

# FastDTI: Drug-Target Interaction Prediction using Multimodality and Transformers

Mathijs Boezer, Maryam Tavakol\*, and Seyedeh Zahra Sajadi

Eindhoven University of Technology, Eindhoven, The Netherlands

## Abstract

Recent advances in machine learning have proved effective in the application of drug discovery by predicting the drugs that are likely to interact with a protein target of a certain disease, leading to prioritizing drug development and re-purposing efforts. State-of-the-art techniques in Drug-Target Interaction (DTI) prediction are often computationally expensive and can only be trained on small specialized datasets. In this paper, we propose a novel architecture, called *FastDTI*, utilizing pre-trained transformers and graph neural networks in a self-supervised manner on large-scale (unlabeled) data, which additionally allows for embedding of multimodal input representations, for both drug and protein properties. Extensive empirical study demonstrates that our approach outperforms state-of-the-art DTI methods on the KIBA benchmark dataset, while greatly improving the computational complexity of training, about 200 times faster, leading to excellent performance results.

## 1 Introduction

The recent COVID-19 pandemic, which is not the first and will likely not be the last pandemic [18], has caused devastating disruptions on health, society, and economy world-wide. It has shown that not being able to rapidly develop effective treatments for a new disease is a significant shortcoming of the world's ability to respond to a pandemic, given that the pipeline of developing new drugs are too slow in most scenarios. One way to efficiently speed up the process of finding treatments for a novel disease is to re-purpose existing drugs [3].

However, this can be highly challenging due to the massive number of chemical compounds that form the candidate drugs for the newly-arrived disease [9]. Hence, we need functional techniques to accurately identify which drugs among the thousands of candidates are the best options for further testing.

Until now, lots of data have been collected on both drugs and drug targets (i.e., proteins), as well as on the interactions between them [23]. Using this data, machine learning models can replace the additional testing in the labs and act as a decision-support system in addressing the problem of Drug-Target Interaction (DTI) prediction, which can speed up the procedure of drug development and hence world's response to pandemics.

The efficacy of machine learning, particularly deep learning approaches, for DTI prediction has been already demonstrated in the literature [2, 27, 32, 1, 30, 21, 22]. Early models utilize techniques similar to those used for recommender systems, with this idea that similar drugs interact with similar targets [9, 10]. However, these techniques fail to cope with the "cold-start" problem, e.g., drugs or proteins that have no known interactions, which is the case in the problem at-hand. Docking [14] is an alternative method in which interactions are determined by a physical simulation using spatial models of the drug and protein. Nevertheless, the 3D data on either drug or their targets are often not available, which leads to limited benefits of these methods in practice [28, 7].

On the other hand, state-of-the-art deep learning techniques employ various architectures, such as convolutional neural networks [2], recurrent neural networks [30], graph neural networks [15], and transformers [4], to automatically discover complex features in the input, and have been shown to be effective for DTI prediction. Although these mod-

---

\*Corresponding Author: m.tavakol@tue.nl

els perform well on small specialized datasets that only deal with a single type of protein, they are greatly limited by their computational complexity and, as a result, are unable to train on general (non-specialized) DTI datasets. Therefore, we aim to get the best of both worlds, by developing a model that is both accurate and fast, and can be trained on large amounts of DTI data, instead of being restricted to a single type of protein (out of the numerous types of proteins [13]). Besides, the existing work rarely use more than one modality for the drug or the protein, while additional modalities may improve the performance, as DeepH-DTA [2] has indicated by using two modalities for the drug.

In this paper, we address these two shortcomings, namely the computational complexity and the lack of multimodality for both drug and protein representations, and propose a novel approach, called *FastDTI*, to improve the time complexity as well as the performance of predictions. We employ the recent developments in natural language processing and graph neural networks to create a model that can leverage different modalities of the drug and protein input, including their properties, in one model. Subsequently, pretrained submodels are utilized to embed their sequences and/or graph representations during the pre-processing step, leading to reduced computational complexity during the training. Additionally, the modalities of drug properties as well as protein properties are introduced, providing valuable information to the model for making more accurate predictions.

Our extensive empirical study illustrates that FastDTI outperforms state-of-the-art DTI methods on the KIBA benchmark dataset, while greatly improves the computational complexity of the training, about 200 times faster, leading to excellent performance results for DTI prediction.

## 2 Related Work

The application of deep learning in the DTI prediction problem has significantly improved upon the traditional machine learning methods [31], and is a notable candidate to train on the large amount of available data [9]. The efficacy of deep learning is evident from many existing work [2, 27, 32, 1, 30, 21, 22], which exert various architectures such as convolutional neural networks [2], recurrent neu-

ral networks [30], graph neural networks [15], and transformers [4] for addressing the problem.

Among the state-of-the-art approaches based on deep learning, DeepH-DTA [2] is the most accurate model, primarily due to using both graph neural networks and three modalities of the input. Almost all other models leverage two modalities: one for the drug, and one for the protein. For the protein, DeepH-DTA uses the protein sequence, and for the drug, both the SMILES sequence [26] and a graph representation of the chemical structure, leading to three total modalities. In addition, it employs a sophisticated graph neural network named heterogeneous graph attention (HGAT) [25], which was not applied to DTI prediction before. However, DeepH-DTA is limited by its computational complexity. The authors report that it would take over four days to train on the small KIBA dataset [24], even if powerful hardware were used.

Other state-of-the-art DTI prediction approaches include GraphDTA [15] and DeepDTA [16]. GraphDTA introduces the use of graph neural networks, achieving decent accuracy at the cost of high complexity. Inversely, DeepDTA [16] is a simple and fast model, which has low complexity at the cost of low accuracy. In this paper, we present an approach that is both accurate and efficient, while incorporating more than one modality for the drugs as well as for the proteins.

## 3 FastDTI

In this section, we propose a novel technique for Drug-Target Interaction (DTI) prediction, called *FastDTI*, that aims to achieve the performance of the state-of-the-arts while significantly decreasing their computational complexity. Hence, it allows to go beyond small specialized datasets and utilize the large amounts of data that were previously inaccessible for DTI prediction problems.

To this end, we take into account two main principles to ensure both high accuracy of the predictions and low computational requirements: (i) to offload the computation to the pre-processing step whenever possible, in particular, for sequence and graph embedding representations and (ii) to incorporate additional modalities for both drugs and proteins. The former follows from the main bottleneck in the existing work, where the processing

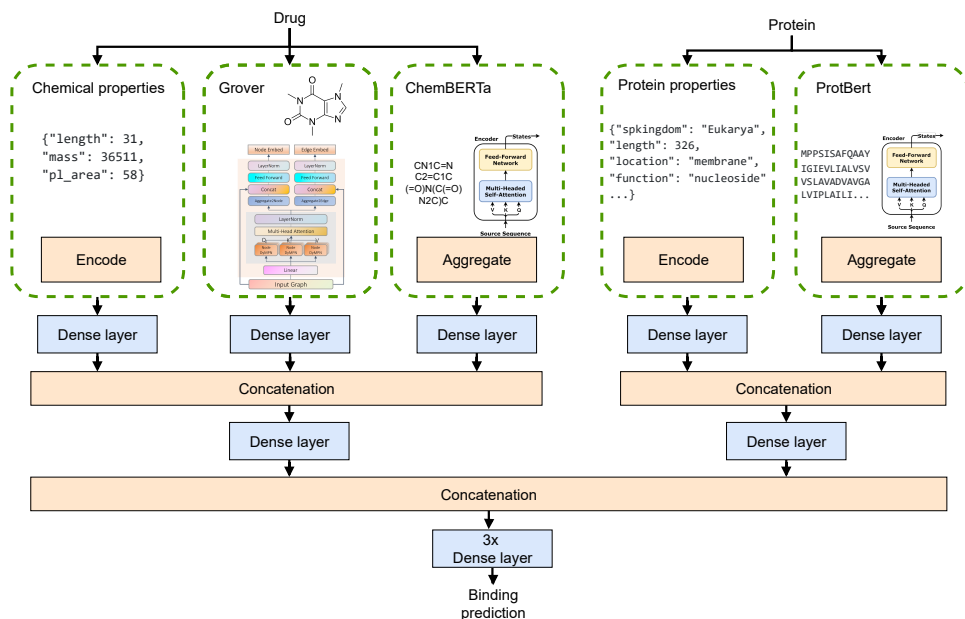


Figure 1: The overall architecture of FastDTI. Grover figure is taken from [20] (Fig. 1).

of the sequences and graphs is very expensive. We address this problem by employing pretrained models to compute embeddings of such complex representations. These embeddings can be computed during the pre-processing step, which dramatically reduces the workload during the training, without sacrificing the performance of the model.

The second principle enhances the accuracy by providing additional information to the model in the form of drug and protein properties. Previous methods typically leverage only one modality for the drug and one for the protein. Instead, we add two more modalities into FastDTI, which compared to DeepH-DTA [2], that incorporate three modalities, leads to a total of five modalities. The properties used by FastDTI are each specifically chosen for giving valuable information for the prediction task, where some are computable, meaning that they can be derived from the protein or drug sequences, while other properties are obtained from measurements. Nevertheless, the properties of both drugs and targets can be expanded by incorporating additional attributes using expert knowledge which is out of the scope of this work.

The overall network architecture of FastDTI is pictured in Figure 1. On the left, is the component that learns a representation for the drugs, consist-

ing of three sub-components, one for each of the drug modalities. The first modality (right) is a pretrained transformer model called ChemBERTa [5] for the SMILES sequence of the drug, which is based on the RoBERTa architecture [12], trained on a large dataset of unlabeled SMILES sequences. Second modality (middle) is a pretrained graph neural network called Grover [20] for the graph structure of the drug, based on a GTransformer architecture, trained on a large dataset of unlabeled drug graphs. The third one (left) is a dense layer for the chemical properties of the drugs, which consist of their **sequence length**, **molecular weight**, and **polar surface area**. Similarly on the right, is the component to learn the protein representations, which leverages a pretrained model, ProtBERT [8], based on BERT [6], trained on a large dataset of protein sequences, to process those sequences (right) and a dense layer for the protein properties (left), which include their **sequence length**, **molecular weight**, **isoelectric point**, **subcellular location**, **superkingdom**, and **function**.

Once the encoding of both the proteins and the drugs from different modalities are computed, the obtained representations are concatenated and fed into the DTI prediction component, with several

Table 1: Datasets used for the verification of the properties of FastDTI.

Name	#Drugs	#Proteins	#Interactions	Purpose
KIBA	2,111	229	118,254	Benchmark against SOTA models
STITCH <sub>full</sub>	116,224,359	9,643,763	~1,600,000,000	Creating STITCH <sub>human</sub>
STITCH <sub>human</sub>	79,047	1,934	9,418,794	Create a general DTI prediction model

dense layers and an output layer to produce the final prediction scores. All the layers incorporate a ReLU activation function. Subsequently, the training is conducted by optimizing a Mean Squared Error loss function. In addition, we train the model to make it robust to partially missing data due to some unknown properties of either proteins or drugs. To do so, some values are replaced by a missing value with a configurable probability during training. This ensures the model is exposed to various combinations of present and missing values, securing robustness to missing data.

## 4 Empirical Study

### 4.1 Experimental Setups

In this section, we evaluate the performance of FastDTI compared to several state-of-the-art techniques for the problem of drug-target prediction on both small- and large-scale data. Accordingly, we first conduct the experiments on the KIBA (Kinase Inhibitor BioActivity) dataset [24], which is a relatively small benchmark data to compare DTI predictions methods, including the models that are computationally expensive and cannot be easily evaluated on larger-scale datasets. We prepare the data according to the procedure described in [10], for a fair comparison of DTI prediction models. In addition, to examine the performance of FastDTI on a larger dataset, we use the STITCH data [23], which is by far the most extensive DTI data that exists, and encompasses a vast number of proteins across numerous organisms, not just humans. Table 1 summarizes the statistics of both datasets.

Subsequently, the data is split into training (80%) and testing (20%) sets, and we employ 5-fold cross-validation on the train set to tune the hyperparameters. After model selection, most dense layers have 1024 neurons, except for the layers after the first two concatenations, which have 512.

Dropout is set to 0.3 for the KIBA dataset, but 0.0 for the STITCH data. In addition, we utilize four metrics to compare the performance of the models: (i) Mean Squared Error (MSE), (ii) Concordance Index (CI), (iii)  $r_m^2$  [19], and (iv) seconds per epoch (for the KIBA dataset and only for training time). Furthermore, we compute the standard deviation for the CI and  $r_m^2$  measures. The standard deviation of the MSE of most models is unfortunately not available, so this is omitted for FastDTI as well. Moreover, a selection of baseline models is made from the best-performing state-of-the-art deep learning techniques, as well as two methods that are not based on deep learning. The references to the baseline methods are indicated in Table 2

### 4.2 Experimental Results

#### 4.2.1 Overall Performance

In the first experiment, we compare FastDTI to the state-of-the-art DTI prediction approaches on the KIBA dataset. This is especially important because most existing work cannot train on the STITCH dataset, due to high computational requirements. However, the results on the time complexity in this experiment can be generalized to a larger data such as STITCH. The evaluation results for the baseline methods are taken from the reported values by their authors (which use the same setup as this experiment), except for “second per epoch”. This metric is independently verified to ensure the comparison is made on the same hardware, which is a cloud-based machine using an NVIDIA RTX A5000 graphics card and AMD Ryzen Threadripper 1950X processor. Moreover, FastDTI is trained 10 times, each for 1000 epochs.

The overall performance of different techniques in terms of four above-mentioned measures is outlined in Table 2. The experimental results illustrate that FastDTI performs well on the KIBA dataset, and outperforms all the competitors on all the

Table 2: FastDTI compared to various state-of-the-art models, as well as traditional methods of SimBoost and KronRLS. Note that seconds per epoch is not a valid metric for KronRLS and SimBoost, due to not training using epochs. Additionally, the authors of C-A DTI did not report the  $r_m^2$  or make their code available for measuring the seconds per epoch.

Model	MSE	CI (std)	$r_m^2$ (std)	Seconds / epoch
(Non-DL) Machine learning methods				
KronRLS [17]	0.411	0.782 (0.001)	0.342 (0.001)	N/A
SimBoost [10]	0.222	0.836 (0.001)	0.629 (0.007)	N/A
Deep learning methods				
<b>FastDTI</b>	<b>0.107</b>	<b>0.931 (0.002)</b>	<b>0.820 (0.004)</b>	<b>6</b>
DeepH-DTA [2]	0.111	0.927 (0.003)	0.799 (0.004)	1412
FusionDTA [30]	0.130	0.906 (0.001)	0.793 (0.002)	258
GraphDTA [15]	0.148	0.891 (0.001)	0.730 (0.015)	1344
C-A DTI [11]	0.175	0.874 (0.001)	N/A	N/A
DeepDTA [16]	0.181	0.868 (0.004)	0.711 (0.021)	162
ML-DTI [29]	0.196	0.862 (0.006)	0.727 (0.012)	25

valuated metrics. Additionally, a t-test is carried out to compare the  $r_m^2$  of FastDTI with DeepH-DTA, which results in a conclusion that FastDTI has a higher  $r_m^2$  with  $p < 0.0001$ . Furthermore, our approach is over 200 times faster than the second most accurate model in the state-of-the-arts, indicating that our approach manages to avoid the trade-off between accuracy and training time.

#### 4.2.2 Performance on STITCH data

The second experiment explores how FastDTI fares on the STITCH data, which is significantly bigger than the KIBA dataset. Consequently, we examine to what extent our approach is able to predict the STITCH interaction values in a reasonable time. To this end, FastDTI is trained on the STITCH data for 50 epochs and the predicted values are compared to the true values in terms of MSE.

The evaluation result demonstrates that FastDTI is able to achieve the MSE of 0.094 on the test set, indicating that the error of prediction is small on average. To better understand this value, the predictions from our model are visually compared against the true values in Figure 2. The plot represents a random subset of 10,000 points from the STITCH data, due to the full data being too large, and the scores are converted from log-odds back to the probabilities for better interpretability. Subsequently, the figure shows that the model appears

to have a harder time making predictions compared to the first experiment, which might be due to the fact that the STITCH data is more complex than the KIBA dataset. However, note that the MSE is not comparable between the KIBA and STITCH datasets, due to expressing the interaction strength in different units. The STITCH set is much less filtered and contains several orders of magnitude more drugs and proteins, including numerous types of proteins, while the KIBA dataset exclusively contains kinases. Nevertheless, the model performs promisingly well on this large-scale data.

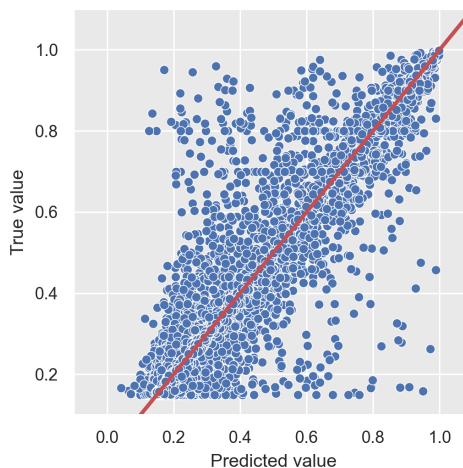


Figure 2: FastDTI Performance on STITCH data.

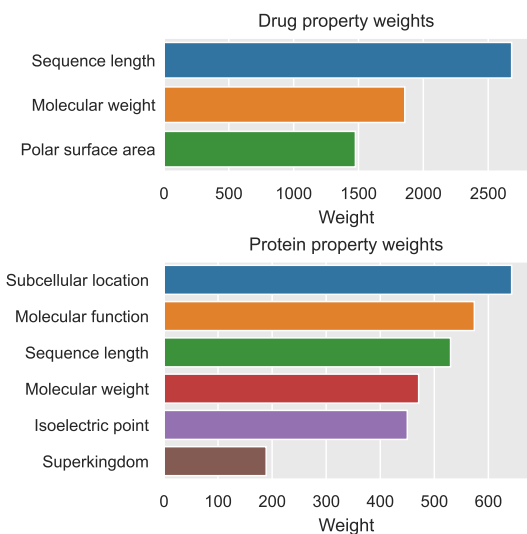


Figure 3: The sum of the absolute values for the weights of the drug and protein properties used by FastDTI on the STITCH dataset.

### 4.2.3 Impact of Multimodality

In this experiment, we study the impact of multimodality to verify whether the additional modalities of drug and protein properties would benefit the performance. Furthermore, the usefulness of each individual property is evaluated using the weights of the model. To do so, we compare the performance of the model with and without properties, and the effect of each property is examined from the weights that the model assigns to them during the training. Hence, we use two configurations of the model for this experiment: FastDTI in its full form with five modalities and FastDTI without drug and protein properties. Each model is trained 10 times for 50 epochs on the STITCH dataset. Consequently, the weights of the properties in the trained model are extracted, by computing the sum of absolute values per each property. For properties using a one-hot encoding, the sum is taken over each dimension of the encoding as well.

The performance in terms of average MSE over the 10 runs on the test set is 0.104 without properties and 0.094 with properties, showing a 0.010 (with a standard deviation of 0.003) advantage of incorporating the properties. In addition, Figure 3 displays the weights of the trained network per property for drugs as well as proteins,

which demonstrates that for the drug properties, the **sequence length** has the highest weight, followed by the **molecular weight**, and finally the **polar surface area**. Note that the **sequence length** for the drug is the length of the SMILES representation. Meanwhile, for the proteins, the **subcellular location** has the highest weight, and the **superkingdom** has the lowest weight by a large margin, which denotes lower importance.

The results illustrate that the model decreases the average error by 0.010 when additional modalities of the drug and protein properties are incorporated. This indicates that using properties enhances the model’s predictive performance, which is a valuable addition to DTI prediction. Nonetheless, the weights for the drug properties show some variation in their usefulness. Interestingly, the **polar surface area** has the lowest weight, even though it is hypothesized to be highly significant for DTI prediction due to indicating the size of the molecule. On the contrary, the **molecular weight** has a significant effect to the model, and the **sequence length** is even more so. Similarly, the results for the protein properties show considerable variance. The **superkingdom** has a relatively low weight, while **subcellular location** and **molecular function** have high weights. This potentially indicates that the **superkingdom** does not provide any significant information that the model can use for DTI prediction, unlike the other properties.

To summarize, it is explicit that the model benefits from additional properties. However, not every property is equally useful, therefore, there is the potential in future work to use expert knowledge to design better properties to further enhance the performance of the model for DTI prediction.

## 5 Conclusion

We presented *FastDTI*, a novel approach to fast but accurate drug-target interaction prediction that leverages pretrained models to compute embeddings of the input while including additional modalities in the form of drug and protein properties. Previous models either achieve a good performance at the cost of time complexity or a good training time at the cost of the performance. FastDTI has a statistically significant improvement in performance compared to the state-of-the-arts while

being several orders of magnitude faster to train. Additionally, FastDTI is the first deep learning DTI model to be trained on the STITCH dataset, which encompasses a far wider range of proteins, allowing for a general DTI prediction model. The empirical results indicate that our approach outperforms the competitors in terms of the predictive performance on the small benchmark dataset, while greatly improving the computational complexity.

In the future, FastDTI can be improved by further fine-tuning the pretrained models for the DTI prediction task, at the cost of extra training time, or by incorporating additional properties for the drugs and proteins. Moreover, the application of FastDTI to other related problems, such as protein-protein prediction, can be explored as future work.

## References

- [1] K. Abbasi, P. Razzaghi, A. Poso, S. Ghanbari-Ara, and A. Masoudi-Nejad. Deep learning in drug target interaction prediction: Current and future perspectives. *Current Medicinal Chemistry*, 28(11):2100–2113, 2021. doi: 10.2174/0929867327666200907141016.
- [2] M. Abdel-Basset, H. Hawash, M. Elhoseny, R. K. Chakraborty, and M. Ryan. DeepHDTA: deep learning for predicting drug-target interactions: a case study of covid-19 drug repurposing. *Ieee Access*, 8:170433–170451, 2020. doi: 10.1109/ACCESS.2020.3024238.
- [3] T. Ashburn and K. B. Thor. Drug repositioning: identifying and developing new uses for existing drugs. *Nature reviews Drug discovery*, 3(8):673–683, 2004. doi: 10.1038/nrd1468.
- [4] L. Chen, X. Tan, D. Wang, F. Zhong, X. Liu, T. Yang, X. Luo, K. Chen, H. Jiang, and M. Zheng. TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*, 36(16):4406–4414, 2020. doi: 10.1093/bioinformatics/btaa524.
- [5] S. Chithrananda, G. Grand, and B. Ramsundar. ChEMBERTA: Large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885*, 2020.
- [6] J. Devlin, M. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [7] A. Dhakal, C. McKay, J. J. Tanner, and J. Cheng. Artificial intelligence in the prediction of protein–ligand interactions: recent advances and future directions. *Briefings in Bioinformatics*, 23(1):bbab476, 2022. doi: 10.1093/bib/bbab476.
- [8] A. Elnaggar, M. Heinzinger, C. Dallago, G. Rehawi, Y. Wang, L. Jones, T. Gibbs, T. Feher, et al. ProtTrans: towards cracking the language of life’s code through self-supervised learning. *bioRxiv*, pages 2020–07, 2021. doi: 10.1101/2020.07.12.199554.
- [9] A. Ezzat. *Challenges and solutions in drug-target interaction prediction*. PhD thesis, 2018.
- [10] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, and M. Ester. Simboost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. *Journal of cheminformatics*, 9(1):1–14, 2017. doi: 10.1186/s13321-017-0209-z.
- [11] K. Koyama, K. Kamiya, and K. Shimada. Cross attention DTI: drug-target interaction prediction with cross attention module in the blind evaluation setup. 2020.
- [12] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.
- [13] H. Lodish, A. Berk, C. A. Kaiser, C. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira, et al. *Molecular cell biology*. Macmillan, 2008.
- [14] G. M. Morris and M. Lim-Wilby. Molecular docking. In *Molecular modeling of proteins*, pages 365–382. Springer, 2008. doi: 10.1007/978-1-59745-177-2\_19.
- [15] T. Nguyen, H. Le, T. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh. GraphDTA: Predicting

- drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8):1140–1147, 2021. doi: 10.1093/bioinformatics/btaa921.
- [16] H. Öztürk, A. Özgür, and E. Ozkirimli. Deep-DTA: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17):i821–i829, 2018. doi: 10.1093/bioinformatics/bty593.
- [17] T. Pahikkala, A. Airola, S. Pietilä, S. Shakyawar, A. Sz wajda, J. Tang, and T. Aittokallio. Toward more realistic drug–target interaction predictions. *Briefings in bioinformatics*, 16(2):325–337, 2015. doi: 10.1093/bib/bbu010.
- [18] J. Piret and G. Boivin. Pandemics throughout history. *Frontiers in microbiology*, 11:631736, 2021. doi: 10.3389/fmicb.2020.631736.
- [19] P. Pratim Roy, S. Paul, I. Mitra, and K. Roy. On two novel parameters for validation of predictive qsar models. *Molecules*, 14(5):1660–1701, 2009. doi: 10.3390/molecules14051660.
- [20] Y. Rong, Y. Bian, T. Xu, W. Xie, Y. Wei, W. Huang, and J. Huang. Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information Processing Systems*, 33:12559–12571, 2020.
- [21] S. Z. Sajadi, M. A. Zare Chahooki, S. Gharaghani, and K. Abbasi. Autodti++: deep unsupervised learning for dti prediction by autoencoders. *BMC bioinformatics*, 22(1):1–19, 2021. doi: 10.1186/s12859-021-04127-2.
- [22] S. Z. Sajadi, M. A. Zare Chahooki, M. Tavakol, and S. Gharaghani. Matrix factorization with denoising autoencoders for prediction of drug–target interactions. *Molecular Diversity*, pages 1–11, 2022. doi: 10.1007/s11030-022-10492-8.
- [23] D. Szklarczyk, A. Santos, C. Von Mering, L. J. Jensen, P. Bork, and M. Kuhn. Stitch 5: augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic acids research*, 44(D1):D380–D384, 2016. doi: 10.1093/nar/gkv1277.
- [24] J. Tang, A. Sz wajda, S. Shakyawar, T. Xu, P. Hintsanen, K. Wennerberg, and T. Aittokallio. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *Journal of Chemical Information and Modeling*, 54(3):735–743, 2014. doi: 10.1021/ci400709d.
- [25] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu. Heterogeneous graph attention network. In *The world wide web conference*, pages 2022–2032, 2019. doi: 10.1145/3308558.3313562.
- [26] D. Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988. doi: 10.1021/ci00057a005.
- [27] M. Wen, Z. Zhang, S. Niu, H. Sha, R. Yang, Y. Yun, and H. Lu. Deep-learning-based drug–target interaction prediction. *Journal of proteome research*, 16(4):1401–1409, 2017. doi: 10.1021/acs.jproteome.6b00618.
- [28] K. Wüthrich. Protein structure determination in solution by nmr spectroscopy. *Journal of Biological Chemistry*, 265(36):22059–22062, 1990. doi: 10.1016/S0021-9258(18)45665-7.
- [29] Z. Yang, W. Zhong, L. Zhao, and C. Y.-C. Chen. Ml-dti: Mutual learning mechanism for interpretable drug–target interaction prediction. *The Journal of Physical Chemistry Letters*, 12(17):4247–4261, 2021. doi: 10.1021/acs.jpcclett.1c00867.
- [30] W. Yuan, G. Chen, and C. Y.-C. Chen. Fusiondta: attention-based feature polymerizer and knowledge distillation for drug–target binding affinity prediction. *Briefings in Bioinformatics*, 23(1):bbab506, 2022. doi: 10.1093/bib/bbab506.
- [31] L. Zhang, J. Tan, D. Han, and H. Zhu. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug discovery today*, 22(11):1680–1685, 2017. doi: 10.1016/j.drudis.2017.08.010.
- [32] P. Zhang, Z. Wei, C. Che, and B. Jin. Deepmgt-dti: Transformer network incorporating multilayer graph information for drug–target interaction prediction. *Computers in Biology and Medicine*, page 105214, 2022. doi: 10.1016/j.combiomed.2022.105214.