

# Deep Neural Networks in the Discovery of Novel Antibiotics Drug Molecule: A Review

Michael Idowu Oladunjoye<sup>1</sup>, Olumide Olayinka Obe<sup>2</sup>

<sup>1,2</sup>Department of Computer Science, Federal University of Technology, Akure, Nigeria

**Abstract** - Machine learning methods have been used in various fields to enhance automation and the ability of computer to learn from experience for decades. Its application in drug discovery is emerging and is now being integrated into the process especially in the early stage for lead compounds screening. The antibiotic discovery process is cumbersome considering the duration and inadvertently the resultant cost. So, the application of deep neural networks will help in efficiency by reducing the duration of process and the overall cost of the process. A deep convolutional neural network successfully predicted new broad-spectrum antibiotic, *Halicin*, with other molecules with distinct structures. The deep learning models have been engaged to learn from known useful chemical compounds (molecules) with their biological activities. The use of the artificial intelligence techniques in the antibiotics discovery will be reviewed with focus on the deep neural networks model as compared with other methods.

**Keywords** - Machine learning, drug discovery, molecules, artificial intelligence, Deep neural networks

## I. INTRODUCTION

Drug discovery is finding a promising molecule known as lead compound that could become a drug [2, 7]. It is also considered as the process of designing molecules that could someday lead to new therapies as part of drug development [3]. Generally, it is a process by which a drug candidate is identified before it is validated having been subjected to series of pre-clinical and clinical trials for the treatment of a specific disease. Before the discovery, efforts must be made to understand the disease to be treated and the underlying cause of the condition by scientist [2].

The first natural antibiotic product, *Penicillin*, was discovered by Alexander Fleming on the observation of a diffusible extract that had antibacterial activity against staphylococci produced by the *Penicillium* molds [4]. Antibiotic that can be effective against a particular biological target causing diseases involves reasonable numbers of experiments through a reasonable numbers of years to get the best (lead) potential molecules and this has been considered extremely cumbersome with advertently high cost [3]. The traditional approach to the discovery of new medicine involves screening of large number of compounds to identify a potential candidate and then the synthesis to optimize the molecular compound [6].

At the early stage of the development of modern medicine, it was generally believed that solutions to all infectious diseases considering the availability of the major classes of

antibacterial agents,  $\beta$ -lactams, quinolone, tetracyclines, and macrolide antibiotics had been accomplished. But the increase in the growth of the methicillin-resistant *Staphylococcus aureus* (MRSA) in the 1980s and 1990s [7] suggested that additional antibacterial agents would be needed to protect humanity.

Also, tons of new diseases as a result of activities of various bacteria demand new antibiotics. The effect of antibiotics resistant pathogens on social and economic life of people is enormous, and it is estimated to be over \$20 billion in health care cost in Europe and the United State of America [8]. Besides, the poor returns on investments in the discovery and development of new antibiotics in drug companies have negatively impacted the interest by various stakeholders [3]. The application of deep neural networks in the process with huge available medical, biological and chemical databases that can be explored will improve the efficiency and reduce the cost.

Drug discovery requires so much information to process to get the lead chemical molecule and such will need what will as quickly as possible point to it and the target. So, the use of machine learning techniques in the process is to shorten the drug pipeline and to reduce the cost especially during the process of lead identification. Also, the *in-silico* pharmacology in context is to apply computer techniques in capturing, analyzing and integrating biological and medical data from diverse sources for the process [10]. This is simply the use available virtual information with a predictive model to make predictions and ultimately produce new drug molecules faster and at a reduced cost.

### A. Antibiotics Discovery Process

In classical pharmacology, huge set of compounds in chemical database of synthetic molecules, natural products or extracts were screened for molecules that had desirable properties of drug and therapeutic effect [11]. This process begins with the disease and target identification, hit identification and lead identification after which the lead molecule is optimized for the drug likeness properties before moving to preclinical testing.

The target identification implies which disease to treat, that is, the identity of a molecular target can be bound by the lead compound. The hit identification is the numbers of molecules that can bind but not necessarily with positive effect while the

lead identification the molecules that have effect on the target [12].

This screening of chemical databases is iterative with repeated operations, consumes time and an expensive process that is sometimes considered searching a needle in haystack, but with machine learning approaches, the process can be performed rapidly and inexpensively.

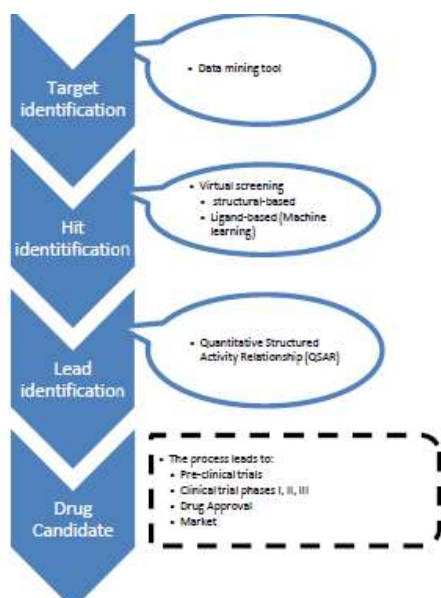


Fig 1: Process flow of drug discovery with possibilities of machine learning techniques

Appropriate data mining techniques can be explored in the target identification and virtual screen methods. Amongst the classical machine learning tools, K-Nearest neighbors has been used to predict a new compound but the method was considered sensitive to noisy data [13]. The structure based or ligand based, for the hit identification and lead identification can be achieved using a machine learning predictive model as Neural network or deep learning.

## II. RELEVANCE OF MACHINE LEARNING TECHNIQUES IN DRUG DISCOVERY PROCESSES

To mitigate the associated problems in the drug discovery and development process, machine learning techniques can be used in the process. This will reduce the cost and speed up the process especially in the screening to determine the lead molecule.

The drug discovery process with AI begins with virtual screening of two major categories, structure-based and ligand-based methods. The structure-based method requires three-dimensional structure of the protein, the biological target, obtained through methods such as x-ray crystallography or NMR spectroscopy [27]. It predicts the binding affinity between the compound and a protein target when 3D structure of the protein receptor is known. It is best for the high-throughput docking [14].

The ligand-based method uses the principle of similarity observed for small molecules by making use of the knowledge of other molecules that bind to the biological target to derive the minimum necessary structural characteristics a molecule must possess in order to bind to the target [14], [13]. The ligand-based method tools include quantitative structure-activity relationship (QSAR), pharmacophore modeling, molecular field analysis and 2D or 3D similarity assessment, and are used in the absence of three-dimensional (3D) structures of potential targets [15].

These methods has been used with deep learning model during virtual screening of existing libraries of chemical structures to classify active and inactive compounds or predicting biological activity based on the values of molecular descriptors [16].

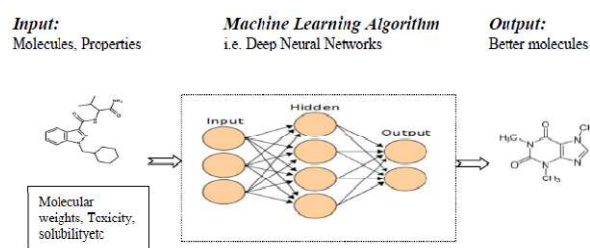


Fig 2: A representation of a neural network model architecture

### A. Datasets for the Machine Learning Algorithms

Data is explored to build machine learning solutions. This data is cleaned, tested with the algorithm desired and have the algorithm tuned to get a good result. The sources of the data for the models is not necessarily sourced directly from the domain experts considering the availability and accessibility of huge biomedical and chemical databases produced at a period of time by various pharmaceutical companies. For example, there are huge amount of annotated and high-quality datasets consisting of more than 700,000 molecules and their property specially designed for testing machine learning methods of molecular properties as a standard benchmarks [29]. Also, the ChEMBL with more than 5.4 million bioactivity measurements for 5,200 protein targets and more than 1 million ligands and host of others are general-purpose datasets that are usually used for general algorithm development and benchmarking in deep learning and machine learning for drug discovery [28].

### III. RECENT APPLICATION OF MACHINE LEARNING TECHNIQUES IN THE DRUG DISCOVERY PROCESSES

There are various challenges in the drug discovery and development processes demand a different method to reduce the time and the cost in the timeline hence the *in silico* pharmacology. It is a computational technique to provide advances in medicine and therapeutics [5]. The traditional approaches are time consuming and costly, so the application AI techniques (e.g. machine learning) to address the problem

has been considered as one of the technologies that have shaped the process [17]. AI has been used to discover new lead compounds that exhibit desired activity and also minimize the hit-to-drug timeline improving costs [1, 5].

Machine learning techniques used in the drug discovery and development processes have been proved to improve the speed of high-throughput screening (HTS) and lead optimization with reduced cost [8]. Neural networks have been considered a useful tool to automate feature extraction to make possible the analysis of large amount of data [18]. Ligand based virtual screening (LBVS) techniques are widely used for hit identification with various types of machine-learning methods [13]. K-nearest neighbor (KNN) and artificial neural networks (ANN) have been used in autonomous knowledge acquisition from the molecular properties of compounds [26]. Also, Artificial Neural Networks has proved to be better in handling non-linear relationship than most classical statistical methods and that made ideally suited for use in drug discovery process [22].

Artificial neural networks (ANN) has many simple units, called neurons, are interconnected by weighted links into larger structures of remarkably high performance [19]. Artificial Neural Networks (ANN) are used to extract information from large datasets as in the case with molecular data [18], [20] and also useful in the prediction of molecular properties of drug but are prone to overtraining, over-fitting, and validation problems [20]. A novel antibacterial 3-hydroxypyridine-4-ones compound known iron chelator to inhibit the proliferation of pathogenic bacteria was designed with Generic Algorithm Partial Least Squares (GA-PLS) and Counter Propagation Artificial Neural Networks (CP-ANN) using QSAR model [21]. Unfortunately, they require a huge number of parameters and very difficult to train when they have more than one or two hidden layers. Nevertheless, an Artificial Neural Network model in a study used a set of 217 antibiotics with multiple classes and targets, divided into a training set (70%) and a testing set (30%) mapped 62 structure-based molecular descriptors to the biological activity with the accuracy of 91.4%, validating the network's utility [8].

Deep learning has been used to accurately predict molecular binding to reduce the time of the lead screening of new molecule. Deep neural networks also called Multi-layer Perceptron (MLP) is made up of multiple hidden layers and are capable of computing non-linear complex data patterns [23]. A deep convolutional neural network successfully predicted new active molecules for targets recently without known modulators with structure-based method [26]. A broad-spectrum antibiotic, *Halicin*, was predicted with other molecules with distinct structures with deep learning [24]. The deep learning model is trained to predict antibiotics based on structure using datasets from the Broad Institute's Drug Repurposing Hub and the *ZINC15* database [24]. Graph convolutional neural network was trained earlier to recognize *E. coli* inhibition as a property of a molecule. This was carried

out using a dataset of molecules known to inhibit or fail to inhibit the growth of *E. coli* represented as a 1 and 0 respectively [24].

The convolutional neural networks (CNN) model is an earlier class of deep neural networks that have been used in image processing and computer vision. It can be an excellent choice for continuous and large data with molecular structures and has been considered to be much easier to train and much less susceptible to over-fitting than most machine learning model [9].

Generally, artificial intelligence (AI) solutions to drug discovery are emerging essential tool for transforming the process and revolutionizing the understanding of how drugs bind to targets [25]. It also has the potentials to reduce timelines for the process, increased accuracy of predictions and also the opportunity to diversify drug pipelines as one of the major contributions. One of the advantages of deep learning models over others is its ability to extract features from data.

#### IV. CONCLUSION

The advances in Artificial Intelligence methods, progressive development in deep learning algorithms, and the availability of large chemical datasets as well as access to high powered computing resources have made the drug discovery process more efficient and interesting. All that is required is well-annotated datasets for the training of the model to recognize important features and characteristics of known drug for the prediction of the new drug candidate. Also, the choice of machine learning algorithm sometimes depends on the knowledge of the problem and the data being used for the training.

In conclusion, the use of Machine learning methods in the discovery of antibiotics drug is neither a replacement of medicinal chemistry nor the wet laboratory experimental approaches but as an aid to reduce time and cost of molecular screening for the lead molecules.

#### REFERENCES

- [1] F. Lake, "Artificial intelligence in drug discovery: what is new, and what is next." *Future Science*. Available at <https://www.future-science.com/>. Accessed on 2020-06-04.
- [2] "Drug Discovery and Development." Available at <http://www.astp4kt.eu/>. Accessed on 2020-05-20.
- [3] "A Scientific Roadmap for Antibiotic Discovery 2016." Available at. <https://www.pewtrusts.org/>. Accessed on 2020-05-20
- [4] L. Rajeev, "Antibiotic Discovery." Available at <https://www.labome.com/method/Antibiotic-Discovery.html>. Accessed on 2020-04-15.
- [5] S. Ortega, L. Cara and M. Salvador. "In silico pharmacology for a multidisciplinary drug discovery process," *Drug Metab Drug Interact*, vol 27, no 4, pp. 199-207, 2012.
- [6] P. Aparoy, K. K. Peddy and P. Reddanna, "Structure and Ligand Based Drug Design Strategies in the Development of Novel 5-LOX Inhibitors," *Current Medicinal Chemistry*, vol 19, pp. 3763-3778, 2012.
- [7] B. Blass, *Basic principles of drug discovery and development*. Elsevier Inc, 2015.

- [8] J. D. Durrant and R. E. Amaro, "Machine-Learning Techniques Applied to Antibacterial Drug Discovery," *Chem Biol Drug Des*, Special issue, pp. 14-21, 2015.
- [9] B. Ramsundar, P. Eastman, P. Walters and V. Pande, *Deep Learning for the Life Sciences: Genomics, Microscopy, Drug Discovery*. USA: O'reilly, 1 ed., 2019.
- [10] S. Ekins, J. Mestres and B. Testa, "In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling," *British Journal of Pharmacology*, vol. 152, pp. 9-20, 2007.
- [11] "Drug discovery." Available at <https://en.wikipedia.org/>. Accessed on 2020-03-21.
- [12] "Machine learning for drug discovery in a nutshell." Available at <https://medium.com/>. Accessed on 2020-03-21
- [13] A. Lavecchia, "Machine-learning approaches in drug discovery: methods and applications," *Drug Discovery Today*, vol. 20, no 3, pp 65-70, 2015.
- [14] B. Meriem and B. Mohamed, "Deep Learning for Ligand-Based Virtual Screening in Drug Discovery," in PAIS *Proceedings of the 3rd International Conference on Pattern Analysis and Intelligent Systems*, pp. 201-209, 2018.
- [15] S. Ou-yang, J. Lu, X. Kong, Z. Liang, C. Luo and H. Jiang, "Computational drug discovery," *Acta Pharmacologica Sinica*. Vol. 33, pp. 1131-1140, 2012.
- [16] I. Baskin, "The power of deep learning to ligand-based novel drug discovery," *Expert Opinion on Drug Discovery*, Vol. no. pp. 2020.
- [17] V. Rees, "Discovering and Designing Drugs with Artificial Intelligence," *Drug Target Review*, 2020.
- [18] G. Schneider, "Neural networks are useful tools for drug design," *Pergamon*, Available at [www.elsevier.com/](http://www.elsevier.com/). Accessed on 2020-03-23.
- [19] M. Kubat, *An Introduction to Machine Learning*. Springer International Publishing. AG 2015, 2017.
- [20] I. Wallach, M. Dzamba and A. Heifets, "AtomNet: A Deep Convolutional Neural Networks for Bioactivity Predictions in Structure-based Drug Discovery," Available at <https://www.arxiv-vanity.com/papers/>. Accessed on 2020-04-23.
- [21] R. Sabet, A. Fassihi, B. Hemmateenejad, L. Saghaei, R. Miri and M. Gholami, "Computer-aided design of novel antibacterial 3-hydroxypyridine-4-ones: application of QSAR methods based on the MOLMAP approach," *Journal Computer Aided Mol Des.*, vol.26, pp 349-361, 2012.
- [22] L. Terfloth and J. Gasteiger, "Neural networks and genetic algorithms in drug design," *Drug Discovery Today*, vol 6, pp 102-108, 2001.
- [23] E. Gawehn, J. A. Hiss and G. Schneider, "Deep Learning in Drug Discovery," Available at <https://onlinelibrary.wiley.com/>. Accessed on 2020-04-23.
- [24] J. Stokes, K. Yang, K. Swanson, T. Jaakkola, R. Barzilay, J. Collins et al., "A Deep Learning Approach to Antibiotic Discovery," *Cell, Elsevier Inc*, vol. 180, pp 688-702, 2020..
- [25] "Intelligent Drug Discovery", Available at <https://www2.deloitte.com/>. Accessed on 2020-05-23.
- [26] F. Lipinski, V. Maltarollo, P. Oliveira, A. da Silva and K. Honorio, "Advances and Perspectives in Applying Deep Learning for Drug Design and Discovery," *Frontiers in Robotics and AI*, vol 6, no 180, 2019.
- [27] "Ligand Based Drug Design" Available at <https://www.ukessays.com/>. Accessed on 2020-06-24.
- [28] A. Gaulton, A. Hersey, M. Nowotka, et al., "The ChEMBL database in 2017," *Nucleic Acids Res.*, Available at <https://pubmed.ncbi.nlm.nih.gov/27899562/> Accessed on 2020-04-22.
- [29] "moleculenet" Available at <http://moleculenet.ai/>. Accessed on 2020-04-20.