DRUG TARGET INTERACTION FOR DRUG REPURPOSING USING COMBINED DEEP NETWORK

A PROJECT REPORT

Submitted by

VIJAYKANTH V

(2019246046)

A report for the phase-II of the project

submitted to the Faculty of

INFORMATION AND COMMUNICATION ENGINEERING

in partial fulfillment

for the award of the degree of

MASTER OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING, GUINDY

ANNA UNIVERSITY

CHENNAI 600 025

MAY 2021

BONAFIDE CERTIFICATE

Certified that this project report titled **DRUG TARGET INTERACTION FOR DRUG REPURPOSING USING COMBINED DEEP NETWORK** for the phase - II of the project, is the bonafide work of VIJAYKANTH V (2019246046) who carried out the work under my supervision. Certified further that to the best of my knowledge and belief, the work reported here in does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

Vidhya

PLACE: CHENNAI	**Dr. K.A.VIDHYA**
DATE:	**TEACHING FELLOW**
	PROJECT GUIDE
	DEPARTMENT OF IST, CEG
	ANNA UNIVERSITY
	CHENNAI 600025

COUNTERSIGNED

Dr. SASWATI MUKHERJEE
HEAD OF THE DEPARTMENT
DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY
COLLEGE OF ENGINEERING, GUINDY
ANNA UNIVERSITY
CHENNAI 600025

ABSTRACT

Drug repurposing aims to repivot an existing drug to a new therapy. There are many factors that can be used to predict new target disease i.e., protein-protein interaction, chemical structure, gene expression and functional genomics, Phenotype and side effect, genetic variation and Machine learning. In today's world a lot of diseases are evolving but at the same time it is very difficult to find a drug for that disease immediately so the purpose of link prediction is to find a new drug from an existing drug using advance machine learning concepts. Drug Discovery is a lengthy process, taking on average 12 years for the drugs to reach the market but the best way to discover a new drug is to start with the old one..The main objective of our project is reducing the time delay and increases the accuracy level for drug-disease technique. The importance of this process is finding all the possible drugs from an existing solution using neural network and learning techniques i.e., Deep Neural Network (DNN) and TRANSFORMER can be used to encode both drug and protein on SMILES, Message Passing Neural Network (MPNN) encode drug in its graph representation, CNN_RNN means a GRU/LSTM on top of a CNN on SMILES, Convolution Neural Network (CNN) are used to extract g Drug-Drug Interaction (DDI) from scientific documents and Drug-Protein Interactions.

சுருக்கம்

இன்றைய காலகட்டத்தில் பலவிதமான கொடிய நோய்கள் மனித இனத்தையே அழித்துக் கொண்டிருக்கின்றன இதில் உயிரிழப்புகளுக்கு மிக முக்கிய காரணம் உடனடி மருந்து இல்லாமையே ஆகும். இது போன்ற கொடிய நோய்களுக்கு உடனடி மருந்துகள் ஏற்கனவே இருக்கும் மருந்துகளில் இருந்து புதிய மருந்தை கண்டுபிடிப்பதே இத்திட்டத்தின் முக்கிய நோக்கமாகும். புதிய இலக்கு நோயைக் கணிக்க பல காரணிகள் பயன்படுத்தலாம் அதாவது, புரத-புரத தொடர்பு, வேதியல் அமைப்பு, மரபணு வெளிப்பாடு மற்றும் செயல்பாட்டு மரபியல், பினோடைப் மற்றும் பக்கவிளைவு மரபணு மாறுபாடு மற்றும் இயந்திர கற்றல். மருந்து மற்றும் நோய் குறித்த முன் அறிவை ஒருங்கிணைக்க கூடிய குறைந்த மேட்ரிக்சை தீர்மானிப்பது ஆகும். எங்கள் திட்டத்தில் மருந்துகள் மற்றும் இலக்குகளுக்கிடையேயான தொடர்புகளை கண்டறிதல் என்பது மருந்து மறுபயன்பாட்டிற்கான பயன்பாட்டிற்கான அங்கீகரிக்கப்பட்ட மருந்துகளில் இருந்து புதிய மருந்துகளை உருவாக்குவதற்கான முதல் படியாகும் மற்றும் மருந்து திரையிடல் மற்றும் மருந்து இயக்கிய தொகுப்புக்கான முக்கிய காரணிகளில் ஒன்றாகும். பின்னர் அனைத்து சாத்தியக்கூறுகளை உருவாக்கி இணைப்பு முன்கணிப்பு பயன்படுத்தி மருந்து மற்றும் நோய்க்கும் இடையிலான மருந்து கண்டுபிடிப்பதற்கான சாத்தியமானவை. எங்கள் திட்டத்தின் முக்கிய நோக்கம் நேரத்தை குறைப்பது மற்றும் மருந்து நோய் உற்பத்திக்கான துல்லியம் அளவை அதிகரிக்கிறது. இந்த செயல்முறையின் முக்கியத்துவம் நரம்பியல் நெட்வொர்க் மற்றும் கற்றல் நுட்பங்களை பயன்படுத்தி ஏற்கனவே இருக்கும் தீர்விலிருந்து சாத்தியமான அனைத்து மருந்துகளையும் கண்டுபிடிப்பது.

ACKNOWLEDGEMENT

I thank the lord Almighty, whose showers of blessings have made this project a reality.

I express my deep sense of appreciation and gratitude to my guide, **Dr.K.A.VIDHYA**, Teaching Fellow, Department of Information Science and Technology, for her valuable support, suggestions, guidance and encouragement. And also I thank her for providing me with the necessary counsel and direction to help me complete this project.

I express my sincere thanks to **Dr.Saswati Mukherjee**, Professor and Head, Department of Information Science and Technology for the prompt and limitless help in providing the excellent computing facilities to do the project and to prepare the thesis.

I wish to record my sincere thanks to the members of review panel, **Dr.S.Sridhar**, Professor, **Dr.N.Thangaraj**, Assistant Professor**,** **Dr. S. Abirami**, Assistant Professor,**Dr. M. Deivamani**, Teaching Fellow, **Dr. L. Sai Ramesh**, Teaching Fellow and **Ms. Tina Esther Trueman**, Teaching fellow, Department of Information Science and Technology for their valuable suggestions and critical reviews throughout the course.

(VIJAYKANTH V)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

DTI	Drug Target Interaction
MPNN	Message Passing Neural Network
DNN	Deep Neural Network
SAE	Sparse Auto Encoder
KEGG	Kyoto Encyclopedia of Genes and Genomes
IUPHAR	International Union of Basic and Clinical Pharmacology
CNN	Convolutional Neural Network
PSSM	Position Specific Scoring Matrix
SPCA	Sparse Principal Component Analysis
LM	Legendre Moment

CHAPTER 1

INTRODUCTION

Drug Target Interaction for drug repurposing using combined deep network is a Deep Learning Based Drug Repurposing and Virtual Screening. Drug Target Interaction (DTI) is one of the most essential steps in drug discovery for locating novel drugs the usage of deep mastering methods. The significance of the drug repositioning objectives to identify the brand new drug from an existing accepted drug. Recently plenty of viruses are spreading over the world however there's no on the spot treatment to protect peoples from the lethal risky virus, for this reason there's a want for drug re-purposing research using deep learning model for drug target interaction. Drug Repurposing factors includes protein-protein interaction, gene expression, chemical structure, functional genomics, genetic variation, Phenotype and side effect. Today's Drug improvement is hard work incentive, consumes time, difficult to infer diverse form of goal interplay and drug repositioning is a completely prolonged technique because the general technique takes on common 12 years for the medication to attain the market.

Therefore, the best way to discover a new drug is to characterize the protein molecule bindings then develop a neural network using the usage of deep gaining knowledge of fashions to lessen all of the conventional risky and prolonged technique for drug goal interaction. However, drug interaction prediction remains a considerable challenge so the unique feature of Deep Purposing is to predict the interaction and find all possible drugs from an existing solution using protein sequence. Identification of drug repurposing by interacting with target proteins to activate the process and biological purpose.

Drug target prediction model to overcome the previous models i.e., SAE – Sparse Auto Encoder and DBN – Deep Belief Network (Deep DTA). So that the process usually takes three years using this DTIs prediction model for find out the local residue patterns target proteins. The importance of DeepConv-DTI is capturing local residue styles of proteins and binding sites.

1.1 OVERVIEW ARCHITECTURE OF DRUG TARGET INTERACTION

The combination of interaction data and target data produce the drug-target interaction for predicting the stable molecule biding and the process of the interactions in protein-protein sequence interaction. In biochemistry and molecular biology, binding site page is a area on a macromolecule which include a protein that binds to every other molecule with specificity. Proteins bind to every different via a mixture of hydrophobic bonding, Van der Waals forces, and salt bridges at unique binding domain names on every protein. These domain names may be small binding clefts or massive surfaces and may be only a few peptides lengthy or span loads of amino acids. Figure 1.1 refers to the drug target interaction.

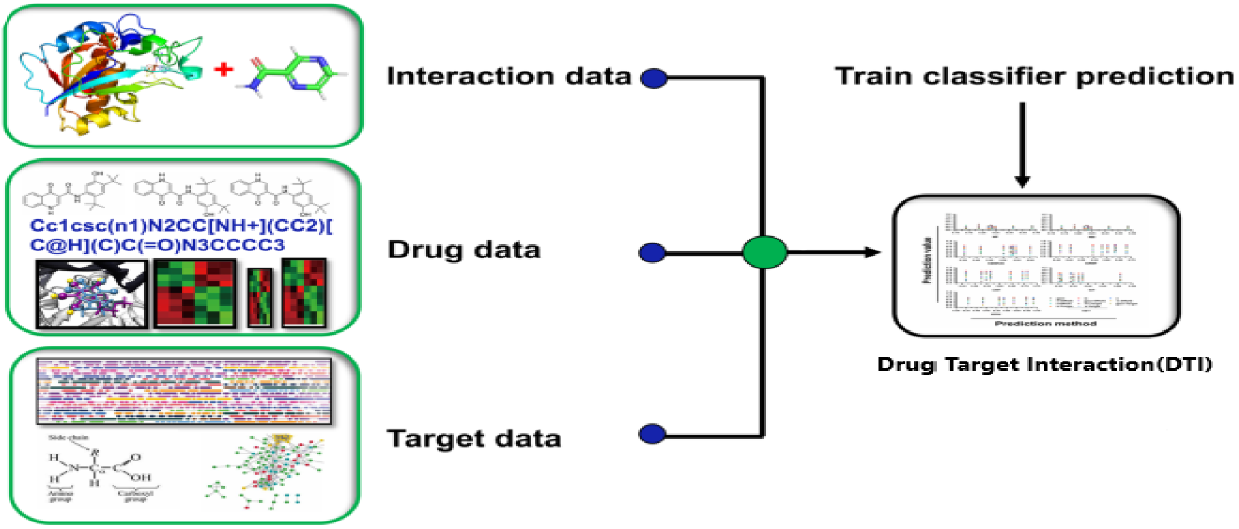

Interaction data
Train classifier prediction
Cc1csc(n1)N2CC[NH+](CC2)[C@H](C)C(=O)N3CCCC3
Drug data
Prediction value
Prediction method
Drug Target Interaction(DTI)
Target data

Figure 1.1: Drug Target Interaction.

1.2 OBJECTIVE

Drug Repurposing is locating a brand new medical use for an permitted drug. Identification of Drug Target Interactions (DTI) is hard to recover lost information using Machine Learning model and the local residue patterns is lost during transformation and the main objective of our project is link prediction for drug repositioning from protein sequences without any loss in better performance compare to previous Deep Learning model with protein independent test dataset. The overall process is collecting large-scale Drug Target Interactions databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG), DrugBank, International Union of Basic and Clinical Pharmacology (IUPHAR) generated data and drug data.

The next step is extraction of the protein sequence from an existing approved dataset and medical valid test dataset. To predict exact novel drug to relevant disease, as of now there is immediate drug for all the disease, the deep leaning model is used to predict the novel drug from the approved drug using SMILES dataset.

Finally, the link Prediction is carried out with the optimized version. However, the proposed version does now no longer depend upon the 3D shape of proteins due to the fact it’s handiest primarily based totally at the protein series unbiased dataset. It lets in very clean utilization for non-computational area researchers so that it will attain a listing of capability tablets the use of deep learning i.e., Deep Neural Network (DNN), TRANSFORMER, Message Passing Neural Network (MPNN), CNN_RNN. Convolution Neural Network (CNN).

CHAPTER 2

LITERATURE SURVEY

2.1 CONVOLUTIONAL NEURAL NETWORK BIDIRECTIONAL LONG SHORT-TERM MEMORY-BASED METHOD FOR PREDICTING DRUG-DIEASE ASSOCIATIONS

Ping Xuan et al.,[6], incorporated noisy novel indicators of approved drugs that speed up the drug improvement procedure and reduce research costs. Most previous studies have used shallow models to prioritize the potential for drug-related illnesses to find a model and have failed to integrate in depth between drug and disease pathways where it may contain additional contact information. An in-depth approach to learn and predict drug associations by combining useful information is needed. The CNN-based framework focuses on studying the first presentation of drug-addicted couples in their similarities to organizations. We evolved a singular approach primarily based totally on the Convolutional Neural Network and then a short bidding memory for guessing drug-related diseases. Our approach includes similarities between the interactions among capsules and sicknesses and the goal among midnight drugs. it is possible that drug relationships in two drugs depends some of goal identification between them, the BiLSTM-based framework studies primarily the representation of the drug disease eradication approach.

2.2 INVERSE SIMILARITY AND RELIABLE NEGATIVE SAMPLES FOR DRUG SIDE-EFFECT PREDICTION

Yi Zheng, et al.,[5]. Therefore, the method of selecting bad bad samples becomes important in improving performance. Most of the prevailing computer-primarily based totally predictions are primarily based totally on the idea that comparable drugs generally tend to proportion comparable effects, which has caused significant results. It additionally makes feel to recall the contradictory idea that one-of-a-kind drugs are much less probable to proportion the same side effects. The importance of the system is to expect the chance of a DTI Based in this view of parallel similarity, the authors have proposed a novel technique for deciding on the maximum dependable bad samples for predicting side effects. The first step in our technique is to create a framework for integrating drug parallels to measure the similarities between drugs in different ways. This step includes the chemical composition of the drug, the targeted proteins, the drug details, and the drug treatment information as components of the integrated framework. Thereafter, the similarity points between each negative drug and the validated drugs were calculated using a similar combination framework. Those illegal drugs with the same pages selected are randomly selected as bad samples. Finally, both proven effective drugs and selected highly reliable samples were used for prediction.

2.3 COMPUTATIONAL DRUG REPOSITIONING: CURRENT PROGRESS AND CHALLENGES

Younhee Ko [1], emphasized that obtaining novel drugs is time-consuming, expensive, and costly due to their high level of attraction.

There are many experiments on drug reuse available to treat depressive disorders, because their safety profiles and pharmacokinetics are already available but drug reuptake is a strategy to identify a new indication of existing or already approved drugs, beyond its original level simultaneously use different computational and diagnostic methods. The available resources have been suggested to gain a higher information of the mechanisms of the disease and to identify recipients of chemotherapy regimens. The database process is information about drug repositioning and summarizes the approaches to drug rehabilitation.

2.4 DLS: A LINK PREDICTION METHOD BASED ON NETWORK LOCAL STARUCTURE FOR PREDICTING DRUG-PROTEIN INTERACTION

Wei Wang et al., proposed in vitro chemical testing that requires a long processing time and high cost to validate as it is difficult to identify appropriate drug-drug interactions but a possible solution is to predict well-controlled drug interactions. Comparing DLS predictive capabilities with improved similarity based on network prediction process and DLS results in the test set is much better but we should suggest a predictive link based on drug protein interactions with local structure to predict DPIs. The DLS method incorporates link forecasting and then a binary network structure to predict DPIs. Therefore, many selected proteins are predicted by three authorized drugs, namely captopril, desferrioxamine and losartan, those predictions also are showed with the aid of using literature. In addition, a mixture of the Common Neighborhood technique and the DLS technique become proposed to offer a brand new angle at the blended use of the hyperlink prediction technique.

2.5 A COMPARATIVE STUDY OF CLUSTER DETECTION ALGORITHMS IN PROTEIN-PROTEIN INTERACTION FOR TARGET DISCOVERY AND DRUG REPURPOSING

Jun Ma et al., [6], proposed topological modules in the calculation method which are defined as functional groups. However, the functions provided by these topological modules are derived from the main cell network and the structures of this active network. The importance of the technique changed into in the end in comparison with the overall performance of these genetic predictive systems, which include genetic mutation data and a cell-based PPI network using two priority approaches to drug targeting, a very short-term approach and distribution correction. The authors further verified the number of genetically modified genes in the groups by finding anti-cancer drugs.

2.6 DRUG-DISEASE ASSOCIATION PREDICTION BASED ON NEIGHBORHOOD INFORMATION AGGREGATION IN NEURAL NETWORK

Yingdong Wang et al., [5], proposed drug-disease asscoaition prediction in which the computational drug re-play plays an important role in predicting drug activity. Many new drugs have been verified compared to traditional drug repositioning, computer-based drug retention reduces time and increases performance. This is achieved primarily based totally on the gathering of neighbour information on neural networks that link similarities and drug concepts but also associations between drugs and diseases. Significant attention has been received in recent years but predictability remains a major challenge. In this paper, a method called HNRD is introduced to predict the link between drugs and disease. Compared to the previous high-tech approach, our approach achieved better results with an excellent AUC of 0.97 on one of the gold databases.

2.7 A DEEP LEARNING-BASED METHOD FOR DRUG TARGET INTERACTION PREDICTION BASED ON LONG SHORT-TERM MEMORY NEURAL NETWORK

Yan-Bin Wang et al.,[7], traditional ones are difficult to use extensively to incorporate this potential collaboration with Drug-Target. There is a need to develop effective calculation methods to ensure the link between drugs and their target. Evolutionary protein proteins are extracted through the Position Specific Scoring Matrix (PSSM) and Legendre Moment (LM) and are associated with cellular fingerprint drugs to form vectors of drug-targeted signals. Sparse Principal Component Analysis (SPCA) is used to press the drug and protein components into the space of the same vector. Lastly, Deep Short-Term Memory (DeepLSTM) is designed to make predictions. Results: Significant improvements in DTI predictive performance can be seen in test results, with AUC of 0.9951, 0.9705, 0.9951, 0.9206, respectively, in four key drug-targeted categories. The addition of preliminary testing proves that the proprietary promotion system has a great advantage in feature presentation and recognition.

2.8 Idrug: INTERACTION OF DRUG REPOSITIONING DRUG-TARGET PREDICTION VIA CROSS-NETWORK EMBEDDING

Huiyuan Chen et al., [3], Proposed a drug rehabilitation techniques and drug-targeted predictions have been important activities in the early stages of drug discovery. In previous studies, these two activities were often viewed differently. Businesses therefore learned from these two activities are naturally related. Drugs, on the other hand, combine to target cells to balance specific functions, which alter biological mechanisms to promote healthy activities and to treat disease. Drug re-enactment and drug-targeted predictions include the same area of medicine, which naturally connects these two problems through the ingenuity of crowds, it is possible to transfer information from one domain to another. The author have introduced a novel approach called iDrug, which combines seamless drug placement and drug-based predictions on the relationship between drug-induced diseases prompting us to collectively consider drug placement and drug prognosis for drug discovery. One of the most compatible models for network embedding.

2.9 DRUG REPOSITIONING USING DRUG-DISEASE VECTORS BASED ON AN INTEGRATED NETWORK

Taekeon Lee et al.,[2], Proposed a method that have Various interactions between biomolecules, such as activation, inhibition, speech, or stress. Previous research based on the drug rehabilitation network has used network communication for a protein-protein communication network without looking at interaction factors. Recently, other drug rehabilitation studies using genetic data have found that organizations communication process predicts that a new drug from the drug pathway exists between genetic pathogens and useful information to identify new drugs to treat disease. However, genetic profiles of drugs and diseases are not always available. Although genetic profiles for drug and disease genes exist, existing methods cannot use drugs or diseases, where the genes expressed differently from profiles are not included in their network. The authors have found a link between drug and disease genes, that creates a targeted network using protein interactions and genetic control information derived from various social data that provides a variety of biological mechanisms. The network incorporates three types of edges depending on the relationship between the biomolecules. To measure the correlation between the target gene and the type of disease, the authors examined shortcuts from targeted genes to the type of disease and calculated the types and weights of the shortest paths. In pairs of diseases, the authors have created a vector that contains the values of each disease-affected disease. Using vectors and well-known drug disease associations, the authors have created new drug identification stages for each disease. The authors have proposed a drug-specific drug testing approach using drug-and-genetic organizations based on a targeted genetic network instead of genetic control data obtained from gene profiles. Compared with existing methods that require knowledge of genetic relationships and gene expression data, our approach can be applied to a large number of drugs and diseases. In addition, to confirm our prediction, we compared the speculation of two drug diseases in clinical trials using hyper geometric tests, which showed significant results.

CHAPTER 3

SYSTEM DESIGN

The importance of this process is to find all the possible novel drugs from an existing solution using Neural Network and automatic identity to do drug target interplay affinity (regression) or drug target interaction prediction (binary) task. i.e., Deep Neural Network (DNN), TRANSFORMER, Message Passing Neural Network (MPNN), CNN_RNN, Convolution Neural Network (CNN).

Easy tracking of schooling method with certain schooling metrics output along with check set figures (AUCs) and assist bloodless target, bloodless drug settings for sturdy version reviews and assist single-goal high throughput sequencing assay facts setup.

3.1 OVERVIEW OF ANTIVIRAL DRUGS REPURPOSING

The general method of the version is a brand new target sequence (e.g. SARSCoV 3CL Pro), training on new data (AID1706 Bioassay), after which retrieve a listing of repurposing drugs from a proprietary library (e.g. antiviral drugs). The version may be trained from scratch or finetuned from the pretraining checkpoint and collecting Training DTI from various databases i.e., KEGG, IUPHAR, DrugBank. Construct the network using convolution. Hyperparameter with an external using validation dataset and Separated the positive and negative drug target interactions using MATADOR. Finally predicted the Drug Target Interaction using independent dataset. Figure 3.1 refer to the find the training Model.

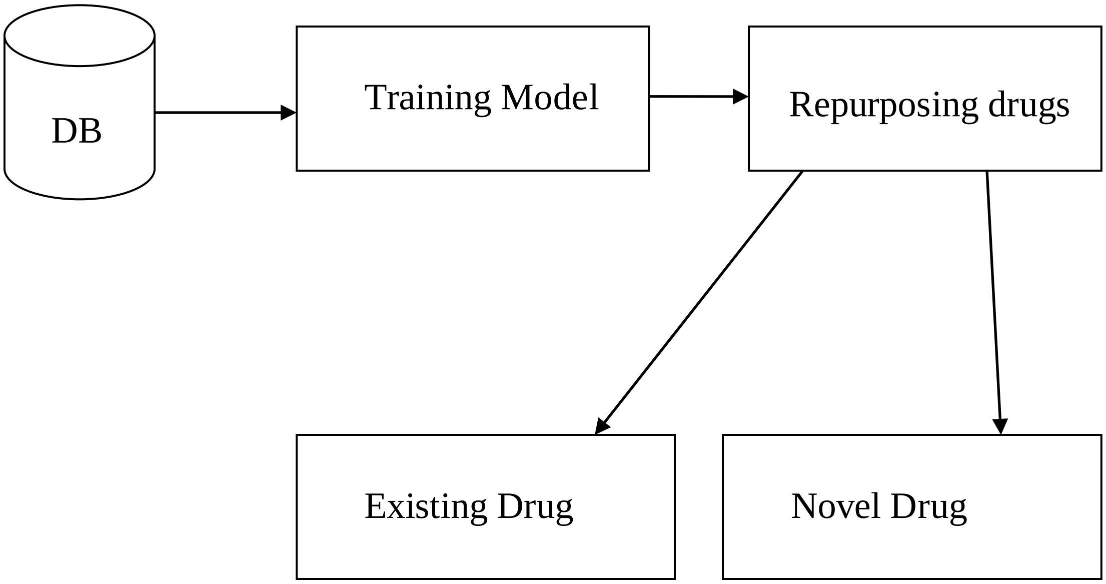


FIGURE 3.1: REFER TO THE FIND THE TRAINING MODEL.

Construct the network using convolution and Retrieve a listing of repurposing tablets from a curated drug library of 81 antiviral drugs. The Binding Score is the Kd values. Results aggregated from five pretrained version on BindingDB dataset. Figure 3.2 refer to the find the Link Prediction.

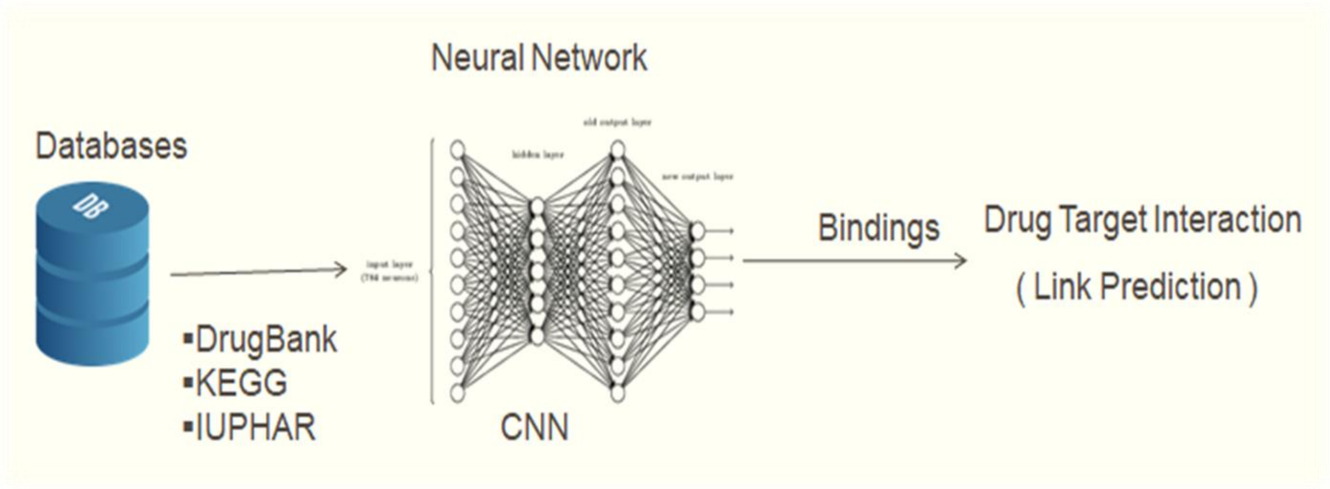


FIGURE 3.2: REFER TO THE FIND THE LINK PREDICTION.

3.2 PERFORMANCE OF Deep-DTI

Constructed the novel DTI prediction model to extract local residue patterns of target protein sequences and CNN based prediction model for Deep DTI but there is no exact method prediction, so that the protein descriptors for DTI prediction and previous model for predicting PubChem independent test dataset. The predicted DTI from bioassays depends on the independent test dataset.

3.3 DRAWBACKS OF EXISTINGSYSTEM

- Imbalance in the dataset.
- Negative samples are very high.
- Time delay is high.
- CTD and SW Score is very less.
- Very low accuracy.

CHAPTER 4

PROPOSED SYSTEM

Link prediction for drug repositioning from protein sequences with better performance compare to previous model and finding a novel drug from an existing approved drug using Deep Learning i.e., Deep Neural Network (DNN) and TRANSFORMER can be used to encode both drug and protein on SMILES, Message Passing Neural Network (MPNN) encode drug in its graph representation, CNN_RNN means a GRU/LSTM on top of a CNN on SMILES, Convolution Neural Network (CNN) are used to extract g Drug-Drug Interaction (DDI) from scientific documents and Drug-Protein Interactions.

4.1 SYSTEM ARCHITECTURE

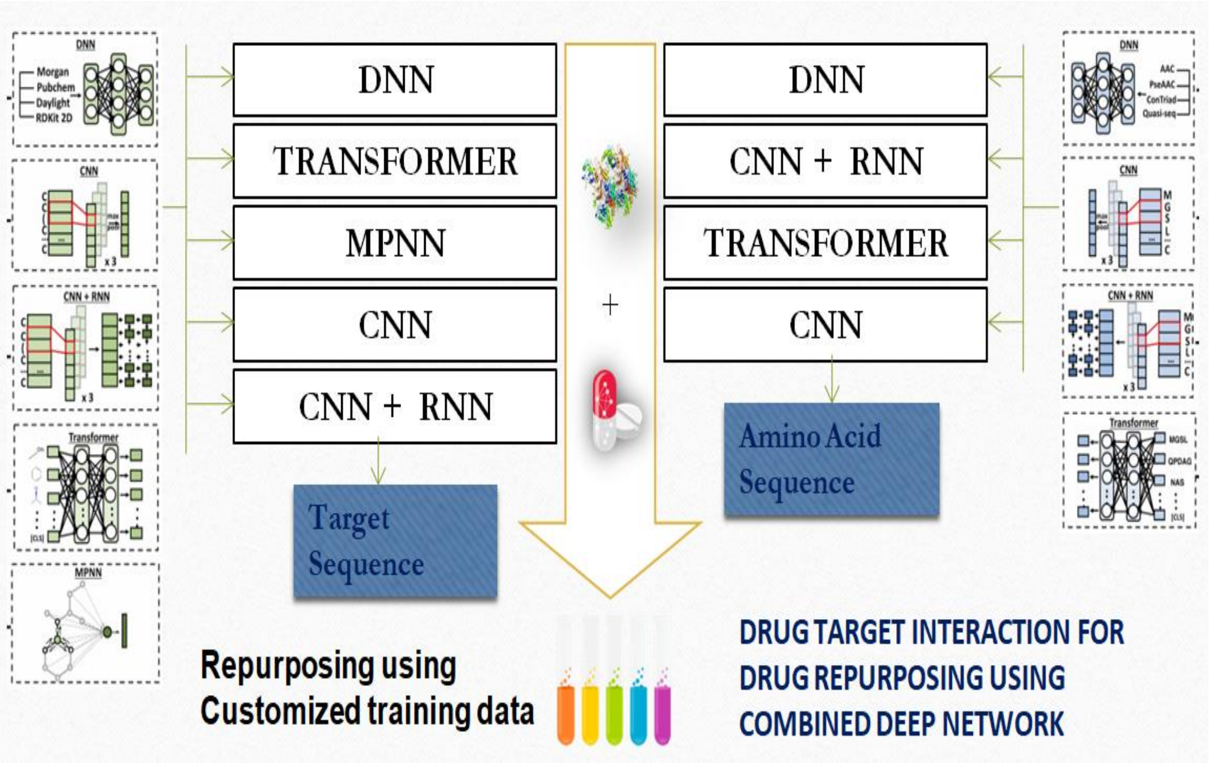


FIGURE 4.1 : NOVEL DRUG INTERACTION

4.2 MODULES OF LINK PREDICTION FOR DRUG REPURPOSEING

1. Training Dataset Generation
2. Targets fed into a decoder to predict DTI binding.
3. Improve The Sequence Interaction Level.
4. New DTI Prediction Using Independent Dataset
5. Interaction between the Drug Target sequence.
6. Finding a Novel drug from an Exiting Approved drug's.
7. Repurposing using Customized training data with probability.
8. Generate ranked lists for repurposing and screening.

4.2.1 TRAINING DATASET GENERATION

The training Drug Target Interaction dataset is collected from the different types of databases i.e., MORGAN, PubChem, IUPHAR, SMILES and imported that all the possible dataset to the input sequence and randomly generated the positive and negative component to the protein sequence.

4.2.2 TARGETS FED INTO A DECODER TO PREDICT DTI

The convolution neural network for training Drug Target Interaction dataset is constructed and encoded with the amino acid sequence and simplified molecular-Input line entry system (SMILES) to Optimization using MORGAN, PubChem, IUPHAR, SMILES and Negative Drug DTI. So that all the possible DTI is separated for increase the test and training level.

4.2.3 IMPROVE THE SEQUENCE INTERACTION LEVEL

The optimization network will generate the SMILES string key for integrate with protein sequence for identifying the training data's should be improved to better performance that again need to redo the process till the interaction level score is satisfied to all the possible prediction.

4.2.4 NEW DTI PREDICTION USING INDEPENDENT DATASET.

A fully Connected Layer is constructed to fetch the fingerprint and predict the negative DTI at the same time protein sequence must be formatted structured for balancing two input sequence so the new DTI will be generated after the training test process using independent test dataset i.e., SMILES, Amino Acid Sequence, Protein Sequence DTI.

4.2.5 INTERACTION BETWEEN THE DRUG TARGET SEQUENCE

To evaluate the drug ratio and disease ratio for improved interaction but there is no reverse process at all and the two interaction between the neural network is extracted for positive interaction then decoding the all the Active binding assays and Inactive binding assays for DTI.

4.2.6 FINDING A NOVEL DRUG FROM AN APPROVED DRUG'S

The local residue patterns of protein sequence captured and new interaction and find out the novel drug from an existing approved drug using independent test dataset for Drug Target Interaction (DTI).

4.2.7 REPURPOSING USING CUSTOMIZED TRAINING DATA WITH PROBABILITY

New target sequence (e.g. SARS-CoV 3CL Pro), training on new data (AID1706 Bioassay), and then retrieve a listing of repurposing capsules from a proprietary library (e.g. antiviral capsules). The version may be skilled from scratch or finetuned from the pretraining checkpoint.

4.2.8 GENERATE RANKED LISTS FOR REPURPOSING AND SCREENING

Virtual screening means to use computer software to automatically screen and a huge space of ability drug-goal pairs to gain a anticipated binding score. Finally DeepPurpose automates this process. By simplest requiring one line of code, it aggregates five pretrained deep mastering fashions and retrieves a listing of ranked ability outcomes. Predicted Novel drugs from an existing approved drug.

4.3 ADVANTAGES

- Better performance compare than previous model
- User friendly
- General framework for DTI

CHAPTER 5
IMPLEMENTATION AND RESULT

5.1 IMPLEMENTATION

DeepPurpose achieve competitive performance against state-of-the-art DL models for DTI prediction tasks on two benchmark datasets, DAVIS and KIBA. DeepPurpose achieves competitive performance on all metrics on both datasets, confirming its powerful performance.

DATASET	NUM OF DRUGS	NUMBER OF PROTEIN	INTERACTION
DAVIS[49]	80	369	31,056
KIBA[36]	2,358	354	138,254
BINDINGDBKD[20]	13,649	1,568	64,641

TABLE 5.1 : DATASET STATISTICS

It has 7.6% increase in MSE, 1.3% increase in concordance index for DAVIS dataset and 2.8% increase in MSE, 1.2% increase in concordance index for KIBA dataset over the best baseline. Due to the small size, models trained on these two datasets are not ideal for generalization over unseen drugs and proteins, especially for what we need for the one-line pretrained model.

	Model	MSE	Concordace
Baseline	GraphDTA	0.263	0.847
	DeepDTA	0.262	0.864
DeepPurpose	CNN	0.254	0.879

TABLE 5.2 : DAVIS DATASET INTERACTION

	Model	MSE	Concordace
Baseline	GraphDTA	0.183	0.682
	DeepDTA	0.196	0.864
DeepPurpose	CNN	0.196	0.856

TABLE 5.3 : KIBA DATASET INTERACTION

The sample 1,000 unseen drug-target pairs from DAVIS and feed them into the pertained models and report the Pearson correlation between the true and predicted binding scores and predicted the drug innovation in requence range the prediction sequence in table [4].

Level 0	index	frequency
1990	CC(C@@)1	2122
1991	NC(=0)c2cn	2119
1992	N(C(C	2118
1993	ClC1ccc(CC1)C2	2113
1994	(c@@h)OC	2111
1995	3CCC(C@h)3	2109
1996	NC=|0=|C@H(0=)C2	2108
1997	C(=0)N(C@H)c	2108
1998	CC@c@H(=0	2107
1999	CC@(C	2106

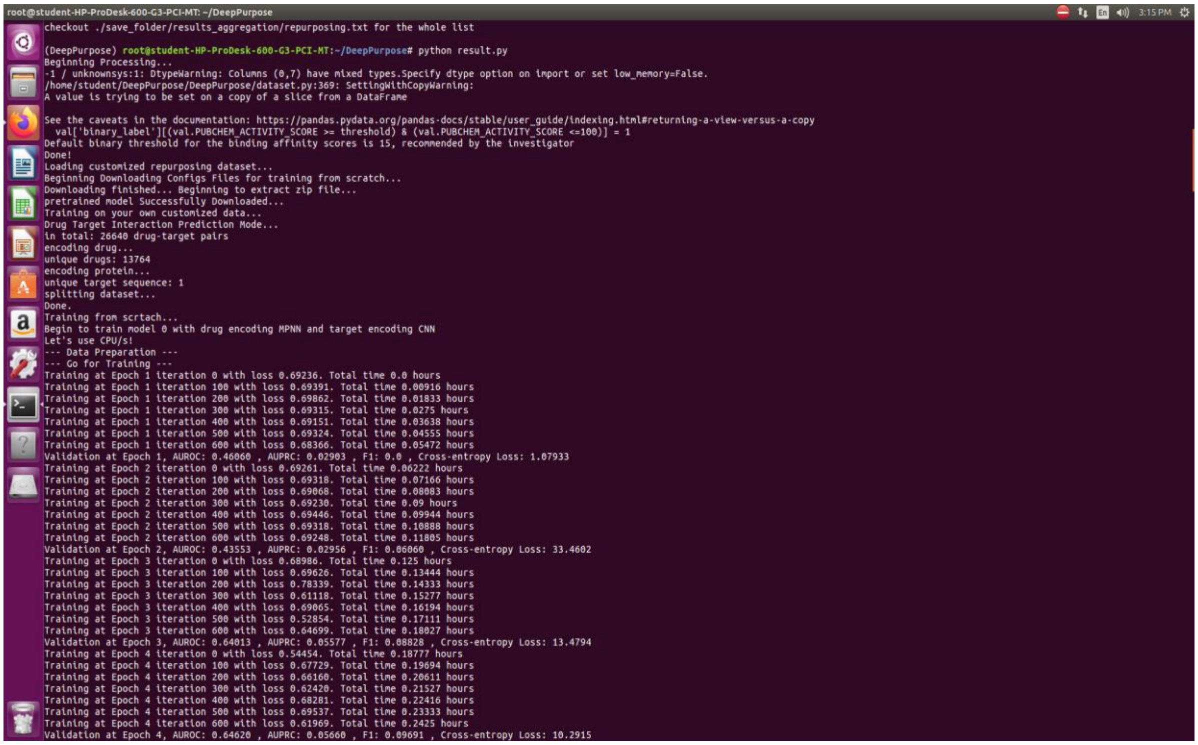
TABLE 5.4: CHEMICAL SEQUENCE

Level 0	index	frequency
1990	KTE	13328
1991	QPL	13302
1992	NR	13297

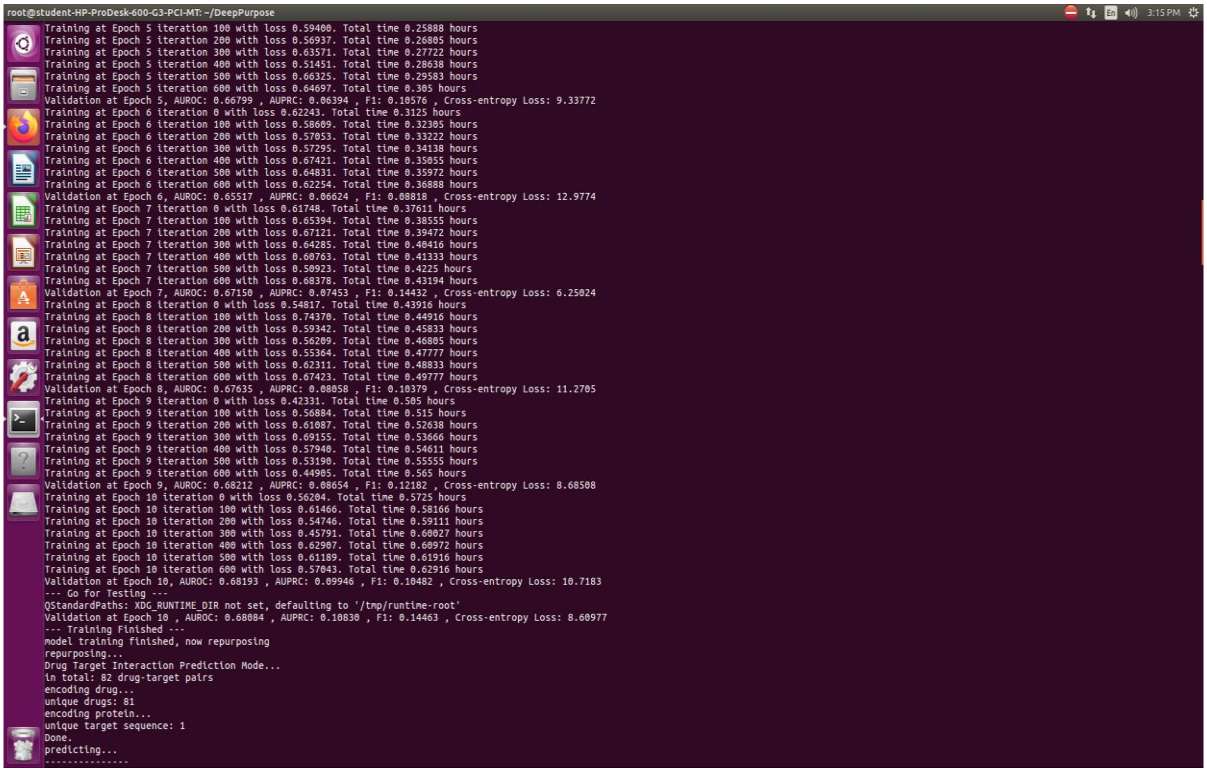
1993	FDG	13284
1994	YAA	13275
1995	KAT	13273
1996	PGL	13217
1997	NS	13207
1998	MKL	13205
1999	MN	13169

TABLE 5.5: PROTEIN SEQUENCE

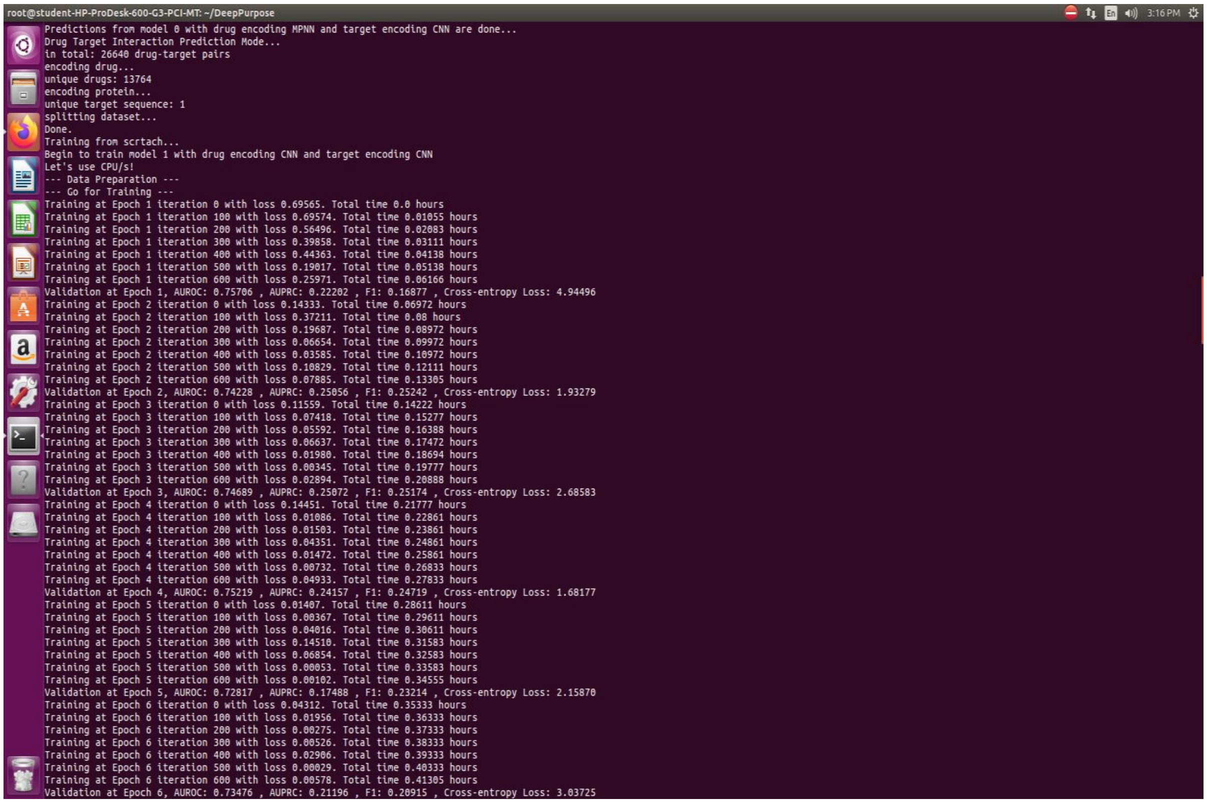
5.2 DRUG REPURPOSEING



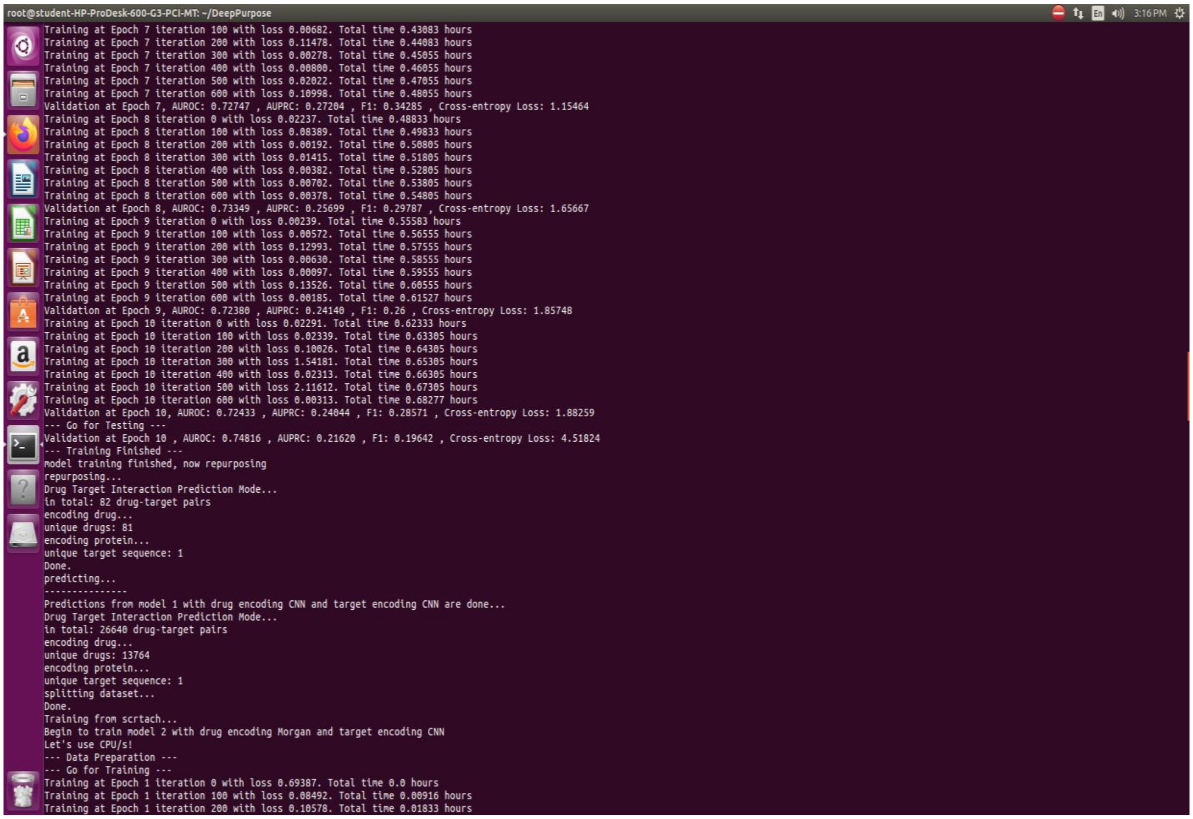
```
checkout ./save_folder/results_aggregation/repurposing.txt for the whole list

(DeepPurpose) root@student-HP-ProDesk-600-G3-PCI-MT:~/DeepPurpose# python result.py
Beginning Processing...
-1 / unknownsys:1: DtypeWarning: Columns (0,7) have mixed types.Specify dtype option on import or set low_memory=False.
/home/student/DeepPurpose/DeepPurpose/dataset.py:369: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  val['binary_label'][(val.PUBCHEM_ACTIVITY_SCORE >= threshold) & (val.PUBCHEM_ACTIVITY_SCORE <=100)] = 1
Default binary threshold for the binding affinity scores is 15, recommended by the investigator
Done!
Loading customized repurposing dataset...
Beginning Downloading Configs Files for training from scratch...
Downloading finished... Beginning to extract zip file...
pretrained model Successfully Downloaded...
Training on your own customized data...
Drug Target Interaction Prediction Mode...
in total: 26640 drug-target pairs
encoding drug...
unique drugs: 13764
encoding protein...
unique target sequence: 1
splitting dataset...
Done.
Training from scrtach...
Begin to train model 0 with drug encoding MPNN and target encoding CNN
Let's use CPU/s!
--- Data Preparation ---
--- Go for Training ---
Training at Epoch 1 iteration 0 with loss 0.69236. Total time 0.0 hours
Training at Epoch 1 iteration 100 with loss 0.69391. Total time 0.00916 hours
Training at Epoch 1 iteration 200 with loss 0.69862. Total time 0.01833 hours
Training at Epoch 1 iteration 300 with loss 0.69315. Total time 0.0275 hours
Training at Epoch 1 iteration 400 with loss 0.69151. Total time 0.03638 hours
Training at Epoch 1 iteration 500 with loss 0.69324. Total time 0.04555 hours
Training at Epoch 1 iteration 600 with loss 0.68366. Total time 0.05472 hours
Validation at Epoch 1, AUROC: 0.46060 , AUPRC: 0.02903 , F1: 0.0 , Cross-entropy Loss: 1.07933
Training at Epoch 2 iteration 0 with loss 0.69261. Total time 0.06222 hours
Training at Epoch 2 iteration 100 with loss 0.69318. Total time 0.07166 hours
Training at Epoch 2 iteration 200 with loss 0.69068. Total time 0.08083 hours
Training at Epoch 2 iteration 300 with loss 0.69230. Total time 0.09 hours
Training at Epoch 2 iteration 400 with loss 0.69446. Total time 0.09944 hours
Training at Epoch 2 iteration 500 with loss 0.69318. Total time 0.10888 hours
Training at Epoch 2 iteration 600 with loss 0.69248. Total time 0.11805 hours
Validation at Epoch 2, AUROC: 0.43553 , AUPRC: 0.02956 , F1: 0.06060 , Cross-entropy Loss: 33.4602
Training at Epoch 3 iteration 0 with loss 0.68986. Total time 0.125 hours
Training at Epoch 3 iteration 100 with loss 0.69626. Total time 0.13444 hours
Training at Epoch 3 iteration 200 with loss 0.78339. Total time 0.14333 hours
Training at Epoch 3 iteration 300 with loss 0.61118. Total time 0.15277 hours
Training at Epoch 3 iteration 400 with loss 0.69065. Total time 0.16194 hours
Training at Epoch 3 iteration 500 with loss 0.52854. Total time 0.17111 hours
Training at Epoch 3 iteration 600 with loss 0.64699. Total time 0.18027 hours
Validation at Epoch 3, AUROC: 0.64013 , AUPRC: 0.05577 , F1: 0.08828 , Cross-entropy Loss: 13.4794
Training at Epoch 4 iteration 0 with loss 0.54454. Total time 0.18777 hours
Training at Epoch 4 iteration 100 with loss 0.67729. Total time 0.19694 hours
Training at Epoch 4 iteration 200 with loss 0.66160. Total time 0.20611 hours
Training at Epoch 4 iteration 300 with loss 0.62420. Total time 0.21527 hours
Training at Epoch 4 iteration 400 with loss 0.68281. Total time 0.22416 hours
Training at Epoch 4 iteration 500 with loss 0.69537. Total time 0.23333 hours
Training at Epoch 4 iteration 600 with loss 0.61969. Total time 0.2425 hours
Validation at Epoch 4, AUROC: 0.64620 , AUPRC: 0.05660 , F1: 0.09691 , Cross-entropy Loss: 10.2915
```



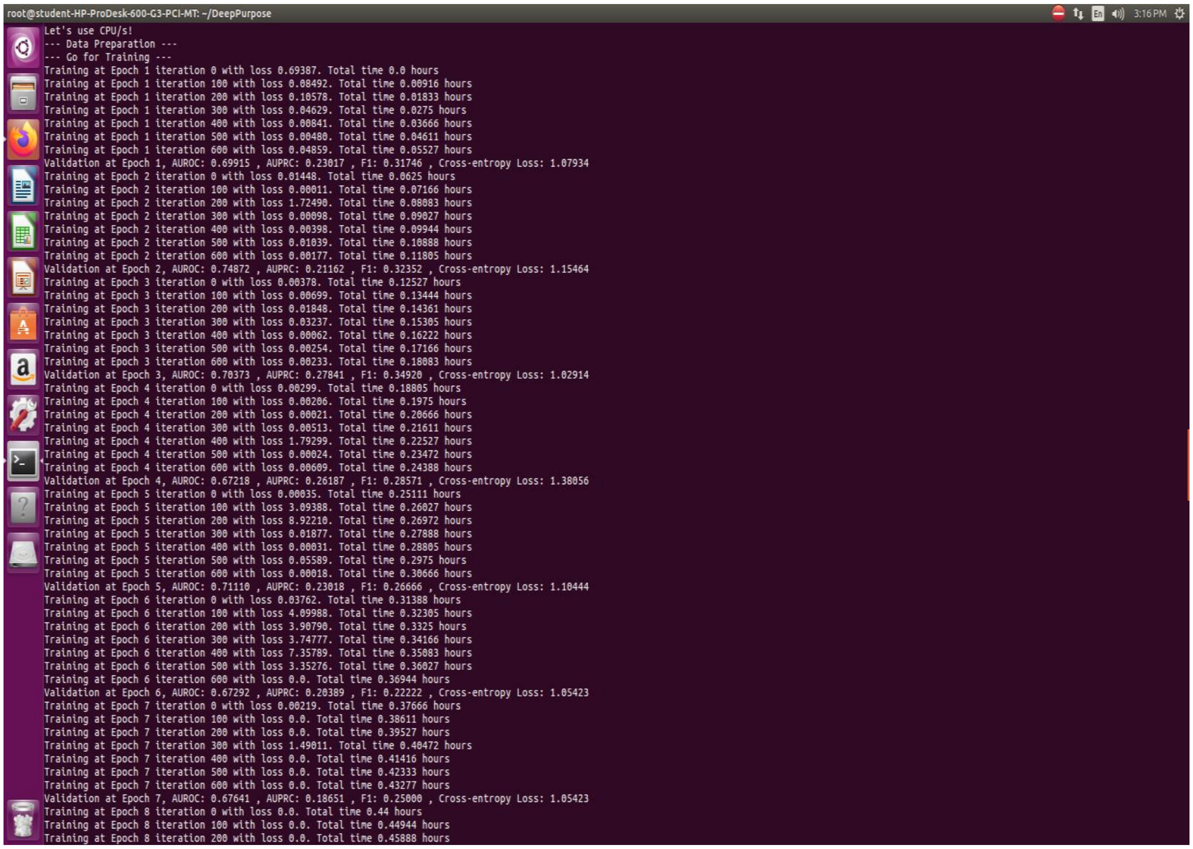
```
Training at Epoch 5 iteration 100 with loss 0.59400. Total time 0.25888 hours
Training at Epoch 5 iteration 200 with loss 0.56937. Total time 0.26805 hours
Training at Epoch 5 iteration 300 with loss 0.63571. Total time 0.27722 hours
Training at Epoch 5 iteration 400 with loss 0.51451. Total time 0.28638 hours
Training at Epoch 5 iteration 500 with loss 0.66325. Total time 0.29583 hours
Training at Epoch 5 iteration 600 with loss 0.64697. Total time 0.305 hours
Validation at Epoch 5, AUROC: 0.66799 , AUPRC: 0.06394 , F1: 0.10576 , Cross-entropy Loss: 9.33772
Training at Epoch 6 iteration 0 with loss 0.62243. Total time 0.3125 hours
Training at Epoch 6 iteration 100 with loss 0.58609. Total time 0.32305 hours
Training at Epoch 6 iteration 200 with loss 0.57053. Total time 0.33222 hours
Training at Epoch 6 iteration 300 with loss 0.57295. Total time 0.34138 hours
Training at Epoch 6 iteration 400 with loss 0.67421. Total time 0.35055 hours
Training at Epoch 6 iteration 500 with loss 0.64831. Total time 0.35972 hours
Training at Epoch 6 iteration 600 with loss 0.62254. Total time 0.36888 hours
Validation at Epoch 6, AUROC: 0.65517 , AUPRC: 0.06624 , F1: 0.08818 , Cross-entropy Loss: 12.9774
Training at Epoch 7 iteration 0 with loss 0.61748. Total time 0.37611 hours
Training at Epoch 7 iteration 100 with loss 0.65394. Total time 0.38555 hours
Training at Epoch 7 iteration 200 with loss 0.67121. Total time 0.39472 hours
Training at Epoch 7 iteration 300 with loss 0.64285. Total time 0.40416 hours
Training at Epoch 7 iteration 400 with loss 0.60763. Total time 0.41333 hours
Training at Epoch 7 iteration 500 with loss 0.50923. Total time 0.4225 hours
Training at Epoch 7 iteration 600 with loss 0.68378. Total time 0.43194 hours
Validation at Epoch 7, AUROC: 0.67150 , AUPRC: 0.07453 , F1: 0.14432 , Cross-entropy Loss: 6.25024
Training at Epoch 8 iteration 0 with loss 0.54817. Total time 0.43916 hours
Training at Epoch 8 iteration 100 with loss 0.74370. Total time 0.44916 hours
Training at Epoch 8 iteration 200 with loss 0.59342. Total time 0.45833 hours
Training at Epoch 8 iteration 300 with loss 0.56209. Total time 0.46805 hours
Training at Epoch 8 iteration 400 with loss 0.55364. Total time 0.47777 hours
Training at Epoch 8 iteration 500 with loss 0.62311. Total time 0.48833 hours
Training at Epoch 8 iteration 600 with loss 0.67423. Total time 0.49777 hours
Validation at Epoch 8, AUROC: 0.67635 , AUPRC: 0.08058 , F1: 0.10379 , Cross-entropy Loss: 11.2705
Training at Epoch 9 iteration 0 with loss 0.42331. Total time 0.505 hours
Training at Epoch 9 iteration 100 with loss 0.56884. Total time 0.515 hours
Training at Epoch 9 iteration 200 with loss 0.61087. Total time 0.52638 hours
Training at Epoch 9 iteration 300 with loss 0.69155. Total time 0.53666 hours
Training at Epoch 9 iteration 400 with loss 0.57940. Total time 0.54611 hours
Training at Epoch 9 iteration 500 with loss 0.53190. Total time 0.55555 hours
Training at Epoch 9 iteration 600 with loss 0.44905. Total time 0.565 hours
Validation at Epoch 9, AUROC: 0.68212 , AUPRC: 0.08654 , F1: 0.12182 , Cross-entropy Loss: 8.68508
Training at Epoch 10 iteration 0 with loss 0.56204. Total time 0.5725 hours
Training at Epoch 10 iteration 100 with loss 0.61466. Total time 0.58166 hours
Training at Epoch 10 iteration 200 with loss 0.54746. Total time 0.59111 hours
Training at Epoch 10 iteration 300 with loss 0.45791. Total time 0.60027 hours
Training at Epoch 10 iteration 400 with loss 0.62907. Total time 0.60972 hours
Training at Epoch 10 iteration 500 with loss 0.61189. Total time 0.61916 hours
Training at Epoch 10 iteration 600 with loss 0.57043. Total time 0.62916 hours
Validation at Epoch 10, AUROC: 0.68193 , AUPRC: 0.09946 , F1: 0.10482 , Cross-entropy Loss: 10.7183
--- Go for Testing ---
QStandardPaths: XDG_RUNTIME_DIR not set, defaulting to '/tmp/runtime-root'
Validation at Epoch 10 , AUROC: 0.68084 , AUPRC: 0.10830 , F1: 0.14463 , Cross-entropy Loss: 8.60977
--- Training Finished ---
model training finished, now repurposing
repurposing...
Drug Target Interaction Prediction Mode...
in total: 82 drug-target pairs
encoding drug...
unique drugs: 81
encoding protein...
unique target sequence: 1
Done.
predicting...
--------------
```



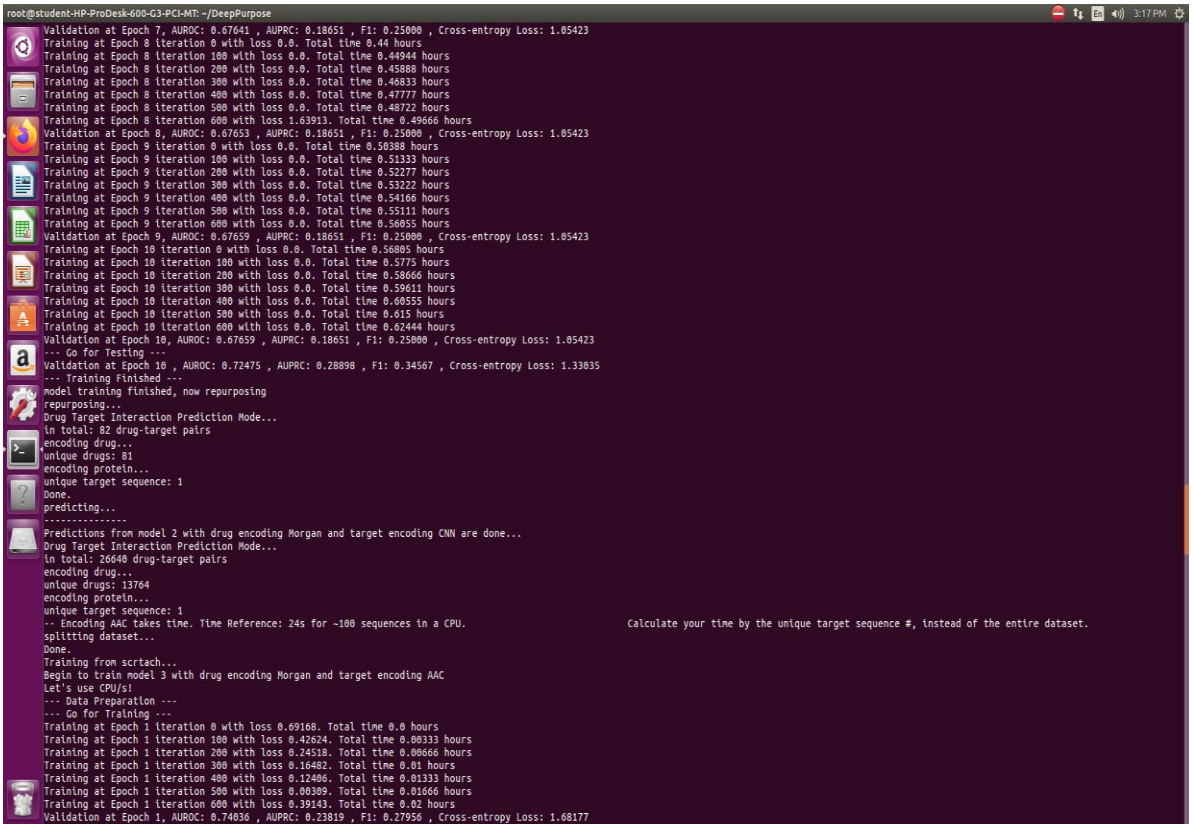
```
Predictions from model 0 with drug encoding MPNN and target encoding CNN are done...
Drug Target Interaction Prediction Mode...
in total: 26640 drug-target pairs
encoding drug...
unique drugs: 13764
encoding protein...
unique target sequence: 1
splitting dataset...
Done.
Training from scrtach...
Begin to train model 1 with drug encoding CNN and target encoding CNN
Let's use CPU/s!
--- Data Preparation ---
--- Go for Training ---
Training at Epoch 1 iteration 0 with loss 0.69565. Total time 0.0 hours
Training at Epoch 1 iteration 100 with loss 0.69574. Total time 0.01055 hours
Training at Epoch 1 iteration 200 with loss 0.56496. Total time 0.02083 hours
Training at Epoch 1 iteration 300 with loss 0.39858. Total time 0.03111 hours
Training at Epoch 1 iteration 400 with loss 0.44363. Total time 0.04138 hours
Training at Epoch 1 iteration 500 with loss 0.19017. Total time 0.05138 hours
Training at Epoch 1 iteration 600 with loss 0.25971. Total time 0.06166 hours
Validation at Epoch 1, AUROC: 0.75706 , AUPRC: 0.22202 , F1: 0.16877 , Cross-entropy Loss: 4.94496
Training at Epoch 2 iteration 0 with loss 0.14333. Total time 0.06972 hours
Training at Epoch 2 iteration 100 with loss 0.37211. Total time 0.08 hours
Training at Epoch 2 iteration 200 with loss 0.19687. Total time 0.08972 hours
Training at Epoch 2 iteration 300 with loss 0.06654. Total time 0.09972 hours
Training at Epoch 2 iteration 400 with loss 0.03585. Total time 0.10972 hours
Training at Epoch 2 iteration 500 with loss 0.10829. Total time 0.12111 hours
Training at Epoch 2 iteration 600 with loss 0.07885. Total time 0.13305 hours
Validation at Epoch 2, AUROC: 0.74228 , AUPRC: 0.25056 , F1: 0.25242 , Cross-entropy Loss: 1.93279
Training at Epoch 3 iteration 0 with loss 0.11559. Total time 0.14222 hours
Training at Epoch 3 iteration 100 with loss 0.07418. Total time 0.15277 hours
Training at Epoch 3 iteration 200 with loss 0.05592. Total time 0.16388 hours
Training at Epoch 3 iteration 300 with loss 0.06637. Total time 0.17472 hours
Training at Epoch 3 iteration 400 with loss 0.01980. Total time 0.18694 hours
Training at Epoch 3 iteration 500 with loss 0.00345. Total time 0.19777 hours
Training at Epoch 3 iteration 600 with loss 0.02894. Total time 0.20888 hours
Validation at Epoch 3, AUROC: 0.74689 , AUPRC: 0.25072 , F1: 0.25174 , Cross-entropy Loss: 2.68583
Training at Epoch 4 iteration 0 with loss 0.14451. Total time 0.21777 hours
Training at Epoch 4 iteration 100 with loss 0.01086. Total time 0.22861 hours
Training at Epoch 4 iteration 200 with loss 0.01503. Total time 0.23861 hours
Training at Epoch 4 iteration 300 with loss 0.04351. Total time 0.24861 hours
Training at Epoch 4 iteration 400 with loss 0.01472. Total time 0.25861 hours
Training at Epoch 4 iteration 500 with loss 0.00732. Total time 0.26833 hours
Training at Epoch 4 iteration 600 with loss 0.04933. Total time 0.27833 hours
Validation at Epoch 4, AUROC: 0.75219 , AUPRC: 0.24157 , F1: 0.24719 , Cross-entropy Loss: 1.68177
Training at Epoch 5 iteration 0 with loss 0.01407. Total time 0.28611 hours
Training at Epoch 5 iteration 100 with loss 0.00367. Total time 0.29611 hours
Training at Epoch 5 iteration 200 with loss 0.04016. Total time 0.30611 hours
Training at Epoch 5 iteration 300 with loss 0.14510. Total time 0.31583 hours
Training at Epoch 5 iteration 400 with loss 0.06854. Total time 0.32583 hours
Training at Epoch 5 iteration 500 with loss 0.00053. Total time 0.33583 hours
Training at Epoch 5 iteration 600 with loss 0.00102. Total time 0.34555 hours
Validation at Epoch 5, AUROC: 0.72817 , AUPRC: 0.17488 , F1: 0.23214 , Cross-entropy Loss: 2.15870
Training at Epoch 6 iteration 0 with loss 0.04312. Total time 0.35333 hours
Training at Epoch 6 iteration 100 with loss 0.01956. Total time 0.36333 hours
Training at Epoch 6 iteration 200 with loss 0.00275. Total time 0.37333 hours
Training at Epoch 6 iteration 300 with loss 0.00526. Total time 0.38333 hours
Training at Epoch 6 iteration 400 with loss 0.02906. Total time 0.39333 hours
Training at Epoch 6 iteration 500 with loss 0.00029. Total time 0.40333 hours
Training at Epoch 6 iteration 600 with loss 0.00578. Total time 0.41305 hours
Validation at Epoch 6, AUROC: 0.73476 , AUPRC: 0.21196 , F1: 0.20915 , Cross-entropy Loss: 3.03725
```



```
Training at Epoch 7 iteration 100 with loss 0.00682. Total time 0.43083 hours
Training at Epoch 7 iteration 200 with loss 0.11478. Total time 0.44083 hours
Training at Epoch 7 iteration 300 with loss 0.00278. Total time 0.45055 hours
Training at Epoch 7 iteration 400 with loss 0.00800. Total time 0.46055 hours
Training at Epoch 7 iteration 500 with loss 0.02022. Total time 0.47055 hours
Training at Epoch 7 iteration 600 with loss 0.10998. Total time 0.48055 hours
Validation at Epoch 7, AUROC: 0.72747 , AUPRC: 0.27204 , F1: 0.34285 , Cross-entropy Loss: 1.15464
Training at Epoch 8 iteration 0 with loss 0.02237. Total time 0.48833 hours
Training at Epoch 8 iteration 100 with loss 0.08389. Total time 0.49833 hours
Training at Epoch 8 iteration 200 with loss 0.00192. Total time 0.50805 hours
Training at Epoch 8 iteration 300 with loss 0.01415. Total time 0.51805 hours
Training at Epoch 8 iteration 400 with loss 0.00382. Total time 0.52805 hours
Training at Epoch 8 iteration 500 with loss 0.00702. Total time 0.53805 hours
Training at Epoch 8 iteration 600 with loss 0.00378. Total time 0.54805 hours
Validation at Epoch 8, AUROC: 0.73349 , AUPRC: 0.25699 , F1: 0.29787 , Cross-entropy Loss: 1.65667
Training at Epoch 9 iteration 0 with loss 0.00239. Total time 0.55583 hours
Training at Epoch 9 iteration 100 with loss 0.00572. Total time 0.56555 hours
Training at Epoch 9 iteration 200 with loss 0.12993. Total time 0.57555 hours
Training at Epoch 9 iteration 300 with loss 0.00630. Total time 0.58555 hours
Training at Epoch 9 iteration 400 with loss 0.00097. Total time 0.59555 hours
Training at Epoch 9 iteration 500 with loss 0.13526. Total time 0.60555 hours
Training at Epoch 9 iteration 600 with loss 0.00185. Total time 0.61527 hours
Validation at Epoch 9, AUROC: 0.72380 , AUPRC: 0.24140 , F1: 0.26 , Cross-entropy Loss: 1.85748
Training at Epoch 10 iteration 0 with loss 0.02291. Total time 0.62333 hours
Training at Epoch 10 iteration 100 with loss 0.02339. Total time 0.63305 hours
Training at Epoch 10 iteration 200 with loss 0.10026. Total time 0.64305 hours
Training at Epoch 10 iteration 300 with loss 1.54181. Total time 0.65305 hours
Training at Epoch 10 iteration 400 with loss 0.02313. Total time 0.66305 hours
Training at Epoch 10 iteration 500 with loss 2.11612. Total time 0.67305 hours
Training at Epoch 10 iteration 600 with loss 0.00313. Total time 0.68277 hours
Validation at Epoch 10, AUROC: 0.72433 , AUPRC: 0.24044 , F1: 0.28571 , Cross-entropy Loss: 1.88259
--- Go for Testing ---
Validation at Epoch 10 , AUROC: 0.74816 , AUPRC: 0.21620 , F1: 0.19642 , Cross-entropy Loss: 4.51824
--- Training Finished ---
model training finished, now repurposing
repurposing...
Drug Target Interaction Prediction Mode...
in total: 82 drug-target pairs
encoding drug...
unique drugs: 81
encoding protein...
unique target sequence: 1
Done.
predicting...
---------------
Predictions from model 1 with drug encoding CNN and target encoding CNN are done...
Drug Target Interaction Prediction Mode...
in total: 26640 drug-target pairs
encoding drug...
unique drugs: 13764
encoding protein...
unique target sequence: 1
splitting dataset...
Done.
Training from scrtach...
Begin to train model 2 with drug encoding Morgan and target encoding CNN
Let's use CPU/s!
--- Data Preparation ---
--- Go for Training ---
Training at Epoch 1 iteration 0 with loss 0.69387. Total time 0.0 hours
Training at Epoch 1 iteration 100 with loss 0.08492. Total time 0.00916 hours
Training at Epoch 1 iteration 200 with loss 0.10578. Total time 0.01833 hours
```

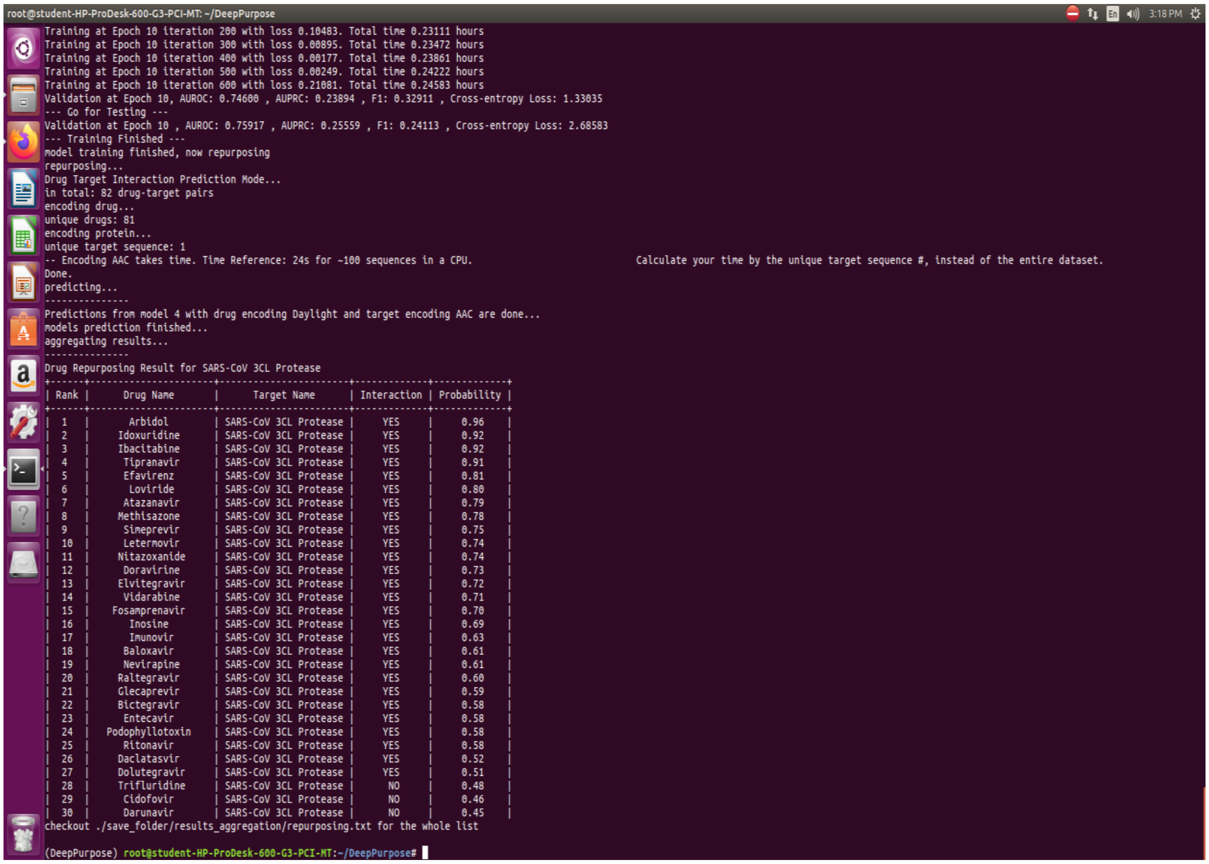


```
Let's use CPU/s!
--- Data Preparation ---
--- Go for Training ---
Training at Epoch 1 iteration 0 with loss 0.69387. Total time 0.0 hours
Training at Epoch 1 iteration 100 with loss 0.08492. Total time 0.00916 hours
Training at Epoch 1 iteration 200 with loss 0.10578. Total time 0.01833 hours
Training at Epoch 1 iteration 300 with loss 0.04629. Total time 0.0275 hours
Training at Epoch 1 iteration 400 with loss 0.00841. Total time 0.03666 hours
Training at Epoch 1 iteration 500 with loss 0.00480. Total time 0.04611 hours
Training at Epoch 1 iteration 600 with loss 0.04859. Total time 0.05527 hours
Validation at Epoch 1, AUROC: 0.69915 , AUPRC: 0.23017 , F1: 0.31746 , Cross-entropy Loss: 1.07934
Training at Epoch 2 iteration 0 with loss 0.01448. Total time 0.0625 hours
Training at Epoch 2 iteration 100 with loss 0.00011. Total time 0.07166 hours
Training at Epoch 2 iteration 200 with loss 1.72490. Total time 0.08083 hours
Training at Epoch 2 iteration 300 with loss 0.00098. Total time 0.09027 hours
Training at Epoch 2 iteration 400 with loss 0.00398. Total time 0.09944 hours
Training at Epoch 2 iteration 500 with loss 0.01039. Total time 0.10888 hours
Training at Epoch 2 iteration 600 with loss 0.00177. Total time 0.11805 hours
Validation at Epoch 2, AUROC: 0.74872 , AUPRC: 0.21162 , F1: 0.32352 , Cross-entropy Loss: 1.15464
Training at Epoch 3 iteration 0 with loss 0.00378. Total time 0.12527 hours
Training at Epoch 3 iteration 100 with loss 0.00699. Total time 0.13444 hours
Training at Epoch 3 iteration 200 with loss 0.01848. Total time 0.14361 hours
Training at Epoch 3 iteration 300 with loss 0.03237. Total time 0.15305 hours
Training at Epoch 3 iteration 400 with loss 0.00062. Total time 0.16222 hours
Training at Epoch 3 iteration 500 with loss 0.00254. Total time 0.17166 hours
Training at Epoch 3 iteration 600 with loss 0.00233. Total time 0.18083 hours
Validation at Epoch 3, AUROC: 0.70373 , AUPRC: 0.27841 , F1: 0.34920 , Cross-entropy Loss: 1.02914
Training at Epoch 4 iteration 0 with loss 0.00299. Total time 0.18805 hours
Training at Epoch 4 iteration 100 with loss 0.00206. Total time 0.1975 hours
Training at Epoch 4 iteration 200 with loss 0.00021. Total time 0.20666 hours
Training at Epoch 4 iteration 300 with loss 0.00513. Total time 0.21611 hours
Training at Epoch 4 iteration 400 with loss 1.79299. Total time 0.22527 hours
Training at Epoch 4 iteration 500 with loss 0.00024. Total time 0.23472 hours
Training at Epoch 4 iteration 600 with loss 0.00609. Total time 0.24388 hours
Validation at Epoch 4, AUROC: 0.67218 , AUPRC: 0.26187 , F1: 0.28571 , Cross-entropy Loss: 1.38056
Training at Epoch 5 iteration 0 with loss 0.00035. Total time 0.25111 hours
Training at Epoch 5 iteration 100 with loss 3.09388. Total time 0.26027 hours
Training at Epoch 5 iteration 200 with loss 8.92210. Total time 0.26972 hours
Training at Epoch 5 iteration 300 with loss 0.01877. Total time 0.27888 hours
Training at Epoch 5 iteration 400 with loss 0.00031. Total time 0.28805 hours
Training at Epoch 5 iteration 500 with loss 0.05589. Total time 0.2975 hours
Training at Epoch 5 iteration 600 with loss 0.00018. Total time 0.30666 hours
Validation at Epoch 5, AUROC: 0.71110 , AUPRC: 0.23018 , F1: 0.26666 , Cross-entropy Loss: 1.10444
Training at Epoch 6 iteration 0 with loss 0.03762. Total time 0.31388 hours
Training at Epoch 6 iteration 100 with loss 4.09988. Total time 0.32305 hours
Training at Epoch 6 iteration 200 with loss 3.90790. Total time 0.3325 hours
Training at Epoch 6 iteration 300 with loss 3.74777. Total time 0.34166 hours
Training at Epoch 6 iteration 400 with loss 7.35789. Total time 0.35083 hours
Training at Epoch 6 iteration 500 with loss 3.35276. Total time 0.36027 hours
Training at Epoch 6 iteration 600 with loss 0.0. Total time 0.36944 hours
Validation at Epoch 6, AUROC: 0.67292 , AUPRC: 0.20389 , F1: 0.22222 , Cross-entropy Loss: 1.05423
Training at Epoch 7 iteration 0 with loss 0.00219. Total time 0.37666 hours
Training at Epoch 7 iteration 100 with loss 0.0. Total time 0.38611 hours
Training at Epoch 7 iteration 200 with loss 0.0. Total time 0.39527 hours
Training at Epoch 7 iteration 300 with loss 1.49011. Total time 0.40472 hours
Training at Epoch 7 iteration 400 with loss 0.0. Total time 0.41416 hours
Training at Epoch 7 iteration 500 with loss 0.0. Total time 0.42333 hours
Training at Epoch 7 iteration 600 with loss 0.0. Total time 0.43277 hours
Validation at Epoch 7, AUROC: 0.67641 , AUPRC: 0.18651 , F1: 0.25000 , Cross-entropy Loss: 1.05423
Training at Epoch 8 iteration 0 with loss 0.0. Total time 0.44 hours
Training at Epoch 8 iteration 100 with loss 0.0. Total time 0.44944 hours
Training at Epoch 8 iteration 200 with loss 0.0. Total time 0.45888 hours
```



```
Validation at Epoch 7, AUROC: 0.67641 , AUPRC: 0.18651 , F1: 0.25000 , Cross-entropy Loss: 1.05423
Training at Epoch 8 iteration 0 with loss 0.0. Total time 0.44 hours
Training at Epoch 8 iteration 100 with loss 0.0. Total time 0.44944 hours
Training at Epoch 8 iteration 200 with loss 0.0. Total time 0.45888 hours
Training at Epoch 8 iteration 300 with loss 0.0. Total time 0.46833 hours
Training at Epoch 8 iteration 400 with loss 0.0. Total time 0.47777 hours
Training at Epoch 8 iteration 500 with loss 0.0. Total time 0.48722 hours
Training at Epoch 8 iteration 600 with loss 1.63913. Total time 0.49666 hours
Validation at Epoch 8, AUROC: 0.67653 , AUPRC: 0.18651 , F1: 0.25000 , Cross-entropy Loss: 1.05423
Training at Epoch 9 iteration 0 with loss 0.0. Total time 0.50388 hours
Training at Epoch 9 iteration 100 with loss 0.0. Total time 0.51333 hours
Training at Epoch 9 iteration 200 with loss 0.0. Total time 0.52277 hours
Training at Epoch 9 iteration 300 with loss 0.0. Total time 0.53222 hours
Training at Epoch 9 iteration 400 with loss 0.0. Total time 0.54166 hours
Training at Epoch 9 iteration 500 with loss 0.0. Total time 0.55111 hours
Training at Epoch 9 iteration 600 with loss 0.0. Total time 0.56055 hours
Validation at Epoch 9, AUROC: 0.67659 , AUPRC: 0.18651 , F1: 0.25000 , Cross-entropy Loss: 1.05423
Training at Epoch 10 iteration 0 with loss 0.0. Total time 0.56805 hours
Training at Epoch 10 iteration 100 with loss 0.0. Total time 0.5775 hours
Training at Epoch 10 iteration 200 with loss 0.0. Total time 0.58666 hours
Training at Epoch 10 iteration 300 with loss 0.0. Total time 0.59611 hours
Training at Epoch 10 iteration 400 with loss 0.0. Total time 0.60555 hours
Training at Epoch 10 iteration 500 with loss 0.0. Total time 0.615 hours
Training at Epoch 10 iteration 600 with loss 0.0. Total time 0.62444 hours
Validation at Epoch 10, AUROC: 0.67659 , AUPRC: 0.18651 , F1: 0.25000 , Cross-entropy Loss: 1.05423
--- Go for Testing ---
Validation at Epoch 10 , AUROC: 0.72475 , AUPRC: 0.28898 , F1: 0.34567 , Cross-entropy Loss: 1.33035
--- Training Finished ---
model training finished, now repurposing
repurposing...
Drug Target Interaction Prediction Mode...
in total: 82 drug-target pairs
encoding drug...
unique drugs: 81
encoding protein...
unique target sequence: 1
Done.
predicting...
---------------
Predictions from model 2 with drug encoding Morgan and target encoding CNN are done...
Drug Target Interaction Prediction Mode...
in total: 26640 drug-target pairs
encoding drug...
unique drugs: 13764
encoding protein...
unique target sequence: 1
-- Encoding AAC takes time. Time Reference: 24s for ~100 sequences in a CPU.                              Calculate your time by the unique target sequence #, instead of the entire dataset.
splitting dataset...
Done.
Training from scrtach...
Begin to train model 3 with drug encoding Morgan and target encoding AAC
Let's use CPU/s!
--- Data Preparation ---
--- Go for Training ---
Training at Epoch 1 iteration 0 with loss 0.69168. Total time 0.0 hours
Training at Epoch 1 iteration 100 with loss 0.42624. Total time 0.00333 hours
Training at Epoch 1 iteration 200 with loss 0.24518. Total time 0.00666 hours
Training at Epoch 1 iteration 300 with loss 0.16482. Total time 0.01 hours
Training at Epoch 1 iteration 400 with loss 0.12406. Total time 0.01333 hours
Training at Epoch 1 iteration 500 with loss 0.00309. Total time 0.01666 hours
Training at Epoch 1 iteration 600 with loss 0.39143. Total time 0.02 hours
Validation at Epoch 1, AUROC: 0.74036 , AUPRC: 0.23819 , F1: 0.27956 , Cross-entropy Loss: 1.68177
```

5.3 NOVEL DRUG OUTPUT



```
Training at Epoch 10 iteration 200 with loss 0.10483. Total time 0.23111 hours
Training at Epoch 10 iteration 300 with loss 0.00895. Total time 0.23472 hours
Training at Epoch 10 iteration 400 with loss 0.00177. Total time 0.23861 hours
Training at Epoch 10 iteration 500 with loss 0.00249. Total time 0.24222 hours
Training at Epoch 10 iteration 600 with loss 0.21081. Total time 0.24583 hours
Validation at Epoch 10, AUROC: 0.74600 , AUPRC: 0.23894 , F1: 0.32911 , Cross-entropy Loss: 1.33035
--- Go for Testing ---
Validation at Epoch 10 , AUROC: 0.75917 , AUPRC: 0.25559 , F1: 0.24113 , Cross-entropy Loss: 2.68583
--- Training Finished ---
model training finished, now repurposing
repurposing...
Drug Target Interaction Prediction Mode...
in total: 82 drug-target pairs
encoding drug...
unique drugs: 81
encoding protein...
unique target sequence: 1
-- Encoding AAC takes time. Time Reference: 24s for ~100 sequences in a CPU.                              Calculate your time by the unique target sequence #, instead of the entire dataset.
Done.
predicting...
---------------
Predictions from model 4 with drug encoding Daylight and target encoding AAC are done...
models prediction finished...
aggregating results...
---------------
Drug Repurposing Result for SARS-CoV 3CL Protease
+------+----------------------+-----------------------+-------------+-------------+
| Rank |      Drug Name       |      Target Name      | Interaction | Probability |
+------+----------------------+-----------------------+-------------+-------------+
|  1   |       Arbidol        | SARS-CoV 3CL Protease |     YES     |     0.96    |
|  2   |      Idoxuridine     | SARS-CoV 3CL Protease |     YES     |     0.92    |
|  3   |      Ibacitabine     | SARS-CoV 3CL Protease |     YES     |     0.92    |
|  4   |      Tipranavir      | SARS-CoV 3CL Protease |     YES     |     0.91    |
|  5   |      Efavirenz       | SARS-CoV 3CL Protease |     YES     |     0.81    |
|  6   |       Loviride       | SARS-CoV 3CL Protease |     YES     |     0.80    |
|  7   |      Atazanavir      | SARS-CoV 3CL Protease |     YES     |     0.79    |
|  8   |      Methisazone     | SARS-CoV 3CL Protease |     YES     |     0.78    |
|  9   |      Simeprevir      | SARS-CoV 3CL Protease |     YES     |     0.75    |
|  10  |      Letermovir      | SARS-CoV 3CL Protease |     YES     |     0.74    |
|  11  |     Nitazoxanide     | SARS-CoV 3CL Protease |     YES     |     0.74    |
|  12  |      Doravirine      | SARS-CoV 3CL Protease |     YES     |     0.73    |
|  13  |     Elvitegravir     | SARS-CoV 3CL Protease |     YES     |     0.72    |
|  14  |      Vidarabine      | SARS-CoV 3CL Protease |     YES     |     0.71    |
|  15  |    Fosamprenavir     | SARS-CoV 3CL Protease |     YES     |     0.70    |
|  16  |       Inosine        | SARS-CoV 3CL Protease |     YES     |     0.69    |
|  17  |       Imunovir       | SARS-CoV 3CL Protease |     YES     |     0.63    |
|  18  |      Baloxavir       | SARS-CoV 3CL Protease |     YES     |     0.61    |
|  19  |      Nevirapine      | SARS-CoV 3CL Protease |     YES     |     0.61    |
|  20  |     Raltegravir      | SARS-CoV 3CL Protease |     YES     |     0.60    |
|  21  |     Glecaprevir      | SARS-CoV 3CL Protease |     YES     |     0.59    |
|  22  |     Bictegravir      | SARS-CoV 3CL Protease |     YES     |     0.58    |
|  23  |      Entecavir       | SARS-CoV 3CL Protease |     YES     |     0.58    |
|  24  |   Podophyllotoxin    | SARS-CoV 3CL Protease |     YES     |     0.58    |
|  25  |      Ritonavir       | SARS-CoV 3CL Protease |     YES     |     0.58    |
|  26  |     Daclatasvir      | SARS-CoV 3CL Protease |     YES     |     0.52    |
|  27  |     Dolutegravir     | SARS-CoV 3CL Protease |     YES     |     0.51    |
|  28  |     Trifluridine     | SARS-CoV 3CL Protease |      NO     |     0.48    |
|  29  |      Cidofovir       | SARS-CoV 3CL Protease |      NO     |     0.46    |
|  30  |      Darunavir       | SARS-CoV 3CL Protease |      NO     |     0.45    |
checkout ./save_folder/results_aggregation/repurposing.txt for the whole list

(DeepPurpose) root@student-HP-ProDesk-600-G3-PCI-MT:~/DeepPurpose#
```

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1CONCLUSION

Thus, the proposed system showed how the drug-target interaction prediction to findout the novel drug from the existing approved drug. Today’s drug development is labor incentive, consumes time, hard to infer various type of target interaction and drug repositioning is a very lengthy process so our method very powerful to analysis and separate the intersctions.The main advantages of our work is drug-drug interaction prediction and protein-protein interaction and drug screening for novel drug with good probability ranking list. Our model is analyzed and interaction between the protein sequence and chemical sequence for better prediction.

6.2 FUTURE WORK

The future recommendations of the proposed framework are the following recommendations. Protein molecule structure and chemical alteration sequence is predicted level. Even more features can be extracted from the raw dataset for analysis. This will increase the accuracy of prediction.

CHAPTER 7

SOFTWARE ENVIRONMENT

6.1 SOFTWARE AND HARDWAREREQUIREMENTS

- **Software**

- PY SCRIPT
- PYCHARM
- ANACONDA 3
- JUPYTER
- COLAB
- DEEP LEARNING PACKAGES

REFERENCES

[1] Drug sensitivity has been represented as a link prediction problem. For example, Turki applies hyperlink prediction to most cancers drug sensitivity prediction, and the proposed hyperlink prediction algorithms are extra predictive and solid than current prediction algorithms (Turki and Wei, 2017).

[2] DTI predictions primarily based totally on similarities among protein sequences or drug structures have boundaries for the reason that its underlying assumption that comparable capsules share comparable objectives isn't always always true (Ding et al., 2014). Lee proposed a method for drug repositioning using integrated networks to achieve excellent performance (Lee and Yoon, 2018).

[3] Application of hyperlink prediction approach in heterogeneous networks overcomes the trouble of high function measurement in conventional gadget learning (Stanfield et al., 2017)..

[4] DPI may be expressed withinside the shape of bipartite network, with pills and proteins forming disjoint units of nodes and the interactions among the medicine and proteins forming the edges (Chen et al., 2018; Wu et al., 2018; Ma et al., 2019).

[5] At present, the bipartite community has made big achievements withinside the studies of drug repositioning, drug-sickness affiliation analysis, drug-protein interaction prediction, and gene-sickness affiliation prédiction (Wang et al., 2014; Sun, 2015; Zhang et al., 2017, 2018a, 2019a; Yue et al., 2019).

[6] Zhang proposed an inference approach primarily based totally on community topology similarity to are expecting unobserved drug-sickness associations (?BR42). Cheng proposed a community-primarily based totally inference (NBI) approach that used best the binary similarity of the target's topological community to deduce novel proteins for recognised drugs (?BR8). Zhang proposed a community hyperlink inference approach primarily based totally on linear neighborhood similarity to are expecting miRNA-sickness associations (Zhang et al., 2019b).

[7] Brown AS, Patel CJ. MeSHDD: Literature-primarily based totally drug-drug similarity for drug repositioning. Journal of the American Medical Informatics Association. 2017;24(3):614–618. doi:10.1093/jamia/ocw142.

[8] Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing Learned Molecular Representations for Property Prediction. Journal of Chemical Information and Modeling. 2019;59(8):3370–3388. doi:10.1021/acs.jcim.9b00237

[9] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. Advances in Neural Information Processing Systems 25. Curran Associates, Inc.; 2012. p. 1097–1105. Available from: http://papers.nips.cc/paper/ 4824-imagenet-classification-with-deep-convolutional-neural-networks. pdf.

[10] Lee I, Keum J, Nam H. DeepConv-DTI: Prediction of drug-goal interactions through deep getting to know with convolution on protein sequences. PLOS ComputationalBiology.2019;15(6):e1007129.

[11] Keizer used chemical two-dimensional (2D) structural Frontiers in Bioengineering and Biotechnology | www.frontiersin.org 1 April 2020 | Volume 8 | Article 330 Wang et al. DLS for Predicting Drug-Protein Interactions similarity to are expecting new objectives for recognized tablets and showed that 5 of the 23 new drug target institutions have been valid (Keiser et al., 2009).