REVIEW ARTICLE

Deep Learning in Drug Design: Progress, Methods, and Challenges

Seyyed Saeid Masoomkhah, Khosro Rezaee*, Mojtaba Ansari*, Hossein Eslami

Department of Biomedical Engineering, Meybod University, Meybod, Iran

*Corresponding Authors: Khosro Rezaee, Mojtaba Ansari
Email: kh.rezaee@meybod.ac.ir, ansari@meybod.ac.ir

Received: 21 January 2023 / Accepted: 13 April 2023

Abstract

Purpose: Artificial Intelligence (AI), which mimics the human brain structure and operation, simulates intelligence. The aim of Machine Learning (ML), which is a branch of artificial intelligence, is to create models by analyzing data. Another type of artificial intelligence, Deep Learning (DL), depicts geometric changes using several layers of model representations. Since DL broke the computational analysis record, AI has advanced in many areas.

Materials and Methods: Contrary to the widespread use of conventional ML methodologies, there is still a need to promote the use and popularity of DL for pharmaceutical research and development. Drug discovery and design have been enhanced by ML and DL in major research projects. To fully realize its potential, drug design must overcome many challenges and issues. Various aspects of medication design must be considered to successfully address these concerns and challenges. This review article explains DL's significance both in technological breakthroughs and in effective medications.

Results: There are numerous barriers and substantial challenges associated with drug design associated with DL architectures and key application domains. The article discusses several elements of medication development that have been influenced by existing research. Two widely used and efficient Neural Network (NN) designs are discussed in this article: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

Conclusion: It is described how these tools can be utilized to design and discover small molecules for drug discovery. They are also given an overview of the history of DL approaches, as well as a discussion of some of their drawbacks.

Keywords: Machine Learning; Deep Learning; Drug Design; Drug Discovery; Neural Network.

DOI: https://doi.org/10.18502/fbt.v11i3.15893

1. Introduction

On average, it takes 15 years to research and develop a drug therapy, and the pharmaceutical industry loses enormous amounts of money for each failed attempt. Over the past two decades, the number of prescription drugs approved has steadily declined despite rising R&D costs. Blockbuster medicines are hampered by regulatory roadblocks [1]. As a result of automation and the growth of information technology, enormous amounts of biological data have been collected and analyzed in the last decade [2, 3]. A classic Machine Learning (ML) technique, Artificial Neural Networks (ANNs) are architectures of artificial neurons designed to function analogously to the central nervous system of humans [4]. In their infancy, ANNs had just one input layer, one hidden layer, and one output layer. Each node's attributes are directly related to its input layer. In the hidden layer, each node receives a weighted linear combination of the outputs from the layer above. A nonlinear activation function modifies the outputs of the units. The hidden layer's efforts were repeated in the visible layer's output. To produce the outcome, the activation function inputs were fed signals from the buried layer. The Feedforward Neural Network (FNN) [5] uses this type of input. Avoid overfitting shallow networks, especially when regularization is used [6]. To uncover more esoteric patterns, more hidden layers must be built from the incoming data. Lower layers learn the basics of these patterns, while upper layers learn the more complex ones [7]. If more nodes and hidden layers are added, the computing cost may increase significantly. Optimization of multilayer NNs with several hidden layers might be challenging if gradient vanishing is an issue [8]. In DL model development, GPU acceleration is commonly used to overcome this limitation. Additionally, regularization and transfer function methods, as well as adjustments to network designs for weight updates and initialization maximization, were used to prevent overfitting [9]. Networks in this category include CNN, RNN [10], Deep Belief Networks (DBN), and Auto-Encoder [11]. The following shallow ML techniques are compared to DL in the age of big data: logistic regression [12], linear regression [13], simple Bayesian models [14], Support Vector Machines (SVM) [15], and decision trees [16]. There is a significant advantage to the random forest algorithm [15]. Due to their superior learnability, these algorithms are preferred over DL [16]. DL requires more human involvement than traditional ML. Although DL may be more difficult to implement, once it's up and running it needs little maintenance. ML systems require less hardware and resources than DL models. Moreover, the ML algorithms are often simple and run on modest hardware. Processing speed has led to GPU popularity. GPUs have high bandwidth memory and can hide latency in memory transfers due to thread parallelism (the ability to perform multiple processes simultaneously). Consequently, although ML systems are easy to set up, they may not produce the desired results. Even though DL systems take longer to set up, they produce near-instant results (although quality is likely to improve over time). In addition, ML relies on structured data and conventional techniques like linear regression. Researchers use neural networks for DL to process massive amounts of unstructured data. DL benefits include best-in-class performance, scalability, and flexibility. With less data, classical ML performs better than DL in several aspects. Hence, computation complexity has been reduced, costs have been reduced, and interpretation has become easier. DL requires a high number of nonlinear transformation layers to effectively extract and integrate information from data and discern more general patterns. In addition, it allows us to learn abilities that break through barriers we currently face and teach us things that scientists haven't discovered yet [17-19]. It is possible to speed up developments in areas such as voice recognition, image classification [20], drug discovery, natural language processing, and medical simulation with deep learning approaches because they can automatically recover CA-relevant information across numerous processing layers [21-25]. Deep learning methods are particularly useful for creating cutting-edge drugs and pharmaceuticals [26]. This study's objective is to educate academics in the fields of computational chemistry and cheminformatics about the wide range of applications of DL. In short, this paper aims to review different DL architectures and their applications in drug design and development.

1.1. Deep Learning Advantages

Without further human input, DL algorithms may generate unique traits from a subset of those already present in the pharmaceutical training dataset [27]. Accordingly, DL can perform tasks that would otherwise require extensive feature engineering [28]. The ability of DL to process semi-structured and unstructured data is one of its many strengths. Because Deep Neural Networks (DNNs) are capable of

acquiring complex properties and performing intensive computations, such as simultaneously executing several complex processes, they are beneficial to drug design models. Because of the large input data and the sufficient learning required for CNN-based structures, DL models may take longer to learn the parameters of a drug design model. To solve this problem, DL models should be trained much faster using parallel and distributed approaches [29]. In the development of medications, models can be trained either locally (on a single machine) or using Graphics Processing Units (GPUs) [30]. The medical industry might eventually save costs by training DL models for drug discovery and design [31].

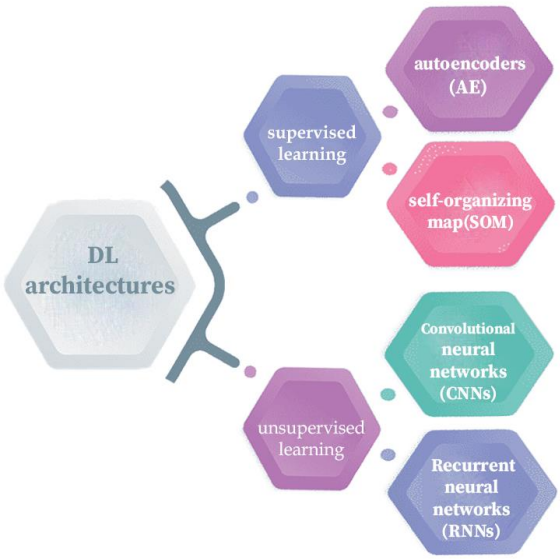

Figure 1. In this figure, a division of the most famous DL structures used in drug design is shown

The cost of inaccurate drug forecasts or defective drugs is relatively high in industries such as medication production [32]. The time and effort invested in developing DL models is often well worth it. When applied to data science, deep learning may produce more accurate and efficient processing models. As a result of its potential for unsupervised learning, precision and efficacy are continuously improved. The results are more actionable and clearer for data scientists as a result [33]. As illustrated in Figure 1, the most famous DL networks are introduced.

As DL can process massive volumes of drug data and perform numerous computations within a short amount of time, it can be used on a large scale. Consequently, productivity (faster rollouts) and modularity and portability (trained models can be applied to a wide range of challenges) are evident.

1.2. Study Motivation

In contrast to classical ML approaches, DL needs to be promoted and supported in the pharmaceutical industry. There is a fundamental difference between machine learning and human-assisted learning: AI needs significant human intervention from the beginning to the end before it can make any meaningful predictions. As deep learning becomes more advanced, it may be able to recognize intricate patterns less supervised once it gets rolling. Learned from raw data even with high dimensions with little guidance [34]. The usage of DL and ML has improved medication discovery and design in extensive investigations [35]. The drug development process must overcome a number of challenges before it can reach its full potential [36]. Predicting drug-target interactions is an important task in drug discovery. In many cases, targets (proteins) have one or more binding sites with substrates or regulatory molecules. It is possible to build predictive models using these data [37]. The inclusion of other protein sites can, however, introduce bias into the analysis [38]. Wang *et al.* [39] used Pair-Input Neural Networks (PINN) to calculate target-ligand interactions based on protein sequences and target profiles. As a result, NNs are better at predicting target-ligand interactions than other methods. There are several prediction algorithms available that can mitigate these issues and save valuable resources when it comes to drug development and assessment. Most often, drugs are discontinued due to toxicities, such as hepatotoxicity. Using computational methods for predicting hepatotoxicity, potentially harmful medications can be avoided. A raw chemical structure can be used to determine compound toxicity using deep learning [40]. In addition to predicting epoxidation, CNN can also predict other properties. Hughes *et al.* used Simple Molecular Input Line Profile Format (SMILES) data to indicate a high degree of reactivity and possible toxicity by using epoxidized molecules and hydroxide molecules as negative controls [41]. To address these concerns and challenges successfully, several aspects of pharmaceutical design must be considered:

- Pharmaceutical designs are generally built in conjunction with automated technologies in industrial

drug trials to facilitate efficient management of screening sets numbering hundreds or even millions of chemicals. Such methods produce enormous amounts of information, which are constantly being used to develop new methods [42].

- Handcrafted features, such as molecular descriptors and fingerprints, were widely used in previous DL drug discovery efforts. These conditions affect DL's capacity to construct molecular properties directly from data. However, this is perhaps one of the most important differences between DNNs and standard ML methods [43].

- The assay procedure is a crucial part of the testing process for compounds. In assessing corresponding data sets, it is important to keep in mind that direct interaction between substances and assay methods can lead to systematic errors [44].

- A given patient population's demographics are influenced by tumor type, illness progression, and co-morbidities. Understanding how various factors affect the drug's efficacy may allow dosing adjustments in larger clinical trials [45].

To achieve optimal results in DL simulation, selecting the proper DL structure and tuning its hyper-parameters are crucial. Hence, the purpose of this study is to shed light on the role that DL has played in both the creation of cutting-edge technology and extremely effective medications [46]. DL structures and significant application domains present a number of limitations and hurdles in the field of drug design [47]. Early research has influenced several aspects of drug development discussed in the article. In drug design and drug discovery, CNNs and RNNS are two common and successful types of NNs based on ML and DL.

2. Principle of DL

DL is a group of approaches under ML that employ ANNs to imitate the higher-level abstractions present in data from several layers of nonlinear processing units. Similar to the architecture of the human brain, ANNs are composed of linked nodes. The interconnectedness of several neurons, each of which may be seen as a processing unit, generates an enormous amount of computing power capable of completing complex computations. In 1943, McCulloch and Pitts presented the first artificial neuron model using a computer model of brain activity [48, 49].

Since 2006, when Geoffrey Hinton, Ian Likan, and others proposed a realistic DL framework, businesses and universities have been swept by an inventive wave of DL and new AI [50]. To solve the problem of DL's abstract data requirements, a unique structure for multilayer NNs used in feature learning was developed. By use of feature learning, DL approaches may automatically extract features from raw format input data, modify those features, and transfer them to higher levels of abstraction [51]. Due to the significant advancements in computer technology and parallel computing strategies, such as the Tensor Processing Unit (TPU) approach, the computationally intensive nature of DNNs may no longer be prohibitively expensive [52, 53]. To distinguish patterns and extract high-level features, DL architectures vary in kind and function according to the training data's structure. Here, we concentrated on the most prevalent architectures, including CNNs, RNNs, DNN, and GAN [54].

2.1. Convolutional Neural Networks

CNNs are commonly utilized in machine learning for image processing in deep learning. CNNs are often called Shift Invariant or Space Invariant ANNs (SIANN) due to their shared-weight structure that slides along input features to generate translation-equivariant outputs. The input-down sampling process of convolutional neural networks does not make them translation-in variants, contrary to popular belief. They can be used for video and image classification, decision-making systems, image understanding and segmentation, time series prediction, Brain-Computer Interfaces (BCI), Natural Language Processing (NLP), and medical image analysis. CNNs are inspired by the human visual cortex [55] since it, too, has neurons that are selectively tuned to distinct areas of the visual field. This notion was examined by Hubel and Wiesel [56]. Fukushima detection in the 1980s led to CNN, which was inspired by Hubel and Wiesel's studies of the cat's visual cortex receptive field [57, 58].

Figure 2 shows the three most common types of CNN layers: convolution, pooling, and full connection. Based on these [59, 60], a multilayer

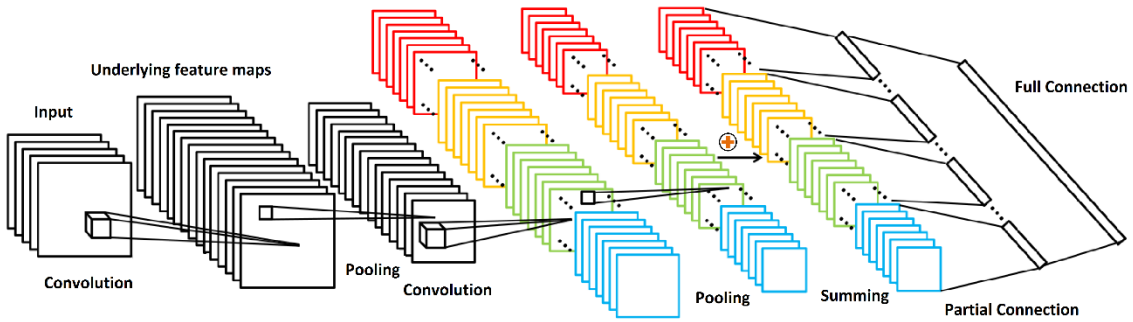

Figure 2. A general convolutional structure for classification that can be used in a variety of applications can be seen in this figure

network was constructed. It is possible to study a variety of layers depending on the data intake strategy. It is possible to form layers for sequential signals, such as language, using one-dimensional arrays. The use of 3D arrays for creating movie layers is possible [61].

2.2. Recurrent Neural Networks

DL can be implemented using RNNs as an alternative architecture. As such, RNNs [62] are a specific type of ANN in which the linkages between nodes always point in the same direction. Therefore, the internal state of the network may be altered, potentially leading to dynamic temporal behavior. Instead of passing signals from one layer to the next, as in feedforward networks, RNNs feature feedback components that suppress them. Furthermore, due to their internal memory, they are the only form of a neural network capable of storing long-term learning data [63].

A variety of RNNs are available. A gated recurrent unit RNN (GRURNN) is a network that utilizes recurrent units of time (RLSTMs or short-term memories), and a clock-RNN is a network that uses a recurrent unit of time (CW-RNN) [64-66]. The most widely accepted and used architectures for natural language processing in recent years have been RNNs and LSTMs. Natural Language Processing (NLP) usually combines LSTM with scattered embedding derived from part-of-speech tagging [67, 68]. Since they employ a custom function to create the transition state in the hidden layer, Long Short-Term Memory (LSTM) architectures outperform conventional RNNs at identifying long-term dependencies. In artificial intelligence, LSTMs are often used alongside CNNs to automatically generate image descriptions, and they have shown to be equally successful and popular as CNNs at retrieving images [69]. Three commonly used RNN architectures are shown in Figure 3.

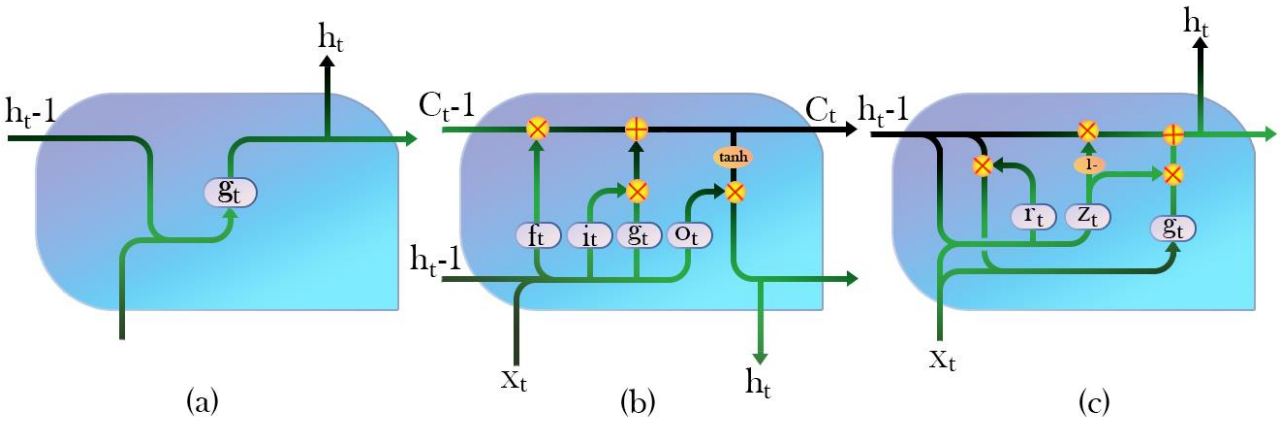

Figure 3. These are the inner architecture of three types of RNNs: (a) RNN model; (b) LSTM architecture; and (c) gated recurrent units (GRU) structure

2.3. Deep Neural Networks

Unsupervised learning uses DNNs for analyzing unlabeled data, in addition to their use in supervised learning. A popular generative network architecture for unsupervised learning is the Deep Autoencoder Network (DEAN) [70, 71]. Hilton *et al.*'s proposed DNN has a mirror image encoder and decoder [72]. DBNs are constructed using Restricted Boltzmann Machines (RBMs) [73], a bipartite network with an undetectable layer and a detectable layer. Symmetric links can only be established between nodes on different levels; links between nodes on the same level are not possible. The BP technique compresses data with the efficiency of an autoencoder, with minimal information loss and full recovery [74]. The ability of DAEN to reduce superfluous information qualifies it as a dimensionality reduction method. By reducing the recovered features, DAEN [74] can be employed to develop a classification structure utilizing supervised learning. The development of DL software can be advanced through this strategy.

2.4. GAN Architecture

Video processing, audio processing, music analysis, image processing, computer vision, and medical imaging are among the numerous applications of Generative Adversarial Networks (GANs) [75, 76]. GAN architectures have also been applied to multi-omics, medical informatics, bioinformatics, medication delivery, and drug discovery in biology [77, 78]. The discriminative network module and the generating network module support a GAN architecture [79]. Both discriminative and generative networks are simultaneously taught as multi-layered ANNs. Unlike the generative network module, which generates fake samples based on the hidden variable, the discriminant network module receives both fake and real samples. The precision of an input is established by verifying that it is verifiable. Whenever a discriminant network module concludes that a sample is more likely to be accurate, it generates more accurate forecasts [80]. The generative network module increases the error probability of the discriminant network module with the assistance of the generative network module. As a result, both positive and destructive networks work together to achieve their goals. In GANs, generating and discriminant structural modules compete for resources [81].

The GAN's generative network module will only outperform other systems if it generates consistent output values (e.g., a vector of numbers creating an image). By using the discriminant network module's loss function gradient, we can train the generative network module and fine-tune its weight. Chemical structures (e.g., molecular compounds) are represented by text strings (e.g., SMILES) or graphical representations (e.g., molecular diagrams). Using GANs for chemistry (or informatics chemistry) poses a significant challenge [29].

There are several types of GAN architectures (e.g., GAN frameworks). Research has examined the Wasserstein GAN, the conditioned GAN, and the deep adversarial autoencoder. Arjovsky *et al.* [81] introduced a Wasserstein GAN structure leveraging Earth-Mover distance (also known as Wasserstein distance) to overcome the previously noted GAN training instability. Originally, GAN relied on Jensen-Shannon divergence, which cannot be used to estimate the distance between distributions that do not overlap [82]. Conditional GANs based on extra input, such as class labels, were proposed by Mirza *et al.* [83]. The conditional GAN architecture (like class labels) uses a conditional variable instead of a conventional GAN design. Generated and discriminative networks both use conditional variables as inputs. A discriminative network module assigns classes to data, while a generative network module generates fake samples using hidden variables. Deep adversarial autoencoders [84] are subsets of GAN-based frameworks [85] that use GAN architecture to transform an autoencoder structure into a generative model. Autoencoders and adversarial networks make up deep adversarial autoencoders [86]. An auto-encoder module consists of both an encoder and a decoder. An encoder and a generative network module make up a deep adversarial encoding structure [87]. In addition, the adversarial network and autoencoder modules can run simultaneously under the deep adversarial autoencoder design. An adversarial autoencoder architecture [88] aims to ensure compatibility between the hidden prior distribution and the hidden data generated by a generative network module.

3. Molecular Properties Prediction

The DL method has been used to predict the efficacy of drugs on several occasions. A number of studies have shown that DL can predict physicochemical characteristics, ADMET properties, and biological activity as well as other ML approaches. The first time DL was used in the drug discovery process was in 2012.

A multitask deep feedback system outperformed Merck's own algorithms by a large margin (15%) in a competition to estimate the properties and activities of many medications. A data fusion network's effectiveness depends on the design of its architecture and the activation function, i.e., the number of hidden layers and the number of neurons in each layer [89, 90].

To classify the toxicity of substances, Mayer *et al.* proposed an ensemble-based DeepTox pipeline model. A dataset of 12,000 environmental chemicals and medicines with up to 12 unique toxicology endpoints was used for 2014 Tox21 toxicology prediction challenge [91]. Our model outperforms previous ML techniques in nine of twelve hazardous outcomes. In several subsequent studies [92-94], ensemble multitask DL architectures were found to be superior to the Random Forest (RF) and single-task models for feature prediction. Furthermore, ChEMBL's temporal cross-validation and random split were used to prove the superiority of DL [95]. To classify drugs into therapeutic categories, Alper *et al.* employed DL methods trained on transcriptional response datasets to examine transcriptional patterns in a range of cell lines [96]. Some ML and chemoinformatics algorithms use graphs (such as molecular structure) instead of predefined chemical descriptors. To precisely classify the solubility of compounds in water, Lusci and collaborators developed UG-RNN (Undirected Recurrent Graph Neural Networks) [97]. As a result of this strategy, the correct chemical descriptors can be automatically learned from the data without requiring extensive selection. The same method was used by Xu *et al.* [98] to model Drug-Induced Liver Injury (DILI). In addition to 475 drugs used to train the model, 198 pharmaceuticals were tested. In recent years, CNN has also been used to replicate chemical data.

A continuous molecular fingerprint was generated from molecular graphs using CNN structure by Duvenaud *et al.* [99]. Using a CNN embedding of attributed molecular graphs (biological toxicity, melting point, and octanol and aqueous solubility), Coley and colleagues improved the predictive power of Duvenaud's structure. For generating the attributed graph representation of molecules, the architecture utilized a molecular tensor composed of bond-level and atom-level features [100].

In the training of RNN models, SMILES strings are used by a variety of methods. To generate predictive models without generating chemical descriptions, Bjerrum proposed using SMILES enumeration, a one-line string identifying a molecule, as the unprocessed input to an LSTM cell-based NN [101]. Quantitative Structure–Activity Relationship (QSAR) models constructed using enumerated SMILES are statistically more reliable, both for predicting individual SMILES and for averaging predictions made using enumerated SMILES. Researchers have used two-dimensional molecule photographs with minimal chemical data to build structures that are, on average, comparable to DNN architecture trained on molecular fingerprints [102, 103].

4. DL in De Novo Design

Based on practical tests, Figure 4 illustrates the stages of drug design using ML models.

The principles of deep learning for the design of de novo drugs are similar to those of conventional de novo drug discovery [104, 105]. DL has been used to build molecules from scratch and retrieve their characteristics. A combination of Variational AE (VAE) and Molecular Logistic Programming (MLP) demonstrated the autonomous production of molecules with the desired attributes by Gomez-Bombarelli *et al.* [106]. Using a collection of anti-tumor drugs with varying potencies, Kadurin *et al.* [107] trained a dual-purpose adversarial AE. Using the developed model, molecular fingerprints with the necessary characteristics were produced. The chemical fingerprints of the generated compounds were startlingly similar to those of well-established, very effective anticancer drugs, such as anthracyclines. By considering other molecular qualities, such as solubility, authors introduced a

superior design called druGAN, which could produce more chemically different molecules [108]. This unique approach may benefit the pharmaceutical industry due to the improvements in feature extraction, generating capacity, and reconstruction error. By using antagonistic AE, Blaschke *et al.* [109] created drugs that target dopamine type 2 receptors. Remember that mode collapse [110, 111] can have negative consequences, resulting in a limited variety of molecules being synthesized. Some models avoid mode collapse [112, 113], but a reliable method still needs to be developed to assess the internal diversity of the molecules generated. RNNs are increasingly used for de novo chemical design. Using SMILES, a string-based representation of chemical graphs, Segler *et al.* [99] developed unique molecule libraries by combining transfer and reinforcement learning. LSTM models were either fine-tuned on a small number of known actives or coupled with external scoring mechanisms to generate new compounds with favorable activity against a particular biological target. An LSTM-based RNN strategy was extended by Gupta *et al.* [114] for both de novo and fragment-based drug discovery. Reinforcement learning has been used by Guimaraes *et al.*, Olivecrona *et al.*, and Jaques *et al.* to impose desirable properties on molecular generative models [115-117].

Drug designs in this section have been developed in conjunction with automated technologies in industrial drug testing [42]. This is to facilitate the efficient management of screening sets containing hundreds or even millions of chemicals. This limitation, which involves a huge amount of drugs and chemical compounds, demonstrates that DL is a method that produces a huge amount of information [118]. By deepening their structures, adjusting different parameters, the option of entering more data, and the use of more powerful hardware, it makes it easier to design drugs. Additionally, according to the results of different approaches in this sector, DL models have been able to produce reliable results with less time and cost [119].

5. CNN Applications in Drug Discovery

Deep CNNs are often used to learn the 3D structures of proteins and other small molecules [120, 121]. Structure-based drug design relies heavily on therapeutic target scoring. A drug-target complex's binding posture and affinity can extend the success rate of virtual screening procedures [122, 123]. In comparison to other well-performing scoring functions built using linear and nonlinear techniques, CNN scoring functions demonstrate higher performance in predicting drug-target complex pose/affinity and detecting active/inactive molecules.

A CNN that learns from protein structure data and can understand proteins as 2D and 3D images was developed by Jimenez *et al.* [124]. It performed better than other approaches when used to find ligand-binding sites. KDEEP-based 3D graph CNN models were used by the same authors to forecast ligand-protein binding affinities in 2018 [125]. A Pearson correlation of 0.82 and an RMS error of 1.27 were obtained for the model. Using CNN-based end-to-end (e2e) models, Zhao *et al.* [126] developed a drug-target affinity prediction model. A model is trained

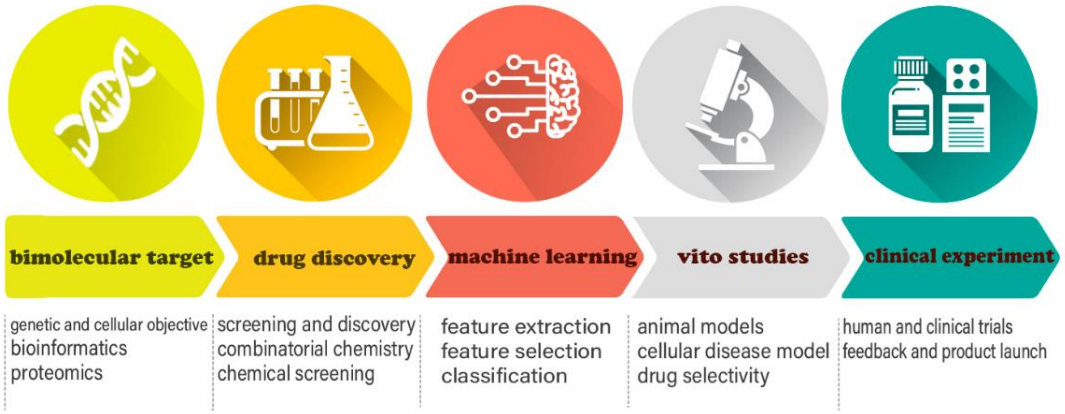

Figure 4. This Figure illustrates how ML models can be used to design drugs based on practical tests

using medical drug SMILES sequences and protein amino acid sequences. A fully connected NN is employed to assess binding affinities, while a one-dimensional Convolutional Neural Network (CNN) and attention processes are utilized to reconstruct representations between proteins and drugs. It improves drug-target affinity prediction by training the attention processes to focus on the most important parts of drug and protein sequences.

By using CNNs with more appropriate architectures and enriched with convolutional learning models, the limitations in the models proposed by the researchers in this field have been overcome. Furthermore, researchers have developed e2e decision-making structures and optimized configurations through classical algorithms, fine-tuned the architectures, applied fusion techniques, and expanded the layers and depth of the models [127].

6. Applications of DL in Drug Chemical Synthesis

Using ML and DL algorithms to predict chemical process outcomes [128-131]. Classification of response types [132, 133] and automated identification of the reaction center are typical statistical learning methods used by these instruments [134]. In retrosynthesis analysis, most DL approaches adopt similar principles; the primary difference lies in the molecular representations used. A sequence-to-sequence (seq2seq) structure developed by Liu *et al.* [135] transforms a SMILES string representing a product into a SMILES string indicating a reactant. An LSTM network was trained on 50,000 responses from US patents, achieving parity with rule-based techniques. The authors of 2018 [136] proposed combining three DNNs with a Monte Carlo Tree Search (MCTS) architecture to predict chemical reactions. Reaxys database was used to train this algorithm, which now generates ideas at a rate and quality comparable to human-driven synthesis. We compared the different deep learning architectures and their uses in drug discovery and design in Table 1.

To evaluate the result of a reaction, Coley *et al.* [137] used a NN to rank the candidates created by their use of reaction templates derived from US patents. Weisfeller-Lehman Network (WLN), a graph CNN (GCNN) that analyzes data about the neighbors of each atom, was used to grade applicants' responses. Due to the model's superior performance over template-based techniques and its greatly increased speed, it was able to apply it to a dataset containing close to 400,000 answers. In the same way, artificial intelligence systems perform admirably when applied to existing chemical processes. As part of an autonomous framework for optimizing chemical reaction conditions, Zhou *et al.* [138] proposed the Deep Reaction Optimizer (DRO), a model based on deep reinforcement learning. After being subjected to real-world reactions, the algorithm performed better.

7. Limitation of DL in Drug Design

DL can achieve high identification accuracy thanks to advances in feature learning, although a large training set is still needed for maximum performance. DL methods may not be as useful as conventional shallow ML techniques on sparse data due to their inability to produce meaningful estimates of generalization [139, 140]. DL techniques require improved hardware capabilities and programming skills, as the complexity of network methods increases the temporal complexity of DL techniques. The majority of data in real-world uses, such as medication research, are unlabeled despite their tremendous informational richness. It remains challenging to adapt algorithms for chemistry-centric modelling in small molecule drug development, especially for CNNs and RNNs, both of which are powerful but have substantial restrictions on input information formats. In contemporary cheminformatics research, molecular fingerprints [141-144], physicochemical properties, topological features, and thermodynamic properties [145, 146] are all considered in traditional ML models. In particular, DL modelling presents two crucial challenges: (1) how to design DL structures that abstract relevant features, and (2) how to interpret those features, given that DL approaches are a sort of representation learning that might automatically abstract features from raw data. The number of chemoinformatics data for DL modelling is much smaller than the amount of massive data employed to train DL architectures such as AlphaGo. It is difficult to develop individual models from big databases such as ChEMBL because there is a paucity of relevant data [147].

Table 1. This table compares the different DL models and their uses in drug discovery and delivery or design

Ref.	DL model	Target	The reason
[89]	DNN	Prediction of drug activities and properties	In this study, it was demonstrated that DNN performance depends on both the network architecture and the activation function.
[90]	DeepTox	toxicity prediction	In DeepTox, models are generated, evaluated, and ensembles of promising results are gathered. A new chemical's toxicity can now be predicted using DeepTox.
[96]	DNN	Classification of different drugs	Training deep learning algorithms on huge transcriptional response datasets using transcriptional patterns in different cell lines.
[97]	UG-RNN	Effective prediction of solubility of medicinal compounds in water	The advantage of this approach is that it can be used to identify suitable molecular descriptors
[98]	CNN	Chemical data modeling of drugs	Using CNN to create continuous molecular fingerprints directly from molecular graphs
[99]	CNN	Prediction of molecular properties	This algorithm employed a molecular tensor that combined bond-level and atomic-level properties to depict the ascribed graph representation of molecules.
[100]	RNN	Prediction without the need to create molecular descriptors	Using SMILES counts as raw input to a cell-based LSTM NN to create predictive architectures
[101]	DNN	Molecular fingerprint	Using two-dimensional molecular images with minimal chemical information to improve algorithms for molecular fingerprint identification
[124]	CNN	Detection of ligand-binding sites	Protein structural identification as 3D images
[126]	CNN	Predicting drug-target preference	A CNN-based algorithm called Attention DTA added a mechanism to the process of predicting drug target liking.
[136]	DNN	Prediction of chemical synthesis	The model was trained on 12 million reactions from the Reaxys dataset and made super-fast recommendations for chemical synthesis predictions.

Table 2. In this table, we compare various widely used deep learning frameworks used in the field of drug design based on their advantages and disadvantages

DL methods	Advantages	Disadvantages
AlexNet	-Having more network depth by adding convolution layers -Faster training of models -Able to extract features	AlexNet is not deep enough compared to later models such as VGGNet, GoogLENet and ResNet.
GoogleNet	-GoogleNet is faster compared to other image classification models such as VGG. -GoogleNet is much more compact compared to other architectures	A drastic change from sequential architectures
Spatial Pyramid Pooling (SPP)	-They remove the limitation of the fixed size of the network, i.e., CNN does not need a fixed size input image.	-It is very heavy in terms of processing
Deep Belief Network (DBN)	- Efficient in using hidden layers - It has a special strength in classification - It can be used on different programs and data types	- Hardware requirements - The need for massive data to perform better techniques - The expensiveness of DBN training - The complexity of the model Difficulty using DBN
Deep Boltzman	-All variables of a layer can be updated in parallel -It is faster than standard Gibbs sampling -Provides examples of sufficient quality for quick learning.	The most important disadvantage of Deep Boltzmann machines is that it has approximate inference.
Sprase Auto	- It can effectively extract data features, easily classify and extract stronger data features. -Compared with traditional BP neural network, sparse autoencoders can be used for effective error detection	It relies heavily on engineering experiments and signal analysis experience.
Sparse coding	-Excellent at classifying images -Total energy consumption decreases with increasing dispersion.	Need a large number of pattern images to recognize the target

There will be an increase in complexity in the creation of small-molecule drugs. A DL system must be capable of handling complex simulations, given its intended application. With DL techniques, it may be possible to integrate all information systems and reach a new level of AI in drug development. We should not limit ourselves to conventional predictions on biological activities, ADMET features, and pharmacokinetic simulations [148-150]. We have compared various widely used deep learning frameworks used in the field of drug design in previous works in Table 2.

8. Conclusion

The pharmaceutical industry has been profoundly impacted by recent advancements in AI, such as the expansion and growth of more complex ML approaches. In the development of drugs, artificial intelligence can alleviate some of the most significant challenges through in silico strategies for de novo drug synthesis, design, classification, and bioactivity categorization. In addition to training models on a single machine, GPUs can also be used to train models. In the future, the medical industry may save costs by training DL models for drug discovery and design. Accurate drug forecasts and defective drugs can be costly in industries such as medication production. Because of their unsupervised learning capability, DL models are often well worth the time and effort they require. In this way, data scientists can make more informed decisions based on the results. The article examines the theoretical underpinnings and recent, practical uses of DL in drug development and chemoinformatics. The success of conventional ML techniques, which rely on feature extraction and engineering, depends on human interaction. It is possible for deep learning to learn most or all of the features automatically if there are enough instances in the training data (many millions). To find more nuanced and consistent characteristics in input signals, DL models employ feature detector units' layer by layer. At lower levels, information is gathered to train the higher levels to recognize increasingly complex characteristics. As opposed to this, traditional ML models often only display a few mapping layers from input characteristics to a problem-specific feature space. As opposed to traditional ML methods, DL methods can be employed across a broad scope of settings. TL can repurpose DL networks that have been trained for multiple tasks in the same domain. We wouldn't be able to say that DL is better than traditional ML even if we enumerated all its benefits. As of right now, there isn't a preferred method for fixing ML problems. If the issue contains mixed sets of chemical or target protein input descriptors, DL performs as well as traditional ML models. DL has been demonstrated to be much more effective at tasks such as classifying and analyzing biomedical images, predicting biological activity, and creating new molecules. This suggests that DL and ML can work together to extend the reach of drug discovery.

Acknowledgments

No funding was provided to the authors during any stage of the article's creation.

References

1- J. W. Scannell, A. Blanckley, H. Boldon, and B. Warrington, "Diagnosing the decline in pharmaceutical R&D efficiency." *Nat Rev Drug Discov*, Vol. 11 (No. 3), pp. 191-200, Mar 1 (2012).

2- D. Krentzel, S. L. Shorte, and C. Zimmer, "Deep learning in image-based phenotypic drug discovery." *Trends Cell Biol*, Jan 7 (2023).

3- S. Kim *et al.*, "PubChem Substance and Compound databases." *Nucleic Acids Res*, Vol. 44 (No. D1), pp. D1202-13, Jan 4 (2016).

4- E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence." *Nat Med*, Vol. 25 (No. 1), pp. 44-56, Jan (2019).

5- B. B. Hansen *et al.*, "Deep Eutectic Solvents: A Review of Fundamentals and Applications." *Chem Rev*, Vol. 121 (No. 3), pp. 1232-85, Feb 10 (2021).

6- E. J. Pérez-Pérez, F. R. López-Estrada, G. Valencia-Palomo, L. Torres, V. Puig, and J. D. Mina-Antonio, "Leak diagnosis in pipelines using a combined artificial neural network approach." *Control Engineering Practice*, Vol. 107(2021).

7- Connor Shorten and Taghi M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning." *Journal of Big Data*, Vol. 6 (No. 1), (2019).

8- C. Ma, L. Wang, and X. Q. Xie, "GPU accelerated chemical similarity calculation for compound library comparison." *J Chem Inf Model*, Vol. 51 (No. 7), pp. 1521-7, Jul 25 (2011).

9- Adam P. Piotrowski, Jaroslaw J. Napiorkowski, and Agnieszka E. Piotrowska, "Impact of deep learning-based dropout on shallow neural networks applied to stream

temperature modelling." *Earth-Science Reviews*, Vol. 201(2020).

10- Qifeng Bai *et al.*, "Application advances of deep learning methods for de novo drug design and molecular dynamics simulation." *WIREs Computational Molecular Science*, Vol. 12 (No. 3), (2021).

11- G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets." *Neural Comput*, Vol. 18 (No. 7), pp. 1527-54, Jul (2006).

12- K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine Learning in Agriculture: A Review." *Sensors (Basel)*, Vol. 18 (No. 8), Aug 14 (2018).

13- X. Yang, Y. Wang, R. Byrne, G. Schneider, and S. Yang, "Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery." *Chem Rev*, Vol. 119 (No. 18), pp. 10520-94, Sep 25 (2019).

14- Yaguo Lei, Bin Yang, Xinwei Jiang, Feng Jia, Naipeng Li, and Asoke K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap." *Mechanical Systems and Signal Processing*, Vol. 138(2020).

15- H. Bennett-Lenane, B. T. Griffin, and J. P. O'Shea, "Machine learning methods for prediction of food effects on bioavailability: A comparison of support vector machines and artificial neural networks." *Eur J Pharm Sci*, Vol. 168p. 106018, Jan 1 (2022).

16- G. B. Goh, N. O. Hodas, and A. Vishnu, "Deep learning for computational chemistry." *J Comput Chem*, Vol. 38 (No. 16), pp. 1291-307, Jun 15 (2017).

17- C. Angermueller, T. Parnamaa, L. Parts, and O. Stegle, "Deep learning for computational biology." *Mol Syst Biol*, Vol. 12 (No. 7), p. 878, Jul 29 (2016).

18- T. Zhang, J. Leng, and Y. Liu, "Deep learning for drug-drug interaction extraction from the literature: a review." *Brief Bioinform*, Vol. 21 (No. 5), pp. 1609-27, Sep 25 (2020).

19- A. Lavecchia, "Deep learning in drug discovery: opportunities, challenges and future prospects." *Drug Discov Today*, Vol. 24 (No. 10), pp. 2017-32, Oct (2019).

20- G. E. Dahl, Yu Dong, Deng Li, and A. Acero, "Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition." *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 20 (No. 1), pp. 30-42, (2012).

21- Y. Yu *et al.*, "A Novel Scalarized Scaffold Hopping Algorithm with Graph-Based Variational Autoencoder for Discovery of JAK1 Inhibitors." *ACS Omega*, Vol. 6 (No. 35), pp. 22945-54, Sep 7 (2021).

22- J. Schmidhuber, "Deep learning in neural networks: an overview." *Neural Netw*, Vol. 61pp. 85-117, Jan (2015).

23- T. Ching *et al.*, "Opportunities and obstacles for deep learning in biology and medicine." *J R Soc Interface*, Vol. 15 (No. 141), Apr (2018).

24- Liuying Wang *et al.*, "Advances of Artificial Intelligence in Anti-Cancer Drug Design: A Review of the Past Decade." *Pharmaceuticals*, Vol. 16 (No. 2), (2023).

25- D. M. Walden, Y. Bundey, A. Jagarapu, V. Antontsev, K. Chakravarty, and J. Varshney, "Molecular Simulation and Statistical Learning Methods toward Predicting Drug-Polymer Amorphous Solid Dispersion Miscibility, Stability, and Formulation Design." *Molecules*, Vol. 26 (No. 1), Jan 1 (2021).

26- R. Gupta *et al.*, "OdoriFy: A conglomerate of artificial intelligence-driven prediction engines for olfactory decoding." *J Biol Chem*, Vol. 297 (No. 2), p. 100956, Aug (2021).

27- O. T. Jones *et al.*, "Artificial Intelligence Techniques That May Be Applied to Primary Care Data to Facilitate Earlier Diagnosis of Cancer: Systematic Review." *J Med Internet Res*, Vol. 23 (No. 3), p. e23483, Mar 3 (2021).

28- L. Chen, X. Pan, Y. H. Zhang, M. Liu, T. Huang, and Y. D. Cai, "Classification of Widely and Rarely Expressed Genes with Recurrent Neural Network." *Comput Struct Biotechnol J*, Vol. 17pp. 49-60, (2019).

29- N. Hampe *et al.*, "Deep learning-based detection of functionally significant stenosis in coronary CT angiography." *Front Cardiovasc Med*, Vol. 9p. 964355, (2022).

30- V. B. Mathema, P. Sen, S. Lamichhane, M. Oresic, and S. Khoomrung, "Deep learning facilitates multi-data type analysis and predictive biomarker discovery in cancer precision medicine." *Comput Struct Biotechnol J*, Vol. 21pp. 1372-82, (2023).

31- S. Luukkonen, H. W. van den Maagdenberg, M. T. M. Emmerich, and G. J. P. van Westen, "Artificial intelligence in multi-objective drug design." *Curr Opin Struct Biol*, Vol. 79p. 102537, Apr (2023).

32- S. Shukar *et al.*, "Drug Shortage: Causes, Impact, and Mitigation Strategies." *Front Pharmacol*, Vol. 12p. 693426, (2021).

33- G. Latif, S. E. Abdelhamid, R. E. Mallouhy, J. Alghazo, and Z. A. Kazimi, "Deep Learning Utilization in Agriculture: Detection of Rice Plant Diseases Using an Improved CNN Model." *Plants (Basel)*, Vol. 11 (No. 17), Aug 28 (2022).

34- A. Churcher *et al.*, "An Experimental Analysis of Attack Classification Using Machine Learning in IoT Networks." *Sensors (Basel)*, Vol. 21 (No. 2), Jan 10 (2021).

35- M. Yang *et al.*, "Prediction of biomarker-disease associations based on graph attention network and text representation." *Brief Bioinform*, Vol. 23 (No. 5), Sep 20 (2022).

36- T. W. Laetsch, S. G. DuBois, J. G. Bender, M. E. Macy, and L. Moreno, "Opportunities and Challenges in Drug Development for Pediatric Cancers." *Cancer Discov*, Vol. 11 (No. 3), pp. 545-59, Mar (2021).

37- A. A. Kalinin *et al.*, "Deep learning in pharmacogenomics: from gene regulation to patient stratification." *Pharmacogenomics,* Vol. 19 (No. 7), pp. 629-50, May (2018).

38- M. M. A. Monshi, J. Poon, and V. Chung, "Deep learning in generating radiology reports: A survey." *Artif Intell Med,* Vol. 106p. 101878, Jun (2020).

39- Y. Chen, Z. Lin, X. Zhao, G. Wang and Y. Gu, "Deep learning-based classification of hyperspectral data.", *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Vol. 7 (No. 6), pp. 2094-2107, Jun (2014).

40- C. Qu, B. I. Schneider, A. J. Kearsley, W. Keyrouz, and T. C. Allison, "Predicting Kovats Retention Indices Using Graph Neural Networks." *J Chromatogr A,* Vol. 1646p. 462100, Jun 7 (2021).

41- T. B. Hughes, G. P. Miller, and S. J. Swamidass, "Modeling Epoxidation of Drug-like Molecules with a Deep Machine Learning Network." *ACS Cent Sci,* Vol. 1 (No. 4), pp. 168-80, Jul 22 (2015).

42- D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, and R. K. Tekade, "Artificial intelligence in drug discovery and development." *Drug Discov Today,* Vol. 26 (No. 1), pp. 80-93, Jan (2021).

43- D. Baptista, J. Correia, B. Pereira, and M. Rocha, "Evaluating molecular representations in machine learning models for drug response prediction and interpretability." *J Integr Bioinform,* Vol. 19 (No. 3), Sep 1 (2022).

44- Z. Qriouet, Z. Qmichou, N. Bouchoutrouch, H. Mahi, Y. Cherrah, and H. Sefrioui, "Analytical Methods Used for the Detection and Quantification of Benzodiazepines." *J Anal Methods Chem,* Vol. 2019, p. 2035492, (2019).

45- R. J. Tyson *et al.*, "Precision Dosing Priority Criteria: Drug, Disease, and Patient Population Variables." *Front Pharmacol,* Vol. 11, p. 420, (2020).

46- W. Chen, X. Liu, S. Zhang, and S. Chen, "Artificial intelligence for drug discovery: Resources, methods, and applications." *Mol Ther Nucleic Acids,* Vol. 31pp. 691-702, Mar 14 (2023).

47- M. Batool, B. Ahmad, and S. Choi, "A Structure-Based Drug Discovery Paradigm." *Int J Mol Sci,* Vol. 20 (No. 11), Jun 6 (2019).

48- Junxiong Wu, Xiaochuan Chen, Wei Fan, Xiaoyan Li, Yiu-Wing Mai, and Yuming Chen, "Rationally designed alloy phases for highly reversible alkali metal batteries." *Energy Storage Materials,* Vol. 48, pp. 223-43, (2022).

49- Md Zahangir Alom *et al.*, "A State-of-the-Art Survey on Deep Learning Theory and Architectures." *Electronics,* Vol. 8 (No. 3), (2019).

50- Caiming Zhang and Yang Lu, "Study on artificial intelligence: The state of the art and future prospects." *Journal of Industrial Information Integration,* Vol. 23(2021).

51- R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges." *Brief Bioinform,* Vol. 19 (No. 6), pp. 1236-46, Nov 27 (2018).

52- L. Maziarka, A. Pocha, J. Kaczmarczyk, K. Rataj, T. Danel, and M. Warchol, "Mol-CycleGAN: a generative model for molecular optimization." *J Cheminform,* Vol. 12 (No. 1), p. 2, Jan 8 (2020).

53- J. Meyers, B. Fabian, and N. Brown, "De novo molecular design and generative models." *Drug Discov Today,* Vol. 26 (No. 11), pp. 2707-15, Nov (2021).

54- G. Q. Sun *et al.*, "Impacts of climate change on vegetation pattern: Mathematical modeling and data analysis." *Phys Life Rev,* Vol. 43, pp. 239-70, Dec (2022).

55- D. Singh, V. Kumar, Vaishali, and M. Kaur, "Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks." *Eur J Clin Microbiol Infect Dis,* Vol. 39 (No. 7), pp. 1379-89, Jul (2020).

56- L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions." *J Big Data,* Vol. 8 (No. 1), p. 53, (2021).

57- Luning Sun, Han Gao, Shaowu Pan, and Jian-Xun Wang, "Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data." *Computer Methods in Applied Mechanics and Engineering,* Vol. 361(2020).

58- I. Castiglioni *et al.*, "AI applications to medical images: From machine learning to deep learning." *Phys Med,* Vol. 83pp. 9-24, Mar (2021).

59- A. C. Turkmen, T. Januschowski, Y. Wang, and A. T. Cemgil, "Forecasting intermittent and sparse time series: A unified probabilistic framework via deep renewal processes." *PLoS One,* Vol. 16 (No. 11), p. e0259764, (2021).

60- Xuan Cao, Yanwei Zhang, Song Lang, and Yan Gong, "Swin-Transformer-Based YOLOv5 for Small-Object Detection in Remote Sensing Images." *Sensors,* Vol. 23 (No. 7), (2023).

61- Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning." *Nature,* Vol. 521 (No. 7553), pp. 436-44, May 28 (2015).

62- R. Thirunavukarasu, G. P. D. C, G. R, M. Gopikrishnan, and V. Palanisamy, "Towards computational solutions for precision medicine based big data healthcare system using deep learning models: A review." *Comput Biol Med,* Vol. 149, p. 106020, Oct (2022).

63- Chenguang Song, Kai Shu, and Bin Wu, "Temporally evolving graph neural network for fake news detection." *Information Processing & Management,* Vol. 58 (No. 6), (2021).

64- A. Fotiadis, S. Polyzos, and T. T. C. Huan, "The good, the bad and the ugly on COVID-19 tourism recovery." *Ann Tour Res,* Vol. 87, p. 103117, Mar (2021).

65- Ha Nguyen, "Role design considerations of conversational agents to facilitate discussion and systems thinking." *Computers & Education*, Vol. 192(2023).

66- Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, and J. Liu, "Human Action Recognition From Various Data Modalities: A Review." *IEEE Trans Pattern Anal Mach Intell*, Vol. PPJun 14 (2022).

67- T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images." *Comput Biol Med*, Vol. 121, p. 103792, Jun (2020).

68- K. Ahmad, A. Rizzi, R. Capelli, D. Mandelli, W. Lyu, and P. Carloni, "Enhanced-Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective." *Front Mol Biosci*, Vol. 9, p. 899805, (2022).

69- P. Yao *et al.*, "Fully hardware-implemented memristor convolutional neural network." *Nature*, Vol. 577 (No. 7792), pp. 641-46, Jan (2020).

70- Qian Lv, Xiaoling Yu, Haihui Ma, Junchao Ye, Weifeng Wu, and Xiaolin Wang, "Applications of Machine Learning to Reciprocating Compressor Fault Diagnosis: A Review." *Processes*, Vol. 9 (No. 6), (2021).

71- Abid Haleem, Mohd Javaid, and Ibrahim Haleem Khan, "Current status and applications of Artificial Intelligence (AI) in medical field: An overview." *Current Medicine Research and Practice*, Vol. 9 (No. 6), pp. 231-37, (2019).

72- Soo Young Lee, Seokyeong Byeon, Hyoung Seop Kim, Hyungyu Jin, and Seungchul Lee, "Deep learning-based phase prediction of high-entropy alloys: Optimization, generation, and explanation." *Materials & Design*, Vol. 197(2021).

73- O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey." *Heliyon*, Vol. 4 (No. 11), p. e00938, Nov (2018).

74- Shota Harada, Hideaki Hayashi, and Seiichi Uchida, "Biosignal Generation and Latent Variable Analysis With Recurrent Generative Adversarial Networks." *IEEE Access*, Vol. 7, pp. 144292-302, (2019).

75- Guillermo Iglesias, Edgar Talavera, and Alberto Díaz-Álvarez, "A survey on GANs for computer vision: Recent research, analysis and taxonomy." *Computer Science Review*, Vol. 48(2023).

76- A. E. Blanchard, C. Stanley, and D. Bhowmik, "Using GANs with adaptive training data to search for new molecules." *J Cheminform*, Vol. 13 (No. 1), p. 14, Feb 23 (2021).

77- V. Thambawita *et al.*, "SinGAN-Seg: Synthetic training data generation for medical image segmentation." *PLoS One*, Vol. 17 (No. 5), p. e0267976, (2022).

78- S. Bond-Taylor, A. Leach, Y. Long, and C. G. Willcocks, "Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models." *IEEE Trans Pattern Anal Mach Intell*, Vol. 44 (No. 11), pp. 7327-47, Nov (2022).

79- Z. Zhou, V. Sodha, J. Pang, M. B. Gotway, and J. Liang, "Models Genesis." *Med Image Anal*, Vol. 67, p. 101840, Jan (2021).

80- Junyi Chai, Hao Zeng, Anming Li, and Eric W. T. Ngai, "Deep learning in computer vision: A critical review of emerging techniques and application scenarios." *Machine Learning with Applications*, Vol. 6(2021).

81- T. Song *et al.*, "DNMG: Deep molecular generative model by fusion of 3D information for de novo drug design." *Methods*, Vol. 211, pp. 10-22, Mar (2023).

82- Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, and Javier Ortega-Garcia, "Deepfakes and beyond: A Survey of face manipulation and fake detection." *Information Fusion*, Vol. 64, pp. 131-48, (2020).

83- Carlo Abate, Sergio Decherchi, and Andrea Cavalli, "Graph neural networks for conditional de novo drug design." *WIREs Computational Molecular Science*, (2023).

84- R. Shi *et al.*, "Generative adversarial network constrained multiple loss autoencoder: A deep learning-based individual atrophy detection for Alzheimer's disease and mild cognitive impairment." *Hum Brain Mapp*, Vol. 44 (No. 3), pp. 1129-46, Feb 15 (2023).

85- E. Lin, S. Mukherjee, and S. Kannan, "A deep adversarial variational autoencoder model for dimensionality reduction in single-cell RNA sequencing analysis." *BMC Bioinformatics*, Vol. 21 (No. 1), p. 64, Feb 21 (2020).

86- Kamran Javed, Nizam Ud Din, Ghulam Hussain, and Tahir Farooq, "Throwaway Shadows Using Parallel Encoders Generative Adversarial Network." *Applied Sciences*, Vol. 12 (No. 2), (2022).

87- Siyu Shao, Pu Wang, and Ruqiang Yan, "Generative adversarial networks for data augmentation in machine fault diagnosis." *Computers in Industry*, Vol. 106, pp. 85-93, (2019).

88- Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu, "Deep Learning-Based Classification of Hyperspectral Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 7 (No. 6), pp. 2094-107, (2014).

89- I. E. Agbehadji, B. O. Awuzie, A. B. Ngowi, and R. C. Millham, "Review of Big Data Analytics, Artificial Intelligence and Nature-Inspired Computing Models towards Accurate Detection of COVID-19 Pandemic Cases and Contact Tracing." *Int J Environ Res Public Health*, Vol. 17 (No. 15), Jul 24 (2020).

90- Z. Wu *et al.*, "Do we need different machine learning algorithms for QSAR modeling? A comprehensive assessment of 16 machine learning algorithms on 14 QSAR data sets." *Brief Bioinform*, Vol. 22 (No. 4), Jul 20 (2021).

91- Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter, "DeepTox: Toxicity prediction using deep learning." *Toxicology Letters*, Vol. 280(2017).

92- G. Palla, D. S. Fischer, A. Regev, and F. J. Theis, "Spatial components of molecular tissue biology." *Nat Biotechnol*, Vol. 40 (No. 3), pp. 308-18, Mar (2022).

93- K. V. Chuang, L. M. Gunsalus, and M. J. Keiser, "Learning Molecular Representations for Medicinal Chemistry." *J Med Chem*, Vol. 63 (No. 16), pp. 8705-22, Aug 27 (2020).

94- B. Ramsundar *et al.*, "Is Multitask Deep Learning Practical for Pharma?" *J Chem Inf Model*, Vol. 57 (No. 8), pp. 2068-76, Aug 28 (2017).

95- E. B. Lenselink *et al.*, "Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set." *J Cheminform*, Vol. 9 (No. 1), p. 45, Aug 14 (2017).

96- A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina, and A. Zhavoronkov, "Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data." *Mol Pharm*, Vol. 13 (No. 7), pp. 2524-30, Jul 5 (2016).

97- A. Lusci, G. Pollastri, and P. Baldi, "Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules." *J Chem Inf Model*, Vol. 53 (No. 7), pp. 1563-75, Jul 22 (2013).

98- Y. Xu, Z. Dai, F. Chen, S. Gao, J. Pei, and L. Lai, "Deep Learning for Drug-Induced Liver Injury." *J Chem Inf Model*, Vol. 55 (No. 10), pp. 2085-93, Oct 26 (2015).

99- M. H. S. Segler, T. Kogej, C. Tyrchan, and M. P. Waller, "Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks." *ACS Cent Sci*, Vol. 4 (No. 1), pp. 120-31, Jan 24 (2018).

100- C. W. Coley, R. Barzilay, W. H. Green, T. S. Jaakkola, and K. F. Jensen, "Convolutional Embedding of Attributed Molecular Graphs for Physical Property Prediction." *J Chem Inf Model*, Vol. 57 (No. 8), pp. 1757-72, Aug 28 (2017).

101- E. J. Bjerrum and B. Sattarov, "Improving Chemical Autoencoder Latent Space and Molecular De Novo Generation Diversity with Heteroencoders." *Biomolecules*, Vol. 8 (No. 4), Oct 30 (2018).

102- Abdulelah S. Alshehri, Rafiqul Gani, and Fengqi You, "Deep learning and knowledge-based methods for computer-aided molecular design—toward a unified approach: State-of-the-art and future directions." *Computers & Chemical Engineering*, Vol. 141(2020).

103- S. D'Souza, K. V. Prema, and S. Balaji, "Machine learning models for drug-target interactions: current knowledge and future directions." *Drug Discov Today*, Vol. 25 (No. 4), pp. 748-56, Apr (2020).

104- Y. C. Lo, S. E. Rensi, W. Torng, and R. B. Altman, "Machine learning in chemoinformatics and drug discovery." *Drug Discov Today*, Vol. 23 (No. 8), pp. 1538-46, Aug (2018).

105- X. Zeng *et al.*, "Deep generative molecular design reshapes drug discovery." *Cell Rep Med*, Vol. 3 (No. 12), p. 100794, Dec 20 (2022).

106- Z. Zhou, S. Kearnes, L. Li, R. N. Zare, and P. Riley, "Optimization of Molecules via Deep Reinforcement Learning." *Sci Rep*, Vol. 9 (No. 1), p. 10752, Jul 24 (2019).

107- A. Kadurin *et al.*, "The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology." *Oncotarget*, Vol. 8 (No. 7), pp. 10883-90, Feb 14 (2017).

108- A. Kadurin, S. Nikolenko, K. Khrabrov, A. Aliper, and A. Zhavoronkov, "druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico." *Mol Pharm*, Vol. 14 (No. 9), pp. 3098-104, Sep 5 (2017).

109- T. Blaschke, M. Olivecrona, O. Engkvist, J. Bajorath, and H. Chen, "Application of Generative Autoencoder in De Novo Molecular Design." *Mol Inform*, Vol. 37 (No. 1-2), Jan (2018).

110- K. Seeliger, U. Guclu, L. Ambrogioni, Y. Gucluturk, and M. A. J. van Gerven, "Generative adversarial networks for reconstructing natural images from brain activity." *Neuroimage*, Vol. 181, pp. 775-85, Nov 1 (2018).

111- Ian Goodfellow *et al.*, "Generative adversarial networks." *Communications of the ACM*, Vol. 63 (No. 11), pp. 139-44, (2020).

112- N. Suresh, N. Chinnakonda Ashok Kumar, S. Subramanian, and G. Srinivasa, "Memory augmented recurrent neural networks for de-novo drug design." *PLoS One*, Vol. 17 (No. 6), p. e0269461, (2022).

113- F. Noe, A. Tkatchenko, K. R. Muller, and C. Clementi, "Machine Learning for Molecular Simulation." *Annu Rev Phys Chem*, Vol. 71, pp. 361-90, Apr 20 (2020).

114- A. Gupta, A. T. Muller, B. J. H. Huisman, J. A. Fuchs, P. Schneider, and G. Schneider, "Generative Recurrent Networks for De Novo Drug Design." *Mol Inform*, Vol. 37 (No. 1-2), Jan (2018).

115- V. Gallego, R. Naveiro, C. Roca, D. Rios Insua, and N. E. Campillo, "AI in drug development: a multidisciplinary perspective." *Mol Divers*, Vol. 25 (No. 3), pp. 1461-79, Aug (2021).

116- B. Tang *et al.*, "AI-Aided Design of Novel Targeted Covalent Inhibitors against SARS-CoV-2." *Biomolecules*, Vol. 12 (No. 6), May 25 (2022).

117- Martha M. Flores-Leonar *et al.*, "Materials Acceleration Platforms: On the way to autonomous experimentation." *Current Opinion in Green and Sustainable Chemistry*, Vol. 25(2020).

118- P. Szymanski, M. Markowicz, and E. Mikiciuk-Olasik, "Adaptation of high-throughput screening in drug discovery-toxicological screening tests." *Int J Mol Sci*, Vol. 13 (No. 1), pp. 427-52, (2012).

119- Khalid Albulayhi, Qasem Abu Al-Haija, Suliman A. Alsuhibany, Ananth A. Jillepalli, Mohammad Ashrafuzzaman, and Frederick T. Sheldon, "IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method." *Applied Sciences*, Vol. 12 (No. 10), (2022).

120- M. Bagherian, E. Sabeti, K. Wang, M. A. Sartor, Z. Nikolovska-Coleska, and K. Najarian, "Machine learning approaches and databases for prediction of drug-target interaction: a survey paper." *Brief Bioinform*, Vol. 22 (No. 1), pp. 247-69, Jan 18 (2021).

121- Y. Zeng, X. Chen, Y. Luo, X. Li, and D. Peng, "Deep drug-target binding affinity prediction with multiple attention blocks." *Brief Bioinform*, Vol. 22 (No. 5), Sep 2 (2021).

122- B. Han, X. H. He, Y. Q. Liu, G. He, C. Peng, and J. L. Li, "Asymmetric organocatalysis: an enabling technology for medicinal chemistry." *Chem Soc Rev*, Vol. 50 (No. 3), pp. 1522-86, Feb 15 (2021).

123- S. Dara, S. Dhamercherla, S. S. Jadav, C. M. Babu, and M. J. Ahsan, "Machine Learning in Drug Discovery: A Review." *Artif Intell Rev*, Vol. 55 (No. 3), pp. 1947-99, (2022).

124- J. Jimenez, S. Doerr, G. Martinez-Rosell, A. S. Rose, and G. De Fabritiis, "DeepSite: protein-binding site predictor using 3D-convolutional neural networks." *Bioinformatics*, Vol. 33 (No. 19), pp. 3036-42, Oct 1 (2017).

125- J. Jimenez, M. Skalic, G. Martinez-Rosell, and G. De Fabritiis, "K(DEEP): Protein-Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks." *J Chem Inf Model*, Vol. 58 (No. 2), pp. 287-96, Feb 26 (2018).

126- Betsabeh Tanoori, Mansoor Zolghadri Jahromi, and Eghbal G. Mansoori, "Drug-target continuous binding affinity prediction using multiple sources of information." *Expert Systems with Applications*, Vol. 186(2021).

127- Edna Chebet Too, Li Yujian, Sam Njuki, and Liu Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification." *Computers and Electronics in Agriculture*, Vol. 161, pp. 272-79, (2019).

128- M. H. S. Segler and M. P. Waller, "Modelling Chemical Reasoning to Predict and Invent Reactions." *Chemistry*, Vol. 23 (No. 25), pp. 6118-28, May 2 (2017).

129- D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, and A. G. Doyle, "Predicting reaction performance in C-N cross-coupling using machine learning." *Science*, Vol. 360 (No. 6385), pp. 186-90, Apr 13 (2018).

130- P. Schwaller, B. Hoover, J. L. Reymond, H. Strobelt, and T. Laino, "Extraction of organic chemistry grammar from unsupervised learning of chemical reactions." *Sci Adv*, Vol. 7 (No. 15), Apr (2021).

131- Tfgg Cova and Aacc Pais, "Deep Learning for Deep Chemistry: Optimizing the Prediction of Chemical Patterns." *Front Chem*, Vol. 7, p. 809, (2019).

132- A. C. Mater and M. L. Coote, "Deep Learning in Chemistry." *J Chem Inf Model*, Vol. 59 (No. 6), pp. 2545-59, Jun 24 (2019).

133- M. H. S. Segler, M. Preuss, and M. P. Waller, "Planning chemical syntheses with deep neural networks and symbolic AI." *Nature*, Vol. 555 (No. 7698), pp. 604-10, Mar 28 (2018).

134- L. Tao, G. Chen, and Y. Li, "Machine learning discovery of high-temperature polymers." *Patterns (N Y)*, Vol. 2 (No. 4), p. 100225, Apr 9 (2021).

135- B. Liu *et al.*, "Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models." *ACS Cent Sci*, Vol. 3 (No. 10), pp. 1103-13, Oct 25 (2017).

136- Y. Li, K. Li, C. Zhang, J. Montoya, and G. H. Chen, "Learning to Reconstruct Computed Tomography Images Directly From Sinogram Data Under A Variety of Data Acquisition Conditions." *IEEE Trans Med Imaging*, Vol. 38 (No. 10), pp. 2469-81, Oct (2019).

137- C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, and K. F. Jensen, "Prediction of Organic Reaction Outcomes Using Machine Learning." *ACS Cent Sci*, Vol. 3 (No. 5), pp. 434-43, May 24 (2017).

138- Z. Zhou, X. Li, and R. N. Zare, "Optimizing Chemical Reactions with Deep Reinforcement Learning." *ACS Cent Sci*, Vol. 3 (No. 12), pp. 1337-44, Dec 27 (2017).

139- L. B. Ayres, F. J. V. Gomez, J. R. Linton, M. F. Silva, and C. D. Garcia, "Taking the leap between analytical chemistry and artificial intelligence: A tutorial review." *Anal Chim Acta*, Vol. 1161, p. 338403, May 29 (2021).

140- M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science." *Nature*, Vol. 566 (No. 7743), pp. 195-204, Feb (2019).

141- Y. Jing, Y. Bian, Z. Hu, L. Wang, and X. Q. Xie, "Deep Learning for Drug Design: an Artificial Intelligence Paradigm for Drug Discovery in the Big Data Era." *AAPS J*, Vol. 20 (No. 3), p. 58, Mar 30 (2018).

142- D. Rogers and M. Hahn, "Extended-connectivity fingerprints." *J Chem Inf Model*, Vol. 50 (No. 5), pp. 742-54, May 24 (2010).

143- K. Gao, D. D. Nguyen, V. Sresht, A. M. Mathiowetz, M. Tu, and G. W. Wei, "Are 2D fingerprints still valuable for drug discovery?" *Phys Chem Chem Phys*, Vol. 22 (No. 16), pp. 8373-90, Apr 29 (2020).

144- Y. Bian and X. Q. Xie, "Generative chemistry: drug discovery with deep learning generative models." *J Mol Model*, Vol. 27 (No. 3), p. 71, Feb 4 (2021).

145- J. Jimenez-Luna, M. Skalic, N. Weskamp, and G. Schneider, "Coloring Molecules with Explainable Artificial Intelligence for Preclinical Relevance Assessment." *J Chem Inf Model*, Vol. 61 (No. 3), pp. 1083-94, Mar 22 (2021).

146- S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, "Molecular graph convolutions: moving beyond

fingerprints." *J Comput Aided Mol Des*, Vol. 30 (No. 8), pp. 595-608, Aug (2016).

147- P. Reiser *et al.*, "Graph neural networks for materials science and chemistry." *Commun Mater*, Vol. 3 (No. 1), p. 93, (2022).

148- A. Gaulton *et al.*, "ChEMBL: a large-scale bioactivity database for drug discovery." *Nucleic Acids Res*, Vol. 40 (No. Database issue), pp. D1100-7, Jan (2012).

149- E. H. B. Maia, L. C. Assis, T. A. de Oliveira, A. M. da Silva, and A. G. Taranto, "Structure-Based Virtual Screening: From Classical to Artificial Intelligence." *Front Chem*, Vol. 8, p. 343, (2020).

150- W. P. Walters and R. Barzilay, "Applications of Deep Learning in Molecule Generation and Molecular Property Prediction." *Acc Chem Res*, Vol. 54 (No. 2), pp. 263-70, Jan 19 (2021).