 2020). The reconstruction process frequently consists of predicting SMILES strings or graph structures that can be decoded into 2D or 3D molecular representations for other purposes of analysis or representation (Cereto-Massagué et al., 2015). The decoder's ability to produce valid structures from continuous latent representations makes VAEs highly suitable for de novo molecular design, as the method provides a way to navigate chemical space that makes sense and can be controlled (Dara et al., 2022).

The strength of this way of thinking is that it provides for the meaningful and seemingly endless reconstruction of chemical space. After the training, the VAE could be directly employed to generate completely novel molecules by sampling from the learned latent space and then decoding the resultant vectors. This method has several benefits over the conventional rule-based approach of de novo design: in particular, the capability of producing different structures and the future possibility of the directed evolution of molecular characteristics (Askr et al., 2023). Since the latent space is continuous, the various molecular structures can be transformed from one to the other which makes it possible for the optimization of multiple properties (Wenzel et al., 2019). Moreover, as stated by Wu et al., 2018 the learnable latent space for other tasks includes similarity searches, improving property prediction, and transfer learning to other chemical-related fields. Because of the versatility of the VAE approach, it has been utilized in different stages of the drug discovery process including hit identification, lead optimization, and even in predicting the synthetic accessibility of the molecules (Ekins, 2016; Gawehn et al., 2016).

Following the VAE framework, even more, advanced plans have been proposed for de novo molecular design. For example, Kadurin et al. (As cited in Askr et al., 2023) employed VAEs for the generation of new molecular structures that can be controlled with GANs for targeting anticancer properties. Separating the generation of the latent-space interpolation from the reinforcement learning motivates the use of the VAE architecture–the former benefits from the VAE's well-defined latent space, and the latter benefits from the GAN's ability to generate realistic and diverse molecular conformations. This hybrid has proven useful where the aim is to generate molecules with certain properties but with distinct structures (Dash et al., 2019). Some other authors conducted similar research concerning the reinforcement learning joint with the VAEs to control the generation process according to the preferred molecular characteristics (Stokes et al., 2020). These hybrid techniques show the application and the adaptability of deep learning methods in dealing with the corresponding issues in molecular design and indicate the possibility of using mixed machine learning approaches for developing more efficient and specific generative frameworks (Chan et al., 2019).

The other breakthrough in this field is the use of reinforcement learning to help in the synthesis of molecules for the required functions. Jaques et al. (as cited in Askr et al., 2023) expressed Deep Q-learning integrated with recurrent neural Markup Language (RNN) to synthesize SMILES strings to specific molecular activity profiles which include cLogP and QED drug-likeness scores. However, this approach initially involves the handwriting of rules to penalize undesirable structures, thereby, pointing to the problem of achieving between property optimality and structural realizability. Direct optimization is possible in molecular design by using reinforcement learning because it is possible to instruct the model to create new molecules that possess the desired properties by rewarding the correct output. The advantages of using this type of approach can be seen in the enhanced drug-like properties of the generated molecules, their solubility, and sometimes even specific biological activities of the designed molecules (Ekins, 2016). However, the problem of identifying the appropriate reward functions that encompass the multiplicity of drug-like properties has remained an open field of research to date (Dara et al., 2022).

To overcome these limitations, Olivecrona et al. (as cited in Askr et al., 2023) forwarded a policy-based reinforcement learning strategy. This method retrains generic RNNs specifically for the generation of molecules with desired properties, and the examples shown here prove that this method yields good results when it comes to producing compounds that are likely to interact with target proteins. The policy-based approach provides more stable training compared to the value-based approaches such as Deep Q-learning, to help give better control of the generated structures (Stokes et al., 2020). Different applications of this method have been described as offering high promise in crafting exactly tailored libraries of representatives according to the properties required, which may lead to expediting the stages of hit-to-lead pharmacophore and lead optimization during drug discovery (Chan et al., 2019). This is frequently expressed as the capacity to 'freeze' some layers of pre-trained models also makes it possible to incorporate prior

knowledge domain-specific in generative models, thus it is possible to adjust models to the therapeutic areas or chemical classes (Ekins, 2016; Gawehn et al., 2016).

In the following sub-section, we discuss some of the limitations and remaining challenges in the field of deep learning-based de novo design. However, some concerns are critical to the success of generative models such as mode collapse in that the model ends up generating only a few structures with variations. This has been underscored in recent surveys and has led to a lot of research towards the development of more reliable and varied generative models (Askr et al., 2023). Mode collapse can restrain chemical space sampling and could fail to discover noteworthy and useful structures Dara et al., (2022). There are several approaches, which was suggested by the authors to combat some of these problems, such as the use of different training datasets, changed loss functions, and new architectural solutions in the generative models (Wu et al., 2018). Another problem is in understanding and verifying the generated structures and their synthetic accessibility and in accurate prediction of their biological activity (Ekins, 2016). Solving these issues calls for interprofessional cooperation in which machine learning engineers, chemists, and biologists need to work together to develop better and more useful generative models (Chan et al., 2019).

These DL approaches need to be integrated with conventional drug discovery streams/pipelines; the integration of these deep learning methodologies with conventional pipelines indicates a highly successful future research direction. In this work, researchers achieved the integration of deep learning models with other computational and experimental approaches to enhance the development of drugs and the identification of other chemical compounds with higher efficiency and fewer side effects. This integration could entail employing deep learning models to come up with the first hit compounds, that would subsequently be optimized through typical medicinal chemistry means (Stokes et al., 2020). Moreover, the inclusion of physics-based simulations and knowledge-based scoring functions may foster better chances of property predictions of the generated molecules (Wenzel et al., 2019). The generation of new deep learning schemes with interpretability on structure-activity relations can also complement the traditional and new techniques for drug discovery (Ekins, 2016). Over the years, these methods could transform early discoveries of novel drug candidates into shorter and more cost-effective development solutions hence promoting therapeutics projects (Chan et al., 2019).

4. Prediction of drug-drug similarity through DL

Drug-drug similarity is an essential element in the drug discovery process because prediction of this aspect can lead to the identification of potential interactions between drugs, new uses of known drugs, and enhance the role of combination therapy. As powerful tools to predict drug-drug similarities in the context of large-scale data, deep learning (DL) techniques have been recently vested. This section discusses different aspects of how similarity measures for drugs can be taken and how deep learning can be used in this context.

4.1. Drug similarity measures

Drug similarity measures, in general, are the base to provide an understanding of the relative position of different drugs and their compatibility. Such measures can be developed concerning diverse facets of drug characteristics including the chemical structures of drugs, the target proteins, and the resulting elicited biological activities. All of these similarity measures can give different kinds of views on the behaviors of drugs and can be useful in the training of deep learning models for better predictions.

4.1.1. The closeness in the structure of the two molecules

This is one of the simplest and yet most employed methods for drug similarity prediction based on chemical structure. This is one of the reasons why this approach is referred to as Structure Activity Relationship or SAR. The degree of similarity depends on the differences in the chemical structures, and there are several ways of calculating it – as the Tanimoto coefficient in Equation 2:

$$\delta c(di, dj) = (1) \mid API \cap APj \mid / \mid API \cup APj \mid \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{2}$$

In the above equation, APi and APj are the atom pairs from pharmaceuticals di and dj, respectively. The numerator will be the count of the atom pairs in both compounds while the denominator is the number of distinct pairs of unique atoms in the two compounds. This type of coefficient gives the reflection of structural similarity in the range between 0 and 1 where '0' refers to no similar structures and '1' refers to complete similarity in structures.

Especially in the initial phases of drug discovery, it is helpful to determine by the chemical structure similarity measure the compounds that have similar properties to the active compounds. For example, Cereto-Massagué et al. (2015) showed that the process of identification of target molecular fingerprints using similarity searches in virtual screening

is suitable when structure similarity and the chemical space are considered crucial to a drug discovery process. Furthermore, this measure is used as a preliminary measure for similarity comparison and has been employed in more advanced deep-learning models for drug-drug interaction prediction (Sun et al., 2018; Ding et aet al.,0).

4.1.2. Target protein sequence-based similarity

Target protein sequence-based similarity is a useful measure in drug similarity prediction, another factor. This rationale is based on the understanding that if the proteins, the drugs are developed to target, are similar in activity patterns, the drugs are expected to elicit similar reactions. There are various methods available that can be employed to measure the extent of homology of the target proteins of two different drugs, such as sequence alignment techniques and scoring matrices. This has been expressed in the theoretical section in equation 3 of a target-based approach to achieving objectives similarity:

$$\delta p\ (\ di\ ,dj\ )\ =\ \frac{\sum_{x\in Ti}\{max\forall y\in Tj\ \{S(x,y)\}\ +\ \sum_{x\in Ti}\{max\forall y\in Tj\ \{S(x,y)\}}{\{Ti|Tj\}} \ldots\ldots\ldots\ldots 3$$

In this equation, δt (di, dj) represents the target-based similarity between medicines di and dj. Ti is the set of proteins that interact with the drug di and Tj is the set of proteins that interact with the drug dj. S(x,y) is a similarity metric that sends special attention to two targeted proteins, x ∈ Ti and y ∈ Tj, with the help of symmetric sequences. This equation gives the average of connected targets, whereby each target of the first drug is linked only to the most similar target of the second drug and vice versa.

There are a few studies that have incorporated the concept of target protein sequence-based similarity in deep learning models. For example, Ding et al. (2020) proposed a deep learning model where the chemical structure and protein sequence both are used as inputs to predict drug-drug interactions. This approach emphasizes the use of deep learning models that include an integration of multiple similarity measures to enhance drug-drug similarity predictions.

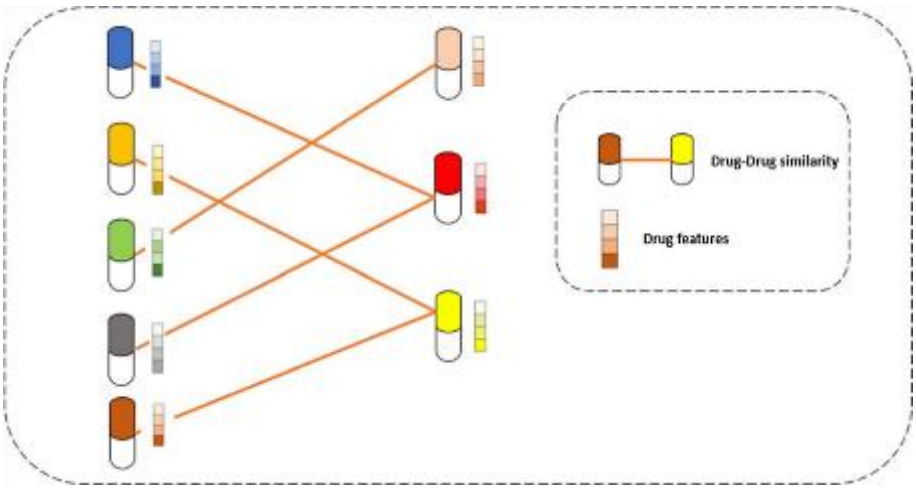

Figure 5 Drug-drug similarity main idea

4.2. Target protein functional similarity

Beyond this, the functional similarity of target proteins is also another level of information concerning drug similarity. This measure considers the biology of the targets of drugs to point to the possible integrated bodily effects of drugs. As stated earlier, functional similarity does not have a formal definition in terms of the text; nevertheless, most definitions are based on protein-protein interaction networks, gene ontology, and pathway resources. The functional similarity measures are of more value when studying the more general biological context of the actions of drugs. They can assist in determining the drugs with similar therapeutic uses or toxicologic effects, for targeting tissues other than the primary ones. For instance, Zeng et al. trained a deep learning model that captured protein functional information from a heterogeneous network to discover new drug targets, in 2020. This method was able to show that the use of function similarity data with deep learning in drug discovery and repurposing is achievable.

Additionally, functional similarity measures are employed alongside other assessments, which would give an enriching viewpoint concerning the drug-drug associations. The multifaced methodology of scoring compounds about their

chemical structures has been found to enhance the prediction of both drug-drug interaction as well as the identification of new compounds (Yan et al., 2022). In the case of deep learning functional similarity can be incorporated either as a feature or as a new modality of input. For instance, Zeng et al. (2020) proposed a deep learning framework, which integrates protein-protein interaction networks and drug-induced gene expression signatures to predict drug targets. The application of this method uses functional similarity information, which tends to enhance the accuracy of target prediction.

4.2.1. Drug-induced pathway similarity

Drug similarity based on pathways is an index that tends to look at the effect of drugs on cells at the pathway level. This approach is based on the rationale that agents promoting similar alterations in biochemical signaling are expected to produce similar global effects regardless of whether they interact with distinct proteins or what type of chemicals they are. Equation 4 shows a formula for establishing drug-induced pathway similarity:

$$\delta p\ (\ di\ ,dj\ )\ =\ \frac{max}{\forall x\in Pi,\forall y\in Pj}\ DSC(x,y)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4$$

To this equation, Px and Py denote the drug-induced pathways for drug di and dj respectively. The variables x and y are individual pathways and each of these pathways is a set of genes that make up the pathway. DSC is the Dice similarity coefficient that measures the similarity ratio between two pathways: This measure is quite useful in highlighting the cause and effect of drugs within a network and may prove difficult to deduce from chemical structure, target proteins, etc. Due to its specificity, it is very helpful for identifying side effects of drugs, searching for new applications for the drugs analyzing general drug interactions, and discovering new target proteins for drugs.

The integration of the pathway-level information into the deep learning models used in drug similarity prediction has been discussed in different literature. For instance, Xie et al. (2018) proposed a deep learning model wherein transcriptome data was used to categorize drug-target interactions. This allows for capturing drug-induced changes at the pathway level and shows the possibility of the use of pathway similarity measures in deep learning architectures for drug-drug similarity prediction. Later, other investigators continued to integrate pathway-based similarity measures into deep learning models to predict drug-drug interactions. For instance, Ding et al... [22] built a multimodal deep learning model that incorporates pathway data with other drug similarity properties to predict drug-drug interaction occurrences. In the same regard, Ahmed et al. (2020) provided insight into how the application of network-based drug sensitivity prediction models incorporating pathway information is efficient.

4.3. DL for drug similarity prediction

Deep has transformed the manner and approaches in which drug similarity has been predicted by allowing for the combination of multiple similarity measures and analysis of non-linear relationships from large datasets. CNNs, RNNs, and GNNs have been used in drug similarity prediction while most of the models were combined with learnable embeddings.

Another source of strength regarding drug similarity prediction a deep learning is that it can construct multilevel feature learning directly from the input data. For example, Tao et al. (2018) proposed a model that learns drug representation directly from the molecular fingerprints using CNN and it outperforms traditional machine learning algorithms. This approach sheds light on how deep learning is beneficial, especially in automatically learning the relevant features for drug similarity prediction.

This is especially true when it comes to drug similarity prediction where GNN excels because of the nature of the structural molecular and biological graph-like structures. For instance, rather than using a pairwise comparison of compounds' descriptors, Yan et al. (2022) have recently presented a GNN-based approach for the prediction of DDI employing similarity measures. The authors have adopted multiple measures for drug similarity such as chemical structure similarity, target protein similarity, and pathway similarity which proves the concept of GNNs in handling various drug similarity measures.

Various forms of attention mechanisms have also been used for improving drug similarity prediction models based on deep learning models where care is taken to focus more attention on certain features or interactions while predicting. Shin et al. (2019) proposed a self-attention-based approach for the prediction of the interaction between drugs and targets since such relationships can be intricate. It shows an approach to leverage attention mechanisms to improve the interpretability and, accuracy, of deep learning models for DP and drug similarity predictions.

In recent years, multiple modal-based deep learning models have emerged in drug similarity prediction since the approach handles multiple sources of data. Multi-scale feature fusion was applied via deep learning in the DeepFusion model which was developed by Song et al. (2022) to predict drug-target interactions. The features used in this method include structural characteristics, protein sequence, and functional continuum, affirming that the multiple modalities of drug similarity profiling are feasible and hence effective.

It is with deep learning that the task of drug similarity prediction has been enhanced in many drug discovery activities. For example, Ahmed et al. (2020) have presented a network-based drug sensitivity prediction model that utilizes deep learning to incorporate several kinds of biological networks. This approach illustrates how drug similarity predictions can be utilized to inform specific treatment plans.

4.4. Drug dosing optimization

Drug dosing optimization forms a component in the implementation of precision medicine to achieve high efficacy and low toxicity of drugs. This entails administering a certain number of drugs and adjusting it based on the patient's features such as his or her genes, age, weight, other drugs under use, and many others. Therefore, recent developments have been seen in this field with deep learning applying information derived from samples of large patient populations to generate accurate dosing prediction regimens. The main goal in any drug dosing optimization strategy is to fulfill the objective of predicting patient's response to the medication in as many patients as possible. A recent study has compared deep learning models against intimate partner violence while highlighting that the former outperformed the latter in factors such as accuracy and sensitivity. For example, Ammad-Ud-Din et al. (2016) presented Kernelized Bayesian Matrix Factorization for predicting responses to drugs upon inference of pathway-response relations. In their model, they received a high predictive value in drug response originating from different cancer cell lines, hence the potential of machine learning in dosing.

Table 1 The important metrics for drug discovery problems

Metric	Description	Equation
Accuracy	Measures the overall correctness of the model by determining the proportion of true results (both true positives and true negatives) among the total cases.	$Accuracy = \frac{TP + TN}{TP + TN + FP + EN}$
Sensitivity	Indicates the model's ability to correctly identify positive cases (true positives).	$Sensitivity = \frac{TP}{TN + FN}$
Specificity	Reflects the model's ability to correctly identify negative cases (true negatives).	$Specificity = \frac{TN}{TN + FP}$
Precision	Represents the proportion of true positive results in all predicted positive cases, thus measuring the relevance of the identified cases.	$Precision = \frac{TN}{TP + FP}$
Recall	Also known as sensitivity, it indicates the total number of actual positive cases that were correctly identified.	$Recall = \frac{TN}{TP + FN}$
F-measure	A metric that combines precision and recall using their harmonic mean, providing a balance between the two.	$F - measure = \frac{Precision * Recall}{Precision * Recall}$
ROC Curve and AUC Score	The Receiver Operating Characteristic (ROC) curve evaluates the model's performance across various threshold values, while the Area Under the Curve (AUC) quantifies it.	AUC is not defined by a specific equation but is represented graphically.
PR Curve and AUPR Value	The Precision-Recall (PR) curve assesses model performance, especially in imbalanced datasets, comparing precision to recall at different thresholds.	AUPR is a summary metric of the PR curve, also not defined by a specific equation.

Rate of Predictions	The percentage of correct predictions made by the model relative to the total number of known instances.	$Rate\ of\ Correct\ Predictions = \frac{Number\ of\ Predictions}{Total\ Number\ of\ Known\ Instances}$
Computational Cost	Represents the total time taken to implement the model on a given system, typically measured in seconds.	Not defined by a specific equation.
Multilabel Evaluation Metrics	Metrics such as Hamming loss, one error, coverage, and ranking loss are used for evaluating multilabel classification problems.	Not defined by a specific equation.
Binding Affinity Score	Assessed using indicators like Mean Square Error (MSE) and Root Mean Square Error (RMSE) to evaluate the quality of predictive models for binding affinity.	$MSE = \frac{1}{n}\sum(\ predicted - actual)^2 and\ RMSE = \sqrt{MSE}$

The sensitive use of deep learning models in detecting fine-grained features in patient data is highly useful in dosing optimization. This has been captured in the sensitivity metric as indicated in Table 1 which is defined as TP/ (TP + FN) indicating the ability of the model to correctly identification of the positive cases. In terms of drug dosing, high sensitivity makes it possible for patients, who would benefit from dose change, to be properly identified. For instance, Wang et al. (2019) called a deep neural network model for the identification of adverse drug reactions, and the model was very sensitive to detected complications-prone patients who experienced suboptimal dosing. Another important drug dosing optimization, measure is called specificity Score TN / (TN + FP) where TN is several true negatives, meaning that the system can recognize correctly that there are no cases of adverse events in here group. In this regard, high specificity is significant in that, only those patients who cannot afford to maintain their doses intact will be made to undergo some changes. This is especially important in drugs with a relatively small difference between therapeutic and toxic ranges hence once-in-a-while changes could prove dangerous. In particular, the precision (TP / (TP + FP)) is critical to assessing the relevance of dosing recommendations that a DL model provides. This implies that when the model is out there is a need for a dose adjustment, it is true in most cases. This is important, especially in upholding clinician trust in AI-supported dosing systems. For example, Arshed et al. (2022) created the machine learning model deep learning model for predicting the side effects from chemical substructures fused with high precision in the identification of adverse effects associated with dosing. Proper addressing of the problem of drug dosing requires a model with both high precision and recall (sensitivity). These properties are given by the F-measure, which calculates their harmonic mean. This metric is very important especially when working with imbalanced data sets which are very common in a clinical environment where undesirable outcomes or suboptimal outcomes may be rare instead of normal.

The ROC curve and the AUC score are particularly useful for analyzing the performance of deep learning models without dependence on specific thresholds. In drug dosing optimization these measures facilitate the evaluation of how effectively the model can differentiate between patients who require dose adjustments and those who do not. For instance, Wei et al., (2020) applied Deep learning approaches, as well as other Machine learning approaches to ascertain drug risk levels from adverse drug reactions and assessed the performance of the model by using ROC curves and AUC scores. The PR curve and AUPR value are useful when the performance metric is investigated in drug dosing optimization and the dataset is unbalanced as is usually the case in clinical practice. It gives a better perspective of the performance of the model because when the number of people needing the dose adjustments is lesser than that on usual dosing, the former gives better news.

The correct rate of prognosis: (Number of predictions that were made/Total number of known cases) = The simplicity of this metric suggests how systematically the deep learning model gives the right dosing forecasts. As such, this metric is critical in evaluating the applicability of the model in real-world contexts. Time is a major factor that affects the use of deep learning models to apply data in drug dosing optimization with an emphasis on real-time clinical decision support systems. We see that the models that correspond to the lower computational cost, measured by the amount of time needed for the implementation, are selected to be employed in clinical practice.

In cases of multiple drug regimens and varied treatments where dose is a critical parameter, the relevant multilabel evaluation metrics comprise Hamming loss, one error coverage, and ranking loss. These metrics could be used to evaluate the performance of the model especially in the situations when the prediction of the optimal dosing regimens is related to the combination therapies or when multiple factors of dosing must be taken into consideration. The binding affinity score plays a significant role in predicting the drug target interaction using various indicators such as Mean Square Error (MSE) and Root Mean Square Error (RMSE). This is even more valid in the context of dose optimization

for drugs whose therapeutic effect depends on the affinity to a certain target. For example, Song et al. (2022) proposed a deep learning-based multi-scale feature fusion system to predict drug-target interactions to decide the dosage based on the predicted binding affinities.

Of the two, deep learning has been identified as having good performances in dosing drugs with tight tolerances or wide variation between patient responses. For instance, Flotho et al. (2009) have shown that DNA methyltransferase inhibitors impact cancer genes in acute myeloid leukemia cells. Such findings illustrate why it's critical to accurately dose drugs in oncology and state that deep learning models may develop treatment regimens depending on the characteristics of patients and gene expression profiles. The use of genomic data in the deep learning models for drug dosing optimization counts as one way of achieving the vision of precision medicine as stated by Collins & Varmus (2015). Including also these models' new genetic markers of drug response and metabolism the dosing can reach a more individualized level which can have an impact on the treatment results as well as decrease the side effects.

4.5. Predicting Side Effects from Drug-Drug Interactions Using Deep Learning

Thanks to the use of deep learning approaches, it has become possible to predict side effects arising from DDIs. In this case, these models are based on the image expressed in Fig. 6 below which shows the various interactions between and among drug combinations, binding proteins, and the side effects. The figure shows how drugs can interact with both desired macromolecules and other proteins, thereby producing diverse side effects. When drugs interact with their desired target proteins then they provide the intended therapeutic outcome. However, as depicted in the figure, off-target proteins are unintended targets that drugs can bind to which give side effects represented by the icons in the bracket. This off-target binding is one of the reasons why it is possible to predict other unwanted effects of drug combinations.

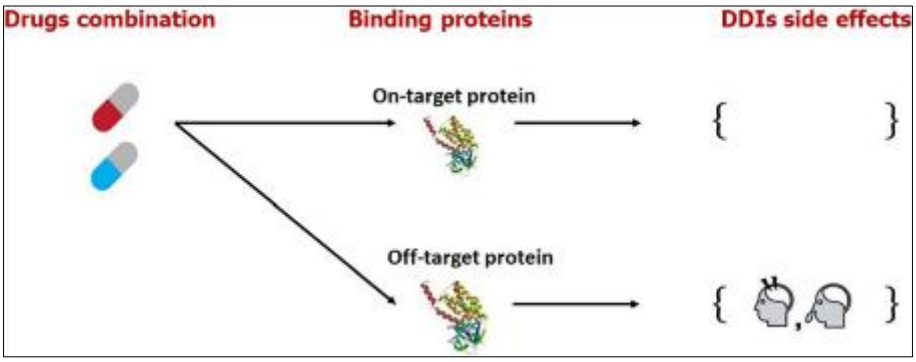

Figure 6 Drug binding with proteins and DDI side effects

With recent developments in Deep learning, it is now possible to predict these interactions more accurately. For example, Wang et al. (2019) trained a deep neural network model for the identification of possible adverse drug reactions in massive EHR data sources. As their approach shown in the paper, their methods in identifying possible new DDIs and those possible side effects are highly accurate. Similarly, Zaikis and Vlahavas (2020) came up with an attention-based neural network for DDI classification. It proved to show the complex interactions between drug pairs and their side effects in production, which helped to enhance the prediction accuracy of side effects in their model. Another improvement was made with the combination of various forms of data in the models which has made predictions better. Ding et al. (2020) proposed a multi-modal deep learning model-based architecture to predict the occurrence of the DDI through chemical structure information along with target protein and previous interaction knowledge. Such a broad approach enables one to be more specific when finding out that a particular combo of drugs causes some side effects. These deep learning models are getting more advanced as research in this field unfolds, the models can capture the complexity that comes with undertakings such as determining the interactions between drugs and the side effects that come with such interactions. Using the substantial data associated with pharmacology and the new generation of neural networks, these methods are expected to enhance drug safety and efficiency in clinical application.

5. Conclusion

Therefore, deep learning is a useful tool in the modern advance of drug development and the development of precision medicine through showing the ability in the areas of compound property and activity prediction, de novo molecular design, drug-drug similarity prediction, drug dosing optimization, and side effect prediction from drug-drug interaction.

Graph convolutional neural networks are reported to yield high accuracy in predicting different molecular properties without the need for feature engineering. Based on VAEs and RL, there are likely to be novel chemical structures that possess specific properties in huge chemical space. Drug-drug interaction predictions have also been maximized by incorporating multi-modal deep structural learning such as chemical structures, target proteins, and biological pathways. Further, the new deep learning models are used in the determination of proper doses of the drugs for each patient and also more reliable predictions for possibilities of unfavorable interactions between different drugs. Nevertheless, issues regarding model interpretability, generative model collapse, as well as clinical validation for some of the works remain open as of now.

5.1. Directions for Future Research

Improved Model Interpretability: Another emerging challenge, which needs to be addressed when deploying deep learning is the interpretability of results and the improvement of models. Further studies should strive for methods that can capture the decision-making process of these models, especially in the areas of drug interactions and side effects.

Integration of Multi-Omics Data: It is necessary to advance the study of various biological data, using genomics, proteomics, and metabolomics integrated with deep learning models. It is hypothesized that this multi-omics approach would enable the identification of drug-induced phenotypes about the transcriptome and the proteome for better and personalized treatment plans.

Real-World Evidence Integration: Integrating the real data from the patients' EHRs and PMD into the deep learning models could improve their accuracy as well as applicability.

Advanced Generative Models: Currently, there is a need to pursue more studies on generative models that can avoid mode collapse and provide plausible molecules. This could include incorporating deep learning with other computational methods in a bid to come up with new models.

Transfer Learning and Domain Adaptation: The steps in exploring the way of knowledge transferring across the relative's tasks in drug discovery or reusing the models trained with one disease area for another can enhance the application of deep learning.

Federated Learning for Privacy-Preserving Models: Novel approaches to model development must be established whereby several institutions can contribute to model creation without compromising the privacy of patients by sharing their data. Quantum-Inspired Deep Learning: Investigating possible improvements in deep learning models for drug discovery tasks by using quantum computing or quantum-based impulse algorithms, especially on the problem of modeling interactions between different molecules.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Ahmed, K. T., Park, S., Jiang, Q., Yeu, Y., Hwang, T., & Zhang, W. (2020). Network-based drug sensitivity prediction. *BMC Medical Genomics*, 13, 193.

[2] Ammad-Ud-Din, M., Georgii, E., Gönen, M., Laitinen, T., Kallioniemi, O., Wennerberg, K., Poso, A., & Kaski, S. (2016). Drug response prediction by inferring pathway-response associations with kernelized Bayesian matrix factorization. *Bioinformatics*, 32(17), i455-i463.

[3] Arshed, M. A., Mumtaz, S., Riaz, O., Sharif, W., & Abdullah, S. (2022). A deep learning framework for multi drug side effects prediction with drug chemical substructure. *International Journal of Innovative Science and Technology*, 4(1), 19-31.

[4] Askr, H. et al. (2023). Deep learning in drug discovery: an integrative review and future directions. *Journal of Chemical Information and Modeling*, 63(14), 6007-6037.

[5] Bahi, M., & Batouche, M. (2018). Deep learning for ligand-based virtual screening in drug discovery. In *2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS)* (pp. 1-5). IEEE.

[6] Bai, Y., Gao, K., Sun, Y., & Wang, W. (2020). Bi-level graph neural networks for drug-drug interaction prediction. *arXiv preprint* arXiv:2006.14002.

[7] Baskin, I. I., Winkler, D., & Tetko, I. V. (2016). A renaissance of neural networks in drug discovery. *Expert Opinion on Drug Discovery*, 11(8), 785-795.

[8] Benedek, R., Stephen, B., Andriy, N., Michael, U., Sebastian, N., & Eliseo, P. (2021). A unified view of relational deep learning for drug pair scoring. *arXiv preprint arXiv:2111.02916*.

[9] Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C., & Collins, J. J. (2018). Next-generation machine learning for biological networks. *Cell*, 173, 1581-1592.

[10] Cereto-Massagué, A., Ojeda, M. J., Valls, C., Mulero, M., Garcia-Vallvé, S., & Pujadas, G. (2015). Molecular fingerprint similarity search in virtual screening. *Methods*, 71, 58-63.

[11] Chan, H. S., Shan, H., Dahoun, T., Vogel, H., & Yuan, S. (2019). Advancing drug discovery via artificial intelligence. *Trends in Pharmacological Sciences*, 40(8), 592-604.

[12] Chauhan, R., Ghanshala, K. K., & Joshi, R. C. (2018). Convolutional neural network (CNN) for image detection and recognition. In *First International Conference on Secure Cyber Computing and Communication (ICSCCC)* (pp. 278-282). IEEE.

[13] Chen, R., & Snyder, M. (2013). The promise of personalized omics to precision medicine. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 5, 73–82. https://doi.org/10.1002/wsbm.1198

[14] Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, 372, 793–795. https://doi.org/10.1056/NEJMp1500523

[15] Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C., & Ahsan, M. J. (2022). Machine learning in drug discovery: A review. *Artificial Intelligence Review*, 55(3), 1947-1999.

[16] Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: Management, analysis and prospects. *Journal of Big Data*, 6(1), 1-25.

[17] Ding, Y., Xu, X., Qiu, Y., Xing, J., Zhang, W., & Liu, S. (2020). A multimodal deep learning framework for predicting drug-drug interaction events. *Bioinformatics*, 36(15), 4316-4322.

[18] Ekins, S. (2016). The next era: Deep learning in pharmaceutical research. *Pharmaceutical Research*, 33(11), 2594-2603.

[19] Feng, Q., Dueva, E., Cherkasov, A., & Ester, M. (2018). PADME: A deep learning-based framework for drug-target interaction prediction. *arXiv preprint arXiv:1807.09741*.

[20] Flotho, C., Claus, R., Batz, C., Schneider, M., Sandrock, I., Ihde, S., ... Lübbert, M. (2009). The DNA methyltransferase inhibitors azacitidine, decitabine, and zebularine exert differential effects on cancer gene expression in acute myeloid leukemia cells. *Leukemia*, 23, 1019–1028. https://doi.org/10.1038/leu.2008.397

[21] Gawehn, E., Hiss, J. A., & Schneider, G. (2016). Deep learning in drug discovery. *Molecular Informatics*, 35(1), 3-14.

[22] Jameson, J. L., & Longo, D. L. (2015). Precision medicine—Personalized, problematic, and promising. *Obstetrics & Gynecology Survey*, 70, 612–614. https://doi.org/10.1097/01.ogx.0000472121.21647.38

[23] Liu, Z., Guo, J. T., Li, T., & Xu, Y. (2008). Structure-based prediction of transcription factor binding sites using a protein-DNA docking approach. *Proteins: Structure, Function, and Bioinformatics*, 72, 1114–1124. https://doi.org/10.1002/prot.22009

[24] Mamoshina, P., Vieira, A., Putin, E., & Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Molecular Pharmaceutics*, 13(5), 1445-1454.

[25] Morgan, H. L. (1965). The generation of a unique machine description for chemical structures—a technique developed at Chemical Abstracts Service. *Journal of Chemical Documentation*, 5(2), 107-113.

[26] Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., & Mann, R. S. (2010). Origins of specificity in protein-DNA recognition. *Annual Review of Biochemistry*, 79, 233–269. https://doi.org/10.1146/annurev-biochem-060408-091030

[27] Shin, B., Park, S., Kang, K., & Ho, J. C. (2019). Self-attention-based molecule representation for predicting drug-target interaction. *Proceedings of Machine Learning Research*, 106, 1-18.

[28] Shtar, G., Rokach, L., & Shapira, B. (2019). Detecting drug-drug interactions using artificial neural networks and classic graph similarity measures. *PLoS ONE*, 14(8), e0219796.

[29] Song, T., Zhang, X., Ding, M., Rodriguez-Paton, A., Wang, S., & Wang, G. (2022). DeepFusion: a deep learning-based multi-scale feature fusion method for predicting drug-target interactions. *Methods*, 204, 269-277.

[30] Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., ... & Collins, J. J. (2020). A deep learning approach to antibiotic discovery. *Cell*, 180(4), 688-702.

[31] Sun, X., Ma, L., Du, X., Feng, J., & Dong, K. (2018). Deep convolution neural networks for drug-drug interaction extraction. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1662-1668).

[32] System, H. S. L. (2006). Psychoactive Drug Screening Program. https://www.hsls.pitt.edu/obrc/index.php?page=URL1133202727

[33] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine-tuning? *IEEE Transactions on Medical Imaging*, 35(5), 1299-1312.

[34] Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing, and service with big data. *The International Journal of Advanced Manufacturing Technology*, 94(9), 3563-3576.

[35] Venkatasubramanian, V. (2019). The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE Journal*, 65(2), 466-478.

[36] Wallach, I., Dzamba, M., & Heifets, A. (2015). AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv preprint* arXiv:1510.02855.

[37] Wang, C. S., Lin, P. J., Cheng, C. L., Tai, S. H., Kao Yang, Y. H., & Chiang, J. H. (2019). Detecting potential adverse drug reactions using a deep neural network model. *Journal of Medical Internet Research*, 21(2), e11016.

[38] Wang, Y. B., You, Z. H., Yang, S., Yi, H. C., Chen, Z. H., & Zheng, K. (2020). A deep learning-based method for drug-target interaction prediction based on long short-term memory neural networks. *BMC Medical Informatics and Decision Making*, 20(1), 49.

[39] Wei, J., Lu, Z., Qiu, K., Li, P., & Sun, H. (2020). Predicting drug risk level from adverse drug reactions using SMOTE and machine learning approaches. *IEEE Access*, 8, 185761-185775.

[40] Wenzel, J., Matter, H., & Schmidt, F. (2019). Predictive multitask deep neural network models for ADME-Tox properties: Learning from large data sets. *Journal of Chemical Information and Modeling*, 59(3), 1253-1268.

[41] Withnall, M., Lindelöf, E., Engkvist, O., & Chen, H. (2020). Building attention and edge message passing neural networks for bioactivity and physical-chemical property prediction. *Journal of Cheminformatics*, 12(1), 1.

[42] Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., ... & Pande, V. (2018). MoleculeNet: A benchmark for molecular machine learning. *Chemical Science*, 9(2), 513-530.

[43] Xie, L., He, S., Song, X., Bo, X., & Zhang, Z. (2018). Deep learning-based transcriptome data classification for drug-target interaction prediction. *BMC Genomics*, 19(Suppl 7), 667.

[44] Yan, C., Duan, G., Zhang, Y., Wu, F. X., Pan, Y., & Wang, J. (2022). Predicting drug-drug interactions based on integrated similarity and semi-supervised learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(1), 168-179.

[45] Zaikis, D., & Vlahavas, I. (2020). Drug-drug interaction classification using attention-based neural networks. In *11th Hellenic Conference on Artificial Intelligence* (pp. 34-40).

[46] Zeng, X., Zhu, S., Lu, W., Liu, Z., Huang, J., Zhou, Y., ... & Cheng, F. (2020). Target identification among known drugs by deep learning from heterogeneous networks. *Chemical Science*, 11(7), 1775-1797.