Artificial Intelligence and Deep Learning in drug discovery, application and future challenges

Alka Singh*, Ravindra Singh Yadav

Department of Computer Science and Engineering
P.K University, Shivpuri (M.P)
*Corresponding Author - **alkasingh@knit.ac.in**

Abstract

This analysis will explore how artificial intelligence and deep learning technologies haveaffected drug discovery. Artificial intelligence in drug discovery has received much attentiondue to its cost effectiveness and it significantly shortens the time. Deep learning basedtechnologies are progressively being used in all stages of drug development as DLtechnology advances and drug-related data grows. This paper presents a Literature reviewthat integrates the recent DL techniques and applications in drug discovery.

Keywords: Deep Learning (DL), Artificial Intelligence (AI), Drug target interaction (DTI), Convolutional neural network (CNN), Recurrent neural network (RNN), Graph neural network (GNN), Generative adversarial network (GAN)

1. Introduction

Drug Discovery is the process of examining of how various drugs interact with the body and how a medication needs to act on body to have a therapeutic impact. Large pharmaceutical corporations migrate toward AI in the wake of the development of DL approaches, eschewing outmoded, ineffective procedures to increase patient profit while also increasing their own. Drug discovery is a systematic process in which new therapeutic drugs are identified using various computers, experimental and clinical methods. The drug development process involves extracting chemicals from sources to make them usable for consumption.After that we have to perform various clinical trials on the medicine, either in-vitro or in-vivo. These clinical trials are used to test whether the drug is effective. The possible side effect needs to be addressed.

Deep Learning is a subfield of AI that creates algorithms with a learning perspective. These algorithmic data are then analysedfrom a multilayer perspective. These multi-layered Algorithms have a unique property. They analyse the data, learn from data, and generate experience as more and more data are subjected to DL models .Machine learning models require large sums of data and complexity to make sense of the raw data gathered by the system.

Deep learning is a sub –area of machine learning and it uses complicated neural networks and algorithms to extract functional patterns from large amounts of data. Artificial neural network are at the very core of deep learning. They are versatile, powerful and scalable, making them ideal to tackle large and highly complex machine learning tasks such as classifying billions of images, powering speech recognition services, learning to beat the world champion at the game of Go or recommending the best videos to watch to hundreds of millions of users every day.

2. Review of related studies

Gupta and Sachdev (2019) presented the various feature based chemogenomic methods for drug target interaction prediction. They presented an overview of the various techniques, tools, datasets and metrics. They presented a current overview of the various feature based methodologies. They described relevant datasets, techniques for determine medication or target properties.

Ali Ezzat, Min We (2019) aim to provide a comprehensive overview and empirical evaluation on the computational DTI prediction techniques. They first described the data used in such computational DTI prediction efforts. Then they categorize and elaborate the state –of art methods for predicting DTIs.

Maryam Bagherian , ElyasSabeti (2021) presented the data required for the task of DTI prediction followed by a comprehensive catalog consisting of machine learning methods and databases which have been proposed and utilized to predict DTIs. The advantages and disadvantages of each set of methods are also briefly discussed. The challenges one may face in prediction of DTI using machine learning approaches are highlighted.

Faranch Haddadi and Mohammad Reza (2021) presented a qualitative analysis in the form of a framework, a drug – target interactions prediction framework .The framework consists three sections. First a classification has been presented for DTI interactions prediction methods based on the link prediction approaches used in these methods. Next general evaluation criteria have been introduced for analyzing approaches. Last a qualitative comparison is made between each approach in term of their advantage and disadvantages.

Ali K. Abdul Raheem and Ban N. Dhannoonin(2024) give a detailed discussion of the initial stage in drug discovery. It focuses on exploring the application of machine learning methods. The prediction of DTI can be categorized into three main computation method : docking simulation approaches , ligand –based methods and machine learning techniques.

Sundaravadivela Sumathi and Suganya(2023) presented various approaches of deep learning in drug discovery like deep generative models for drug discovery, synergy prediction and precision medicine.

3. Introduction to artificial intelligence and deep learning

Artificial intelligence confers cognitive abilities upon machines, enabling them to perform tasks requiring human intelligence. Machine Learning, a subset of AI, emphasizes algorithms that allow computers to advance from data without explicit programming. Deep learning represents an even further specialization in machine learning. Deep learning utilizes

artificial neural networks with many layers to process complex information. Artificial neural network are at the very core of deep learning.

ANNs were first introduced back in 1943 by the neurophysiologist Warren McCulloch and the mathematician Walter Pitts. Structures of artificial neurons are inspired by biological neurons. It's composed of a cell body containing the nucleus and most of the cell's complex components, many branching extensions called dendrites, plus one very long extension called the axon. The axon's length may be just a few times longer than the cell body or up to tens of thousands of time longer. Near its extremity the axon splits off into many branches called telodentria,and at the tip of these branches are minuscule structures called synaptic terminals , which are connected to the dendrites or cell bodies of the neurons. Biological neurons produce short electrical impulses called action potentials, which travel along the axons and make the synapses release amount of these neurotransmitters within a few milliseconds.Individual biological neurons seem to behave in a simple way, but they are organized in a vast network of billions, with each neurons typically connected to thousands of other neurons.

McCulloch and Pitts proposed a very simple model of the biological neurons which has known as artificial neuron. It has one or more binary inputs and one binary output. The artificial neuron activates its output when more than a certain number of its inputs are active.
The perceptron is one of the simplest ANN architecture, invented in 1957 by Frank Rosen-blatt. It is based on a slightly different artificial neuron called a threshold unit (LTU). The inputs and outputs are numbers (instead of binary on/off values) and each input connection is associated with a weight. The TLU first computes a linear function of its inputs:

$$Z = W_1X_1 + W_2 X_2 + \_ \_ \_ \_ + W_nX_n + b = W^T X + b$$

Then it applies a step function to the result

$$H_w(X) = \text{step}(Z)$$

The most common step function used in perceptions is the Heaviside step function. Sometimes the time function is used instead.

$$\text{Heaviside}(Z) = \{0 \text{ if } Z = 0 \text{ and } 1 \text{ if } Z \geq 0\}$$

A single TLU can be used for simple linear binary classification. It computes a liner functions of its inputs and if the result exceeds a threshold, its output the positive class. Training such a TLU would require finding the right values for w_1 ,w_2 and b.

A perceptron is composed of one or more TLUs organised in a single layer, where every TLU is connected to every input. Such a layer is called a fully connected layer, or a dense layer. The input constitutes the input layer and since the layer of TLU produces the final outputs, it is called the output layer. A perceptron with two input and three outputs is presented in this figure

An MLP is composed of one input layer, one or more layers of TLUs called hidden layers and one final layer of TLUs called the output layer. The layers close to the input layer

are usually called the lower layers and the one close to the outputs are usually called the upper layers .When an ANN contains a deep stack of hidden layers it is called a deep neural network. The field of deep learning studies DNNs, and more generally it is interested in models containing deep stack of computations.

4. Deep Learning Techniques

Deep learning techniques are subset of neural network algorithms that contains multiple hidden layers to enabling automatic extraction and selection of high level features from large databases. In this section we will review the various deep learning models.

4.1. Classic neural network

Multilayer perceptron are frequently employed to recognize fully connected neural network. For training MLPs back propagation which is combination of reverse mode autodiff and gradient descent most popular. Back propagation makes predictions for a mini- batch (forward pass), measures the error, then goes through each layer in reverse to measure the error contribution from each parameter (reverse pass) and finally modify the connection weights and biases to reduce the error .MLPs can be used for classification tasks and regression task both.For binary classification problem we just need a single output neuron using the sigmoid function and the output will be a number between 0 and 1, which can be predicted as the estimated probability of the positive class. The estimated probability of negative class is equal to one minus that number. MLP can easily handle multilevel binary classification task.For regression task if we want to predict a single value, we need a single output neuron and itsoutput is the predicted value. For multivariate regression we need one output neuron per output dimension. Classic neural networks are best for regression and categorization with real valued data.

4.2 Convolutional Neural Network

CNN are also known as ConvNet. They are specialized type of deep learning algorithm designed for task that necessitate image classification, object recognition and image segmentation .CNNs were motivated by the layered architecture of human brain. The key components of CNN are:

- Convolution layer: The main mathematical task performed is called convolution. Convolution is the application of sliding window function to a matrix of pixels which Represent an image. Sliding function applied to matrix is called a filter. Several filters of equal size are applied in convolutional layer .Each filter is used to observe a specific patternfrom the image such as edges , the curving of digits and the whole shape of digits. In Convolutionlayer , we use small grids that move over the image. Grids are like a mini magnifying glass that focuses specific patterns in the image like curves, shapes and lines.

- Activation Function : After each convolution operation a ReLu activation function is applied .It helps the network to learn non-linear relationship between the features in the image makingthe network more robust.
- Pooling Layer: Pooling layer is used to pool most significant features from the convolutionmatrix .For pooling some aggregation operations are applied, which reduce the dimension ofthe convoluted matrix.
- Fully connected layers: Last layer of the CNN and their inputs correspond to the flattenedone-dimensional matrix activation functions are applied. At last a softmax prediction layer is used to generate probability values for every possible output labels. One with the highest probability score will be the final label.

4.3 Generative adversarial Networks

GAN use two neural network a Generator and a Discriminator to create new realistic data. GAN take a creative way by generating new content that resembles real world data.GAN combines two main models.A generator model which is a deep neural network that takes random noise as input to generate realistic data samples like images or songs. A discriminator which acts as a classifier.It try to distinguish between real data from the original data and the fake data which are produced by the generator. The generator model tries to mislead the discriminative model into classifying fake data as real, and discriminators continuously improves its ability to differentiate real and fake data. This is performed by a class of function that measures network performance. If the generator is successfully creating realistic data loss will be low. A discriminator loss is measure of how well the discriminator can differentiate between real and fake data.GANs can be used for image generation, computer vision ,image to image translation, text to image generation and more.

4.4 Recurrent neural network(RNNs)

RNN differ from regular neural networks in processing information standard neural networks pass information in one direction. RNN feeds information back into the network at each step. Recurrent unit is fundamental processing unit of RNN. They have a hidden state that maintains information about previous inputs in a sequence. RNN unrolling is the process of expanding recurrent structure over time steps. RNNs are used in Speech recognition, natural machine translation, handwriting recognition and natural language processing. RNN have a very high level of computational power. It could be used to model non- linear dynamic system.

4.5 Boltzmann Machine

Boltzman Machines are deep neural networks used in unsupervised learning, feature extraction and optimization. They play a key role in probabilistic graphical models. For modelling data distributions it uses a stochastic recurrent neural network. Boltzman Machine consists of a set of neurons connected in a network. In this each neuron can take binary states 0 or 1. It works by updating their neurons states based on probability distribution.

Boltzmanmachine have applications in Image recognition, recommendation system and financial forecasting.

4.6 Autoencoders

Autoencoders are most popular in the field of unsupervised machine learning. Autoencoders are special neural networks that learn to compress data into a compact form and afterthat reconstruct it to closely match the original input. They consist, an encoder that captures important features by reducing dimensionality. Decoder that reconstruct the data from this compressed representation. Auto encoders are trained by minimizing reconstruction error using loss functions like Binary Cross Entropy or Mean Squared error.

There are different types of Autoencoder

- Convolution Autoencoder: This encoder extracts features using Convolutional layer, decoder reconstruct the image through deconvolution.
- Denoising Autoencoder: Denoising Autoencoder is trained to handle noisy inputs. It helps in reconstruction of clean data. It encourages learning the core features.
- Sparse Autoencoder: Sparse Autoencoder allows a few neurons to be active concurrently. It contains more hidden layers than input features.

4.7 Self- organizing maps (SOM)

Self-organizing maps are unsupervised deep learning networks. They are used to reducedimensionality and data visualization. They are useful tools for clustering and mapping highdimensional data into a lower dimensional grid for easier understanding of complex patternsby clustering similar data points together. Applications of SOM are topology preservation,anomaly detection.

5 Application of deep learning in drug discovery

Drug discovery is the problem of finding suitable drugs to treat a target protein. It is a time-consuming and costly process. Significant steps involved in drug discovery include designing and identifying compounds that are specific to their targeted biological molecules and cellular processes. Drug discovery can be forecast using deep learning (DL) models. These relations represent critical molecular correlations that involve specific drugs and their target proteins in the body. Estimation of drug–target interactions (DTIs) is important for understanding the drug’s mode of action, identifying specific drug targets, and developing safe and effective therapeutic doses. Traditionally, DTIs are predicted using molecular docking, which evaluates the interaction between the protein structure and the drug molecule by determining their binding affinity. However, molecular docking has limitations, such as high computational cost, long processing time, and the need for deep understanding of complex DTI mechanisms. Different characteristics of drug–protein interactions can be evaluated by DL models, including molecular features, chemical patterns, and protein sequence information.

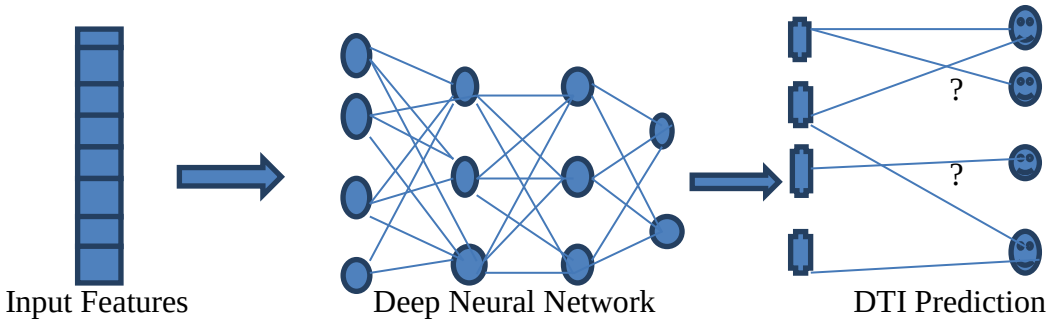

Fig 1. DTIs prediction using deep learning models

6. Conclusion

Recent progresses in AI, including the development and growth of more complex DL techniques have a huge impact on drug target interaction process. DL algorithms have demonstrated advantages in feature extraction and recognition from huge and complex raw data. Deep learning models can learn highly nonlinear and complex patterns from data.Deep learning can be adapted to various types of training regimes facilitating the development of tailored models for diverse applications. Deep learning techniques optimize millions of neural network parameters .It makes process more robust by training on large datasets.One major challenge in model interpretation in drug design using deep learning is the complexity of model architecture.

References

[1] J. Rantanen and J. Khinast, "The future of pharmaceutical manufacturing sciences," *J. Pharm. Sci.*, vol. 104, no. 11, pp. 3612–3638, 2015.
[2] A. T. Greenhill and B. R. Edmunds, "A primer of artificial intelligence in medicine," *Tech. Innov. Gastrointest. Endosc.*, vol. 22, no. 2, pp. 85–89, 2020.
[3] E. Gawehn, J. A. Hiss, and G. Schneider, "Deep learning in drug discovery," *Mol. Inf.*, vol. 35, no. 1, pp. 3–14, 2016.
[4] J. Kim, S. Park, D. Min, and W. Kim, "Comprehensive survey of recent drug discovery using deep learning," *Int. J. Mol. Sci.*, vol. 22, no. 18, Art. 9983, 2021.
[5] A. Lavecchia, "Deep learning in drug discovery: opportunities, challenges and future prospects," *Drug Discov. Today*, vol. 24, no. 10, pp. 2017–2032, 2019.
[6] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, and T. Blaschke, "The rise of deep learning in drug discovery," *Drug Discov. Today*, vol. 23, no. 6, pp. 1241–1250, 2018.
[7] J. Vamathevan*et al.*, "Applications of machine learning in drug discovery and development," *Nat. Rev. Drug Discov.*, vol. 18, no. 6, pp. 463–477, 2019.
[8] K. Sachdev and M. K. Gupta, "A comprehensive review of feature-based methods for drug target interaction prediction," *Comput. Biol. Med.*, vol. 93, Art. 103159, 2019.

[9] A. Ezzat, M. Wu, X.-L. Li, and C.-K. Kwoh, "Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey," *Brief. Bioinform.*, vol. 20, no. 4, pp. 1337–1357, 2019.
[10] L. Patel, T. Shukla, X. Huang, D. W. Ussery, and S. Wang, "Machine learning methods in drug discovery," *Molecules*, vol. 25, no. 22, Art. 5277, 2020.
[11] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, 2021.
[12] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D*, vol. 404, Art. 132306, 2020.
[13] J. Zhou *et al*., "Graph neural networks: a review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.
[14] R. Gupta, D. Srivastava, M. Sahu, S. Tiwari, R. K. Ambasta, and P. Kumar, "Artificial intelligence to deep learning: machine intelligence approach for drug discovery," *Mol. Divers.*, vol. 25, pp. 1315–1360, 2021.
[15] L. Zhang, J. Tan, D. Han, and H. Zhu, "From machine learning to deep learning: progress in machine intelligence for rational drug discovery," *Drug Discov. Today*, vol. 22, no. 11, pp. 1680–1685, 2017.
[16] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," *J. Pharm. Biomed. Anal.*, vol. 22, no. 5, pp. 717–727, 2000.
[17] B. Ramsundar, P. Eastman, P. Walters, N. Matlock, V. Pande, and K. Leswing, "Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more," *O'Reilly Media*, 2019.
[18] M. Zitnik, B. Zupan, and J. Leskovec, "Predicting drug–target interactions from heterogeneous network data," *Bioinformatics*, vol. 34, no. 17, pp. 3037–3045, 2018.
[19] B. Aliper*et al.*, "Deep learning applications for predicting pharmacological properties of drugs and drug repurposing opportunities," *Mol. Pharm.*, vol. 13, no. 7, pp. 2524–2530, 2016.