Drug–Target Interaction Prediction via Deep Learning and Model Ensembling

Radito Dhali [24141150]
Department of Computer Science and Engineering
BRAC University

***Abstract*—Precise prediction of drug–target binding affinity is a critical task in computational drug discovery, enabling early-stage screening of candidate compounds against target proteins. In this project, we evaluate deep learning approaches for Drug–Target Interaction (DTI) prediction using the Davis dataset. Two distinct frameworks, *DeepPurpose* and *DeepChem*, are employed to train regression models for estimating the log-transformed dissociation constant (pK_d).**

***DeepPurpose* utilizes convolutional neural networks (CNNs) to encode both SMILES strings and protein sequences, while *DeepChem* relies on Morgan fingerprints for drug representation and a feedforward neural network for affinity prediction. Our results demonstrate that *DeepPurpose* achieves superior performance, with a mean squared error (MSE) of 0.4120 and R^2 of 0.4680, compared to *DeepChem* (MSE $\approx$ 0.7336). To enhance robustness, we propose an ensemble model that averages predictions from both frameworks, yielding improved accuracy. This approach highlights the benefits of integrating complementary molecular representations.**

Overall, this project underscores the potential of multimodal deep learning and model fusion to advance the reliability of DTI predictions in bioinformatics. Future directions may explore hybrid architectures and larger datasets to further refine predictive performance.

***Index Terms*—Drug-Target Interaction (DTI) prediction, deep learning, computational drug discovery, binding affinity prediction, DeepPurpose, DeepChem, ensemble learning, convolutional neural networks (CNN), Extended-Connectivity Fingerprints (ECFP), Davis dataset, cheminformatics, bioinformatics**

I. Introduction

Drug-target interaction (DTI) prediction is one of the most crucial tasks in computational drug discovery, enabling the determination and quantification of binding affinity between small molecules (drugs) and biological macromolecules (typically proteins). Traditional laboratory methods for measuring binding affinities are both time-consuming and expensive. Consequently, machine learning techniques have emerged as scalable and efficient alternatives for early-stage drug screening.

In recent years, deep learning has become a powerful tool for modeling complex biochemical interactions. By training on large drug-protein datasets, these models can learn nonlinear mappings from chemical structures to binding affinities, with potential generalization to novel compounds and targets. However, the performance of such models heavily depends on how drugs and proteins are encoded as features, as well as the neural architecture used to interpret these representations.

This project focuses on two prominent frameworks for DTI prediction: *DeepPurpose* and *DeepChem*. *DeepPurpose* is a versatile deep learning toolkit that supports multiple strategies for encoding drugs and proteins. In our implementation, we employ CNN-based encoders for both drug SMILES strings and protein sequences. In contrast, *DeepChem* utilizes Extended-Connectivity Fingerprints (ECFP) to represent drug molecules and a feed-forward neural network (`MultitaskRegressor`) for predictions, without incorporating protein features.

We evaluate both models on the Davis dataset, a benchmark DTI dataset containing kinase inhibitors and their binding affinities. To further enhance predictive accuracy, we implement a simple ensemble strategy that averages the outputs of *DeepPurpose* and *DeepChem*. This integration aims to leverage the complementary strengths of their molecular representations and generalization capabilities.

Our results demonstrate the advantages of multimodal feature encoding and ensemble learning in improving the reliability of DTI prediction systems. These advancements contribute to more accurate and efficient virtual screening methods in bioinformatics and cheminformatics, paving the way for accelerated drug discovery pipelines.

II. Literature Review

DeepPurpose, proposed by Huang *et al.* [1], represents a significant advancement in deep learning frameworks for drug–target interaction (DTI) prediction. The development of DeepPurpose addresses two critical challenges: the growing complexity of biochemical data and the need for accessible, high-performance models in computational drug discovery. Traditional DTI prediction tools often require specialized expertise, creating barriers for non-specialists. DeepPurpose bridges this gap through a user-friendly Python library that enables training, evaluation, and deployment of deep learning models for DTI tasks with minimal configuration.

The framework supports over 50 neural architectures and includes 15 compound and protein encoders, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and graph neural networks (GNNs). It processes compound SMILES strings and protein sequences, transforming them into learned embeddings that are decoded through multilayer perceptrons to predict interaction scores. DeepPurpose accommodates both regression and classification tasks, with additional capabilities for drug repurposing and virtual screening.

Empirical evaluations demonstrate that DeepPurpose achieves state-of-the-art performance on benchmark datasets such as DAVIS and KIBA. The library's modular design accelerates prototyping and enhances accessibility for interdisciplinary researchers. By lowering the barrier to entry for deep learning in bioinformatics, DeepPurpose promotes scalable and reproducible DTI modeling, ultimately accelerating drug discovery and democratizing access to advanced computational tools.

Öztürk *et al.* [2] developed DeepDTA, a deep learning model for drug–target binding affinity prediction that addresses limitations of traditional binary classification methods in quantifying interaction strength. Unlike existing approaches relying on 3D protein–ligand complexes or 2D molecular representations, DeepDTA processes raw drug SMILES strings and protein sequences through convolutional neural networks (CNNs) to predict binding affinities.

The architecture employs two parallel CNN blocks to extract high-level features from both molecular and protein sequences, which are then combined through fully connected layers for affinity prediction. Benchmark evaluations on Davis and KIBA datasets demonstrated DeepDTA's superiority over traditional methods like KronRLS and SimBoost. Notably, it achieved a Concordance Index (CI) of 0.863 on KIBA, significantly outperforming KronRLS (CI: 0.782) and SimBoost (CI: 0.836) with statistical significance ($p < 0.0001$).

These results suggest CNNs can effectively capture interaction patterns from sequence data alone. However, the authors acknowledge that CNN architectures may not optimally model sequential dependencies in proteins, proposing Long Short-Term Memory (LSTM) networks as potential improvements. Future directions include integrating ligand-based protein representations and exploring alternative deep learning architectures to enhance prediction accuracy.

Zhao *et al.* [3] proposed MSI-DTI, a novel framework for drug–target interaction (DTI) prediction that integrates multi-source information, including sequence-based biological features and network-based features from bioinformatics knowledge graphs. The model employs a comprehensive drug representation module combining five distinct molecular descriptors: MACCS keys, Morgan fingerprints, Avalon fingerprints, Mol2Vec, and Graph2Vec. This multimodal integration captures chemical substructures, molecular topology connectivity patterns, and comprehensive molecular representations.

The architecture constructs a Drug–Target Knowledge Graph (DTKG) and utilizes multi-head self-attention with residual connections to capture both low- and high-order interaction patterns. For biological context, the framework incorporates BioBERT, a pre-trained language model, to extract entities and relationships. The protein representation layer implements two graph-based embedding approaches:

- AttentionWalk algorithm for random walk-based node embedding
- Composition-based multi-relation graph convolutional networks (CompGCN) for edge relationship analysis

Following feature extraction, principal component analysis (PCA) reduces dimensionality while preserving critical information. The multi-head self-attention mechanism enables detection of complex feature relationships, with separate attention heads focusing on distinct feature subspaces. The model combines head outputs through residual connections that maintain primary features while integrating learned patterns.

Benchmark evaluations demonstrate MSI-DTI's superior performance in capturing heterogeneous biological information and interaction patterns compared to single-modality approaches.

III. Methodology

In this project, two separate deep learning frameworks; *DeepPurpose* and *DeepChem*, are used for predicting affinity between drug molecules and protein targets. The Davis dataset is used as a ground for both training and evaluation. Apart from that, ensemble models are developed using the averaged output of both frameworks to estimate the potential beneficial effect on performance of model fusion.

The data set used is the Davis benchmark that consists of binding affinities between 72 kinase inhibitors and 442 protein targets, therefore 30,056 unique drug-target pairs. Each entry has a SMILES string defining a small molecule, a protein amino acid sequence, and log-transformed dissociation constant (pK_d). The CSV file is loaded and parsed into three lists: drugs (SMILES), targets (Sequence), affinity values (Affinity).

The *DeepPurpose* pipeline starts with preprocessing the input using its `data_process()` function. A Convolutional Neural Network (CNN) is used in encoding drugs and proteins for both components. The dataset is randomly partitioned into 80% (training), 10% (validation) and 10% (test) sets. Model hyperparameters are programmed using `generate_config()`, which indicates the batch size, learning rate, and number of epochs as 256, $1e^{-4}$ and 50 respectively. The model is then initialized and trained using a `net.train()` function that has the training, validation and test data as inputs. Post training, the performance of the model is evaluated using mean squared error (MSE) and coefficient of determination (R^2). These are calculated by comparing the true and the predicted affinities on the test set.

At the same time, a *DeepChem* model is built based on the drug features, disregarding the protein sequence data. SMILES strings are transformed to molecular fingerprints using RDKit's Morgan fingerprinting algorithm having a radius of 2 and 1024 bit length. Splitting obtained feature vectors are carried out using standard 80:20 train–test ratio through scikit-learn's `train_test_split()` function. *DeepChem*'s `MultitaskRegressor` is used to implement a deep feedforward neural network which consists of two hidden layers and is equipped with 1024 and 512 neurons respectively. The model is taught on a training set for 200 epochs and predictions are made for the test set.

As two distinct splitting mechanisms are used by *DeepPurpose* and *DeepChem*, it was essential to re-align their test set so as to be compatible for ensemble evaluation. This was done by featuring *DeepPurpose*'s test set SMILES strings using the Morgan fingerprinting technique the same way, with appropriate test data generation for *DeepChem*. Then both models made predictions for this common input. The ensemble prediction was obtained by averaging the outputs of both the models for each of the test samples.

This can be mathematically defined as:

$$\hat{y}_{\text{ensemble}} = \frac{1}{2}\left(\hat{y}_{\text{DP}} + \hat{y}_{\text{DC}}\right)$$

where $\hat{y}_{\text{DP}}$ denotes the prediction from *DeepPurpose* and $\hat{y}_{\text{DC}}$ denotes the prediction from *DeepChem* on the aligned test set.

The ensemble model was assessed with the same metrics which were used for the individual models– MSE and R^2—on the ground-truth affinities of the common test set. This approach was meant to create complementary strengths of both models and better robustness of prediction.

IV. Result

In order to assess the performance of the *DeepPurpose*, *DeepChem*, and ensemble models, their MSE and R^2 were computed on the test set. The *DeepPurpose* model performed best with the MSE and R^2 scores of 0.4120 and 0.4680, respectively, thanks to the dual CNN encoders that were able to capture drug–protein interactions well.

By comparison, the MSE from the *DeepChem* model—which only used drug fingerprints and did not use protein data—was 0.7336. This reflects the minimized capacity to predict the interaction-specific patterns due to the lack of the target input.

To enhance robustness, an ensemble model was created by combining the predictions from *DeepPurpose* and *DeepChem*. The ensemble had an MSE of 0.4723 and R^2 of 0.3903. While slightly less accurate than *DeepPurpose* alone, the ensemble showed more balanced predictions and greater stability across the test set.

A summary of the results is presented in Table I.

TABLE I
Model Evaluation Metrics

Model	**MSE**	R^2
DeepPurpose	0.4120	0.4680
DeepChem	0.7336	—
Ensemble	0.4723	0.3903

V. Discussion

This project evidences the power of deep learning models in predicting drug-target binding affinity; significant variations are noted for the two frameworks applied. The *DeepPurpose* model demonstrated better performance than *DeepChem*, with an MSE value of **0.4120** and an R^2 score of **0.4680**. This could primarily be attributed to its adoption of CNN-based encoders, which can learn sequence-specific interactions that play a key role in the accurate prediction of affinities.

However, *DeepChem*, which used only Morgan fingerprints of drug molecules and omitted protein data, had a higher MSE of **0.7336**. The lack of target information limited its ability to capture interaction-specific patterns, resulting in less precise predictions. Nevertheless, its performance underscores the utility of generic drug-centric representations in predictive tasks.

An ensemble model was constructed by averaging the predictions of *DeepPurpose* and *DeepChem* to enhance robustness. While this combined approach slightly reduced accuracy (MSE: **0.4723**, R^2: **0.3903**), it demonstrated that integrating models with complementary strengths could yield more stable and balanced predictions.

Overall, the results highlight the importance of incorporating features intrinsic to both drugs and proteins for detailed drug-target interaction (DTI) modeling. While drug-only approaches like *DeepChem* serve as a viable starting point, multimodal frameworks like *DeepPurpose* enable finer-grained affinity predictions. Future work could explore alternative encoders, larger datasets, or transformer-based architectures to further improve performance.

References

[1] K. Huang, T. Fu, L. M. Glass, M. Zitnik, C. Xiao, and J. Sun, "DeepPurpose: A Deep Learning Library for Drug-Target Interaction Prediction," *Bioinformatics*, 2020.

[2] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: deep drug–target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, Sep. 2018, doi: 10.1093/bioinformatics/bty593.

[3] W. Zhao, Y. Yu, G. Liu, Y. Liang, D. Xu, X. Feng, and R. Guan, "MSI-DTI: predicting drug–target interaction based on multi-source information and multi-head self-attention," *Briefings in Bioinformatics*, vol. 25, no. 3, p. bbae238, May 2024, doi: 10.1093/bib/bbae238.