



Graph Neural Network for Drug Molecular Structures using Multiplex Graph

Sneha Khairé¹, Dr. Pawan Bhaladhare²

^{1,2} School of Computer Sciences & Engineering, Sandip University, Nashik, India.

¹ khairesneha69@gmail.com

Abstract

One of the most important tasks for artificial intelligence-assisted molecular design is the prediction of physicochemical qualities from molecular structures. To meet this problem, an increasing number of Graph Neural Networks (GNNs) have been proposed. By including more information in molecules, these models expand their expressive power while unavoidably increasing their computational complexity. In this work, we seek to create a powerful and effective GNN for molecular structures. By first representing each molecule as a two-layer multiplex graph, one layer of which only contains local connections that primarily capture covalent interactions and the other layer of which contains global connections that can simulate non-covalent interactions, we propose a molecular mechanics-driven approach to accomplishing this goal. Then, in order to balance the trade-off between expression strength and computing complexity, a corresponding message passing module is proposed for each layer. We construct the Multiplex Molecular Graph Neural Network based on these two modules. When it was verified using a dataset for big protein-ligand complexes and tiny compounds.

101216

Keywords Neural network, GNN, AI, deep learning, drug discovery, molecules.

DOI Number: 10.48047/NQ.2022.20.6.NQ22996 NeuroQuantology2022;20(6): 101216-101229

INTRODUCTION

From COVID-19 vaccines to solar cells, the discovery and design of novel compounds with desired features profoundly improves human society. The molecular discovery process is speeding up thanks in large part to artificial intelligence (AI). Predicting the physicochemical characteristics of molecules based on their structures is one of the most important tasks in AI-assisted molecular design. Discovering safe and effective treatments for human ailments is the goal of drug discovery. From target selection to meticulous clinical studies, drug development is time- and money-consuming. The optimal pharmaceutical

options for the following phase of treatment must be chosen at each checkpoint. The "hit-to-lead" approach is required to identify lead compounds from hits and confirm their medicinal potential. The polypharmacology idea [1] asserts that single or multiple medications typically interact with many targets, which contributes to clinical trials' adverse effects and lack of in vivo efficacy. In-vivo investigations would be ideal for each sickness model, but they would take a lot of time and effort, which is not feasible. Modern pharmaceutical R & D has relied significantly on computer-aided drug discovery or design technologies for the hit-to-lead process since the 1980s [2-4].



Despite the usage of this in silico approach, the pharmaceutical industry's R&D output has been declining since the middle of the 1990s.

Recent developments in drug discovery have made it possible for academics and the pharmaceutical industry to apply AI in important and cost-effective ways. Artificial Intelligence (AI) in drug discovery was made possible by the advent of high-performance processors like graphics processing unit computing and the huge volumes of chemical and biological data accumulated over decades [4-6]. Various AI-driven drug development pipelines and frameworks are being created in addition to the implementation of cutting-edge AI technologies in the drug development process [7-9].

AI is widely employed in both the corporate and academic world. Machine learning (ML), AI's fundamental component, has been used in a variety of domains, including data collecting and analysis. Applied mathematics and computational theory are essential to algorithm-based techniques like machine learning. Deep learning aided self-driving cars, improved speech recognition, and smarter search engines are just a few of the possible applications of ML models [10-13].

1.1 Motivation

Drug discovery and medicine that started with Ayurveda in India are now benefiting from an altogether different artificial intelligence (AI). Due to its ability to drastically cut the time and money required to develop new medicines, artificial intelligence-based drug discovery has recently attracted a lot of attention. The fields of drug research and development have made use of machine learning and deep learning technologies to develop new medication prospects. Machine learning and deep learning-based techniques are emerging at every level of the drug

development process as a result of the proliferation of drug-related data.

1.2 Scope of work

The paper involving drug discovery have benefited from artificial intelligence thanks to rapid advancements in computing power and the generation of enormous amounts of chemical and biological data. There has been an increase in the number of studies that use deep learning, particularly in the last few decades. Virtual screening, drug repositioning, and in silico prediction of pharmacokinetic properties (absorption, distribution, metabolism, and excretion-ADME) and toxicity are some examples of deep learning (DL) applications. The inherent characteristics of biological and chemical data, such as complexity, uncertainty, diversity, and high dimensionality, have made DL outperform traditional techniques in some applications related to drug design. To build robust and predictive models, deep learning relies on large and complex neural networks, both of which have the flexibility to adapt to a variety of problems.

For the purpose of lowering the computing cost of quantum chemistry calculations (DFT) and molecular dynamics simulations (MD) [14], various machine learning algorithms have been proposed for the representational learning of molecules in recent years. By modelling the molecule as a graph and applying a message-passing system to it, Graph Neural Networks (GNNs) have outperformed competing approaches [15].

The chemical characteristics, pairwise distances between atoms, and angular information have all been utilised by prior GNNs to better represent the interactions in molecules and boost the expressive capacity of approaches [16-22]. But incorporating this data into GNNs will raise the computational complexity.

We propose a new, powerful, and efficient GNN to overcome this restriction. Taking a

101217



cue from molecular mechanics approaches [23], we model only the local connections utilising angular information to save time and money. Reducing computational complexity, the innovative GNN moduler ignores angular information in nonlocal interactions. As a last step, we employ a Multiplex Graph to build a Graph Neural Network specifically for drug molecular structures.

RELATED WORK

2.1 Machine Learning (ML) Algorithms Used in Drug Discovery

Machine learning approaches have had a significant impact on drug discovery. Pharmaceutical companies have reaped major benefits from the application of machine learning algorithms in the process of drug discovery. Multiple models for predicting the chemical, biological, and physical properties of molecules have been constructed using machine learning methods [14–22]. It is possible to use machine learning techniques at any stage of the drug discovery process. When it comes to discovering new drug uses, for example, machine learning algorithms have been used to predict drug-protein interactions, identify medicine efficacy, ensure safety biomarkers, and optimize molecular bioactivity [23-27]. Examples of machine learning algorithms used in drug development include support vector machine (SVM), Naive Bayesian (NB) and Random Forest (RF) etc. [28-30].

A subset of artificial intelligence known as machine learning is not a homogenous category (AI). Supervised algorithms and unsupervised algorithms are the two main categories of machine learning algorithms. In supervised learning, previously labelled samples are used to predict fresh samples' labels. Unsupervised learning may be used to detect patterns in a collection of unlabeled examples. The reduction of a high-dimensional input to a lower-

dimensional input is a common goal of unsupervised learning techniques. Even if unsupervised learning is more successful in a low-dimensional space, the pattern that emerges is also more understandable. With semi- and reinforcement-learning approaches, a wide range of data sets may be utilised [31]. The development and expansion of successful machine learning algorithms during the drug discovery process are dependent on the availability of large amounts of data. In precision medicine and drug development, the necessity of high-quality data and well-defined training sets is substantially larger. For the creation of really effective tailored medications, it is necessary to describe all related pan-omic data, such as genomes, transcriptomics, and proteomics. There has been a boom in data development, gathering, and maintenance for drug research as online multi-omic databases and machine learning methods have become more commonly employed over the last two decades. To some extent, attempts at deciphering freshly created data have been effective thanks to analytical advancements. Drug research is currently helped by machine learning methodologies and networked databases via different software/web-tools, such Software. For example, the capacity of new data analytics to mix with known methodologies and current hypotheses to produce new hypotheses has helped with re-positioning, target identification, small molecule discovery, synthesis, and other applications [32–34]. It's possible to get a wide range of information from medical and multi-omics domains. Having data that is inconsistent and derived from a variety of sources is not unusual. Generalized linear models with non-negative reverse inference may be useful when dealing with multidimensional data (NB). There are a number of different ML techniques and models that are extensively utilized in

101218



various fields of research [34-35]. These include regression, clustering, regularization, neural networks, decision trees, and dimensionality reduction.

2.2 Deep Learning (DL) Methods Used in Drug Discovery

If you're interested in cutting-edge research and development, deep learning algorithms are an excellent place to start. Deep learning relies on the translation of artificial neural networks (ANNs) from theoretical and predicted applications to workable algorithms, which were first developed in the 1950s, as a fundamental pillar. A multidimensional data representation can be learned through abstraction using DL methods [37]. For example, picture identification and speech recognition have been addressed thanks to deep learning, which is a more advanced kind of machine learning (ML). Drug activity prediction, target discovery, and lead compound discovery have all benefited from the use of DL approaches [38–40]. In NN systems, the DL foundations are widely used to construct systems that can recognise, understand, and produce complicated data.

Three main types of artificial neural networks used in drug discovery are convolutional neural networks (CNNs), deep neural networks (DNNs) and recurrent neural networks (RNNs). The parameters that influence the selection of NNs from the subset's variations are numerous. Data flows from the input layer to the hidden layer(s) and then to the output layer (in DNNs) in a single path. To identify the outputs, trained supervised learning algorithms are typically used. Other machine learning techniques can be utilised to train neural networks in deep learning algorithms. Supervised and reinforcement learning approaches can be used to teach DNNs difficult tasks. Based on existing libraries and training sets, it is possible for a predictive neural network (DNN) like this one to forecast the chemical properties of

novel compounds [41,42]. QSAR models are currently being employed in the field to find the link between the chemistry and the activity of these medications. DL-based AI in drug discovery and development is now using QSAR analysis, one of the most advanced versions of DL. Researchers have used 2D chemical structures to discover the physicochemical parameters relevant to the action of the molecule. 3D-QSAR has helped researchers better understand how ligand-target interactions are influenced by the structure of the environment [43,44]. Predicting whether newly synthesised lead compounds will act on or miss specified targets has been done using QSAR in the pharmaceutical sector. Algorithmic methods for discovering and developing new products are not without flaws.

Research using these AI algorithms has yielded numerous errors and inaccuracies over time. In QSAR studies, it was discovered that NNs have a few drawbacks when compared to other ML techniques. NN redundancy and output blockage result from the existence of unnecessary descriptors. There is a risk that this redundancy will decrease the NN's efficiency and produce results that aren't ideal. The usage of unidentified descriptors is also a problem because they could affect the outcome. In order to acquire a smaller number of higher quality descriptors, these concerns have been addressed by using more specialised feature selection methods; however, NN-based QSAR will continue to face this difficulty. These NN-based assays face another challenge in implementing optimal network parameters without sacrificing accuracy [44]. These issues are not going away, despite the fact that workable remedies have been proposed and put in place [45]. Chemical synthesis and identification are prioritised only when significant research into target-molecule interactions has been completed.

101219



As part of the de novo drug design process, descriptive simplified molecular input line-entry system nomenclature (SMILES) nomenclature is employed widely. RNNs are self-learning neural networks that use generational input processing and the formation of hidden layers. For the creation of new chemical structures, an RNN-type long short-term memory has become the norm. RNNs differ from feed-forward neural networks and DNNs in that they use neurons connected in the same hidden layer to establish a working cycle of processing inputs and outputs, rather than connecting across layers. However, the initial SMILE training sets did not include these RNNs, which have shown promising results in producing unique SMILE structures that are logical, structurally accurate, and viable [46–49]. With the help of generative RNN models, Segler et al. were able to identify possible chemical structures that could be effective against *S. aureus* and *P. falciparum* (*P. falciparum*). Chemical structures with known efficacy against these target organisms were supplied to their models, and they were able to generate a total of 14 percent of the 6051 potential *S. aureus* molecules. In addition, 28% of the present *P. falciparum* compounds were produced by the model [50]. Chemical synthesis routes have typically been developed and implemented by chemists alone. However, as artificial intelligence (AI) advances, this position will increasingly entail computer-aided synthesis planning (CASP). Researchers have used Monte Carlo tree search (MCTS) methods to build CASP processes in recent studies. A good method for finding the best conditions and solutions is the MCTS methodology [52,53], which uses a random step search without branching. Three neural networks (NNs) and 12.4 million transformation rules were utilised by researchers Segler and Waller [54] to develop the first real CASP process, which

was built using this technology. One of the first NNs, an expansion node, searches the past and predicts if the 12.4 million transformation rules may be used to make the chemical. A better selection of transformations, such as those that are feasible and high-yielding, can be made by the expansion node as a result. With the rollout node, inputs are filtered so that only the most often reported transformation rules are included. This increases transformation success rates. In order to incorporate the new path into the search tree, the update node must be triggered. It was able to solve 80% of retrosynthesis questions in 5 seconds and more than 90% in 60 seconds using this technique. Chemical synthesis and reaction pathways based on artificial intelligence (AI) have been improved in several studies [55–57]. AI-based chemical synthesis and characterisation will be able to transfer drug discovery from the bench into in silico by increasing its use across the entire drug discovery process. Discovery and management will be more efficient as a result.

2.3 Graph Neural Networks (GNNs)

Graph Neural Networks (GNNs) have been proposed [16, 68, 69] and are gaining popularity as a means of learning the representations of graph-structured data using neural networks. Researchers started using GNNs to predict different molecular properties because of their superior performance in many tasks. Initial works treat the chemical bonds in molecules as edges and atoms as nodes to create graphs for molecules [16, 17, 18]. In addition, these GNNs incorporate a large number of carefully selected chemical features to boost performance. However, they ignore the importance of molecules' three-dimensional structures to their physiochemical properties. Therefore, subsequent works [70, 19, 71, 20] turn to account for atomic positions and

101220



interatomic distances to generate edges and interatomic features at the edges. Usually, a cutoff distance is used to find the neighbours in molecules instead of creating a complete graph to reduce the computational complexity and overfitting. However, depending on the cutoff value chosen, GNNs may not be able to differentiate between two molecules [21]. To solve this issue, angular information is further used in GNNs to achieve higher expressive power [21, 22]. The time and space complexity of those angle-aware GNNs, however, is much higher than that of the prior works. These methods cannot be applied to larger systems, such as macromolecules or batch learning.

Graph Multiplexing. Multiple types of edges connect a collection of nodes in the multiplex graph (also known as the multi-view graph). A group of graphs, where each type of edge connects to the same group of nodes, is one way to think about it. Each node's representation requires careful consideration of both intra-layer and cross-layer relationships. The multiplex graph has many potential applications [80, 81, 82], and numerous methods have been proposed to learn its embedding [71–75]. Previous work [84] implicitly represents molecular graphs as multiplex graphs and passes messages according to the edge types, which is useful for representation learning on molecules. In this work, we use the geometric information contained in molecules to create an explicit representation of molecules as multiplex graphs. Furthermore, we suggest a variety of message passing schemes for use across the multiplex graph's various layers.

PROPOSED APPROACH

Our Multiplex Molecular Graph Graph Neural Network module is described in this section.

3.1 Multiplex Molecular Graphs

The local, covalent interactions, including Ebond, which depends on bond lengths, Eangle, which depends on bond angles, and Edihedral, which depends on dihedral angles, are modelled in molecular mechanics methods [10] as $E = E_{\text{local}} + E_{\text{nonlocal}}$, where $E_{\text{local}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}$. The nonlocal model predicts the properties of atomic-pair interactions that are not local or covalent. By analysing the molecular mechanics method from a geometric standpoint, we learn that the local interactions capture both the angles θ_{local} and the pairwise distances d_{local} , while the nonlocal interactions only capture the pairwise distances d_{nonlocal} . As a result, we are motivated to model only the local interactions, rather than all interactions, using the angular information. In order to accomplish this, we separate the GI used in molecular mechanics techniques into two categories: local GI, which includes local and d_{local} , and global GI, which includes d_{local} and d_{nonlocal} . Next, we use the 3D model of the molecule to build interaction graphs with various GI. The local GI's edges can be generated in two different ways: via chemical bonds between nodes or by identifying the neighbours of each node within a small cutoff distance. We define the neighbours of each node within a large cutoff distance to generate the edges for the global GI. As if they were layers of an onion, the interaction graphs are used to construct a multiplex molecular graph $G = fGI; G_{\text{gg}}$, which is divided into a local layer G_{l} and a global layer G_{g} . Our model will take the resulting G as an input.

101221

3.2 Graph Neural Network using Multiplex Molecular Graph

Molecular properties are predicted using a neural network constructed using the Multiplex Molecular Graph module. In the Embedding section, the input node



embeddings are representations of the atomic numbers Z that are randomly initialised and trainable. Cartesian coordinates (r) of atoms are used to determine pairwise distances and angles in the RBF & SBF subroutine. The representations of eRBF and aSBF are built using the basis functions proposed in [8]. Then, we send messages by stacking modules of a Multiplex Molecular Graph. To obtain the output at the node level, an Output module is used in each Multiplex Molecular Graph module. All of the outputs from each node and layer are added together to form the final prediction y .

METRICS OF EVALUATION

Expressive Power: We examine how the inclusion of captured geometric information impacts the accuracy with which molecular structures are represented using Multiplex Molecular Graph. Comparatively, Multiplex Molecular Graph outperforms GNNs that only capture the pairwise distance information because it also accounts for the angular information in local connections. Since Multiplex Molecular Graph does not account for angular information in nonlocal connections, it theoretically has less expressive power than GNNs, which account for both pairwise distance information and angular information in global connections. We will conduct experiments to demonstrate that Multiplex Molecular Graph has excellent generalisation ability and state-of-the-art performance, despite the fact that expressive power does not directly speak about the generalisation ability of GNNs.

Complexity in Computation: We examine the time and space requirements for message passing in a Multiplex Molecular Graph to gain insight into the computational complexity of this model. When making the edges, we use the notation d_g for Group G and d_l for Group L to indicate the minimum distance required to create the edge. Average k -neighbors per

node in G_g is k_g and in G_l it's k_l . We have $k_g / d_g \geq 3$ and $k_l / d_l \geq 3$ for three-dimensional molecules. We can deduce that $k_g \geq k_l$ since $d_g > d_l$. As was mentioned previously, our MXM module's message passing operations require the computation of $O(2Nk_g + 2Nk_l + Nk_l + 2N)$ messages. Given that GNNs that capture angular information in global connections require $O(Nk_g^2)$ messages, Multiplex Molecular Graph is a much more efficient alternative.

PLATFORM AND TOOLS USED

System requirements as high as Windows 7, with Windows 10 being the recommended minimum.

It is possible to use the open-source Anaconda package and environment management system on Windows, macOS, and Linux. Packages and their dependencies can be quickly installed, executed, and updated using Conda. It also allows for simple local environment creation, saving, loading, and switching. Although it was designed for Python code, it can be used to distribute applications written in any language.

Python is a powerful programming language that is also very accessible. Object-oriented programming is made easy while still being highly effective, and it has efficient high-level data structures. Python's interpretive nature, together with its dynamic typing and elegant syntax, make it a great language for rapid application development in a wide variety of domains and on a wide variety of platforms. A Python 3.8 update. Only Python 3.7 and 3.9 are compatible with PyTorch on Windows. There is no support for Python 2.x.

Database of Results from Experiments

Exploring the chemical compound space rigorously and objectively is essential for computational de novo design of new drugs and materials. Its size scales combinatorially with molecular size, however, so there are still vast unexplored regions. We report on

101222



the calculated geometric, energetic, electronic, and thermodynamic properties of 134k CHONF-based small organic molecules that are stable under standard conditions. From the 166 billion organic molecules in the GDB-17 chemical universe, this set corresponds to the 133,885 species with up to nine heavy atoms (CONF). Energy-efficient geometries are reported, as are the corresponding harmonic frequencies, dipole moments, polarizabilities, energies, enthalpies, and free energies of atomization. All properties were determined using a quantum chemical method that operates at the level of B3LYP/6-31G(2df,p). In addition, out of the total of 134,000 molecules, 6,095 are constitutional isomers of the most common stoichiometry (C₇H₁₀O₂). All of their atomization energies, enthalpies, and free energies are reported at the more precise G4MP2 level of theory. For a meaningful, consistent, and all-encompassing chemical space of small organic molecules, this data set provides quantum chemical properties. This database has the potential to aid in the systematic identification of structure-property relationships, the benchmarking of existing methods, and the development of new methods like hybrid quantum mechanics/machine learning.

Methodology for establishing Experimental Parameters

In our experiments, we test our novel GNN's ability to generalise and its performance on the QM9 dataset for predicting molecular properties. For further evaluation, several state-of-the-art baseline models are also provided. In our subsequent scholarly article, we will show the outcomes of our proposed method.

CONCLUSION

Here we propose Multiplex Molecular Graph, a robust and effective GNN for predicting molecular properties. Our model can greatly enhance GNNs for molecules'

capacity for expression and memory efficiency. The innovation of GNN is found in its molecular-mechanics-based representation of molecules as a multiplex graph. The next steps of our research will involve conducting experiments on datasets that will show how effective and powerful novel GNN are in comparison to the current state-of-the-art baselines. Furthermore, since molecules can exist in a variety of shapes. To what extent these conformations affect our model and related GNNs is still unknown.

Reference

- [1] Reddy, A.S.; Zhang, S. Polypharmacology: Drug discovery for the future. *Expert Rev. Clin. Pharmacol.* 2013, 6, 41–47.
- [2] Sachdev, K.; Gupta, M.K. A comprehensive review of feature-based methods for drug target interaction prediction. *J. Biomed. Inform.* 2019, 93, 103159.
- [3] Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* 2019, 18, 463–477.
- [4] Kimber, T.B.; Chen, Y.; Volkamer, A. Deep learning in virtual screening: Recent applications and developments. *Int. J. Mol. Sci.* 2021, 22, 4435.
- [5] Lipinski, C.F.; Maltarollo, V.G.; Oliveira, P.R.; da Silva, A.B.F.; Honorio, K.M. Advances, and Perspectives in Applying Deep Learning for Drug Design and

101223



- Discovery. *Front. Robot. AI* 2019, 6, 108.
- [6] Rifaioğlu, A.S.; Atas, H.; Martin, M.J.; Cetin-Atalay, R.; Atalay, V.; Doğan, T. Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools, and databases. *Brief. Bioinform.* 2019, 20, 1878–1912.
- [7] Zhavoronkov, A.; Ivanenkov, Y.A.; Aliper, A.; Veselov, M.S.; Aladinskiy, V.A.; Aladinskaya, A.V.; Terentiev, V.A.; Polykovskiy, D.A.; Kuznetsov, M.D.; Asadulaev, A.; et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* 2019, 37, 1038–1040.
- [8] Feng, Q.; Dueva, E.; Cherkasov, A.; Ester, M. PADME: A Deep Learning-Based Framework for Drug-Target Interaction Prediction. *arXiv* 2018, arXiv:1807.09741.
- [9] Skalic, M.; Varela-Rial, A.; Jiménez, J.; Martínez-Rosell, G.; De Fabritiis, G. LigVoxel: Inpainting binding pockets using 3Dconvolutional neural networks. *Bioinformatics* 2019, 35, 243–250.
- [10] Fayyad, J.; Jaradat, M.A.; Gruyer, D.; Najjaran, H. Deep Learning Sensor Fusion for Autonomous Vehicle Perception and Localization: A Review. *Sensors* 2020, 20, 4220.
- [11] Deng, L.; Li, X. Machine Learning Paradigms for Speech Recognition: An Overview. *IEEE Trans. Audio Speech Lang. Process.* 2013, 21, 1060–1089.
- [12] Afouras, T.; Chung, J.S.; Senior, A.; Vinyals, O.; Zisserman, A. Deep Audio-visual Speech Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2018, 1.
- [13] Joachims, T.; Radlinski, F. Search Engines that Learn from Implicit Feedback. *Computer* 2007, 40, 34–40.
- [14] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.
- [15] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1263–1272, 2017.
- [16] KT Schutt, Pan Kessel, Michael Gastegger, KA Nicoli, Alexandre Tkatchenko, and K-R Muller. Schnetpack: A deep learning toolbox for atomistic systems. *Journal of chemical theory and computation*, 15(1):448–455, 2018.
- [17] Oliver T Unke and Markus Meuwly. Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *Journal of chemical theory and computation*, 15(6):3678– 3693, 2019.
- [18] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9):3564–3572, 2019.
- [19] Chengqiang Lu, Qi Liu, Chao Wang, Zhenya Huang, Peize Lin, and Lixin

101224



- He. Molecular property prediction: A multilevel quantum interactions modeling perspective. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 1052–1060, 2019.
- [20] Felix A Faber, Luke Hutchison, Bing Huang, Justin Gilmer, Samuel S Schoenholz, George E Dahl, Oriol Vinyals, Steven Kearnes, Patrick F Riley, and O Anatole Von Lilienfeld. Prediction errors of molecular machine learning models lower than hybrid dft error. *Journal of chemical theory and computation*, 13(11):5255–5264, 2017.
- [21] Johannes Klicpera, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020.
- [22] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.
- [23] Tamar Schlick. *Molecular modeling and simulation: an interdisciplinary guide*, volume 21. Springer Science & Business Media, 2010.
- [24] Lo, Y.C.; Rensi, S.E.; Tornig, W.; Altman, R.B. Machine learning in chemoinformatics and drug discovery. *Drug Discov. Today* 2018, 23, 1538–1546.
- [25] Talevi, A.; Morales, J.F.; Hather, G.; Podichetty, J.T.; Kim, S.; Bloomingdale, P.C.; Kim, S.; Burton, J.; Brown, J.D.; Winterstein, A.G.; et al. Machine Learning in Drug Discovery and Development Part 1: A Primer. *CPT Pharmacomet. Syst. Pharmacol.* 2020, 9, 129–142.
- [26] Gertrudes, J.C.; Maltarollo, V.G.; Silva, R.A.; Oliveira, P.R.; Honório, K.M.; da Silva, A.B. Machine learning techniques and drug design. *Curr. Med. Chem.* 2012, 19, 4289–4297.
- [27] Agarwal, S.; Dugar, D.; Sengupta, S. Ranking chemical structures for drug discovery: A new machine learning approach. *J. Chem. Inf. Model.* 2010, 50, 716–731.
- [28] Rodrigues, T.; Bernardes, G.J.L. Machine learning for target discovery in drug development. *Curr. Opin. Chem. Biol.* 2020, 56, 16–22.
- [29] Gao, D.; Chen, Q.; Zeng, Y.; Jiang, M.; Zhang, Y. Application of Machine Learning on Drug Target Discovery. *Curr. Drug Metab.* 2020.
- [30] Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* 2019, 18, 463–477.
- [31] Zo_mann, S.; Vercruysse, M.; Benmansour, F.; Maunz, A.; Wolf, L.; Marti, R.B.; Heckel, T.; Ding, H.; Truong, H.H.; Prummer, M.; et al. Machine learning-powered antibiotics phenotypic drug discovery. *Sci. Rep.* 2019, 9, 5013.
- [32] Ekins, S.; Puhl, A.C.; Zorn, K.M.; Lane, T.R.; Russo, D.P.; Klein, J.J.; Hickey, A.J.; Clark, A.M. Exploiting machine



- learning for end-to-end drug discovery and development. *Nat. Mater.* 2019, 18, 435–441.
- [33] Khamis, M.A.; Gomaa, W.; Ahmed, W.F. Machine learning in computational docking. *Artif. Intell. Med.* 2015, 63, 135–152.
- [34] Leelananda, S.P.; Lindert, S. Computational methods in drug discovery. *Beilstein J. Org. Chem.* 2016, 12, 2694–2718.
- [35] Maia, E.H.B.; Assis, L.C.; de Oliveira, T.A.; da Silva, A.M.; Taranto, A.G. Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Front. Chem.* 2020, 8, 343.
- [36] Talambedu, U.; Shanmugarajan, D.; Goyal, A.K.; Kumar, C.S.; Middha, S.K. Recent Updates on Computer-aided Drug Discovery: Time for a Paradigm Shift. *Curr. Top. Med. Chem.* 2017, 17, 3296–3307.
- [37] Réda, C.; Kaufmann, E.; Delahaye-Duriez, A. Machine learning applications in drug development. *Comput. Struct. Biotechnol. J.* 2020, 18, 241–252.
- [38] Breiman, L. Random Forests. *Mach. Learn.* 2001, 45, 5–32.
- [39] Webb, G.I. Naïve Bayes. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2010.
- [40] Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* 1995, 20, 273–297.
- [41] Rifaioğlu, A.S.; Atas, H.; Martin, M.J.; Cetin-Atalay, R.; Atalay, V.; Doğan, T. Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools, and databases. *Brief Bioinform.* 2019, 20, 1878–1912.
- [42] Dugger, S.A.; Platt, A.; Goldstein, D.B. Drug development in the era of precision medicine. *Nat. Rev. Drug Discov.* 2018, 17, 183–196.
- [43] Hulsen, T.; Jamuar, S.S.; Moody, A.R.; Karnes, J.H.; Varga, O.; Hedensted, S.; Spreafico, R.; Hafler, D.A.; McKinney, E.F. From Big Data to Precision Medicine. *Front. Med.* 2019, 6, 34.
- [44] Liu, B.; He, H.; Luo, H.; Zhang, T.; Jiang, J. Artificial intelligence and big data facilitated targeted drug discovery. *Stroke Vasc. Neurol.* 2019, 4, 206.
- [45] Cirillo, D.; Valencia, A. Big data analytics for personalized medicine. *Curr. Opin. Biotechnol.* 2019, 58, 161–167.
- [46] Kaiser, T.M.; Burger, P.B. Error Tolerance of Machine Learning Algorithms across Contemporary Biological Targets. *Molecules* 2019, 24, 2115.
- [47] LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* 2015, 521, 436–444.
- [48] Dana, D.; Gadhiya, S.V.; St Surin, L.G.; Li, D.; Naaz, F.; Ali, Q.; Paka, L.; Yamin, M.A.; Narayan, M.; Goldberg, I.D. Deep Learning in Drug Discovery and Medicine; Scratching the Surface. *Molecules* 2018, 23, 2384.
- [49] Korotcov, A.; Tkachenko, V.; Russo, D.P.; Ekins, S. Comparison of Deep Learning with Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets. *Mol. Pharm.* 2017, 14, 4462–4475.



- [50] Ekins, S. The Next Era: Deep Learning in Pharmaceutical Research. *Pharm. Res.* 2016, 33, 2594–2603.
- [51] D’Souza, S.; Prema, K.V.; Balaji, S. Machine learning models for drug-target interactions: Current knowledge and future directions. *Drug Discov. Today* 2020, 25, 748–756.
- [52] Baskin, I.I.; Winkler, D.; Tetko, I.V. A renaissance of neural networks in drug discovery. *Expert Opin. Drug Discov.* 2016, 11, 785–795.
- [53] Verma, J.; Khedkar, V.M.; Coutinho, E.C. 3D-QSAR in drug design—A review. *Curr. Top. Med. Chem.* 2010, 10, 95–115.
- [54] Ghasemi, F.; Mehridehnavi, A.; Pérez-Garrido, A.; Pérez-Sánchez, H. Neural network and deep-learning algorithms used in QSAR studies: Merits and drawbacks. *Drug Discov. Today* 2018, 23, 1784–1790.
- [55] Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* 2014, 15, 1929–1958.
- [56] Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* 2018, 23, 1241–1250.
- [57] Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* 2018, 4, eaap7885.
- [58] Yuan, W.; Jiang, D.; Nambiar, D.K.; Liew, L.P.; Hay, M.P.; Bloomstein, J.; Lu, P.; Turner, B.; Le, Q.T.; Tibshirani, R.; et al. Chemical Space Mimicry for Drug Discovery. *J. Chem. Inf. Model.* 2017, 57, 875–882.
- [59] Gupta, A.; Müller, A.T.; Huisman, B.J.H.; Fuchs, J.A.; Schneider, P.; Schneider, G. Generative Recurrent Networks for De Novo Drug Design. *Mol. Inf.* 2018, 37, 1700111.
- [60] Segler, M.H.S.; Kogej, T.; Tyrchan, C.; Waller, M.P. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Cent. Sci.* 2018, 4, 120–131.
- [61] De Almeida, A.F.; Moreira, R.; Rodrigues, T. Synthetic organic chemistry driven by artificial intelligence. *Nat. Rev. Chem.* 2019, 3, 589–604.
- [62] Chan, H.C.S.; Shan, H.; Dahoun, T.; Vogel, H.; Yuan, S. Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol. Sci.* 2019, 40, 592–604.
- [63] Segler, M.H.S.; Preuss, M.; Waller, M.P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* 2018, 555, 604–610.
- [64] Segler, M.H.S.; Waller, M.P. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chem. Eur. J.* 2017, 23, 5966–5971.
- [65] Zhou, Z.; Li, X.; Zare, R.N. Optimizing Chemical Reactions with Deep Reinforcement Learning. *ACS Cent. Sci.* 2017, 3, 1337–1344.
- [66] Coley, C.W.; Barzilay, R.; Jaakkola, T.S.; Green, W.H.; Jensen, K.F. Prediction of Organic Reaction

101227



- Outcomes Using Machine Learning. *ACS Cent. Sci.* 2017, 3, 434–443.
- [67] Lee, A.A.; Yang, Q.; Sresht, V.; Bolgar, P.; Hou, X.; Klug-McLeod, J.L.; Butler, C.R. Molecular Transformer unifies reaction prediction and retrosynthesis across pharma chemical space. *Chem. Commun.* 2019, 55, 12152–12155.
- [68] Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. Learning convolutional neural networks for graphs. In *International conference on machine learning*, pages 2014–2023, 2016.
- [69] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.
- [70] Kristof T Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus R Müller, and Alexandre Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature communications*, 8(1):1–8, 2017.
- [71] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9):3564–3572, 2019.
- [72] Weiyi Liu, Pin-Yu Chen, Sailung Yeung, Toyotaro Suzumura, and Lingli Chen. Principled multilayer network embedding. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 134–141. IEEE, 2017.
- [73] Hongming Zhang, Liwei Qiu, Lingling Yi, and Yangqiu Song. Scalable multiplex network embedding. In *IJCAI*, volume 18, pages 3082–3088, 2018.
- [74] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, pages 593–607. Springer, 2018.
- [75] Yukuo Cen, Xu Zou, Jianwei Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. Representation learning for attributed multiplex heterogeneous network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1358–1368, 2019.
- [76] Yao Ma, Suhang Wang, Chara C Aggarwal, Dawei Yin, and Jiliang Tang. Multi-dimensional graph convolutional networks. In *Proceedings of the 2019 SIAM International Conference on Data Mining*, pages 657–665. SIAM, 2019.
- [77] Ta-Chu Kao and Mason A Porter. Layer communities in multiplex networks. *Journal of Statistical Physics*, 173(3-4):1286–1302, 2018.
- [78] Bohyun Lee, Shuo Zhang, Aleksandar Poleksic, and Lei Xie. Heterogeneous multi-layered network model for omics data integration and analysis. *Frontiers in Genetics*, 10:1381, 2020.
- [79] Duo Wang, Mateja Jamnik, and Pietro Lio. Abstract diagrammatic reasoning with multiplex graph networks. In *International Conference on Learning Representations*, 2020.



- [80] Chence Shi, Minkai Xu, Hongyu Guo, Ming Zhang, and Jian Tang. A graph to graphs framework for retrosynthesis prediction. arXiv preprint arXiv:2003.12725, 2020.

