OXFORD BROOKES UNIVERSITY

Explainable Deep Learning for Drug-Target Interaction Prediction Using Protein Structures

Done by: Kyrie Liu

Chengdu University of Technology, CDUT Sino-British Collaborative Education Supervised By Dr. Grace Ugochi Nneji

Abstract

Drug–target interaction (DTI) prediction is a vital component of the drug discovery pipeline, yet traditional similarity-based and laboratory methods face challenges in handling high-dimensional features, nonlinear relationships, and data imbalance. To overcome these limitations, a new hybrid deep learning model is introduced, which combines FCS mining module, Transformer-based encoders, and CNNs technologies. It achieved a test AUC of 0.8779, AUPRC of 0.8851, and an F1 score of 0.8046.

Dataset & Data Process

• About dataset: In this project, three datasets are used BindingDB, BIOSNAP, and DAVIS. Each dataset consists of drug-protein pairs labeled as either interacting or non-interacting. The drug molecules are represented by their SMILES strings, and the target proteins are represented by their amino acid sequences.

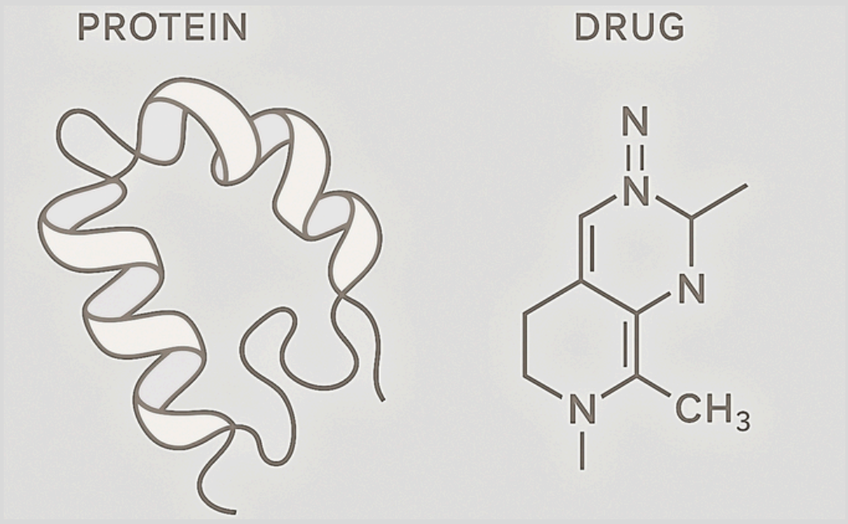

Figure1: stucture of protrin and drug

Implement & Result

Proposed modle architecture is shown by figure 2

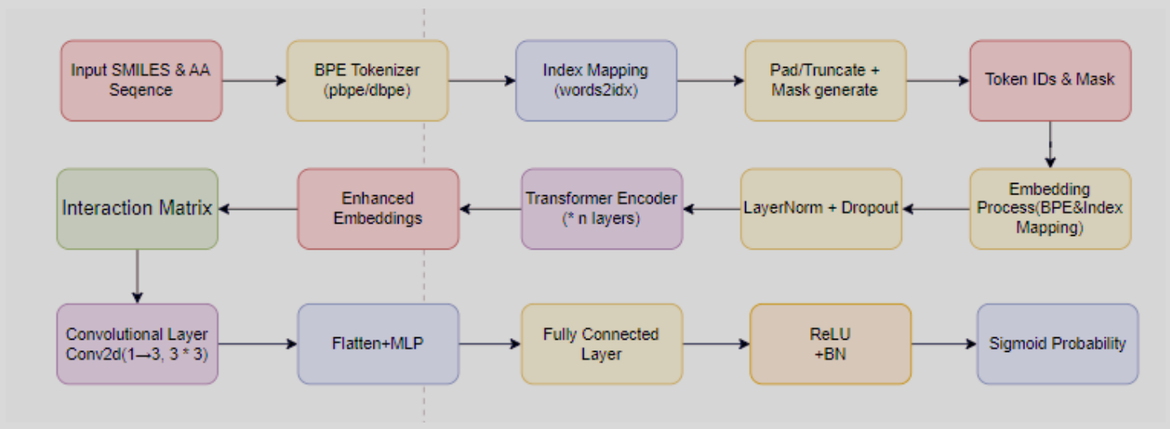

Figure 2: Model architecture of DTIprediction Model

The performance of the model is shown in the following table.

	AUC	AUPRC	F1	Loss
Training	0.9755	0.9757	0.9258	0.1864
Test	0.8853	0.8881	0.8287	0.4683
Validation	0.8958	0.8998	0.8223	0.4374

Table 1: Performance of the modle

GUI Design

The GUI interface provides two main function for users:

1. Predicting the possibility of interaction for a specific pair of drug and protein information.
2. Predicting the overall prediction results of the complete dataset.

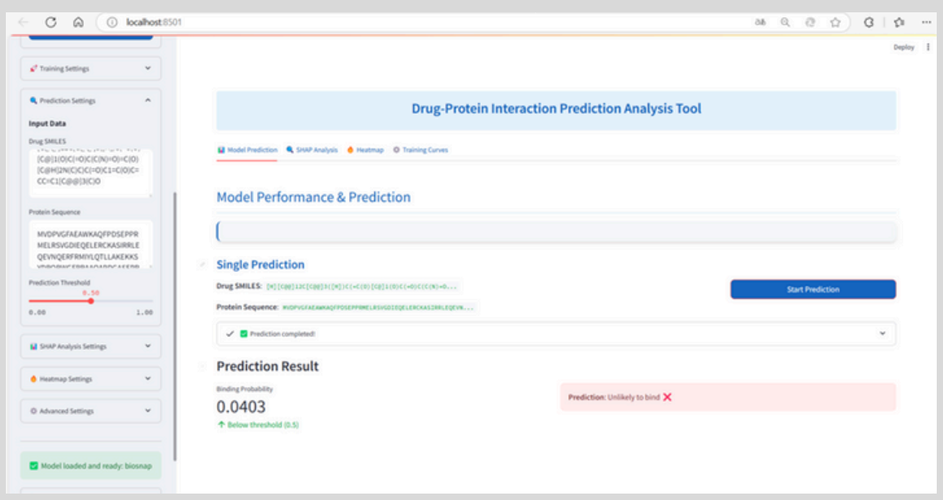

Figure 3: Single prediction

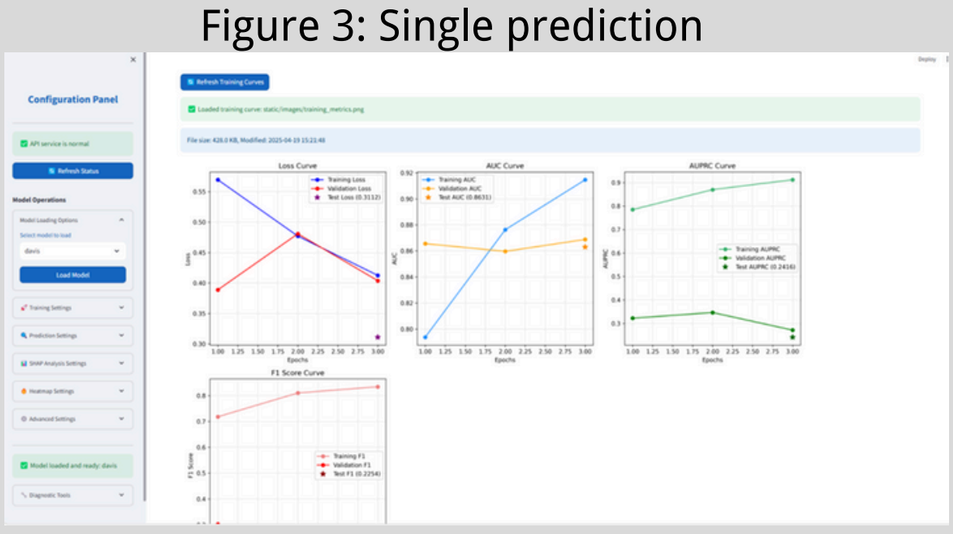

Figure 4: Overall prediction

Conclusion

- The model in this work achieved good performance in AUC, AUPRC, and F1 scores, indicating higher prediction capability in drug-protein interaction prediction.
- A few limitations of the existing model remain, for example loss of important contextual information and data imbalance.
- For future work, this project will focus on enhancing the model’s ability to generalize across different datasets and improving its computational efficiency.