Group 4 - Project Report: Enhancing Drug-Drug Interaction Prediction with Graph and Transformer Models

Shraddha Bharadwaj Akkshaya Rajesh Stephen Abhishek Raj

Abstract

Understanding adverse drug reactions and combinations is imperative for the development of pharmaceutical drugs. This project explores the application of Bidirectional Recurrent Neural Network model to predict the effect of the interaction of drugs on the basis of their structure and biological contexts. We employ a multimodal machine learning framework to concatenate target/enzyme relations and chemical structure (via SMILES representations) to predict the biological consequence of the interaction of drugs: serum concentration changes, and metabolic alterations. By selecting DrugBank data (37264 drug pair interactions across 572 drugs), we tested three neural architectures : a simple Multi-Layer Perceptron (MLP), a Bidirectional RNN (BiRNN), and a BiRNN-based Variational Autoencoder (BiRNN-VAE). Feature engineering combines one-hot encoding for protein interactions with structure based ChemBERTa embeddings in our work. The MLP model achieved 0.904 accuracy, dominating in performance over graoh attention networks, and BiRNN models, throwing light on the importance of data scaling and pre-processing for complex models especially in biological contexts. This work demonstrates that protein interaction features along with transformer-based chemical representations of drugs facilitate more confident DDI prediction without the need for model complexity.

1 Introduction:

1.1 Problem Statement:

Drug-drug interactions, or DDIs, occur when two or more drugs interact in a manner that alters their effectiveness, leading to reduced or adverse effects. For this reason, classifying these interactions is imperative in ensuring therapeutic safety and results. They are also a crucial step in the process of drug development, since off-site interactions are a common side effect of drug combinations. Due to the large number of drug combinations, it is not feasible to experimentally test all potential interactions. Computational approaches in the form of ML models have shown promising results in this field, taking advantage of existing pharmacokinetic data.

Transformer models are widely used in all domains of life sciences. Traditional representations of small molecules, such as fingerprints and graph-based embeddings, have been widely used in drug prediction tasks, but they cannot always fully capture the physicochemical properties of molecules. Our work addresses the following gap in research:

- **Multi-modal Integration:** Biological data is not simple enough to be represented by numerical conventions in a single dimension. Taking interactions of multiple biologically potent compounds the complexity skyrockets, since we are presented with multiple physico-chemical properties that act in that instance. In the case of drugs, biological effects the drug has on other proteins is combined with its structural information to predict the consequences of interactions, especially in ex-vivo experimental settings.

1.2 Related Work:

Prior research has been performed on DDI prediction with neural frameworks varying in prediction accuracies due to differences in input filtering and feature processing in a multi-modal biological context. [3]

Deng et. al.'s DDIMDL presented a multimodal deep learning framework whic integrates drug feature selection, extraction concatenating them to predict DDIs. [2] Our work extended this reserach direction by employing transformer based SMILES embeddings and diversifying the classification to three types of interactions. Wang et al.'s BiRNN-DDI model, integrates graph embeddings feeding them into bidirectional RNNs, inspired the BiRNN-VAE architectures in our work. [4]

Label	Interaction Type
0	Adverse Effects
1	Serum Concentration Changes
2	Metabolic Interactions

Table 1: Classification labels and interaction types for drug-drug interactions in this study.

2 Materials and methods:

2.1 Dataset Description

In this work, we use a curated dataset collected from the DrugBank v.5.1.3 database for initial training, hyperparameter tuning, and validation of the model, due to their comprehensive catalog of drug-target information for FDA-approved small molecule, biotech drugs, and experimental drugs.[5] The dataset was originally curated by Deng et al. with tabulated information on 572 drugs and their 74528 interactions along with the corresponding 65 types of events. [2]

The dataset comprises of interrelated tables, including a drug feature table and an event table listing known drug-drug interactions. The drug feature table contains, for each drug, their unique identifiers, names, protein targets, enzymes, biological pathways that it affects, and SMILES representations of chemical structure.

The event table records pairwise drug interactions, specifying the drugs involved, the mechanism of interaction, and a categorical label indicating the type of interaction, such as adverse effects, serum concentration changes, or metabolic effects.

2.2 Feature Engineering

2.2.1 SMILES Reformating

This pipeline started with the generation of SMILES for each drug using pubchempy. The representation the dataset originally came with was of high complexity and hence we resorted to the obtaining the standard notation from the PubChem database.SMILES, or Simplified Molecular Input Line Entry System, is a text-based notation system used to denote the structure of a molecule in the form of a string, using symbols to denote atoms and bond types apart from rings and branches. It is a powerful representation of a molecule in that it can denote 2-D structure of a molecule despite being in 1-D form. Converting SMILES representations into formats that can be understood by machine learning models is a very useful way of categorising structure-function relationship of drugs and drug-like molecules. [1]

2.2.2 Drug Feature Encoding

Three parallel encoding strategies were employed to prepare the data for ML tasks. The target and enzyme ID columns underwent hot encoding using the scikit-learn MultiLabelBinarizer,generating binary vectors to represent drug-protein interactions. SMILES sequences were encoded using ChemBERTa-zinc-base-v1 transformer which was pretrained with 100k SMILES from the ZINC database for chemical compounds, in turn generating 768 contextual embeddings.

2.2.3 Quantification of Similarity

Pairwise Gaussian similarity matrices were calculated and generated separately for all targets,enzymes and SMILES corresponding to all the drugs. The following kernel yielded three distinct similarity matrices for targets,enzymes and the drug SMILE embeddings.

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

Graphical representations of the embeddings were constructed with NetworkX: Three undirected graph2seq representations where nodes represent the drugs and edge weights correspond to the similarity scores from the matrices generated from the feature-space proximity.

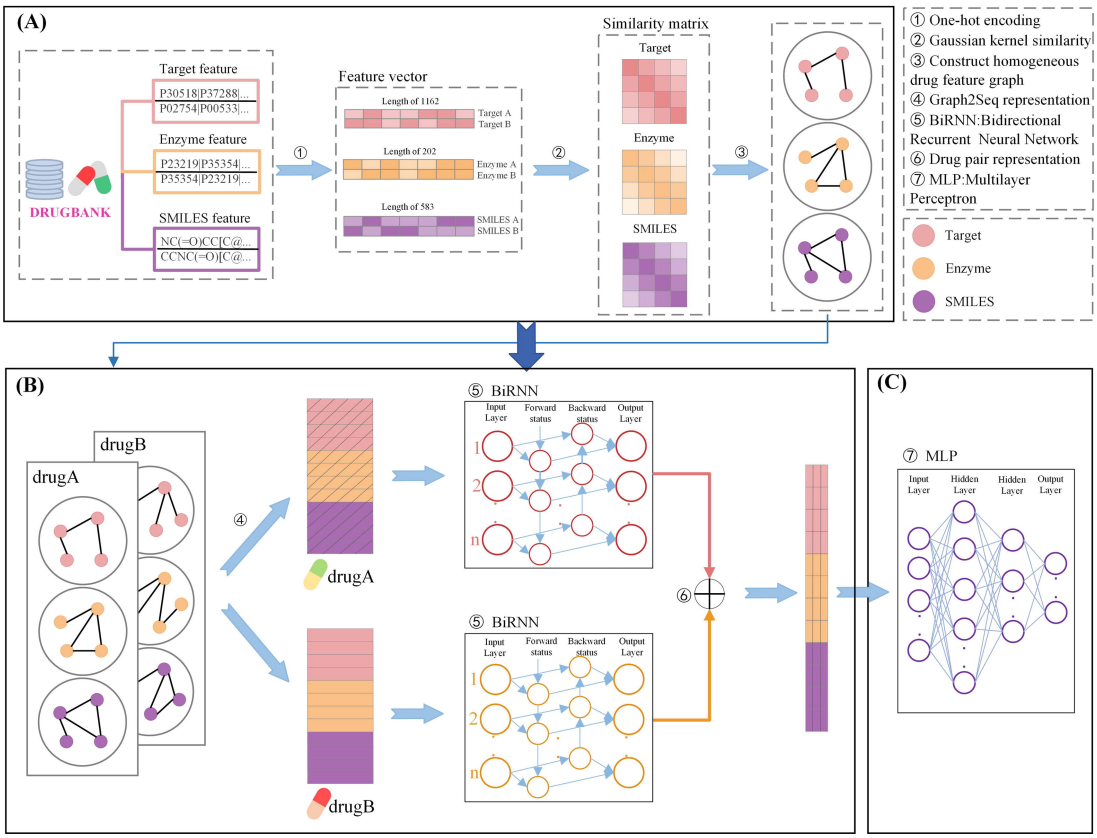


Figure 1: Workflow of BiRNN-DDI model. (A) Drug feature graph construction. (B) Graph2Seq representation learning. (C) Drug-drug interaction event type prediction [4]+.

2.3 Methodology for BiRNN and MLP model

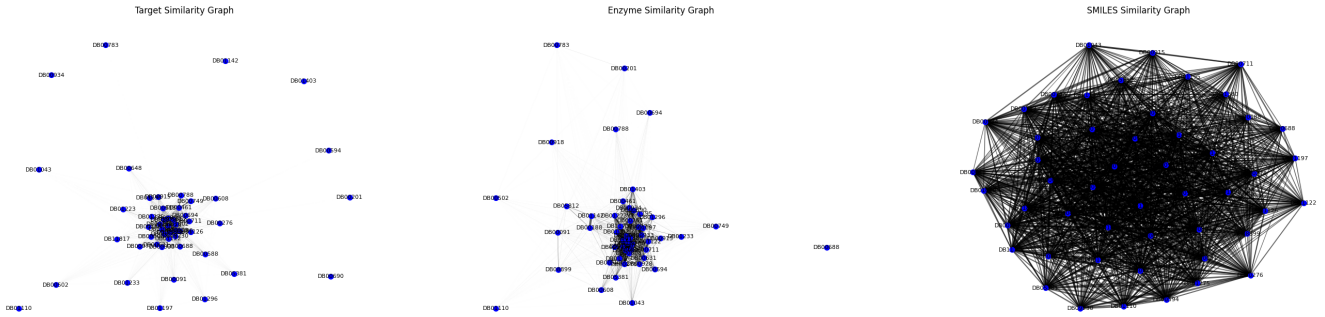
2.3.1 Data Processing and Cleaning

- The reference paper suggested that an important feature to represent a drug is a list of targets, a list of enzymes, and the SMILES of the drug, which was one-hot encoded in the reference paper.
- In data processing, we converted the string mechanism of the event to labels that will further be used for classification, and converted features to numerical enumerations

2.3.2 Feature Extraction and Feature Encoding

- **Targets and Enzymes:** Both were one-hot encoded to create binary vectors where sparseness represents the presence or absence of the molecule in the interaction.

- **SMILES Representations:** The SMILES for each drug was encoded using the "seyonec/ChemBERTa-zinc-base-v1" ChemBERTa Transformer model creating molecular context centric embeddings.Comparative analysis with transformer-based SMILES encoding revealed that this method of feature engineering for SMILES produced denser similarity graphs than one-hot encoding, potentially implying better molecular representation.



(a) Target (b) Enzyme (c) SMILES

Figure 2: Similarity Graphs for target, enzyme and SMILES features generated through NetworkX

2.3.3 Similarity matrix and feature graph construction

For each of the unique 572 drugs in the database, three seperate similaity matrices of dimension 572×572 were constructed with the target, enzymes, and SMILES encoding.From the similarity matrix we filtered out similar pairs with a similarity threshold greater than 0. For each feature,we generated homogenous graph with node as the drug and edge connecting pairs of drugs that have a similarity score greater than 0. This visualization can be found in the implementation of the notebook *Similarity_matrices_and_graphs_alternate.ipynb*

2.3.4 Graph2Seq to create single representation

The aim of this process is to accurately represent a drug with all its features, which also depends on the drug associated with the feature graphs. To capture this relation,we created a vector of dimensions 572×3 for each feature. Edge weight of every neigbouring node was populated in the Graph2Seq encoding of the drug. Finally, each drug was encoded using its relation to other drugs using the homogenous graph representations of target, encoding, and SMILES features, producing a vector of length 1716. This vector was later used downstream by the classification model for every paired input of drugs for supervised learning.

2.3.5 Classification Task

Two architectures were evaluated for the interaction classification task:

- **BirNN Model:**

  Each drug from a pair is an input to a 2-layered bidirectional Gated Recurrent Unit and the output is concatenated which is then an input to an MLP that is used to classify the event type. The GRU outputs (2 × hidden size per drug) were concatenated into a 4 × hidden size vector and fed into an MLP classifier with ReLU-activated layers (512 → 256 → output classes).Backward pass updates losses of GRU and MLP units. Despite k-fold cross-validation, the model plateaued in training and validation accuracy, suggesting that the architecture could be underfitting the data.

- **MLP Model:** We employed a simpler architecture which generated concatenations of paired drug embeddings (2 * hidden_size) which were then passed through the same Multi-Layer Perceptron network included in the previous architecture.This model trained and performed better achieving higher classification accuracy without any repetitive processing units. This could imply that feature-rich context based embeddings, reduce the necessity for more processing.

2.4 Methodology for Alternate GAT model

To compare the performance of the proposed new model, we used real drug similarity graphs(with the edges having biological/chemical meaning), and applied a Graph Attention Network(GAT)-based GNN. This pipeline started with generating SMILES for each drug using pubchempy. Target and enzyme IDs were one-hot encoded, and SMILES were encoded using ChemBERTa (seyonec/ChemBERTa-zinc-base-v1) Transformer embeddings. Pairwise Gaussian similarity matrices were calculated and generated separately for all targets, enzymes and SMILES. We then constructed undirected graph2seq graphs(using networkx) for each of the three columns with weighted edges to gauge their similarity. Following this, we used a GATConv - based GNN with RelU onto the graph2seq vectors in order to effectively train the model on the node features.

The architecture stacks two GAT layers with 512 and 256 hidden units, and attempts to avoid overfitting using dropout regularization. This model processes node embeddings from the three feature graphs.The vectors are then concatenated and fed into a MLP classifier ($512 \rightarrow 256$) with ReLU activation,predicting the type of effect the drug pair interaction produces. (e.g., ”increase” or ”decrease” in serum concentration).

2.5 Experimental parameters setting

We tested different parameter settings in our work, such as training epoch, learning rate, batch size along with K-fold validation and early stopping.

2.6 Evaluation Metrics

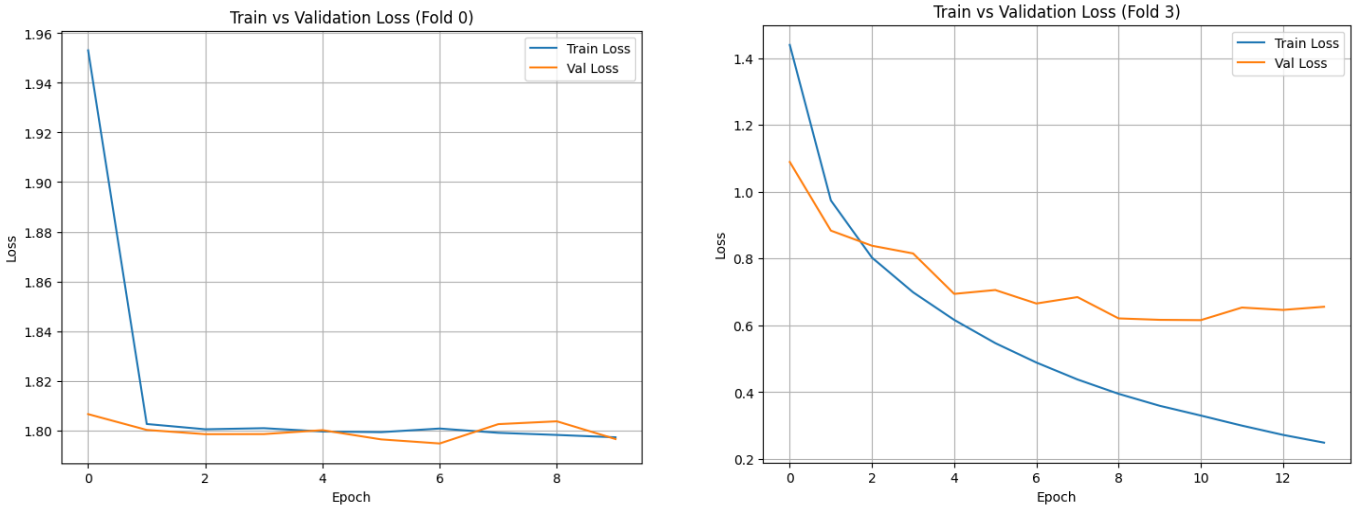
To quantify the effects of the prediction task, we evaluated model performance using Accuracy (ACC), Area Under the Precision-Recall Curve (AUPR), Area Under the ROC Curve (AUC), F1 Score, Precision, and Recall. We will compare the traditional deep learning approach, the BiRNN-DDI model, and the proposed representation learning method integrated into BiRNN-DDI. This comparison will highlight the impact of enhanced drug representation on DDI event-type prediction, demonstrating improvements in capturing structural and contextual relationships.

3 Results:

3.1 Evaluation of model performance:

The MLP model trained on the concatenated drug embeddings produced using one-hot encoded target/enzymes and ChemBERT-a encoded SMILES features graphs demonstrated superior performance in DDI classification. The results can be found in mlp.ipynb files. The performance noted is comparable to the BiRNN reference paper, implying that the representation learning methodology borrowed from the paper has resulted in good results even with a simple MLP model.

In contrast, the alternate GAT model trained on the Gaussian similarity matrices with ChemBERTa-SMILES embeddings underperformed only exhibiting a testing accuracy of 0.4771.



(a) BiRNN model (Fold 0)

(b) MLP model (Fold 3)

Figure 3: **Comparison of training and validation loss curves between BiRNN (left) and MLP (right) models for Drug-Drug Interaction (DDI) type prediction.** The BiRNN model shows overfitting with divergence between training and validation loss as epochs progress, with training loss decreasing to approximately 0.25 while validation loss plateaus around 0.65. In contrast, the MLP model demonstrates better generalization with closely aligning training and validation curves that settle around 1.80, showing better predictive performance on unknown data despite higher absolute loss.

Metric	Value
Accuracy	0.9042
Precision	0.9042
Recall	0.9016
F1 Score	0.9013

Table 2: MLP Model Performance

Metric	Value
Accuracy	0.4771
ROC AUC	0.6025
AUPR	0.0251
Precision	0.0085
Recall	0.0179
F1 Score	0.0115

Table 3: GAT Model Performance

3.2 Discussion:

3.2.1 Limitations:

- We observed that the model from the reference paper was not able to learn from the representation. The divergence of performance from the ideal metrics could be due to:
  - **Data Limitations:** If the subset of data we chose did not have enough complex patterns and hence the data was underfitting on the complex model.
  - **Graph Sparsity:** Since the reference paper used one hot encoding to encode all features it might have created a more balanced representation vector as opposed to our method where the SMILES similarity graph was much dense, in the future, we would want to validate this theory.
  - The size of the data we used was much less which also might have resulted in a more complex model failing. Results can be seen in BIRNN.ipynb. A combination of representation learning was tested on the model.

3.2.2 Methodological Insights

- **Class Imbalance:** Metabolism-related interactions (59 percent of labels) dominated predictions, producing a 12 percent F1 drop in minority classes (adverse effects).
- **GPU Bottlenecks:** The BiRNN-VAE demanded 3.8 hrs/epoch on an A100 GPU vs. 1.2 hrs for MLP. This begs the question of prioritizing computational cost of latent space regularization compared to other processing steps.

3.2.3 Future Work

- **Data Augmentation:** Employing less computational generative VAEs to synthesize minority-class interactions.
- **Unified Embeddings:** Trying multi-view fusion to harmonize biological, chemical, and graph features as opposed to individual graphs.
- **Simplified Architectures:** Exploring MFPS-based models for scenarios with limited chemical diversity.

4 Code submission:

Code Repository - https://github.com/sbharadwaj5/drug_drug_interaction_classification

5 Contributions:

Each team member contributed as follows (see Table 4)

Team Member	Contributions
Akkshaya Rajesh	Dataset preprocessing and feature engineering
Shraddha Bharadwaj	Transformer Models and Representation Learning Integration, BiRNN implementation , MLP model implementation
Stephen Abhishek Raj	SMILES generation, batch encoding and GAT model

Table 4: Contributions of Team Members

References

[1] Chithrananda, S., Grand, G., and Ramsundar, B. (2020). Chemberta: Large-scale self-supervised pretraining for molecular property prediction.

[2] Deng, Y., Xu, X., Qiu, Y., Xia, J., Zhang, W., and Liu, S. (2020). A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics*, 36(15):4316–4322.

[3] Gan, Y., Liu, W., Xu, G., Yan, C., and Zou, G. (2023). Dmfddi: deep multimodal fusion for drug–drug interaction prediction. *Briefings in Bioinformatics*, 24(6):bbad397.

[4] Wang, G., Feng, H., and Cao, C. (2025). Birnn-ddi: A drug-drug interaction event type prediction model based on bidirectional recurrent neural network and graph2seq representation. *Journal of Computational Biology*, 32(2):198–211. PMID: 39049806.

[5] Wishart, D., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., and Hassanali, M. (2008). Drugbank: a knowledgebase for drugs, drug actions and drug targets. nucleic acids res 36:d901-d906. *Nucleic acids research*, 36:D901–6.