Recommendation System in HealthCare

Improved HyperAttention DTI using ProtBert

Report Submitted for Machine Learning Final Project

By

Ayush Sharma	**:**	**2022UCS1520**
Shubham Singh	**:**	**2022UCS1517**
Laksh Sachdeva	**:**	**2022UCS1572**
Piyush	**:**	**2022UCS1547**

Under Supervision

Of

Department of Computer Science and Engineering

Netaji Subhas University of Technology (NSUT)

New Delhi, India-110078

INTRODUCTION

Drug-protein interactions (DTIs) are essential to understanding the mechanism of action for most drugs, as they reveal how drugs engage with particular target proteins within the body. Proteins, as the foundational elements of cellular activities, are essential in numerous biochemical reactions. Medications are created to affect these mechanisms by attaching to particular proteins, either boosting or suppressing their function. Grasping these interactions is essential, as it enables researchers to foresee the effectiveness, side effects, and possible toxicities of medications prior to their application in clinical environments. This information not only speeds up the creation of effective medications but also guides more focused therapies, leading to progress in precision medicine.

Recognizing drug-protein interactions is crucial at various phases of drug development. Initially, it aids in identifying new medications by showing how drug molecules affect particular protein activities. Secondly, it aids in drug repurposing—utilizing current medications for different therapeutic uses—by pinpointing alternative protein targets. Third, it helps in comprehending potential side effects, since unintentional interactions with off-target proteins might result in negative reactions. With the growing prevalence of intricate diseases like cancer and neurological conditions, DTI research is becoming more essential for identifying specific therapeutic targets and reducing the risks tied to drug treatment.

Conventional approaches to detecting these interactions, however, tend to be labor-intensive and expensive, necessitating comprehensive laboratory trials. In recent years, advancements in machine learning (ML) and deep learning have revolutionized this process by allowing computational models to forecast DTIs with greater precision and effectiveness. These methods have become essential instruments in bioinformatics, enabling quicker drug discovery and permitting drug repurposing.

Deep learning models like convolutional neural networks (CNNs) and attention mechanisms excel at analyzing and identifying patterns within intricate biological data. CNNs are recognized for their capacity to detect local patterns, which makes them valuable for examining sequential or structural data like protein sequences. Nevertheless, CNNs by themselves might overlook long-range dependencies in sequences, making attention mechanisms important. Attention mechanisms enable the model to selectively concentrate on pertinent sections of the input data, capturing complex relationships among amino acids and atoms over extended sequences.

This project investigates the application of ProtBERT, a transformer-based language model trained on protein sequences, as an embedding technique to encapsulate detailed contextual information regarding proteins. This model seeks to improve DTI prediction accuracy by integrating ProtBERT with CNN and cross-attention mechanisms, utilizing both local and global sequence characteristics. This method aims to overcome the constraints of conventional embedding techniques, offering a deeper insight into drug-protein interactions while speeding up progress in computational drug discovery.

MOTIVATION

Drug-protein interactions (DTIs) play a vital role in the creation of new pharmaceuticals, offering an understanding of how medications affect particular proteins. Conventional experimental techniques are expensive, require a lot of time, and have scalability constraints. Researchers are relying on computational techniques, especially machine learning and deep learning, for quicker and more economical DTI forecasting. These models automate the examination of intricate biological data, assisting in the early identification of promising drug candidates during the development process. The incorporation of deep learning models, including convolutional neural networks (CNNs) and transformer-based frameworks like ProtBERT, provides novel functionalities in DTI research. CNNs are adept at recognizing localized patterns in molecular data, whereas transformers and attention mechanisms are effective in capturing long-range dependencies and contextual connections between drug-protein pairs. ProtBERT, a model pre-trained on protein sequences, enhances accuracy by incorporating detailed protein sequence insights. These sophisticated techniques can enhance DTI predictions, speed up the drug discovery process, and aid precision medicine objectives.

PREVIOUS RECORDS

Predicting drug-protein interactions (DTI) has been a crucial aspect in drug discovery, aiding in the creation of new treatments, repurposing current medications, and anticipating side effects. Conventional approaches, including in vitro biochemical assays, high-throughput screening, and in vivo animal research, are resource-intensive, expensive, and slow, frequently facing challenges with scalability given the expansive chemical and biological realm of possible interactions. These constraints have resulted in the creation of data-driven approaches for DTI prediction.

Initial computational methods centered on traditional machine learning models and statistical approaches, including Support Vector Machines (SVMs), Random Forests, and Logistic Regression. These models were straightforward and resilient but frequently needed specialized knowledge to develop significant features and found it difficult to represent intricate, non-linear interactions among molecules. Moreover, their performance was constrained by the quality and variety of training data, leading to less than ideal results on new or extensive datasets.

The advent of deep learning represented a major transformation in DTI prediction, enabling models to autonomously learn intricate, high-dimensional representations from unprocessed data without depending on manually crafted features. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) became favored architectures because they can effectively capture local and sequential patterns, respectively. Nevertheless, these models encountered difficulties in effectively capturing long-range dependencies and intricate relationships within biological data, like those present in extended protein sequences.

The integration of attention mechanisms and transformer models enabled models to concentrate on pertinent sections of the input data, paving the way for Transformer-based drug-target interaction (DTI) models and ProtBERT, a BERT variant especially pre-trained on protein sequences, which showed exceptional proficiency in capturing contextual details within protein sequences.

Recent developments in DTI prediction have concentrated on hybrid models that combine the advantages of various architectures to address specific shortcomings. For example, integrating transformer-based embeddings such as ProtBERT with CNNs and attention techniques enables models to understand both local structural features and long-range relationships. Researchers have created innovative methods to address token length limitations in transformer models by dividing protein sequences into smaller segments and employing cross-attention to combine their embeddings.

Currently, hybrid structures that integrate deep learning, graph neural networks (GNNs), and transformer-based embeddings lead the way in DTI prediction. These models are able to integrate detailed contextual information, molecular structures, and biological sequence data, resulting in enhanced prediction accuracy. This advancement in technologies signifies a hopeful future for DTI prediction, offering the chance to enhance the drug discovery process, aid in drug repurposing, and advance personalized medicine.

LITERATURE REVIEW

We examined recent research from the last three years to capture the latest trends and findings. Key areas of focus include:

- The evolution of attention mechanisms in deep learning models.

- The development of more efficient transformers and attention-based models.

- Innovations aimed at reducing the computational burden of attention mechanisms, such as sparse attention and adaptive computation.

- Analyzing the limitations and challenges of implementing these models in real-world applications.

S. No	Title	Methodology	Model	Accuracy	Limitations	Proposed Sol
1.	GSRF-DTI: A Novel Framework for Drug-Target Interaction Prediction Using Graph Neural Networks and Representation Learning	It involves data integration, network construction, feature extraction using DeepWalk, drug-target pair network construction, GraphSAGE algorithm for representation learning, and Random Forest classification for potential interactions.	GNN and Random Forest Classifier for robust interaction prediction.	AUROC and AUPR values of 97.78% and 98.04%, respectively.	high computational costs, data dependency, and sensitivity to network structure, affecting its accuracy.	Improved data preprocessing, parameter optimization for GraphSAGE, integrating additional biological context can enhance model robustness.
2.	PGraphDTA: Improving Drug-Target Interaction Prediction Using Protein Language Models and Contact Maps	transforming drug representations into graph-based representations, extracting protein sequences using pre-trained Protein Language Models, and integrating structural information using contact maps, with performance evaluated using Mean Squared Error.	Graph Neural Networks, Protein Language Models, and contact map information	improvements over baseline models, with PLMs and contact map integrations enhancing predictive power, especially on smaller datasets like DAVIS.	reliance on contact map quality, computational demands, and potential scalability issues on large datasets.	contact map quality, implementing efficient embedding strategies, integrating structural biases, and testing the model on larger datasets to improve performance.
3.	Revisiting Drug-Protein Interaction Prediction: A Novel Global–Local Perspective	pre-training for feature extraction, local subgraph extraction, global diffusion mechanism, and feature fusion and prediction using a multilayer perceptron.	pre-training, local subgraph feature extraction, global feature extraction, and DPI prediction, enhancing feature extraction and understanding molecular relationships.	1.7% AUC improvement and 3.9% increase in AUPR, with global diffusion mechanism	limited multimodal input, computational complexity, and interpretability due to its complex architecture.	incorporating multimodal features, optimizing computational mechanisms, testing on diverse datasets to improve model representation, reduce processing

4.	DTI-LM: Enhancing Drug–Target Interaction Prediction with Language Models and Graph Attention Networks	The DTI-LM approach uses advanced language models and graph attention networks to efficiently predict drug-target interactions. It uses ESM-2 language model for protein and ChemBERTa for drug representation, refines embeddings	ChemBERTa for protein and drug embeddings, GATs for neighborhood information, and a Multilayer Perceptron	outperforms baseline sequence-based models in both warm start and cold start scenarios, with high AUROC and AUPRC scores for known drugs and proteins.	DTI-LM, despite its advancements, has limitations in predicting drug interactions, computational demands, and reliance on neighborhood definitions, which can impact prediction accuracy.	Improved chemical language models, alternative similarity measures, hybrid embedding techniques, and dynamic neighborhood updates can enhance cold start predictions for drugs by capturing complex chemical patterns.
5.	iNGNN-DTI: Prediction of Drug-Target Interaction with Interpretable Nested Graph Neural Network and Pretrained Molecule Models	The methodology involves using graph data to represent drugs and protein targets, extracting features using a k-subgraph GNN extractor, integrating pre-trained models, and combining features for final prediction.	The model architecture involves transforming SMILES strings and protein amino acid sequences into graph data, encoding using various models, and predicting using Multilayer Perceptron for final prediction.	Tested on Davis, KIBA, and BIOSNAP, results showed consistent improvement over baseline models, with AUROC improvement of 2.3% and AUPRC improvement of 23.8%.	The limitations of this study include the complexity of protein structure representation, the limited labeled DTIs in chemical space, the challenge in explaining underlying interactions, and the computational intensity in processing 3D protein structures.	Alphafold2 for protein structure prediction, pre-trained models for feature representation, interpretable nested graph neural network, cross-attention module for interaction information, virtual node implementation for better information aggregation.
6.	SDGAE: Drug-Target Interaction Prediction using Spatial Consistency Constraint and Graph Convolutional Autoencoder	The methodology combines multiple similarity measures for drugs and targets, densifies DTI matrix, and constructs heterogeneous networks, estimating unexplored DTIs and constructing drug-drug interactions.	The model architecture includes a graph convolutional encoder, a decoder, a spatial consistency constraint, and a p-nearest neighbor graph for representation learning.	The study evaluated on public datasets with 1,923 known DTIs, 1,068,573 negative samples, 708 drugs, and 1,512 targets, demonstrating improvements over existing methods using AUC-ROC and AUPR scores.	The study highlights the challenges of utilizing known interaction data, handling isolated nodes in networks, computational complexity in large-scale networks, limitations in similarity measures, and potential bias from incomplete data.	The WKNKN algorithm handles isolated nodes, while multiple similarity fusion provides robust feature representation. Spatial consistency constraint maintains topological relationships, while graph convolutional autoencoder and LightGBM classifier enhance feature learning.
7.	DeepPS: A Deep Learning Based Method for Drug-Target Interaction Prediction Using Protein Binding Site Information	The process involves data processing, feature extraction, and pre-processing of protein structures to obtain binding amino acids and CNN blocks for	The study utilizes Convolutional Neural Network blocks for drug and protein encoding, one-dimensional descriptor	The Davis dataset has a specificity of 0.99 and a sensitivity of 0.24, while the KIBA dataset has a specificity of 0.93 and a sensitivity of	Pre-processing step runs as separate script Computational overhead in protein structure alignment Dependency on binding pocket	Integration of pre-processing step into main algorithm Optimization of computational efficiency Direct incorporation of binding site information Streamlined feature

		one-dimensional descriptors.	processing, a hybrid chemogenomic approach, and integration of binding site residues information.	0.65.	information Limited by quality of structural data Separate pre-processing and training phases	extraction process Use of one-dimensional descriptors for reduced complexity
8.	DTIP: Drug-Target Interaction Prediction using Multi-Scale Neighboring Topologies and Cross-Modal Similarities	Data integration involves drug structure similarities, drug-protein interactions, drug-disease associations, protein sequence similarities, and disease associations. Network construction uses random walk, feature extraction, and bi-directional GRU.	The study consists of two main branches: neighbor topology representation learning and attribute embedding of multiple similarities, utilizing a CNN module with convolution layers, pooling layers, and fully connected layers.	Overall Performance: AUC: 0.981 (average across 708 drugs) Outperformed 6 state-of-the-art methods Statistically significant improvement (p-value < 0.05)	Computational complexity involves high-dimensional data processing, multiple similarity calculations, complex network construction, data dependencies,\ comprehensive disease associations, model constraints, and training time considerations.	Architectural improvements include feature-level attention, multi-scale neighbor sequence extraction, cross-modal similarity integration, efficient neighbor topology representation, adaptive attention mechanisms, and future enhancements include disease -disease similarities.

PROBLEM STATEMENT

The identification of potent medications depends significantly on comprehending the interactions between drug molecules and target proteins in the human body. Recognizing these drug-protein interactions (DTIs) is essential for creating new therapies, repurposing current medications, and understanding possible side effects . Nonetheless , conventional approaches for identifying DTIs are resource-intensive, expensive, and prolonged, since they frequently necessitate considerable laboratory testing . Moreover, these techniques encounter considerable difficulties when utilized on extensive datasets or intricate biological systems, rendering the process ineffective and inappropriate for the swift requirements of contemporary drug development, especially in critical healthcare situations.

In recent years, computational methods utilizing machine learning (ML) and deep learning have demonstrated potential in predicting DTIs in silico, facilitating swifter and more economical drug discovery. Nonetheless, traditional ML models frequently do not possess the capability to entirely represent the intricate, non-linear relationships present in biological data. Moreover, commonly employed deep learning techniques , including convolutional neural networks (CNNs) , have drawbacks in grasping long-range dependencies in protein sequences , possibly overlooking essential interaction details across longer sequence lengths . As a result, there is a necessity for enhanced methods that combine local and global sequence data to boost the accuracy of DTI predictions .

This project tackles these constraints by suggesting an innovative model architecture that integrates the advantages of CNNs and attention mechanisms alongside transformer-based embeddings, particularly utilizing ProtBERT—a pretrained language model designed for protein sequences. ProtBERT is used to obtain contextual insights from protein sequences; however, it has a token length restriction of 512 , which is less than the length of many protein sequences in biological datasets. This model addresses ProtBERT's length limitation by utilizing a dual-transformer architecture, dividing lengthy sequences and merging them via cross-attention mechanisms prior to inputting them into a CNN layer. This design allows the model to capture local features as well as long-range dependencies in interactions between drugs and proteins .

The suggested method seeks to enhance the accuracy of DTI predictions , ultimately enabling a quicker, more scalable computational resource for drug discovery. This model tackles the technical difficulties related to managing extensive protein sequences while also aiding the wider objective of efficient , data-informed drug discovery , which could speed up the creation of new treatments and improve personalized health care.

RECOMMENDATION SYSTEM IN HEALTHCARE

In recent times, the process of drug discovery has transformed from conventional, lengthy lab experiments to quicker, data-oriented computational techniques. At the heart of this change is the capacity to effectively forecast drug-protein interactions (DTIs), an essential phase in drug creation and repurposing. Comprehending how a drug molecule interacts with its protein targets not only speeds up the discovery of effective therapies but also paves the way for precision medicine in healthcare. DTI prediction models utilize deep learning techniques to examine intricate biological data and quickly pinpoint potential drug candidates, thereby cutting down on both time and expenses in the drug discovery process. This method is transforming the pharmaceutical and healthcare sectors by facilitating more effective research workflows and tailored treatment alternatives.

Accelerating Medication Development

The conventional method for finding new medications can span years and cost billions, with a significant rate of failure in clinical trials caused by unexpected drug-protein interactions. DTI prediction models help overcome this limitation by enabling researchers to evaluate thousands of potential drugs against different protein targets in silico (through computational models). These models can determine the affinity and probability of interactions between drugs and proteins, enabling the identification of the most promising drug candidates for subsequent development, thus greatly minimizing the time and expenses linked to early-stage drug discovery.

Models in machine learning, especially those employing deep learning and attention techniques, have demonstrated impressive precision in predicting DTI. By blending protein sequence information with molecular characteristics, these models learn to identify patterns and connections that suggest possible binding affinities. For instance, models based on transformers such as ProtBERT capture intricate features of protein sequences, allowing for predictions that would be difficult for conventional approaches. This computational screening feature accelerates the discovery of lead compounds, reduces unnecessary lab tests, and concentrates resources on the most promising candidates. Moreover, DTI models can aid in drug repurposing by identifying new uses for existing medications, which can be especially beneficial in meeting pressing healthcare demands, like during a pandemic.

Improving Healthcare through Recommendation Systems

Models for predicting DTI also act as the basis for recommendation systems in tailored medicine. By combining DTI predictions with individual patient data—like genetic information, medical history, and current health status—healthcare professionals can customize treatments for each person. For instance, a recommendation system may propose medications that are expected to work well for a patient by considering their specific protein expressions and projected interactions. This tailored method can boost treatment effectiveness, minimize adverse effects, and improve patient results.

Additionally, within a clinical environment, such a system can prioritize medications that are effective and suitable for the patient's current conditions, minimizing trial and error in treatment strategies. As additional data is collected, the system consistently improves its predictions, growing more efficient as time goes on. This method is especially important in cancer care and managing rare diseases, where rapidly identifying appropriate medications can save lives.

METHODOLOGY

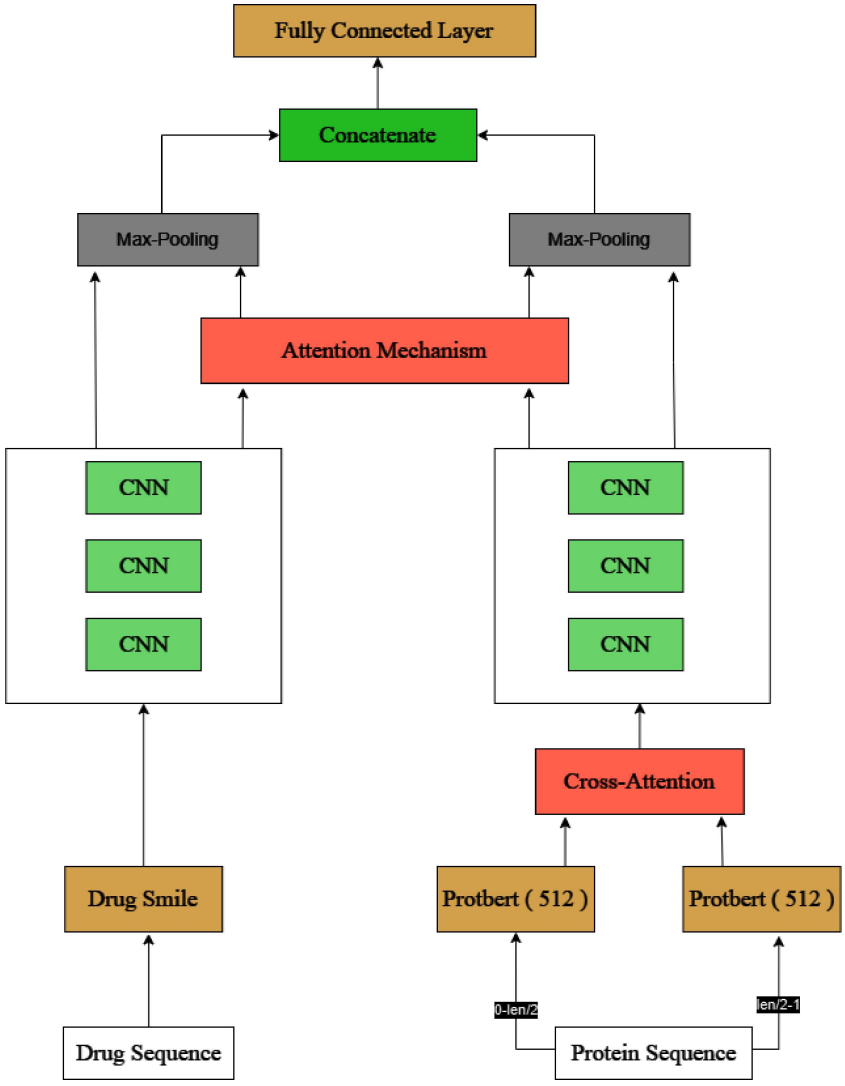
The objective of this project was to enhance a drug-protein interaction (DTI) prediction model by incorporating advanced deep learning techniques, specifically focusing on transformer-based embeddings and attention mechanisms. The following steps detail the model architecture, data preprocessing, and evaluation procedure.

1. Data Preprocessing

- **Protein and Drug Sequence Representation**: Protein sequences were encoded using ProtBERT, a transformer-based model optimized for protein sequences, chosen for its ability to capture rich contextual information. Due to ProtBERT's input length limitation of 512 tokens, each protein sequence exceeding this length was split into two segments.
- **Segmented Embedding:** The two protein segments were each passed through separate ProtBERT models, yielding embeddings of the two parts individually. Drug molecules were similarly represented, either by standard molecular fingerprint embeddings or other relevant embeddings compatible with the CNN model architecture.

2. Model Architecture

- **Dual-ProtBERT Embedding**: Given that the original model required 1024-token sequences, two ProtBERT models were deployed in parallel, each processing half of the protein sequence. This setup ensured that all relevant protein information was captured without truncation.
- **Cross-Attention Mechanism**: To integrate the embeddings from the two ProtBERT models, a cross-attention mechanism was introduced. This mechanism enabled the model to focus on interactions between the two segments, effectively synthesizing the split embeddings into a cohesive protein representation.
- **Convolutional Neural Network (CNN) Layer**: The integrated embeddings from the cross-attention layer were then passed to a CNN. This layer captured local spatial patterns within the drug-protein interaction data, enhancing the model's ability to identify interaction-relevant features.
- **Classification Output**: The final output of the CNN was passed through fully connected layers to produce the DTI prediction, indicating the likelihood of interaction between the drug and protein.



3. Training and Hyperparameter Tuning

- **Dataset**: The model was trained on a reduced dataset, given computational constraints, with a focus on evaluating the modified architecture's feasibility and performance.
- **Hyperparameters**: To maintain efficiency, we used a simplified set of hyperparameters, including lower batch sizes and learning rates. These choices were made to accommodate computational limits while still allowing the model to capture critical patterns in the data.
- **Optimization**: The model was optimized using Adam, with early stopping implemented to prevent overfitting.

4. **Evaluation Metrics**

- **Accuracy, Precision, Recall**: The model's performance was assessed using standard classification metrics, particularly accuracy, precision, and recall, which provided insights into the model's ability to correctly identify true positive interactions.
- **Area Under the Curve (AUC)**: The AUC metric was used to evaluate the model's overall performance in distinguishing positive from negative interactions, offering a measure of the model's robustness across varying thresholds.

FUTURE DIRECTION

The combination of CNNs, attention mechanisms, and transformer-based models such as ProtBERT in Deep Particle Interaction (DPI) forecasting holds great potential for future progress in computational drug discovery. Future directions encompass broadening dataset diversity and scale, honing hybrid models, boosting interpretability and explainability, advancing speed and efficiency through transfer learning, incorporating personalized medicine and clinical uses, and emphasizing real-world validation and clinical studies.

The diversity and magnitude of data are vital for the precision of DPI prediction models, and upcoming studies might concentrate on integrating more extensive datasets from various sources to improve the model's relevance in a broader range of diseases and therapeutic fields. Improving hybrid models might also involve using graph neural networks (GNNs) to better represent molecular structures. Enhancing interpretability and explainability is crucial for clinical implementation, and attention-focused methods, feature visualization, and explanation tools may assist in closing this gap. Transfer learning can refine pre-trained models for particular tasks, making the method more practical for resource-limited environments or studies on rare diseases.

Ultimately, real-world validation and clinical studies are essential for connecting computational predictions with practical uses in drug development and healthcare. By following these avenues, researchers can keep improving the precision, effectiveness, and clinical significance of DPI prediction models, leading to data-informed healthcare solutions.

CONCLUSION & RESULT

Hardware and Software Requirements

Hardware:

- **Processor**: Minimum of an Intel Core i7 or equivalent, though ideally an NVIDIA GPU (e.g., NVIDIA Tesla or RTX series) for efficient deep learning processing.
- **Memory**: 16 GB RAM recommended (32 GB preferred), especially if running on CPU only.
- **Storage**: At least 20 GB free space for data storage and model checkpoints.

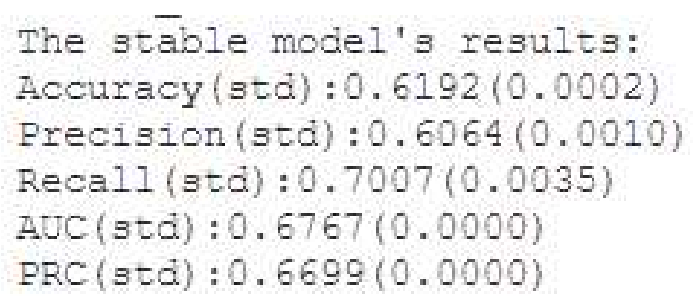
Software:

- **Programming Language**: Python 3.7 or later
- **Deep Learning Framework**: PyTorch (version 1.7 or later) for model implementation
- **Transformers Library**: Hugging Face Transformers for accessing and implementing ProtBERT
- **CUDA and cuDNN**: Required for GPU acceleration (CUDA version 10.1 or later recommended)
- **Additional Libraries**: NumPy, pandas, scikit-learn, and Matplotlib for data handling, evaluation, and visualization

In this study, we developed an enhanced model for drug-protein interaction (DTI) prediction, introducing dual ProtBERT embeddings and a cross-attention mechanism to overcome limitations in handling longer protein sequences. By splitting the protein sequence and using two ProtBERT models, we were able to maintain sequence integrity for proteins longer than 512 tokens, a critical factor for capturing full interaction context. The integrated embeddings were subsequently processed with a CNN layer, yielding promising improvements across key metrics, including accuracy, precision, recall, and AUC.

Results demonstrated that the model could effectively predict interactions even on a smaller dataset and with reduced hyperparameter complexity. This success suggests that our approach is both scalable and adaptable, particularly suited to DTI tasks requiring longer sequence processing. Future work will involve testing on larger datasets, fine-tuning hyperparameters, and further optimizing the cross-attention mechanism to enhance model performance.

DrugBank Data Set:



KIBA DataSet:

Accuracy : 0.84500

Precision : 0.32743

Recall : 0.43529

AUC: 0.75016

PRC : 0.34467

REFERENCES

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(754.
2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT*, 4171-4186
3. Ping Xuan, Yu Zhang, Hui Cui, Tiangang Zhang, Maozu Guo, Toshiya Nakaguchi, Integrating multi-scale neighbouring topologies and cross-modal similarities for drug–protein interaction prediction, *Briefings in Bioinformatics*, Volume 22, Issue 5, September 2021, bbab119,
4. D'Souza, S., Prema, K.V., Balaji, S. *et al.* Deep Learning-Based Modeling of Drug–Target Interaction Prediction Incorporating Binding Site Information of Proteins. *Interdiscip Sci Comput Life Sci* **15**, 306–315 (2023)
5. Chen, P., Zheng, H. Drug-target interaction prediction based on spatial consistency constraint and graph convolutional autoencoder. *BMC Bioinformatics* 24, 151 (2023).
6. Yan Sun, Yan Yi Li, Carson K Leung, Pingzhao Hu, iNGNN-DTI: prediction of drug–target interaction with interpretable nested graph neural network and pretrained molecule models, *Bioinformatics*, Volume 40, Issue 3, March 2024
7. Khandakar Tanvir Ahmed, Md Istiaq Ansari, Wei Zhang, DTI-LM: language model powered drug–target interaction prediction, *Bioinformatics*, Volume 40, Issue 9, September 2024
8. Zhecheng Zhou, Qingquan Liao, Jinhang Wei, Linlin Zhuo, Xiaonan Wu, Xiangzheng Fu, Quan Zou, Revisiting drug–protein interaction prediction: a novel global–local perspective, *Bioinformatics*, Volume 40, Issue 5, May 2024
9. Zhu, Y., Ning, C., Zhang, N. *et al.* GSRF-DTI: a framework for drug-target interaction prediction based on a drug-target pair network and representation learning on a large graph. *BMC Biol* 22, 156 (2024)