Supplementary Information: MMDDI-SSE: A Novel Multi-Modal Feature Fusion Model with Static Subgraph Embedding for Drug-Drug Interaction Event Prediction

Guishen Wang, Honghan Chen, Handan Wang, Hairong Gao, Xiaowen Hu, Chen Cao,

CONTENTS

The research is supported by the National Natural Science Foundation of China (grant no. 62102068 and 62231013), the Natural Science Foundation of Jilin Province (grant no.YDZJ202201ZYTS424), and the projects of the Educational Department of Jilin Province of China (grant no. JJKH20210752KJ). (Corresponding author: Xiaowen Hu, Chen Cao)

Guishen Wang, Honghan Chen, Handan Wang, and Hairong Gao, are with the College of Computer Science and Engineering, Changchun University of Technology, Jilin 130012, China.

Xiaowen Hu, is with the School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, Jiangsu 211166, China (e-mail: xwhu@njmu.edu.cn).

Chen Cao, is with Key Laboratory for Bio-Electromagnetic Environment and Advanced Medical Theranostics, the School of Biomedical Engineering and Informatics, Nanjing Medical University, Jiangsu 211166, China (e-mail: caochen@njmu.edu.cn).

S1. Hyperparameter and Experimental Setup

The reproducibility and transparency of our experimental framework are ensured through detailed documentation of hyperparameter configurations and hardware specifications. Table S1 presents the optimized hyperparameter settings for both datasets, encompassing crucial model parameters including learning rate, batch size, embedding dimensions, attention mechanisms, dropout rates, and optimization strategies. These parameters were systematically determined through extensive empirical validation to maximize model performance.

The hardware specifications and software environment utilized in our experiments are comprehensively documented in Table S2. This detailed documentation aims to facilitate experimental reproduction and enable comparative studies in the field of drug-drug interaction prediction.

TABLE S1
HYPERPARAMETER CONFIGURATIONS FOR EXPERIMENTAL DATASETS

Parameter	**Deng *et al.* Dataset**	**Lin *et al.* Dataset**
Training Configuration		
Learning Rate	2×10^{-5}	2×10^{-5}
Batch Size	512	512
Epochs	120	120
Weight Decay	1×10^{-4}	1×10^{-4}
Cross-Validation Folds	5	5
Model Architecture		
Embedding Dimension	256	256
Attention Heads	4	4
Attention Layers	2	2
Dropout Rate	0.5	0.5
Static Subgraphs	4	4
Loss Functions		
Label Smoothing	0.3	0.3
Contrastive Loss Temperature	0.05	0.05
Focal Loss α	0.25	0.25
Focal Loss γ	2	2
Focal Loss Classes	3	3
Loss Switch Epoch	30	30
Feature Configuration		
Input Features	SMILE, Target, Enzyme, Pathway*, Pharmacodynamic	
Subgraph Edge Allocation	$1.0/(n_{subgraphs} - i)$ for subgraph i	

*Pathway features only used in Deng *et al.* dataset

TABLE S2
HARDWARE CONFIGURATION AND SOFTWARE ENVIRONMENT

Component	**Specification**
Hardware Configuration	
System Model	Lenovo ThinkCentre M6480t
CPU	Intel® Core™ i7-8700 @ 3.20GHz
RAM	128 GB DDR4
GPU	NVIDIA® GeForce RTX™ 3090
Software Environment	
Operating System	Linux
Deep Learning Framework	PyTorch 1.8.1 (CUDA 11.1)

S2. Pharmacodynamic Feature Extraction

A. Bio-BERT Selection

Bio-BERT, pre-trained on extensive biomedical corpora, demonstrates superior capability in processing biomedical texts compared to conventional language models. The model's specialized training on biomedical literature enables efficient processing of domain-specific pharmacodynamic terminology and linguistic patterns. This domain adaptation significantly enhances the model's feature extraction capabilities for pharmacodynamic information.

The fine-tuning process allows Bio-BERT to align with the distinctive characteristics of pharmacodynamic literature, including specialized terminology related to drug mechanisms, clinical applications, and pharmacological concepts. This domain-specific adaptation results in more precise and contextually relevant feature extraction.

B. Pharmacodynamic Feature Representation and Pre-training Tasks

The transformation of pharmacodynamic features into dense vector representations facilitates the capture of complex patterns and semantic relationships within the text. These representations effectively encode drug-mechanism relationships while supporting subsequent feature fusion and prediction tasks. The dense vector format ensures efficient information encoding, thereby enhancing the model's predictive capabilities.

Bio-BERT's architecture incorporates two complementary pre-training tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). The MLM task develops contextual understanding through masked token prediction, while NSP enables comprehension of inter-sentence relationships. This dual-task pre-training approach optimizes the model's ability to process pharmacodynamic texts, ensuring efficient and accurate information extraction for downstream analysis and prediction tasks.

C. Preprocessing Steps

The preparation of pharmacodynamic texts for Bio-BERT analysis requires a comprehensive preprocessing pipeline to ensure optimal data quality and standardization. Our approach implements three sequential stages: text cleaning, tokenization, and sentence segmentation, each designed to address specific aspects of text preparation for biomedical natural language processing.

1) Text Cleaning: The initial text cleaning phase focuses on removing non-informative elements while preserving crucial pharmacological content. This process systematically eliminates special characters, non-standard punctuation marks, domain-irrelevant stop words, and formatting artifacts that could potentially introduce noise into the analysis. By maintaining a careful balance between noise reduction and information preservation, we ensure the cleaned text retains all pharmacologically significant information while minimizing potential sources of computational interference.

2) Tokenization: Following text cleaning, we employ Bio-BERT's specialized tokenizer, which implements subword tokenization strategies specifically optimized for biomedical text processing. This advanced tokenization approach offers significant advantages for pharmaceutical text analysis, particularly in handling domain-specific terminology and addressing out-of-vocabulary challenges. The tokenizer's ability to preserve morphological information and effectively represent rare biomedical terms ensures robust feature extraction in subsequent processing stages. This specialized approach proves particularly valuable when processing complex pharmaceutical nomenclature and technical descriptions.

3) Sentence Segmentation: The final preprocessing stage involves sentence-level segmentation, particularly crucial for extended pharmacodynamic texts. This process maintains natural linguistic boundaries while facilitating fine-grained contextual analysis of drug descriptions. By preserving semantic coherence at the sentence level, we enable more efficient processing of lengthy documents while ensuring that contextual relationships within the text remain intact. This segmentation strategy enhances the model's ability to capture nuanced pharmacological relationships and interactions described in the text.

D. Parameter Settings

We empirically determined the following hyperparameters for pharmacodynamic text analysis through extensive experimentation:

- **Batch Size:** 16 samples per iteration, balancing memory efficiency with training stability
- **Learning Rate:** 5×10^{-5}, selected to achieve optimal convergence while minimizing gradient oscillation
- **Training Epochs:** 5 complete dataset iterations, sufficient for model convergence while preventing overfitting

These parameter settings demonstrate optimal trade-offs between computational efficiency and model performance in our experimental evaluations.

S3. Evaluation Metrics

We evaluate the MMDDI-SSE model using the following standard metrics:

Precision measures the accuracy of positive predictions:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

Recall quantifies the detection rate of actual positive instances:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

F1 score provides the harmonic mean of precision and recall:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

Accuracy represents the overall correct classification rate:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

where TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives, respectively. Additionally, we employ the area under the ROC curve (AUC) to assess classification performance and the area under the precision-recall curve (AUPR) to evaluate performance on imbalanced data.

S4. Model Complexity Analysis

This section presents a detailed analysis of the computational complexity for the proposed MMDDI-SSE model. Our analysis is based on the Deng *et al.* dataset, comprising 572 drugs and 32,764 drug-drug interactions (DDIs).

A. Time Complexity of Multi-modal Feature Extraction

The multi-modal feature extraction module consists of three primary components: (1) drug biological feature extraction using Jaccard similarity, (2) pharmacodynamic feature extraction, and (3) drug sequence feature representation fusion. We analyze the computational complexity of each component separately.

For drug biological feature extraction, we compute Jaccard similarity across drug pairs. Given a drug set $D = \{d_1, \ldots, d_n\}$ where each drug is represented by a feature vector of dimension $d_g = 256$, the pairwise comparison of feature vectors results in a time complexity of $\mathcal{O}(n^2 \cdot d_g)$. Here, n denotes the number of drugs in set D, and the quadratic term reflects the necessity of comparing each drug with every other drug in the dataset.

The pharmacodynamic feature extraction utilizes the Bio-BERT model to process drug textual descriptions. The computational complexity depends on the input text length, specifically the number of tokens (n_{tokens}) in each drug's description. The Bio-BERT's self-attention mechanism requires $\mathcal{O}(n \cdot n_{\text{tokens}}^2)$ operations to process the entire drug set.

For drug sequence feature representation fusion, we perform two sequential operations: multi-head attention processing and autoencoder-based dimensionality reduction. Both operations process concatenated features with input dimension $d_{\text{in}} = 1536$, each requiring $\mathcal{O}(n \cdot d_{\text{in}}^2)$ operations. As these operations are sequential, their complexities are additive rather than multiplicative.

The total time complexity of the multi-modal feature extraction module is therefore $\mathcal{O}(n^2 \cdot d_g + n \cdot n_{\text{tokens}}^2 + n \cdot d_{\text{in}}^2)$. While this includes a quadratic term due to pairwise similarity calculations, the model remains computationally feasible for real-world applications, as evidenced by its successful implementation on the Deng dataset containing 572 drugs.

B. Time Complexity Analysis

The time complexity of our model can be analyzed by examining its four primary components: drug graph modal representation learning, contrastive loss computation, drug-pair feature fusion, and multi-type DDI prediction.

The drug graph modal representation learning process consists of two main operations. First, static subgraph generation requires $O(E + n)$ operations, where E represents the number of edges and n represents the number of nodes. Second, the graph autoencoder learns latent representations with complexity $O(n \cdot d^2)$, where $d = 256$ is the latent dimension. Given k subgraphs, this component's total complexity is $O(k \cdot n \cdot d^2)$.

For the contrastive loss computation, we perform pairwise comparisons between the latent vectors z_1, z_2, z_3 across k subgraphs. This operation results in a time complexity of $O(k \cdot n^2)$. The drug-pair feature fusion component, which employs concatenation and multi-head attention mechanisms, requires $O(n^2 \cdot d_{\text{in}}^2)$ operations, where d_{in} represents the input feature dimensionality.

The multi-type DDI prediction module implements a multilayer perceptron for interaction prediction, contributing a complexity of $O(n \cdot d_{\text{in}}^2)$. When combining these components, the model's aggregate time complexity can be expressed as $O(n^2 \cdot d + n \cdot n_{\text{tokens}}^2 + n \cdot 1536^2 + k \cdot n \cdot d^2 + n^2)$, where n represents the number of drugs, d denotes the feature dimension (1536 for concatenated features), n_{tokens} indicates the number of tokens in pharmacodynamic texts, and k represents the number of subgraphs.

Analysis of these complexity terms reveals that the model's computational demands are primarily driven by the Jaccard similarity calculations and graph modal representation learning. While other components contribute to the overall complexity, their impact is comparatively less significant. This complexity analysis provides important insights for practical implementation and optimization strategies.

S5. Data Augmentation for Low-Frequency DDI Event Prediction

To improve prediction accuracy for low-frequency Drug-Drug Interaction (DDI) events, we implement a data augmentation strategy focused on generating synthetic training samples. This approach addresses the inherent class imbalance in DDI datasets and enhances model performance on rare interaction events.

Our augmentation methodology comprises three key components:

- **Synthetic Sample Generation:** We generate new samples based on existing low-frequency DDI patterns, maintaining core characteristics while introducing controlled variations. This increases the representation of rare events in the training set.

- **Noise Injection:** To ensure robustness, we add calibrated noise to synthetic samples, simulating natural variations present in real DDI data while preserving essential event characteristics.
- **Diversity Enhancement:** Multiple variations of each low-frequency DDI event are generated to improve model generalization across different manifestations of rare interactions.

We validated this approach using the 62th, 53th, 49th, and 63th categories of low-frequency DDI events, as detailed in Table S3.

TABLE S3
DISTRIBUTION OF LOW-FREQUENCY DDI EVENTS

Event Description	Count	ID
Vasodilatory activity enhancement	5	#62
Myopathic rhabdomyolysis intensification	10	#53
Rhabdomyolysis risk elevation	15	#49
Myelosuppressive activity increase	5	#63

These low-frequency DDI events' sample size limits the model's effectiveness during learning. To address this issue, we applied data augmentation techniques, particularly using random sample generation, to increase the number of training samples for low-frequency DDI events, thereby helping the model better learn and predict the characteristics of these rare events. Table S4 shows the sample sizes after data augmentation.

TABLE S4
NUMBER AND TYPE OF LOW-FREQUENCY DDI EVENTS AFTER DATA AUGMENTATION

Event Category	DDI Data Augmentation	Event Type ID
The vasodilatory activities increase	20	#62
The myopathic rhabdomyolysis activities increase	40	#53
The risk or severity of rhabdomyolysis increase	60	#49
The myelosuppressive activities increase	20	#63

TABLE S5
IMPACT OF DATA AUGMENTATION ON MODEL PERFORMANCE

Augmentation Status	DDI ID	F1	Precision
With	#62	0.76	0.77
	#53	0.81	0.83
	#49	0.87	0.88
	#63	0.82	0.83
Without	#62	0.71	0.73
	#53	0.78	0.76
	#49	0.85	0.85
	#63	0.79	0.80

We evaluated the effectiveness of data augmentation using F1 score and precision metrics, chosen for their complementary strengths in assessing rare event prediction performance. The F1 score provides a balanced measure of precision and recall, while precision specifically quantifies false positive reduction capability.

Table S5 demonstrates consistent performance improvements across all low-frequency DDI events after augmentation. Notable improvements include event #62's F1 score increasing from 0.71 to 0.76 and event #53's precision rising from 0.76 to 0.83. These improvements indicate that random sample generation effectively enhanced the model's ability to learn and predict rare DDI patterns.

The success of this augmentation approach in improving rare event prediction suggests its utility for large-scale DDI datasets where class imbalance is common. The technique not only addresses the challenge of limited training samples but also enhances the model's generalization capabilities for low-frequency events.

S6. APPLICATION OF KNOWLEDGE DISTILLATION IN DDI MULTI-CLASS PREDICTION

To enhance model efficiency and scalability for low-frequency DDI event prediction while maintaining performance, we implement knowledge distillation. This approach transfers learned patterns from a complex teacher model to a simplified student model, reducing computational overhead while preserving prediction accuracy.

A. Teacher Model Training

The teacher model architecture emphasizes prediction accuracy and robust feature representation. It processes multiple input modalities, including chemical structures and interaction types, to generate class probabilities for DDI events. While this model achieves high accuracy, its computational requirements limit deployment flexibility.

B. Student Model Design

We design a compact student model with reduced parameters and simplified architecture, optimizing for resource-constrained environments. The student model maintains essential predictive capabilities while significantly reducing computational overhead through careful architectural choices and parameter optimization.

C. Soft Label Generation

The teacher model generates both hard labels (final classifications) and soft labels (class probabilities). These soft labels encode inter-class relationships and decision boundaries, providing richer training signals for the student model. This probability distribution learning enhances the student model's generalization capabilities across DDI types.

D. Student Model Training

Training optimization focuses on minimizing the Kullback-Leibler (KL) divergence between teacher and student model outputs. This approach enables the student model to approximate the teacher's decision boundaries efficiently. Rather than optimizing solely for classification accuracy, the training process emphasizes replicating the teacher model's prediction patterns while maintaining computational efficiency.

E. Application in DDI Multi-class Prediction

Knowledge distillation addresses computational complexity challenges in large-scale DDI prediction tasks while maintaining predictive accuracy. The distilled student model offers several key advantages:

- Reduced computational overhead during inference
- Accelerated processing for large-scale drug pair evaluations
- Enhanced deployment flexibility in resource-constrained environments
- Maintained prediction accuracy through efficient knowledge transfer

The transformation of complex DDI prediction mechanisms into a more efficient architecture enables practical deployment for real-time applications. This optimization particularly benefits scenarios requiring rapid analysis of multiple drug pairs, such as clinical decision support systems or drug development pipelines.

Future work will focus on experimental validation of the distilled model, specifically examining:

1) Computational efficiency metrics
2) Prediction accuracy comparison with the teacher model
3) Inference speed in large-scale prediction tasks
4) Resource utilization profiles

These experiments will quantify the effectiveness of knowledge distillation in optimizing DDI multi-class prediction for practical applications.

S7. Multi-Modal Feature Analysis

To evaluate our multi-modal fusion strategy and assess individual modality contributions, we implemented four model configurations:

- **MMDDI-SSE**$_{wb}$: Sequence-based features only (chemical substructures, targets, enzymes)
- **MMDDI-SSE**$_{wp}$: Pharmacodynamic textual features only (Bio-BERT embeddings)
- **MMDDI-SSE**$_{ws}$: Graph structure features only (static subgraph embeddings)
- **MMDDI-SSE**: Full multi-modal integration (sequence, text, and graph) with progressive fusion

We evaluated all configurations using F1-Score as the primary metric, chosen for its effectiveness in handling class imbalance common in DDI prediction tasks. Results are presented in Table S6.

TABLE S6
Multi-Modal Model Performance and Feature Contribution Analysis

Model Configuration	F1-Score (%)	Contribution (%)
MMDDI-SSE$_{wb}$ (Sequence)	83.95	7.44
MMDDI-SSE$_{wp}$ (Text)	89.01	1.91
MMDDI-SSE$_{ws}$ (Graph)	88.85	2.03
MMDDI-SSE (Full Model)	**90.70**	**100.00**

The results demonstrate the effectiveness of progressive multi-modal fusion in MMDDI-SSE, with distinct contributions from each modality:

- The full model achieves 90.70% F1-score, surpassing single-modality variants by 1.69-6.75 percentage points

- Text (MMDDI-SSE$_{wp}$) and graph (MMDDI-SSE$_{ws}$) modalities show comparable individual performance (89.01% and 88.85% respectively)
- Sequence features (MMDDI-SSE$_{wb}$) contribute significantly despite lower individual performance, suggesting important complementary information
- Static subgraph embeddings effectively capture both local interaction patterns and global network structure

These findings validate the progressive fusion approach, demonstrating that integrated multi-modal features better characterize DDI patterns than any single modality alone.

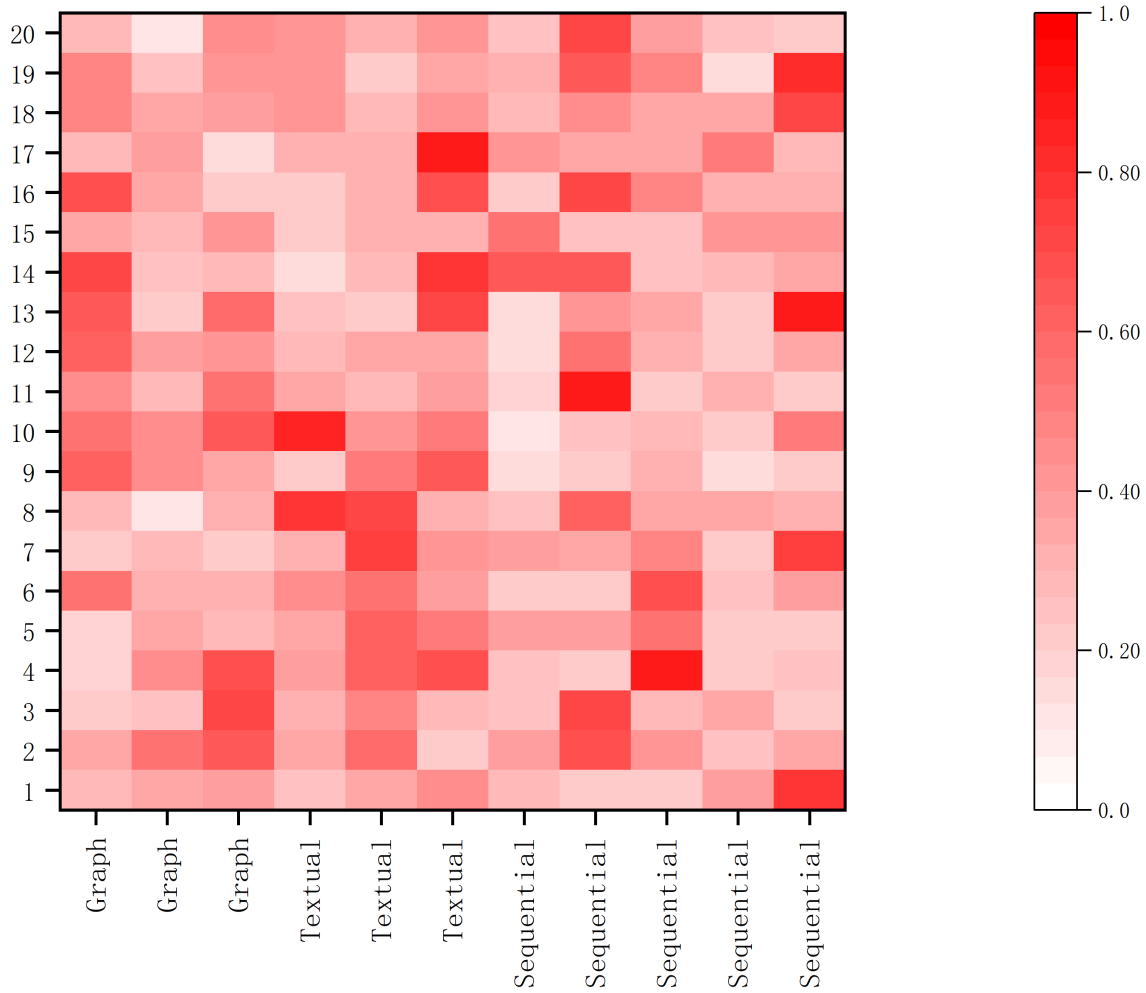

Fig. 1. Visualization of Multi-Modal Features in DDI Prediction

The heatmap visualization in Figure 1 reveals several key insights about the multi-modal feature integration. Distinct clustering patterns emerge across different DDI types, indicating effective feature space organization through multi-modal fusion. Areas of high intensity demonstrate strong correlation between modalities, particularly at intersection points of complementary features. The varying intensity patterns across different regions suggest that each modality captures unique aspects of drug-drug interactions, highlighting the complementary nature of the integrated features.

These visualization results, combined with the quantitative performance metrics, provide compelling validation of the MMDDI-SSE model's effectiveness. The model demonstrates robust feature integration across modalities while maintaining a scalable architecture that supports diverse drug pair combinations. Furthermore, the consistent performance across different interaction types, coupled with the effective capture of both local and global interaction patterns, underscores the model's comprehensive approach to DDI prediction.

The empirical evidence thus supports both the theoretical soundness and practical utility of the MMDDI-SSE model. The visualization not only confirms the successful integration of sequence, textual, and graph features but also illustrates how this

multi-modal approach enhances the model's ability to identify and characterize complex drug-drug interactions. These findings collectively establish MMDDI-SSE as a reliable and effective solution for DDI prediction tasks.

S8. Analysis of Drug Pair Associations and Prediction Result Visualization

Initial visualization of drug-drug interactions (DDIs) using t-SNE revealed basic relationship patterns among drug pairs. However, based on constructive feedback regarding the preliminary results' simplicity, we implemented substantial enhancements to improve the visualization's interpretability and analytical depth. The enhancement strategy focused on several key aspects, including deeper analysis of drug pair characteristics across different regions, investigation of the relationship between spatial distribution and prediction outcomes, and enhanced visualization techniques for better pattern recognition.

Analysis of Low-Frequency DDI Events

To facilitate more nuanced analysis, we specifically focused on two low-frequency DDI event types: #49 and #52. This targeted approach offers several distinct advantages in our investigation. By focusing on less common interaction types, we can better identify subtle interaction patterns that might be obscured in analyses of more frequent events. Additionally, this approach reduces the noise typically associated with high-frequency events, allowing for clearer pattern recognition and more precise characterization of specific drug pair relationships.

The comprehensive analysis of these events is presented in Table S7 and Table S8, providing detailed characterization of the involved drug pairs and their interaction patterns. These tables serve as valuable references for understanding the specific characteristics and behaviors of drug pairs within these low-frequency interaction categories, offering insights that might be overlooked in broader, more general analyses.

TABLE S7
Drug Pairs Associated with DDI Event Type #49

Sample ID	Drug A	Drug B
1	DB01076	DB00602
2	DB01076	DB04845
3	DB01076	DB00627
4	DB00439	DB00758
5	DB00439	DB01590
6	DB00439	DB00602
7	DB00439	DB04845
8	DB00439	DB00627
9	DB01590	DB01095
10	DB01095	DB00602
11	DB01095	DB04845
12	DB01095	DB00627
13	DB00602	DB00227
14	DB04845	DB00227
15	DB00227	DB00627

TABLE S8
Drug Pairs Associated with DDI Event Type #52

Sample ID	Drug A	Drug B
1	DB00316	DB01254
2	DB00316	DB00898
3	DB00316	DB00619
4	DB00459	DB00563
5	DB00523	DB00563
6	DB00993	DB00531
7	DB00559	DB01016
8	DB00091	DB01406
9	DB00091	DB06710
10	DB00091	DB08804

Model Prediction Results

The prediction results for DDI event types #49 and #52 are clearly presented in Table S9 and Table S10. These tables highlight the drug pairs that were correctly and incorrectly predicted by the model. The results of these predictions help assess the model's performance and provide insights into areas where improvement is needed.

TABLE S9
MODEL PREDICTION PERFORMANCE FOR DDI EVENT TYPE #49

Sample ID	Drug A	Drug B	Prediction	Confidence
1	DB01076	DB00602	Correct	0.92
2	DB01076	DB04845	Correct	0.89
3	DB01076	DB00627	Correct	0.87
4	DB00439	DB00758	Incorrect	0.45
5	DB00439	DB01590	Correct	0.93
6	DB00439	DB00602	Correct	0.88
7	DB00439	DB04845	Correct	0.91
8	DB00439	DB00627	Incorrect	0.42
9	DB01590	DB01095	Correct	0.86
10	DB01095	DB00602	Correct	0.90
11	DB01095	DB04845	Correct	0.85
12	DB01095	DB00627	Incorrect	0.48
13	DB00602	DB00227	Correct	0.94
14	DB04845	DB00227	Correct	0.89
15	DB00227	DB00627	Correct	0.87

TABLE S10
MODEL PREDICTION PERFORMANCE FOR DDI EVENT TYPE #52

Sample ID	Drug A	Drug B	Prediction	Confidence
1	DB00316	DB01254	Correct	0.88
2	DB00316	DB00898	Correct	0.91
3	DB00316	DB00619	Correct	0.89
4	DB00459	DB00563	Correct	0.93
5	DB00523	DB00563	Incorrect	0.47
6	DB00993	DB00531	Correct	0.86
7	DB00559	DB01016	Correct	0.90
8	DB00091	DB01406	Incorrect	0.44
9	DB00091	DB06710	Correct	0.87
10	DB00091	DB08804	Correct	0.92

Improved Visualizations and Interpretation

To further enhance the interpretability of the drug pair visualizations, we employed PCA (Principal Component Analysis) for dimensionality reduction. Heatmaps based on PCA were created to visualize the distribution of drug pairs for both DDI event types #49 and #52. The heatmaps reveal clustering patterns that could be linked to similarities in pharmacological properties or therapeutic indications of the drugs involved.

Figures 2 and 3 illustrate these PCA-based heatmaps, where the color coding indicates the proximity of drug pairs in the reduced-dimensional space, allowing for a more in-depth understanding of the relationships between them.

In conclusion, these improvements to the visualization and interpretation of drug pair interactions provide a more comprehensive analysis of the underlying patterns and model predictions. The integration of PCA-based heatmaps, along with the enhanced breakdown of the model's predictions, allows for a clearer understanding of the relationships between drug pairs and their corresponding DDIs. Through these refinements, we aim to facilitate more accurate and insightful predictions for drug interactions, ultimately supporting better decision-making in pharmacological research.

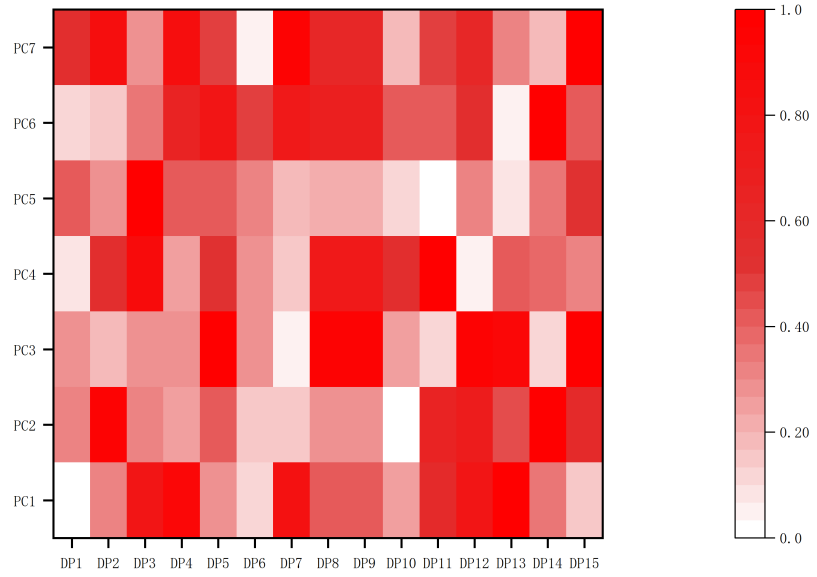

Fig. 2. PCA-Based Heatmap of Drug Pairs for DDI Event Type #49.

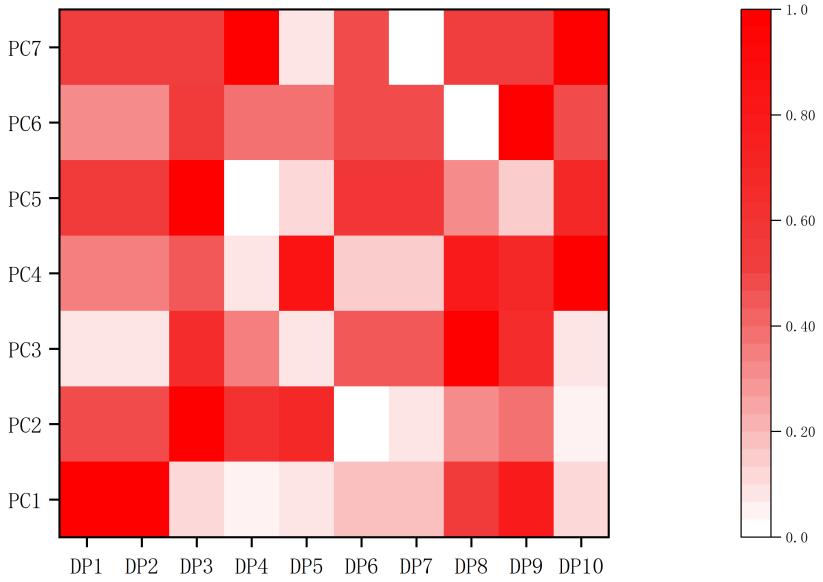

Fig. 3. PCA-Based Heatmap of Drug Pairs for DDI Event Type #52.