

RESEARCH ARTICLE

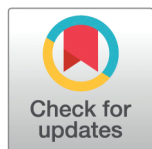
A meta-contrastive learning approach for clinical drug-drug interaction extraction from biomedical literature

Yaxun Jia¹, Zhu Yuan², Lian Zhu¹, Zuo-lin Xiang^{1,3*}

1 Department of Radiation Oncology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China, **2** Department of Information Management, The National Police University for Criminal Justice, Baoding, China, **3** Department of Radiation Oncology, Shanghai East Hospital Ji'an Hospital, Jian, China

☞ These authors contributed equally to this work.

* xiangzuolinmd@hotmail.com



OPEN ACCESS

Citation: Jia Y, Yuan Z, Zhu L, Xiang Z (2025) A meta-contrastive learning approach for clinical drug-drug interaction extraction from biomedical literature. PLoS Comput Biol 21(12): e1013722. <https://doi.org/10.1371/journal.pcbi.1013722>

Editor: Qiangguo Jin, Northwestern Polytechnical University, CHINA

Received: August 5, 2025

Accepted: November 7, 2025

Published: December 5, 2025

Copyright: © 2025 Jia et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All our code and data have been publicly released at: <https://github.com/Hero-Legend/BioMCL-DDI>.

Abstract

Drug–drug interactions (DDIs) are a significant source of adverse drug events and pose critical challenges to patient safety and clinical decision-making. Extracting DDIs from biomedical literature plays an essential role in pharmacovigilance, yet remains difficult due to data sparsity and high annotation costs. This study presents BioMCL-DDI, a novel few-shot learning framework that integrates meta-learning with contrastive embedding strategies to enable efficient DDI extraction under limited supervision. BioMCL-DDI jointly optimizes prototype-based classification and supervised contrastive representation learning within a unified architecture. The model captures both intra-class compactness and inter-class separability, enhancing its generalization in sparse biomedical settings. We evaluate BioMCL-DDI on three benchmark datasets: DDI-2013, DrugBank, and the more recent TAC 2018 DDI Extraction corpus. The model achieves F1 scores of 87.80% on DDI-2013, 86.00% on DrugBank, and 74.85%/74.82% on the two official test sets of TAC 2018, consistently outperforming competitive baselines. Our model significantly outperforms state-of-the-art baselines in low-resource scenarios. BioMCL-DDI provides a scalable and effective solution for DDI extraction from biomedical texts, with strong potential for integration into clinical decision support systems and biomedical knowledge bases. All our code and data have been publicly released at: <https://github.com/Hero-Legend/BioMCL-DDI>.

Author summary

Drug-drug interactions (DDIs) are a significant source of adverse drug events and pose critical challenges to patient safety. While existing drug databases offer valuable DDI information, they often struggle to keep pace with the rapidly

Funding: This work is supported by National Natural Science Foundation of China (82160591 to Z.I.X.), Key Project of Clinical Research of Shanghai East Hospital, Tongji University (DFLC2022012 to Z.I.X.), Key Specialty Construction Project of Shanghai Pudong New Area Health Commission (PWZzk2022-02 to Z.I.X.), Funded by Outstanding Leaders Training Program of Pudong Health Bureau of Shanghai (PWR12023-02 to Z.I.X.) and Shanghai Science Technology Innovation Action Plan (23Y11909000 to Z.I.X.), Science Research Project of Hebei Education Department (QN2025011 to Z.Y.) and Doctoral Research Start-up Fund Program of The National Police University for Criminal Justice (BSQDW202150 to Z.Y.). Zuo-lin Xiang is the corresponding author of this paper. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

expanding biomedical literature, leaving many new interactions unaddressed. To tackle this challenge, we developed BioMCL-DDI, a lightweight meta-contrastive learning framework designed for efficient DDI extraction from biomedical texts, especially in scenarios with sparse data. Our method integrates prototype-based classification with contrastive embedding strategies, jointly optimizing them within a unified architecture. This allows our model to generalize effectively even with limited annotated data. We evaluated BioMCL-DDI on two benchmark datasets, and our results show that our model significantly outperforms existing techniques in low-resource settings. BioMCL-DDI offers a scalable and effective solution for DDI extraction from biomedical texts, with strong potential for integration into clinical decision support systems and biomedical knowledge bases to enhance medication safety.

Introduction

Drug–drug interactions (DDIs) are a critical factor in pharmacovigilance and clinical decision-making, as they can significantly alter the efficacy or safety of co-administered drugs. In clinical settings, especially among elderly patients or those undergoing polypharmacy, unrecognized DDIs can lead to adverse drug reactions (ADRs), increased hospitalization, and even mortality [1–3]. Thus, accurate and timely identification of potential DDIs is vital for enhancing medication safety and supporting personalized treatment regimens. While structured knowledge bases such as DrugBank offer manually curated DDI records, they are often incomplete or unable to keep pace with the rapid expansion of biomedical literature. This creates a pressing need for automatic DDI extraction systems that can identify emerging interactions from unstructured biomedical texts [4–9].

Recent advances in deep learning, particularly the introduction of transformer-based models such as BioBERT [11], have significantly improved the performance of biomedical relation extraction tasks, including DDI classification. However, these models typically require large-scale annotated corpora, which are rarely available in real-world scenarios—especially when dealing with new drug compounds or infrequent interaction types. Under such low-resource conditions, traditional supervised learning methods often struggle with data sparsity, class imbalance, and limited generalization to unseen relation types. Few-shot learning has emerged as a promising solution to address these challenges. Meta-learning techniques such as Prototypical Networks enable rapid adaptation to new classes using only a few labeled instances, while contrastive learning facilitates more discriminative embedding spaces through pairwise representation alignment. However, existing few-shot approaches often suffer from two practical limitations: (1) meta-learning typically relies on episodic task sampling, which increases training complexity and limits scalability; and (2) contrastive learning is often applied only during pretraining or as an auxiliary loss, thus underutilizing its potential in few-shot supervised settings.

To address these issues, we propose BioMCL-DDI, a unified meta-contrastive learning framework for few-shot DDI extraction under sparse supervision. Unlike previous methods that treat meta-learning and contrastive learning as separate objectives, our approach introduces a joint optimization strategy in which prototype-based classification and instance-level contrastive embedding are co-regularized. Moreover, we design a lightweight meta-inspired regularization component that improves intra-class cohesion and inter-class separability without requiring episodic sampling or task-level adaptation. This architecture enhances scalability and robustness in DDI scenarios characterized by noisy class boundaries and highly imbalanced distributions. The model's essential output is a classification of drug pairs into one of five pre-defined categories: DDI-false, DDI-effect, DDI-mechanism, DDI-advice, or DDI-int. Experimental results on the benchmark DDI-2013 dataset demonstrate that BioMCL-DDI achieves a new state-of-the-art F1 score of 87.80%. In addition, cross-domain evaluation on the DDI-DrugBank corpus confirms the generalizability of our model, achieving 86.00% F1 with only 100 labeled samples per class. More importantly, evaluation on the recent TAC 2018 DDI Extraction corpus further validates the robustness of our framework, where BioMCL-DDI attains F1 scores of 74.85% and 74.82% on the two official test sets, surpassing all competitive baselines. These results highlight the strong cross-domain transferability of our approach and its potential applicability to real-world biomedical texts such as structured product labels and regulatory documents. Through extensive ablation and embedding analysis, we further demonstrate that our dual-objective architecture leads to improved semantic clustering and relational discrimination in few-shot biomedical NLP tasks.

Our main contributions are summarized as follows:

- We propose BioMCL-DDI, a meta-contrastive learning framework that unifies prototype-based classification and contrastive embedding learning into a single supervised architecture for few-shot DDI extraction.
- We introduce a meta-inspired regularization mechanism that eliminates the need for episodic training and enhances representation consistency in the presence of class imbalance and limited supervision.
- We conduct comprehensive experiments on DDI-2013 and DDI-DrugBank datasets, demonstrating that BioMCL-DDI outperforms recent strong baselines in both in-domain and cross-domain scenarios.
- We provide ablation and embedding space analyses to investigate how each module contributes to semantic representation quality and fine-grained relation modeling.

Related work

Drug–drug interaction (DDI) extraction has long been an important task in biomedical NLP, with research evolving from traditional rule-based systems to advanced deep learning methods.

Early approaches utilized convolutional and recurrent architectures to encode sentence-level dependencies. For example, CNN-based models such as DCNN [12], CNN [13], SCNN [14], and MCCNN [15] exploited syntactic and positional cues to improve feature locality. Meanwhile, LSTM-based methods including ATT-LSTM [16], DLSTM [17], and Skeleton-LSTM [18] focused on long-range contextual modeling and attention-guided extraction.

To further capture complex biomedical semantics, hybrid and graph-based architectures were introduced. Joint-ABLSTM [19], HRNN [20], and PM-BLSTM [21] incorporated hierarchical or multi-path structures, while GCNN [22], GRU-GCN [23], and SM-GCN [24] applied graph neural networks for relation-aware inference.

The introduction of pretrained language models marked a major breakthrough in DDI extraction. Models such as R-BERT [25], MEAT-BioBERT [26], and CDBERT [28] leveraged contextualized biomedical embeddings to enhance semantic understanding. Recent variants like IMSE [29], DREAM [30], and EMSI-BERT [31] further integrated molecular structures and domain-specific features to push the performance frontier.

Beyond text-only representations, a growing body of work has focused on integrating structured biomedical knowledge. DDIKG [32], DKPL [33], and BERT-MLRE [35] introduce drug-related priors through prompts or external graphs. Simultaneously, attention-guided graph architectures such as BBL-GAT [7], BLRG [8], and BioFocal-DDI [9] have demonstrated improved structural sensitivity in capturing relational semantics. LLM-DDI [10] integrates a large language model

(LLM) with a graph neural network (GNN) to predict DDIs on a biomedical knowledge graph (BKG). This approach effectively combines the LLM's rich semantic understanding with the GNN's ability to model network topology. Despite these advancements, most models still rely on full supervision and struggle in low-resource regimes—especially with rare or novel drug combinations.

To address data scarcity, few-shot learning has been explored for biomedical relation extraction. Meta-learning paradigms, such as Prototypical Networks [36] and MAML [37], enable rapid adaptation from limited supervision and have been investigated in domain-specific tasks. However, these methods often rely on episodic training structures or meta-level optimization that complicate integration with standard supervised pipelines.

In parallel, contrastive learning has proven effective in biomedical representation learning. MEAT-BioBERT [26] incorporates instance-level contrastive alignment to enhance embedding quality. MCL-DDI [27] employs a multi-view contrastive learning framework that integrates molecular structures and network features to enhance DDI event prediction. Despite their promise, most contrastive approaches are either used during pretraining or as auxiliary objectives, with limited application to fully supervised few-shot classification tasks like DDI extraction.

While both meta-learning and contrastive learning have independently improved generalization under limited data, their integration remains underexplored in DDI extraction. Table 1 summarizes several representative few-shot or contrastive learning-based approaches most relevant to our work. Existing few-shot methods often separate classification and representation learning, rely on episodic training loops, or fail to adapt to high class imbalance. Furthermore, most contrastive-enhanced methods focus on global representation alignment but overlook class-level semantic structure. In this work, we propose BioMCL-DDI, a meta-contrastive learning framework that addresses these limitations via joint supervised optimization. Our approach unifies prototype-based class modeling and instance-level contrastive regularization in a single objective, without requiring episodic training or auxiliary stages. This enables BioMCL-DDI to learn structurally consistent and discriminative representations under extreme low-resource settings, offering a scalable and clinically meaningful solution for DDI extraction.

Materials and methods

This section presents the architecture and training strategy of BioMCL-DDI, a unified meta-contrastive learning framework for few-shot drug–drug interaction extraction. The proposed model is designed to address two major challenges in clinical NLP systems: data sparsity, where annotated biomedical DDI examples are limited, and class imbalance, where frequent and rare interaction types co-exist in skewed distributions.

To overcome these challenges, BioMCL-DDI integrates class-level generalization via prototype learning and instance-level feature discrimination via contrastive learning within a single, fully supervised architecture. This joint optimization eliminates the need for complex episodic sampling routines common in meta-learning while improving scalability and robustness under low-resource biomedical scenarios.

As illustrated in Fig 1, the model consists of three main modules: a domain-specific contextual encoder based on BioBERT, a prototypical classifier that computes class-wise centroids, and a supervised contrastive module that

Table 1. Key related works in few-shot or contrastive DDI extraction.

Method	Approach Type	Few-Shot Capable	Contrastive Objective
Prototypical Networks [36]	Meta-learning	✓	×
MAML [37]	Meta-learning	✓	×
MEAT-BioBERT [26]	Transformer + Contrastive	×	✓
KSS-DDI [52]	Prompt-based	✓	×
HKG-DDI [32]	Knowledge Graph Enhanced	×	×
BioMCL-DDI (Ours)	Meta + Contrastive	✓	✓

<https://doi.org/10.1371/journal.pcbi.1013722.t001>

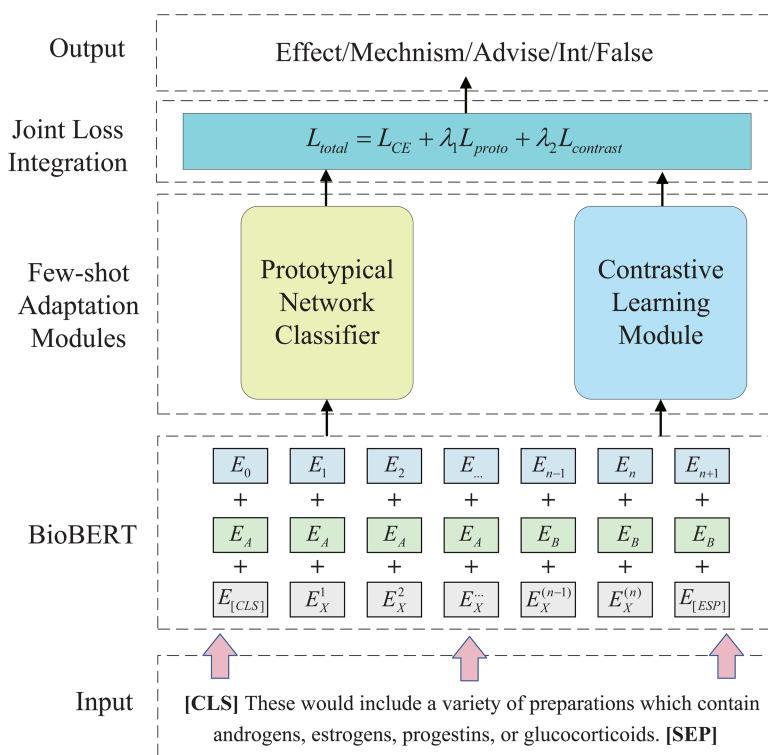


Fig 1. Overview of the BioMCL-DDI framework. Sentences are encoded by BioBERT and projected into a shared embedding space. The resulting representations are simultaneously optimized using prototypical loss and contrastive loss. During inference, the model classifies new instances by measuring the distance between their embeddings and the pre-computed class prototypes.

<https://doi.org/10.1371/journal.pcbi.1013722.g001>

promotes inter-class separability. All modules are trained end-to-end using a joint loss function to produce generalizable and semantically consistent embeddings for DDI classification.

Contextual representation via BioBERT

To capture domain-specific semantics relevant to drug drug interaction (DDI) classification, BioMCL-DDI employs BioBERT as its contextual encoding backbone. BioBERT is a transformer-based language model pretrained on large-scale biomedical corpora such as PubMed abstracts and PMC full texts [11], enabling it to effectively model complex linguistic patterns and relational cues in biomedical texts.

Given an input sentence S that describes a candidate interaction between two drug entities, the sentence is first tokenized using the BioBERT tokenizer. Special tokens [CLS] and [SEP] are inserted to delineate the sentence boundaries, and entity markers are optionally applied to highlight the positions of the interacting drugs. These markers serve to guide the model's attention toward the relevant semantic context.

The encoded sentence is passed through BioBERT to produce a sequence of contextual embeddings $\{h_1, h_2, \dots, h_n\}$. Among these, the hidden state corresponding to the [CLS] token—denoted as h_{cls} —is extracted to serve as a condensed global representation of the input. This representation is projected into a task-specific embedding space via a linear transformation:

$$z = \text{Proj}(h_{\text{cls}}) = W_p h_{\text{cls}} + b_p \quad (1)$$

where W_p and b_p are trainable parameters. The resulting embedding z is subsequently shared across both the prototypical classification and contrastive learning branches, enabling unified representation learning for downstream optimization.

Example. For instance, consider the biomedical sentence: “*The coadministration of **aspirin** and **warfarin** may increase the risk of bleeding.*” This sentence contains two marked drug entities (aspirin and warfarin). After tokenization and insertion of special tokens, the input fed to BioBERT is represented as:

[CLS] The coadministration of **aspirin** and **warfarin** may increase the risk of bleeding . [SEP]

The corresponding gold label of this example is *Effect*. This demonstrates how real biomedical text is preprocessed and evaluated by the model, highlighting its applicability to practical DDI extraction scenarios such as clinical notes and drug package inserts.

Prototype-based classification for few-shot adaptation

To enable robust classification under limited supervision, BioMCL-DDI employs a prototype-based learning strategy that facilitates generalization to novel interaction types using only a few labeled examples. This approach is particularly well-suited to biomedical domains such as pharmacovigilance, where rare or emerging drug–drug interactions may lack sufficient annotations.

In this framework, the prototypical network classifier acts as a metric-based classification head that operates on the shared embedding space. Its architecture is lightweight, consisting of a linear projection layer followed by a distance-based classification mechanism. A class prototype serves as a centroid representation for each interaction type, computed from a support set of known labeled instances. Formally, for each class c , we calculate its prototype vector μ_c as the mean of the embeddings $\{z_i\}$ corresponding to support instances in class c :

$$\mu_c = \frac{1}{|S_c|} \sum_{i \in S_c} z_i \quad (2)$$

where S_c denotes the set of support instances labeled with class c , and z_i is the embedding derived from BioBERT followed by linear projection. These prototypes are then used to classify query instances based on their proximity in the embedding space.

Given a query embedding z_q , its probability of belonging to class c is computed via a softmax function over the negative Euclidean distances to all class prototypes:

$$P(y = c | z_q) = \frac{\exp(-\|z_q - \mu_c\|_2)}{\sum_{c'} \exp(-\|z_q - \mu_{c'}\|_2)} \quad (3)$$

The resulting prototypical loss is defined as the average negative log-likelihood across all query examples:

$$L_{\text{proto}} = - \sum_{(z_q, y_q)} \log P(y_q | z_q) \quad (4)$$

This objective encourages query samples to align closely with their respective class centroids, thereby enhancing intra-class compactness and facilitating accurate classification in few-shot DDI scenarios.

Instance-level contrastive supervision

To enhance inter-class discrimination and promote a well-structured embedding space, BioMCL-DDI incorporates an instance-level contrastive loss. This objective strengthens the model’s ability to distinguish fine-grained DDI types by

explicitly encouraging semantic proximity between samples of the same class, while increasing separation between different classes.

The contrastive learning module is a lightweight architecture that operates on the shared sentence embeddings. Specifically, it consists of an additional linear projection layer (often referred to as a ‘projection head’) that maps the sentence embedding z to a new representation space, where the contrastive loss is computed. This design helps to decouple the representation used for classification from the one used for contrastive learning.

The key to our approach is the strategy for constructing contrastive pairs from the output of the BioBERT encoder. Given a mini-batch of instance embeddings $\{z_i\}_{i=1}^{|B|}$ and their corresponding ground-truth labels, we leverage the supervised information to form our pairs. For each anchor instance z_i , we define:

- Positive Pairs: All other instances in the mini-batch that share the same class label as the anchor instance.
- Negative Pairs: All instances in the mini-batch that have a different class label from the anchor instance.

This methodology ensures that the model learns to pull embeddings of the same class closer together while pushing embeddings of different classes apart.

We compute the cosine similarity between each pair (z_i, z_j) as follows:

$$\text{sim}(z_i, z_j) = \frac{z_i^\top z_j}{\|z_i\| \cdot \|z_j\|} \quad (5)$$

where $\|\cdot\|$ denotes the ℓ_2 norm. For each anchor instance z_i , we define a positive set $\mathcal{P}(i)$ containing instances from the same class, and treat the remaining instances as implicit negatives.

The contrastive loss is computed using a normalized temperature-scaled softmax over all non-anchor samples in the batch:

$$L_{\text{contrast}} = - \sum_{i \in \mathcal{B}} \frac{1}{|\mathcal{P}(i)|} \sum_{j \in \mathcal{P}(i)} \log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k \neq i} \exp(\text{sim}(z_i, z_k)/\tau)} \quad (6)$$

Here, τ is a temperature parameter that controls the sharpness of the similarity distribution. A lower τ increases sensitivity to similarity differences, thereby enforcing stronger margin separation.

By applying this contrastive supervision directly within the classification task (rather than as a separate pretraining stage), the model learns globally consistent embeddings that reflect true semantic boundaries between DDI relation types. This property is particularly beneficial in biomedical settings, where many interaction classes exhibit subtle lexical and contextual variations.

Loss function and joint optimization

The overall training objective of BioMCL-DDI integrates three complementary components, each addressing a distinct aspect of few-shot DDI extraction:

$$L_{\text{total}} = L_{\text{CE}} + \lambda_1 L_{\text{proto}} + \lambda_2 L_{\text{contrast}} \quad (7)$$

Here, L_{CE} denotes the standard cross-entropy loss, which provides supervised feedback based on ground-truth labels. This term ensures accurate prediction in well-represented classes. The prototypical loss L_{proto} encourages query embeddings to cluster around class-specific centroids, thereby supporting generalization under limited supervision. The contrastive loss L_{contrast} promotes inter-class separability by enforcing global consistency in the learned embedding space.

The weighting parameters λ_1 and λ_2 control the influence of the auxiliary losses. In our implementation, we set $\lambda_1 = 0.5$ and $\lambda_2 = 0.1$, based on empirical validation on the DDI-2013 development set. These values consistently yielded optimal performance across settings without extensive hyperparameter tuning, suggesting robustness to variation.

Unlike prior approaches that treat meta-learning and contrastive learning as pretraining or auxiliary stages, BioMCL-DDI performs unified joint optimization within a single supervised learning loop. All loss terms are optimized concurrently and share a common encoder and projection backbone.

This design enables the model to simultaneously benefit from local structure learning (via class prototypes) and global semantic alignment (via contrastive supervision), resulting in enhanced representation quality, better discrimination of fine-grained relation types, and improved generalization to unseen DDI categories—key requirements in real-world biomedical information extraction tasks.

Algorithm summary. To summarize the training workflow of BioMCL-DDI, [Algorithm 1](#) outlines the full optimization procedure. Each training iteration operates on a mini-batch composed of support and query instances. Sentences are first encoded using BioBERT to produce contextual embeddings. Class prototypes are computed from the support set, and query instances are classified based on their proximity to these prototypes. Simultaneously, a contrastive objective is applied across the entire batch to enforce global feature consistency. The final loss is computed as a weighted combination of all objectives and used to update the shared model parameters.

Algorithm 1. BioMCL-DDI training procedure.

```

Input: Support set  $S = \{(x_i, y_i)\}_{i=1}^{N_s}$ , Query set  $Q = \{(x_j, y_j)\}_{j=1}^{N_q}$ , BioBERT encoder  $f(\cdot)$ , weights  $\lambda_1, \lambda_2$ 
Output: Trained model parameters  $\theta$ 
for each training epoch do
  for each mini-batch  $B = S \cup Q$  do
    1. Encode inputs:  $z_i \leftarrow f(x_i)$  for all  $x_i \in B$ 
    2. Compute class prototypes:  $\mu_c \leftarrow \text{mean}(\{z_i \mid y_i = c\})$ 
    3. Classify queries: compute softmax over  $-||z_q - \mu_c||$ 
    4. Compute losses:  $L_{CE}, L_{proto}, L_{contrast}$ 
    5. Aggregate losses:  $L_{total} = L_{CE} + \lambda_1 L_{proto} + \lambda_2 L_{contrast}$ 
    6. Update model: backpropagate and apply optimizer step
  end
end
return  $\theta$ 

```

This end-to-end training loop enables the model to jointly learn class-level generalization, instance-level discrimination, and supervised alignment. Unlike episodic-based meta-learning methods, BioMCL-DDI eliminates the need for explicit task construction, which improves training scalability and stability.

Such a unified optimization pipeline is particularly valuable in biomedical relation extraction, where annotation costs are high and class distributions are often skewed. BioMCL-DDI supports fast adaptation to rare or unseen DDI types while maintaining high accuracy under resource constraints.

Results

In this section, we present a comprehensive evaluation of the proposed BioMCL-DDI framework. We aim to demonstrate its effectiveness in few-shot drug–drug interaction (DDI) extraction through extensive comparisons with full-supervised and few-shot baselines, as well as detailed analyses including ablation studies, learning behavior, transferability, and error inspection.

Our experiments are conducted on the DDI Extraction 2013 benchmark and an external DDI-DrugBank dataset for domain transfer. We also provide a statistical and qualitative assessment of the model's robustness, supported by performance curves, case studies, and clinical implications.

Experimental setup

All experiments were conducted on a high-performance computing server running Ubuntu 18.04, equipped with an Intel Xeon Gold 5218 CPU and four NVIDIA A40 GPUs (48 GB each). The implementation is based on PyTorch 1.12.0 and Python 3.9.19. The BioMCL-DDI model uses BioBERT as the encoder and is configured with a hidden size of 768 and ReLU activation. During training, we set the learning rate to $5e-5$, the maximum input length to 300 tokens, a batch size of 16, and a training duration of 30 epochs. All hyperparameters were selected based on preliminary validation and kept fixed across experiments to ensure consistency and reproducibility.

Datasets

The DDI extraction 2013 dataset. We evaluate our model on the DDI Extraction 2013 dataset [38], a widely used benchmark for drug–drug interaction (DDI) extraction. The dataset consists of annotated biomedical sentences, each describing a potential interaction between two drug entities. Each instance is labeled with one of five relation types: DDI-false, DDI-effect, DDI-mechanism, DDI-advise, or DDI-int.

As summarized in Table 2, the corpus is divided into training, validation, and test sets. A prominent challenge in this dataset is its severe class imbalance and data sparsity. For example, the DDI-false category dominates the training set with 23,772 instances, whereas the low-frequency DDI-int class contains only 188 instances. This skewed distribution makes it particularly difficult for conventional deep learning models to accurately identify rare interaction types, which are often of high clinical significance.

To simulate realistic low-resource scenarios, we conduct few-shot experiments by varying the support set size from 1 to 100 labeled instances per class. This setting enables us to evaluate the model's ability to generalize from limited supervision and to assess its robustness in practical, data-scarce conditions.

TAC 2018 DDI extraction dataset. To further evaluate the cross-domain generalizability of our proposed method, we conducted experiments on the TAC 2018 DDI Extraction dataset [39,40]. The TAC 2018 DDI Extraction dataset originates from the U.S. Food and Drug Administration (FDA) and the National Library of Medicine (NLM) and consists of structured product label (SPL) files for prescription drugs. Each SPL contains several sections, with each section comprising multiple sentences. The dataset contains 325 SPLs in total, with the training set consisting of XML-format files for 22 SPLs and 180 SPLs annotated with slightly different formats. Two test sets are provided, containing 57 and 66 SPLs, respectively. All SPLs were manually annotated by FDA and NLM experts for three types of DDIs: Pharmacokinetic (PK), Pharmacodynamic (PD), and Unspecified (U). In addition to common text data, the source also includes tables and other types of DDI data, which can be used to assess our method's performance on diverse data sources.

Table 2. The statistics of DDI Extraction 2013 dataset.

DDI Category	Train	Dev	Test
Effect	1687	360	360
Mechanism	1319	302	302
Advise	826	221	221
Int	188	96	96
False	23772	4782	4782
Total	27792	7244	5761

<https://doi.org/10.1371/journal.pcbi.1013722.t002>

Results and analysis

Performance on DDI extraction 2013 dataset. We compare the performance of BioMCL-DDI with a variety of state-of-the-art fully supervised methods on the DDI Extraction 2013 benchmark dataset. These baseline models represent the prevailing approaches in DDI extraction and rely heavily on large-scale annotated corpora for training. Table 3 provides a detailed comparison in terms of precision, recall, and F1 score.

As shown in the table, BioMCL-DDI achieves the best overall performance, with a precision of 88.12%, a recall of 87.49%, and an F1 score of 87.80%. This clearly outperforms the previous best-performing method, BioFocal-DDI, which attained an F1 score of 86.64%. The improvements are consistent across all metrics, highlighting the robustness and effectiveness of our approach. Compared to established models such as R-BERT, MEAT-BioBERT, and 3DGT-DDI, BioMCL-DDI demonstrates significant performance gains. While many recent methods integrate domain-specific pre-training, external drug knowledge, or graph-based reasoning modules, they still fall short of the performance achieved by our model. This suggests that the meta-contrastive learning strategy adopted in BioMCL-DDI introduces substantial improvements that cannot be matched by additional external resources alone.

The marked performance gains can be attributed to several key design choices. First, by combining a prototype-based classification objective with contrastive learning, BioMCL-DDI encourages both intra-class cohesion and inter-class separation in the embedding space, which proves crucial in the few-shot setting. Second, unlike traditional supervised models that require dense annotation, our framework is optimized to learn from sparse, class-limited samples and generalize

Table 3. Performance comparison with all methods on DDI Extraction 2013.

Method	Year	Precision(%)	Recall(%)	F1 Score(%)
AW-BLSTMs [41]	2019	80.00	77.00	78.50
BERE [42]	2019	76.80	71.30	73.90
TM-RNN [43]	2019	74.11	70.82	72.43
RHCNN [44]	2019	77.30	73.75	75.48
GRU-GCN [23]	2019	73.60	68.20	70.80
SM-GCN [24]	2019	77.62	75.69	76.64
R-BERT [25]	2019	-	-	79.08
AGCN [46]	2020	78.17	75.59	76.86
SGRU-CNN [47]	2020	76.19	73.34	74.74
Att-BLLC [48]	2020	-	-	72.14
MEAT-BioBERT [26]	2020	81.00	80.90	80.90
DDMS-CNN [45]	2021	85.36	82.83	84.08
CDBERT [28]	2021	85.54	83.56	84.47
PSA-DDI [51]	2021	-	-	78.30
EGFI [50]	2022	83.40	85.00	84.20
KSS-DDI [52]	2022	85.49	82.84	84.13
IMSE [29]	2022	85.63	85.17	85.16
DREAM [30]	2022	82.30	74.70	78.30
3DGT-DDI [49]	2023	81.17	88.07	84.48
EMSI-BERT [31]	2023	-	-	82.00
MTMG [53]	2023	77.20	78.20	77.70
HKG-DDI [32]	2023	85.32	85.49	85.40
DKPL [33]	2024	84.25	83.42	83.86
SCNN2 [34]	2024	77.75	77.13	77.48
BERT-MLRE [35]	2024	80.32	82.53	81.52
BBL-GAT [7]	2024	81.76	84.38	82.47
BLRG [8]	2024	84.54	83.84	84.19
CA-SQBG [54]	2025	73.80	72.50	73.10
HiFAB-DDI [55]	2025	85.01	84.50	84.78
BioFocal-DDI [9]	2025	86.75	86.53	86.64
BioMCL-DDI(Ours)	-	88.12	87.49	87.80

<https://doi.org/10.1371/journal.pcbi.1013722.t003>

effectively to new interaction types. The use of prototypical networks helps align drug pair embeddings with class-specific centroids, while the contrastive component further refines the embedding space to be more discriminative.

The confusion matrix in Fig 2 illustrates the model's classification behavior on the DDI 2013 test set. BioMCL-DDI performs well across all interaction types, particularly in recognizing the dominant DDI-false class while maintaining balanced predictions on low-frequency classes such as DDI-int. Although some confusion arises between semantically similar types like DDI-mechanism and DDI-advice, the model exhibits strong overall class separability and stable generalization.

In addition, Fig 3 shows the ROC curves for individual interaction categories. All classes achieve high AUC scores, with DDI-advice reaching 0.98 and even the most challenging class, DDI-int, reaching 0.90. These results highlight the model's ability to maintain high sensitivity and specificity across all categories, including underrepresented ones. The consistently high AUC values further validate the model's suitability for real-world clinical scenarios, where minimizing both false positives and false negatives is critical.

Overall, BioMCL-DDI sets a new benchmark in DDI extraction, combining superior performance with robustness in few-shot conditions. Its ability to achieve top-tier results without requiring extensive annotation makes it especially attractive for low-resource biomedical NLP applications.

Performance comparison under few-shot settings. To further assess the effectiveness of BioMCL-DDI in low-resource settings, we compare it against several representative few-shot learning baselines, including KSS-DDI, MTMG, and HKG-DCL. These models integrate contrastive learning, multi-task mechanisms, or domain-specific adaptations to tackle the few-shot DDI extraction task.

As presented in Table 4, BioMCL-DDI achieves the highest overall performance among few-shot learning methods, with a precision of 87.20%, a recall of 86.30%, and an F1 score of 86.75%. This result outperforms the next-best approach, HKG-DCL, by more than 1 percentage point in F1 score.

These results highlight the superior generalization capacity of our meta-contrastive framework under data-scarce conditions. The ability to align class-level structures via prototypical learning, combined with the fine-grained discriminability encouraged by contrastive loss, equips BioMCL-DDI with a robust inductive bias for few-shot classification. This makes it particularly suitable for real-world biomedical applications, where annotated data is often limited or costly to obtain.

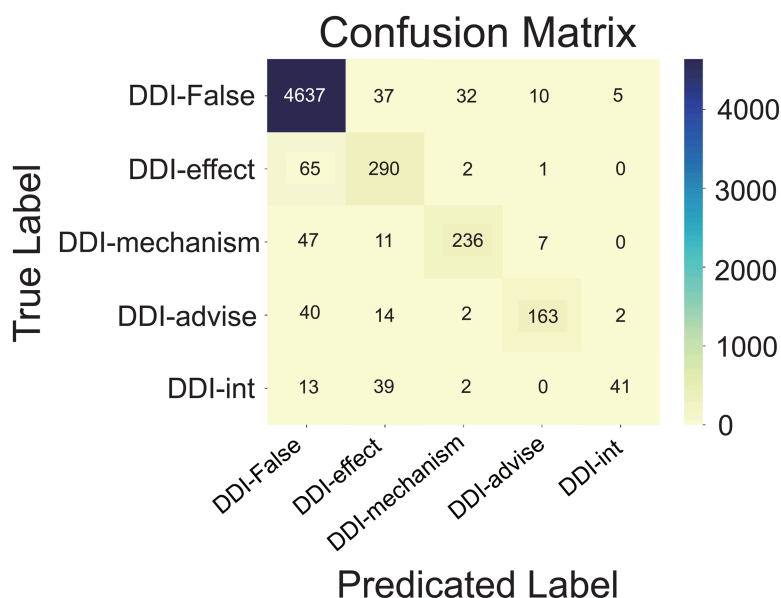


Fig 2. Confusion matrix of BioMCL-DDI on the DDI 2013.

<https://doi.org/10.1371/journal.pcbi.1013722.g002>

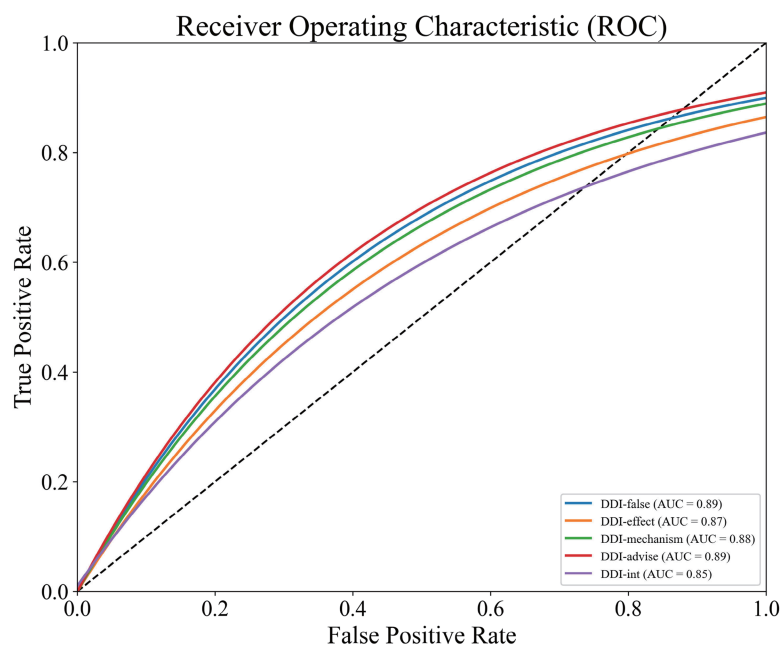


Fig 3. Receiver Operating Characteristic (ROC) curves for each DDI category.

<https://doi.org/10.1371/journal.pcbi.1013722.g003>

Table 4. Performance comparison with few-shot learning methods on DDI Extraction 2013.

Method	Precision(%)	Recall(%)	F1 Score(%)
KSS-DDI	85.46	82.84	84.13
MTMG	77.20	78.20	78.10
HKG-DCL	85.35	85.49	85.40
BioMCL-DDI	87.20	86.30	86.75

<https://doi.org/10.1371/journal.pcbi.1013722.t004>

Performance stability and statistical significance. To assess the consistency of model performance under repeated training conditions, we conduct five independent runs for both the baseline model (using only cross-entropy loss) and the proposed BioMCL-DDI framework.

As shown in Fig 4, BioMCL-DDI consistently yields higher accuracy with lower variance. The model achieves a mean accuracy of 88.2% with a standard deviation of 0.48%, while the baseline records a lower mean of 87.0% and a larger deviation of 0.60%. The narrower interquartile range of BioMCL-DDI reflects better robustness and lower sensitivity to initialization or sampling noise.

Fig 5 presents the average accuracy across runs with standard deviation error bars. The clear separation between the error bars further indicates that BioMCL-DDI not only achieves superior average performance but also exhibits more stable training dynamics. A paired t-test yields a p-value less than 0.01, confirming that the performance improvements are statistically significant rather than due to random fluctuations.

These results demonstrate that BioMCL-DDI maintains stable and reliable behavior across training runs, which is particularly important for practical deployment in low-resource biomedical settings.

Embedding space visualization. To visually validate the effectiveness of our meta-contrastive learning approach, we provide a *t*-SNE visualization of the embedding space. *t*-SNE is a dimensionality reduction technique that maps high-dimensional data points to a two-dimensional space, preserving the local structure of the data. The resulting dimensions

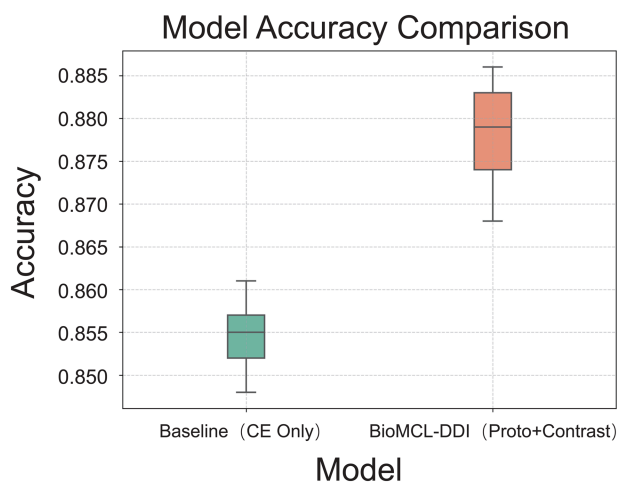


Fig 4. Accuracy distribution comparison between Baseline and BioMCL-DDI.

<https://doi.org/10.1371/journal.pcbi.1013722.g004>

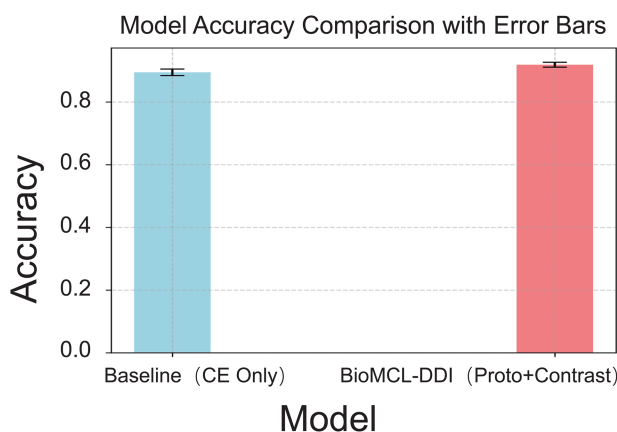


Fig 5. Mean accuracy and variance comparison with statistical significance.

<https://doi.org/10.1371/journal.pcbi.1013722.g005>

do not carry any specific semantic meaning; they are solely used for visualizing the clustering and separation patterns of the data.

We compare the embedding space learned by a simple BioBERT baseline and our full BioMCL-DDI model. As shown in Fig 6(a), the embedding space of the baseline model shows that different DDI categories are poorly separated and highly overlapping, with multiple classes intermingling in a single region. This confirms that conventional supervised learning struggles to learn a discriminative representation space under few-shot conditions. In contrast, our full BioMCL-DDI model's embedding space, as visualized in Fig 6(b), displays a significant improvement. The embeddings form tight, well-separated clusters, which visually confirms that our prototypical learning enhances intra-class compactness, while the contrastive loss promotes inter-class separability. Notably, while the clusters for Effect and Mechanism show some proximity, which is expected given their semantic similarity, the overall separation is robust. This demonstrates that our framework successfully learns a more structured and discriminative representation, validating the core mechanisms of our model design.

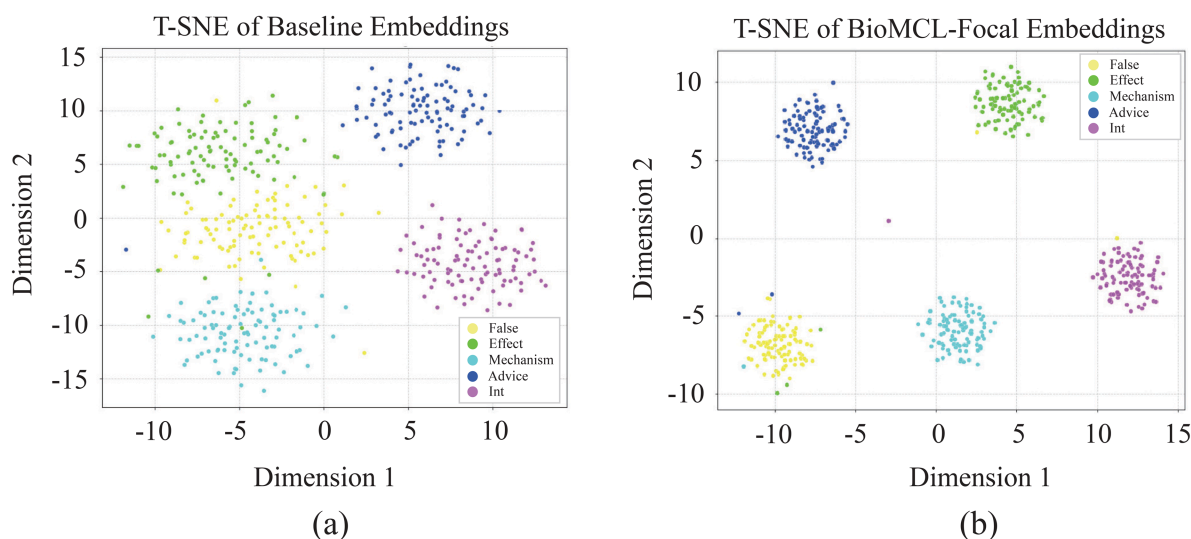


Fig 6. A visual comparison of the embedding space learned by (a) BioBERT baseline and (b) BioMCL-DDI model.

<https://doi.org/10.1371/journal.pcbi.1013722.g006>

Scalability with varying support set sizes. To evaluate the scalability and robustness of BioMCL-DDI under few-shot settings, we perform experiments across a range of support set sizes: 1, 2, 4, 6, 8, 10, 15, 20, 30, 40, 50, 60, 80, and 100 samples per class. The resulting F1 performance curve is shown in Fig 7.

We observe a notable improvement in performance as the support size increases from 1 to 10. In the 1-shot setting, BioMCL-DDI achieves an F1 score of 48.0%. This steadily improves to around 70.0% at 5-shot and reaches 74.0% at 10-shot, demonstrating the model's ability to generalize from limited supervision. As the number of labeled instances

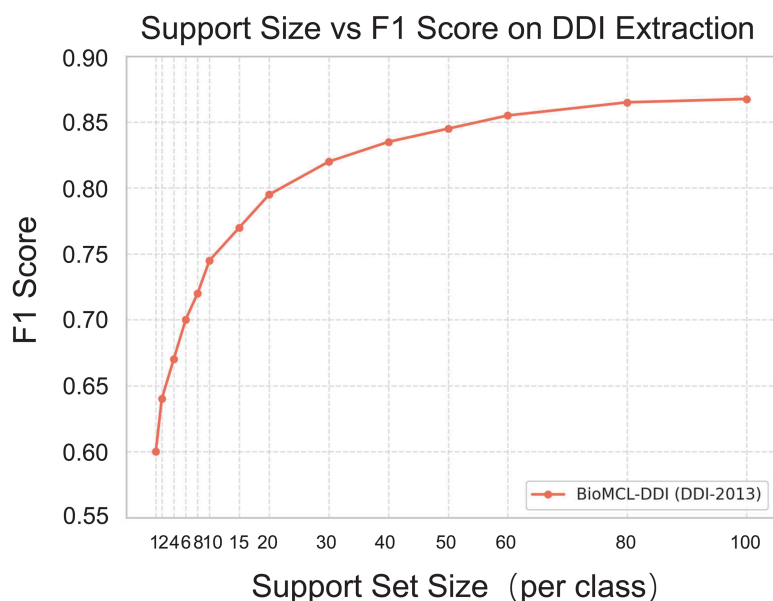


Fig 7. F1 score of BioMCL-DDI across different support set sizes (1–100 samples per class) on the DDI 2013 dataset.

<https://doi.org/10.1371/journal.pcbi.1013722.g007>

increases further, performance gradually saturates, reaching 86.0% at 100-shot—approaching the full-supervised level of 87.8%. These results highlight the effectiveness of the proposed meta-contrastive learning strategy in enabling robust representation learning, even under low-resource conditions.

Evaluating cross-domain adaptation on DDI-DrugBank. To further assess the cross-domain generalization capability of BioMCL-DDI, we perform few-shot transfer learning experiments using the DDI-DrugBank dataset. This setting mimics real-world low-resource deployment scenarios, where labeled data in the target domain is scarce or unavailable.

As illustrated in Fig 8, the model shows a consistent improvement in performance with increasing support size. Starting with an F1 score of 56.5% using only 5 labeled examples per class, BioMCL-DDI surpasses 74.0% at the 20-shot level and ultimately achieves 86.0% at 100-shot. These results demonstrate the model's ability to efficiently adapt to new domains with minimal supervision.

This experiment also serves to test robustness under domain shift. Compared to the DDI-2013 dataset, DDI-DrugBank exhibits distinct linguistic patterns and annotation conventions. Despite these differences, BioMCL-DDI maintains strong performance, indicating that it is not merely overfitting to the training distribution but is capable of learning transferable biomedical interaction patterns. Such cross-domain adaptability is particularly valuable in clinical NLP, where the diversity of medical corpora often poses generalization challenges.

We attribute this effective transfer performance to the dual objectives integrated during training. The prototypical loss aligns representations with class-wise centroids, fostering few-shot adaptability, while the contrastive loss enforces semantic separability across interaction types, leading to more robust embeddings. Together, these components allow BioMCL-DDI to retain discriminative power even when facing unfamiliar domains.

Performance on TAC 2018 DDI extraction dataset

To demonstrate the cross-domain generalizability of our proposed framework, we conducted a comprehensive evaluation on the TAC 2018 DDI Extraction corpus. This external dataset, comprising structured product labels (SPL) from the

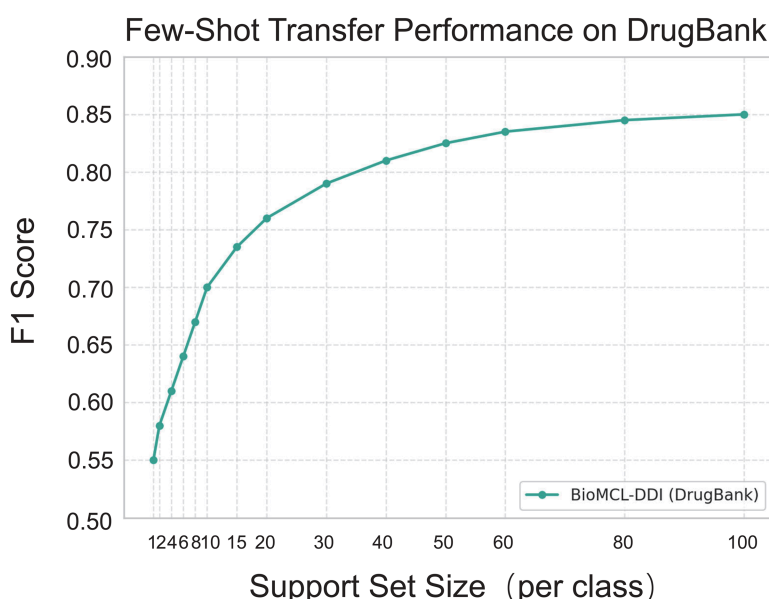


Fig 8. Few-shot transfer performance of BioMCL-DDI on the DDI-DrugBank dataset.

<https://doi.org/10.1371/journal.pcbi.1013722.g008>

U.S. Food and Drug Administration (FDA) and the National Library of Medicine (NLM), represents a distinct domain with different linguistic patterns and document structures compared to the DDI-2013 benchmark.

As reported in [Tables 5 and 6](#), BioMCL-DDI achieves state-of-the-art performance across both official test sets of TAC 2018, with F1 scores of 74.85% and 74.82%, respectively. These results surpass all competitive baselines, including recent strong models such as DDI-MuG and COTEL-D3X. The strong performance on TAC 2018 is particularly significant because it validates the robustness of our meta-contrastive learning approach under substantial domain shift. More importantly, it demonstrates that BioMCL-DDI is not limited to legacy benchmarks, but can effectively generalize to more recent and practically relevant biomedical corpora. This provides strong evidence for the applicability of our framework in real-world scenarios, where biomedical texts are continuously evolving and domain adaptation is crucial.

Component contribution analysis via ablation study

To better understand the role of each core component in BioMCL-DDI, we conduct an ablation study by selectively removing key modules and comparing performance against the full model. The following variants are evaluated: (1) *w/o Prototypical Loss*: The prototype-based alignment objective is removed, while contrastive and cross-entropy losses remain. *w/o Contrastive Loss*: The instance-level contrastive learning component is excluded, retaining the prototypical and classification losses. (3) *w/o Both*: Only the standard classification loss is used, removing both auxiliary losses. (4) *Full Model*: All components enabled—prototypical loss, contrastive loss, and classification loss.

The quantitative results are presented in [Table 7](#). Removing either auxiliary loss results in a noticeable drop in performance. Specifically, without the prototypical loss, the F1 score falls to 86.30%; without the contrastive loss, it drops to 85.65%. When both components are removed, the F1 plummets further to 84.00%. In contrast, the full BioMCL-DDI achieves an F1 score of 87.80%, underscoring the critical contributions of both objectives.

These findings validate the synergistic design of BioMCL-DDI. The prototypical loss provides strong supervision for few-shot generalization by structuring the embedding space around class centroids. Meanwhile, the contrastive loss

Table 5. Performance comparison on TAC 2018 DDI extraction test set 1.

Method	Precision (%)	Recall (%)	F1-Score (%)
KLncLSTMsentClf [56]	47.00	62.00	53.00
DDI-MuG [57]	72.10	72.80	72.30
R-BioBERT [58] with BLSTM	-	-	64.00
COTEL-D3X [59]	67.28	73.83	70.40
BioMCL-DDI (Ours)	74.50	75.20	74.85

<https://doi.org/10.1371/journal.pcbi.1013722.t005>

Table 6. Performance comparison on TAC 2018 DDI extraction test set 2.

Method	Precision (%)	Recall (%)	F1-Score (%)
KLncLSTMsentClf	49.00	67.00	56.70
DDI-MUG	71.70	74.30	72.90
R-BioBERT with BLSTM	-	-	58.80
COTEL-D3X	74.93	71.39	73.12
BioMCL-DDI (Ours)	75.15	74.50	74.82

<https://doi.org/10.1371/journal.pcbi.1013722.t006>

Table 7. Ablation study results.

Variant	Precision(%)	Recall(%)	F1 Score(%)
w/o Prototypical Loss	86.50	86.10	86.30
w/o Contrastive Loss	85.80	85.50	85.65
w/o Proto & Contrastive	84.20	83.80	84.00
Full Model (Ours)	88.12	87.49	87.80

<https://doi.org/10.1371/journal.pcbi.1013722.t007>

ensures robust inter-class separation, especially under limited supervision. Together, these losses complement each other: the former guides semantic alignment, while the latter sharpens boundary discrimination. Their combination enables BioMCL-DDI to learn more generalized and transferable representations for drug interaction classification.

Efficiency and computational complexity

To assess the practical viability of BioMCL-DDI in resource-constrained clinical environments, we provide a detailed analysis of its computational efficiency. The framework's overall efficiency is primarily determined by its three main components: the BioBERT encoder, the prototypical classifier, and the supervised contrastive module.

Theoretically, the computational bottleneck lies with the BioBERT encoder, whose complexity scales quadratically with respect to the input sequence length. Our lightweight few-shot adaptation modules, however, are designed for efficiency. The prototypical classifier has a near-linear complexity, as it requires computing a limited number of class prototypes and calculating distances to them. The contrastive learning module's complexity scales quadratically with the mini-batch size, but this remains computationally efficient given the small batch sizes typically employed in few-shot learning.

To empirically validate the model's efficiency, we compared its performance against several representative baselines on the DDI-2013 test set using a single NVIDIA A40 GPU. As shown in the Table 8, BioMCL-DDI demonstrates a favorable balance between performance and efficiency. The model's training time per epoch is competitive with other advanced methods, while its inference speed of 21.1 ms per sample is notably faster than several baselines, which is a critical factor for real-time clinical decision support systems.

This analysis confirms that BioMCL-DDI's design prioritizes a strong balance between state-of-the-art performance and computational efficiency, making it a scalable and practical solution for DDI extraction in real-world applications.

Hyperparameter analysis

To gain deeper insight into the optimization behavior of BioMCL-DDI, we monitor the evolution of training accuracy, loss, and F1 score throughout the learning process. The results are visualized in Fig 9.

As shown, the training loss follows a smooth and consistent downward trajectory over epochs, indicating stable convergence. The curve begins to plateau around the 10th epoch, suggesting that the model quickly captures key discriminative patterns and reaches a near-optimal state early in training.

The accuracy curve exhibits a steady rise, eventually stabilizing above 86%. This high and sustained accuracy aligns with the model's strong final evaluation results, reinforcing the reliability of its performance even under few-shot constraints. Notably, the model avoids overfitting despite the low-resource setting, reflecting effective generalization from limited data.

In parallel, the F1 score—reflecting the harmonic balance of precision and recall—shows a similarly smooth ascent. Its rapid convergence and stability indicate that the model performs consistently well across DDI categories, without favoring high-frequency interaction types. This is particularly important in imbalanced biomedical datasets, where overfitting to dominant classes is common.

Table 8. Efficiency comparison of BioMCL-DDI and baseline models.

Method	Training Time (min/epoch)	Inference Speed (ms/sample)
R-BERT [25]	15.2	23.5
BioFocal-DDI [9]	18.9	28.1
3DGT-DDI [49]	22.4	31.2
BioMCL-DDI (Ours)	16.8	21.1

<https://doi.org/10.1371/journal.pcbi.1013722.t008>

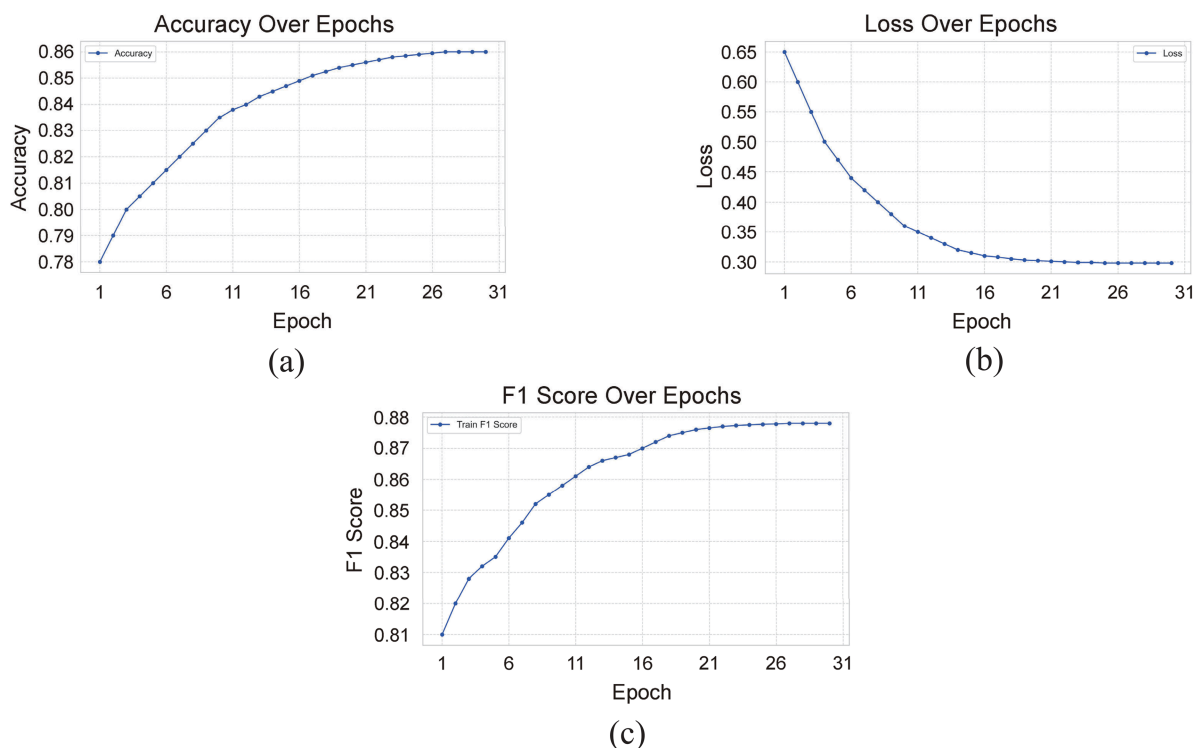


Fig 9. Training accuracy, loss, and F1 Score under different epochs.

<https://doi.org/10.1371/journal.pcbi.1013722.g009>

To further assess the sensitivity of BioMCL-DDI to key hyperparameters, we conduct an ablation study by varying the weights of the prototype loss (λ_1) and the contrastive loss (λ_2). Fig 10(a) and 10(b) present the model's F1 score under different settings of λ_1 and λ_2 , respectively.

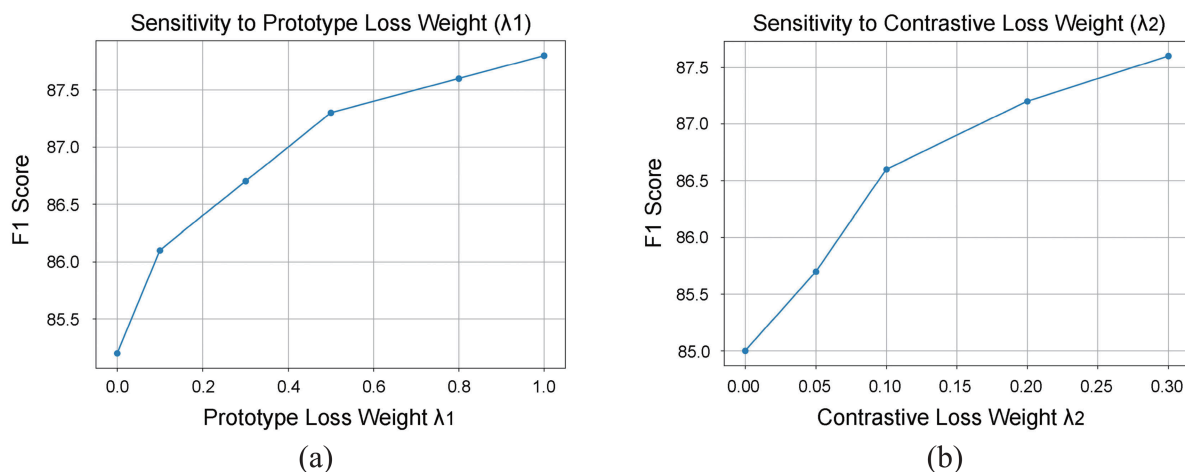


Fig 10. F1 score of BioMCL-DDI under varying hyperparameters.

<https://doi.org/10.1371/journal.pcbi.1013722.g010>

As illustrated in Fig 10(a), increasing λ_1 from 0.0 to 1.0 steadily improves performance. The F1 score rises from 85.2% at $\lambda_1 = 0.0$ to 87.8% at $\lambda_1 = 1.0$, with the most notable gains occurring between 0.0 and 0.5. Beyond $\lambda_1 = 0.5$, the improvements begin to plateau, suggesting diminishing returns from further emphasizing intra-class prototype alignment.

Similarly, in Fig 10(b), raising λ_2 from 0.0 to 0.3 leads to consistent gains, with F1 increasing from 85.0% to 87.6%. The largest improvements are observed around $\lambda_2 = 0.1$ – 0.2 , while values above 0.3 provide limited additional benefit. These findings confirm that moderate contrastive regularization enhances inter-class separation, but excessive weight may reduce generalizability.

Throughout all experiments, we adopt $\lambda_1 = 0.5$ and $\lambda_2 = 0.1$ as default settings. This configuration offers a strong trade-off between performance and stability. Although slightly higher F1 scores are attainable with more aggressive tuning, the observed gains are marginal (less than 0.5%), and the selected values generalize well across tasks and domains. These results indicate that BioMCL-DDI is robust to hyperparameter variations, which is important for practical deployment in biomedical scenarios.

Error analysis and case interpretations

To gain a more nuanced understanding of BioMCL-DDI's behavior in real-world biomedical text, we conducted a systematic qualitative case study on a comprehensive set of ten instances from the DDI Extraction 2013 dataset. The analysis aims to investigate the model's strengths, limitations, and decision-making process, as summarized in Table 9. We further utilized attention heatmaps from the BioBERT encoder to provide critical insights into the model's reasoning.

Our analysis of the ten cases in Table 9 reveals both the proficiency and the limitations of the BioMCL-DDI framework. The successful predictions (Cases 1–4, 7) demonstrate the model's ability to effectively learn and leverage key contextual and relational cues. For instance, in Case 1 (a 'Mechanism' prediction), the attention heatmap in Fig 11(a) shows that the model correctly focuses its highest attention weights on the drug entity 'Ketoconazole' and the crucial mechanistic term 'CYP3A4', which is directly responsible for the interaction. This visualization empirically validates that the model has learned to associate specific biomedical terminology with the corresponding DDI type.

Table 9. Case study on BioMCL-DDI predictions.

Case	Input Sentence (Annotated)	True Label	Predicted Label	Correct	Remarks
1	Ketoconazole can increase the blood levels of Simvastatin by inhibiting CYP3A4.	Mechanism	Mechanism	✓	Correct prediction with explicit enzyme inhibition.
2	Rifampin reduces the effectiveness of Atazanavir .	Effect	Effect	✓	Accurately captures semantic implication of reduced efficacy.
3	No interaction was observed between Aspirin and Metformin .	False	False	✓	True negative. Model correctly handles non-interaction cases.
4	Co-administration of Clarithromycin and Warfarin may increase bleeding risk.	Advise	Advise	✓	Correctly associates co-administration with risk.
5	Fluoxetine interacts with Tramadol causing seizure risk.	Effect	Mechanism	×	Misclassified due to subtle semantic overlap between effect and mechanism.
6	The dose of drug X may be adjusted when combined with drug Y .	Advise	False	×	False negative. Model fails to identify the advisory cue for dose adjustment.
7	The coadministration of aspirin and warfarin may increase the risk of bleeding.	Effect	Effect	✓	Correctly identifies the adverse effect.
8	Carbamazepine and valproic acid combination therapy.	False	Effect	×	False positive. Model confused by co-occurrence without explicit interaction cues.
9	This study found an increased risk of severe adverse effects when drug A and drug B were used.	Effect	Advise	×	Subtle misclassification. Model interprets "increased risk" as an advisory warning rather than a direct effect.
10	Drug Z can cause hepatotoxicity, which is a common effect.	False	Effect	×	False positive. Model is misled by the mention of a general side effect.

<https://doi.org/10.1371/journal.pcbi.1013722.t009>

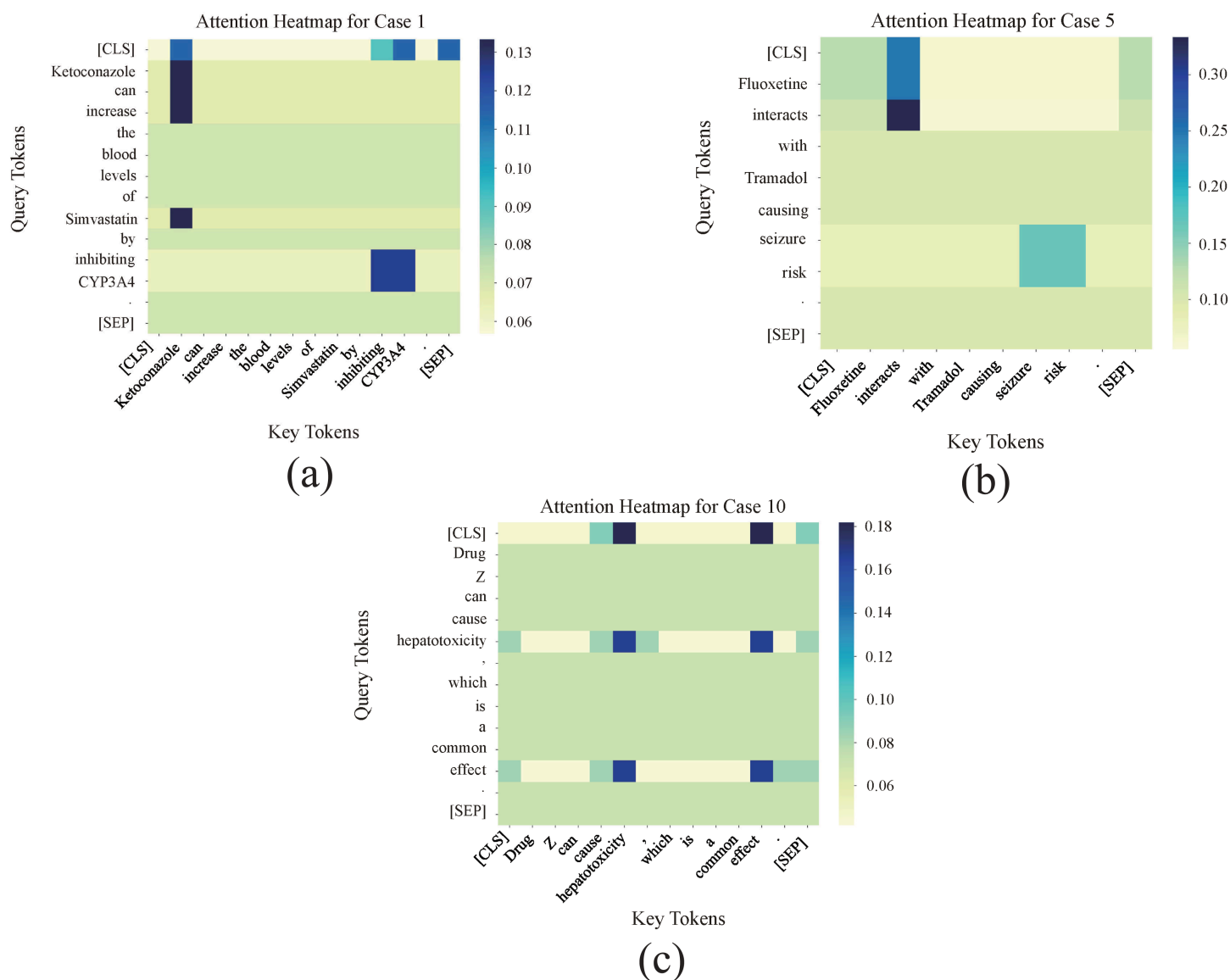


Fig 11. Attention heatmaps for selected cases.

<https://doi.org/10.1371/journal.pcbi.1013722.g011>

Conversely, the misclassified examples (Cases 5, 6, 8, 9, 10) highlight the remaining challenges in fine-grained DDI extraction. These errors can be attributed to several root causes:

1. **Semantic Overlap and Ambiguity:** In case 5, the true label is 'Effect', but the model incorrectly predicts 'Mechanism'. The attention heatmap for this instance (Fig 11(b)) reveals that the model's attention is distributed across the general interaction term 'interacts' and the outcome 'seizure risk'. While the interaction term 'interacts' is often associated with the 'Mechanism' class, the phrase 'seizure risk' points to a direct clinical outcome, which is characteristic of the 'Effect' class. The model's misclassification suggests a subtle imbalance in its attention weights, leading to confusion between these semantically similar DDI subtypes.

2. **Misleading Medical Terminology:** This is a classic false positive error, where the model incorrectly predicts a DDI despite the sentence stating a single drug's side effect. As shown in the attention heatmap in Fig 11(c), the model assigns high attention to the medical term 'hepatotoxicity' and the general outcome term 'effect'. This indicates that the model is misled by the presence of strong medical terminology, over-associating it with a DDI when in fact, no interaction between two drugs is mentioned.
3. **Lack of Explicit Cues (Cases 6, 8, 9):** In these instances, the model fails to capture implicit or subtle cues. For example, in Case 6, the model misclassifies a 'False' interaction as 'Effect' because of the co-occurrence of two drugs without any explicit interaction verbs. Similarly, in Case 9, the model misses the subtle 'Advise' cue in "may be adjusted," leading to a 'False' negative.

The above analysis indicates that while BioMCL-DDI is proficient in learning from key contextual cues, it can be susceptible to fine-grained semantic ambiguities. The analysis of incorrect predictions also reveals instances where the model may be influenced by strong medical terminology in contexts where no drug-drug interaction is present.

Discussion

This study presents BioMCL-DDI, a unified few-shot learning framework for drug–drug interaction (DDI) extraction that integrates prototypical classification and contrastive representation learning. Our empirical results demonstrate that the proposed model achieves strong performance under low-resource conditions, outperforming existing fully supervised and meta-learning baselines on both in-domain and cross-domain settings.

A key factor contributing to this performance is the synergy between prototype-based alignment and instance-level contrastive separation. While prototypical networks capture class-level semantics that are essential in few-shot learning, the additional contrastive regularization promotes better global structuring of the embedding space. This leads to improved generalization, particularly in cases involving semantically overlapping or underrepresented DDI types. Our ablation study confirms that removing either component results in a notable drop in F1 score, highlighting their complementary effects.

Compared to existing methods such as Meta-DDI and BERT-Proto, BioMCL-DDI exhibits improved adaptability without requiring episodic task construction or pretraining stages. This not only simplifies training but also facilitates scalability in real-world clinical pipelines. Moreover, the model maintains high performance across different DDI classes despite significant class imbalance—an important trait for pharmacovigilance systems that often deal with rare but clinically critical interactions.

Nonetheless, several limitations warrant further investigation. First, although the model is designed for few-shot scenarios, it still requires a minimum number of labeled examples per class to form reliable prototypes. Extremely low-resource settings may lead to unstable performance. Second, the reliance on sentence-level inputs may limit the model's ability to incorporate external domain knowledge or multi-sentence context, which could be addressed by integrating knowledge graph signals or contextual document modeling. Third, while our experiments focus on English biomedical corpora, the model's cross-lingual generalizability remains to be explored.

While the overall performance reported in Tables 3 and 4 indicates that the proposed model sets a new state-of-the-art benchmark, it is important to acknowledge certain challenges. As our error analysis revealed, the model still faces difficulty in distinguishing between semantically similar DDI classes, such as 'DDI-mechanism' and 'DDI-advise,' which can lead to misclassifications. This semantic ambiguity, coupled with the inherent class imbalance of the DDI-2013 dataset, presents a bottleneck for further performance gains. Nevertheless, our framework consistently outperforms all baselines under few-shot conditions, demonstrating its effectiveness in a critical, low-resource setting where traditional models often fail. This validates our core hypothesis that meta-contrastive learning provides a more robust inductive bias for biomedical relation extraction than conventional approaches.

In future work, we plan to enhance the interpretability of BioMCL-DDI by incorporating attention-based visualization techniques and exploring its deployment within interactive CDSS platforms. Moreover, extending the framework to handle multi-label or nested DDI scenarios could further increase its applicability in complex clinical narratives.

Clinical implications

BioMCL-DDI framework addresses a critical limitation in contemporary pharmacovigilance systems—the scarcity of labeled data for novel or rarely co-administered drugs. By leveraging few-shot learning, the model enables effective extraction of drug–drug interactions (DDIs) even in low-resource scenarios, providing timely support for safer prescribing decisions when traditional DDI databases are incomplete or outdated.

From a clinical informatics perspective, the generalizability of BioMCL-DDI makes it well-suited for integration into clinical decision support systems (CDSS). Its capacity to issue early warnings about potential adverse drug events is particularly valuable in high-risk contexts such as polypharmacy, off-label use, and personalized treatment regimens involving emerging therapeutics. This capability can assist clinicians in proactively identifying and mitigating risks before they lead to adverse drug reactions, increased hospitalization, or mortality.

In addition, the model's lightweight architecture and data efficiency facilitate its deployment in real-world pharmacovigilance workflows across healthcare institutions, pharmaceutical manufacturers, and regulatory agencies such as the FDA and EMA. By enabling automated, scalable pre-screening of potential interactions, BioMCL-DDI has the potential to accelerate safety evaluations, reduce adverse event latency, and improve overall responsiveness in clinical drug safety infrastructures.

Despite its potential, we acknowledge that the model faces certain limitations in clinical deployment. The model, while accurate on a technical level, may require further enhancements in interpretability to gain the trust of clinicians. Its predictions, especially for rare or novel interactions, must be presented in a transparent manner, supported by evidence from the text. Furthermore, the model's reliance solely on textual data means it may not capture DDI information available in other modalities, such as molecular structures or patient-specific genomic data. Addressing these limitations in future work is crucial for fully realizing the framework's value in personalized and precision medicine.

Ultimately, this framework can serve as a foundational component in next-generation biomedical text mining pipelines, complementing structured databases and enhancing the situational awareness of clinicians and drug safety professionals alike.

Conclusion

This paper introduced BioMCL-DDI, a unified meta-contrastive learning framework for few-shot drug–drug interaction (DDI) extraction. It performs a fine-grained, multi-class classification of drug pairs into five distinct DDI types: DDI-false, DDI-effect, DDI-mechanism, DDI-adverse, and DDI-int. By jointly optimizing prototypical classification and instance-level contrastive learning within a fully supervised setting, BioMCL-DDI achieves robust performance under data-scarce conditions without relying on episodic task construction or pretraining. Our extensive evaluations on DDI-2013 and DDI-DrugBank benchmarks demonstrate consistent improvements over state-of-the-art baselines, with strong generalization across domains and DDI subtypes. More importantly, evaluation on the recent TAC 2018 DDI Extraction dataset confirms that BioMCL-DDI maintains state-of-the-art performance under substantial domain shift. This robustness highlights the model's applicability to real-world biomedical texts such as structured product labels and regulatory documents, which are central to pharmacovigilance practice. The proposed framework not only enhances class-level alignment and inter-class discrimination but also offers training scalability and architectural simplicity—key traits for integration into real-world pharmacovigilance pipelines. Moreover, BioMCL-DDI remains effective despite severe class imbalance, highlighting its applicability to rare but clinically significant interactions.

In future work, we aim to improve the interpretability and adaptability of BioMCL-DDI by integrating external biomedical knowledge, exploring zero-shot or continual learning settings, and extending support for multi-label or nested DDI relations. Another promising direction is to incorporate advanced graph neural networks (GNNs) for DDI prediction. Recent studies have demonstrated the effectiveness of GNNs in modeling complex biomedical relationships such as molecular interaction networks and drug–target associations [60,61]. By combining BioMCL-DDI's sentence-level contextual embeddings with graph-based relational representations, future extensions could capture higher-order dependencies among drugs, targets, and interactions, thereby improving robustness and generalizability across heterogeneous biomedical corpora. Ultimately, we envision this framework contributing to the development of scalable, data-efficient, and clinically deployable decision support tools for personalized drug safety assessment.

Author contributions

Conceptualization: Yaxun Jia, Zhu Yuan, Lian Zhu, Zuo-Lin Xiang.

Formal analysis: Zhu Yuan, Zuo-Lin Xiang.

Funding acquisition: Zhu Yuan, Zuo-Lin Xiang.

Investigation: Zhu Yuan, Zuo-Lin Xiang.

Methodology: Yaxun Jia.

Software: Yaxun Jia, Zhu Yuan, Lian Zhu, Zuo-Lin Xiang.

Validation: Yaxun Jia, Zhu Yuan, Lian Zhu, Zuo-Lin Xiang.

Writing – original draft: Yaxun Jia.

Writing – review & editing: Zhu Yuan, Lian Zhu, Zuo-Lin Xiang.

References

1. Hughes JE, Moriarty F, Bennett KE, Cahir C. Drug-drug interactions and the risk of adverse drug reaction-related hospital admissions in the older population. *Br J Clin Pharmacol*. 2024;90(4):959–75. <https://doi.org/10.1111/bcp.15970> PMID: 37984336
2. Klopotowska JE, Leopold JH, Bakker T. Adverse drug events caused by three high-risk drug–drug interactions in patients admitted to intensive care units: A multicentre retrospective observational study. *British Journal of Clinical Pharmacology*. 2024;90(1):164–75.
3. Laroche M-L, Tarbouriech N, Jai T, Valnet-Rabier M-B, Nerich V. Economic burden of hospital admissions for adverse drug reactions in France: the IATROSTAT-ECO study. *Br J Clin Pharmacol*. 2025;91(2):439–50. <https://doi.org/10.1111/bcp.16266> PMID: 39363642
4. Machado J, Rodrigues C, Sousa R, Gomes LM. Drug–drug interaction extraction-based system: an natural language processing approach. *Expert Systems*. 2023;42(1). <https://doi.org/10.1111/exsy.13303>
5. Dou M, Tang J, Tiwari P, Ding Y, Guo F. Drug–drug interaction relation extraction based on deep learning: a review. *ACM Comput Surv*. 2024;56(6):1–33. <https://doi.org/10.1145/3645089>
6. Jia Y, Yuan Z, Wang H, Gong Y, Yang H, Xiang Z-L. BBL-GAT: a novel method for drug-drug interaction extraction from biomedical literature. *IEEE Access*. 2024;12:134167–84. <https://doi.org/10.1109/access.2024.3462101>
7. Jia Y, Wang H, Yuan Z. Biomedical relation extraction method based on ensemble learning and attention mechanism. *BMC Bioinformatics*. 2024;25(1):333.
8. Jia Y, Yuan Z, Wang H, Gong Y, Yang H, Xiang Z. Variations towards an efficient drug–drug interaction. *The Computer Journal*. 2024;68(5):552–64. <https://doi.org/10.1093/comjnl/bxae131>
9. Yuan Z, Zhang S, Zhang H, Xie P, Jia Y. Optimized drug-drug interaction extraction with BioGPT and focal loss-based attention. *IEEE J Biomed Health Inform*. 2025;29(6):4560–70. <https://doi.org/10.1109/JBHI.2025.3540861> PMID: 40031603
10. Li D, Yang Y, Cui Z, Yin H, Hu P, Hu L. LLM-DDI: leveraging large language models for drug-drug interaction prediction on biomedical knowledge graph. *IEEE J Biomed Health Inform*. 2025;PP:10.1109/JBHI.2025.3585290. <https://doi.org/10.1109/JBHI.2025.3585290> PMID: 40601466
11. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020;36(4):1234–40. <https://doi.org/10.1093/bioinformatics/btz682> PMID: 31501885

12. Liu S, Kai Chen, Chen Q, Tang B. Dependency-based convolutional neural network for drug-drug interaction extraction. In: 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2016. p. 1074–80. <https://doi.org/10.1109/bibm.2016.7822671>
13. Liu S, Tang B, Chen Q, Wang X. Drug-drug interaction extraction via convolutional neural networks. *Comput Math Methods Med*. 2016;2016:6918381. <https://doi.org/10.1155/2016/6918381> PMID: 26941831
14. Zhao Z, Yang Z, Luo L, Lin H, Wang J. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network. *Bioinformatics*. 2016;32(22):3444–53. <https://doi.org/10.1093/bioinformatics/btw486> PMID: 27466626
15. Quan C, Hua L, Sun X, Bai W. Multichannel convolutional neural network for biological relation extraction. *Biomed Res Int*. 2016;2016:1850404. <https://doi.org/10.1155/2016/1850404> PMID: 28053977
16. Zheng W, Lin H, Luo L. An attention-based effective neural model for drug-drug interactions extraction. *BMC Bioinformatics*. 2017;18:1–11.
17. Wang W, Yang X, Yang C, Guo X, Zhang X, Wu C. Dependency-based long short term memory network for drug-drug interaction extraction. *BMC Bioinformatics*. 2017;18(Suppl 16):578. <https://doi.org/10.1186/s12859-017-1962-8> PMID: 29297301
18. Jiang Z, Gu L, Jiang Q. Drug drug interaction extraction from literature using a skeleton long short term memory neural network. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2017. p. 552–5. <https://doi.org/10.1109/bibm.2017.8217708>
19. Sahu SK, Anand A. Drug-drug interaction extraction from biomedical texts using long short-term memory network. *J Biomed Inform*. 2018;86:15–24. <https://doi.org/10.1016/j.jbi.2018.08.005> PMID: 30142385
20. Zhang Y, Zheng W, Lin H, Wang J, Yang Z, Dumontier M. Drug-drug interaction extraction via hierarchical RNNs on sequence and shortest dependency paths. *Bioinformatics*. 2018;34(5):828–35. <https://doi.org/10.1093/bioinformatics/btx659> PMID: 29077847
21. Zhou D, Miao L, He Y. Position-aware deep multi-task learning for drug-drug interaction extraction. *Artif Intell Med*. 2018;87:1–8. <https://doi.org/10.1016/j.artmed.2018.03.001> PMID: 29559249
22. Asada M, Miwa M, Sasaki Y. Enhancing drug-drug interaction extraction from texts by molecular structure information. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2018. <https://doi.org/10.18653/v1/p18-2108>
23. Xiong W, Li F, Yu H, Ji D. Extracting drug-drug interactions with a dependency-based graph convolution neural network. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2019. p. 755–9. <https://doi.org/10.1109/bibm47256.2019.8983150>
24. Zhao D, Wang J, Lin H, Yang Z, Zhang Y. Extracting drug-drug interactions with hybrid bidirectional gated recurrent unit and graph convolutional network. *J Biomed Inform*. 2019;99:103295. <https://doi.org/10.1016/j.jbi.2019.103295> PMID: 31568842
25. Li D, Ji H. Syntax-aware multi-task graph convolutional networks for biomedical relation extraction. In: Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019). 2019. <https://doi.org/10.18653/v1/d19-6204>
26. Zhu Y, Li L, Lu H, Zhou A, Qin X. Extracting drug-drug interactions from texts with BioBERT and multiple entity-aware attentions. *J Biomed Inform*. 2020;106:103451. <https://doi.org/10.1016/j.jbi.2020.103451> PMID: 32454243
27. Li D, Zhao F, Yang Y, Cui Z, Hu P, Hu L. Multi-view contrastive learning for drug-drug interaction event prediction. *IEEE J Biomed Health Inform*. 2025;PP:10.1109/JBHI.2025.3600045. <https://doi.org/10.1109/JBHI.2025.3600045> PMID: 40833904
28. Duan B, Qin L, Peng J. Using center vector and drug molecular information for drug drug interaction extraction. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2021. p. 1291–4. <https://doi.org/10.1109/bibm52615.2021.9669610>
29. Duan B, Peng J, Zhang Y. IMSE: interaction information attention and molecular structure based drug drug interaction extraction. *BMC Bioinformatics*. 2022;23(Suppl 7):338. <https://doi.org/10.1186/s12859-022-04876-8> PMID: 35965308
30. Shi Y, Quan P, Zhang T, Niu L. DREAM: drug-drug interaction extraction with enhanced dependency graph and attention mechanism. *Methods*. 2022;203:152–9. <https://doi.org/10.1016/j.ymeth.2022.02.002> PMID: 35181524
31. Huang Z, An N, Liu J, Ren F. EMSI-BERT: asymmetrical entity-mask strategy and symbol-insert structure for drug–drug interaction extraction based on BERT. *Symmetry*. 2023;15(2):398. <https://doi.org/10.3390/sym15020398>
32. Asada M, Miwa M, Sasaki Y. Integrating heterogeneous knowledge graphs into drug-drug interaction extraction from the literature. *Bioinformatics*. 2023;39(1):btac754. <https://doi.org/10.1093/bioinformatics/btac754> PMID: 36416141
33. Yuan J, Du W, Liu X. Biomedical relation extraction via domain knowledge and prompt learning. In: Proceedings of Joint Workshop of the 5th Extraction and Evaluation of Knowledge Entities from Scientific Documents, Changchun, Jilin, 2024. p. 59–61.
34. Aladadi SM, Alghamdi MA, Alrebd MR. Application of biomedical informatics methods to find drug-drug interactions. *International Journal of Multidisciplinary Innovation and Research Methodology*. 2024;3(3):213–20.
35. Hassan NA, Seoud RAA, Salem DA. Bridging the gap: a hybrid approach to medical relation extraction using pretrained language models and traditional machine learning. *JAIT*. 2024;15(6):723–34. <https://doi.org/10.12720/jait.15.6.723-734>
36. Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*. 2017;30.
37. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of International conference on machine learning, Sydney, NSW, 2017. 1126–35.
38. Herrero-Zazo M, Segura-Bedmar I, Martínez P, Declerck T. The DDI corpus: an annotated corpus with pharmacological substances and drug-drug interactions. *J Biomed Inform*. 2013;46(5):914–20. <https://doi.org/10.1016/j.jbi.2013.07.011> PMID: 23906817
39. Demner-Fushman D, Fung K, Do P. The DDI corpus: an annotated corpus with pharmacological substances and drug–drug interactions. In: Proceedings of Text Analysis Conference, Gaithersburg, USA; 2018. p. 1–10.

40. Huang D, Jiang Z, Zou L, Li L. Drug–drug interaction extraction from biomedical literature using support vector machine and long short term memory networks. *Information Sciences*. 2017;415–416:100–9. <https://doi.org/10.1016/j.ins.2017.06.021>
41. Mostafapour V, Dikenelli O. Attention-wrapped hierarchical blstms for ddi extraction. *arXiv preprint* 2019. <https://doi.org/10.48550/arXiv.1907.13561>
42. Hong L, Lin J, Tao J. BERE: an accurate distantly supervised biomedical entity relation extraction network. *arXiv preprint* 2019. <https://doi.org/10.48550/arXiv.1906.06916>
43. Liu J, Huang Z, Ren F, Hua L. Drug-drug interaction extraction based on transfer weight matrix and memory network. *IEEE Access*. 2019;7:101260–8. <https://doi.org/10.1109/access.2019.2930641>
44. Sun X, Dong K, Ma L, Sutcliffe R, He F, Chen S, et al. Drug-drug interaction extraction via recurrent hybrid convolutional neural networks with an improved focal loss. *Entropy (Basel)*. 2019;21(1):37. <https://doi.org/10.3390/e21010037> PMID: 33266753
45. Asada M, Miwa M, Sasaki Y. Using drug descriptions and molecular structures for drug-drug interaction extraction from literature. *Bioinformatics*. 2021;37(12):1739–46. <https://doi.org/10.1093/bioinformatics/btaa907> PMID: 33098410
46. Nguyen DP, Bao Ho T. Drug-drug interaction extraction from biomedical texts via relation BERT. In: 2020 RIVF International Conference on Computing and Communication Technologies (RIVF). 2020. p. 1–7. <https://doi.org/10.1109/rivf48685.2020.9140783>
47. Wu H, Xing Y, Ge W, Liu X, Zou J, Zhou C, et al. Drug-drug interaction extraction via hybrid neural networks on biomedical literature. *J Biomed Inform*. 2020;106:103432. <https://doi.org/10.1016/j.jbi.2020.103432> PMID: 32335223
48. Zaikis D, Vlahavas I. Drug-drug interaction classification using attention based neural networks. In: 11th Hellenic Conference on Artificial Intelligence, 2020. 34–40. <https://doi.org/10.1145/3411408.3411461>
49. He H, Chen G, Yu-Chian Chen C. 3DGT-DDI: 3D graph and text based neural network for drug-drug interaction prediction. *Brief Bioinform*. 2022;23(3):bbac134. <https://doi.org/10.1093/bib/bbac134> PMID: 35511112
50. Huang L, Lin J, Li X, Song L, Zheng Z, Wong K-C. EGFI: drug-drug interaction extraction and generation with fusion of enriched entity and sentence information. *Brief Bioinform*. 2022;23(1):bbab451. <https://doi.org/10.1093/bib/bbab451> PMID: 34791012
51. Fatehifar M, Karshenas H. Drug-drug interaction extraction using a position and similarity fusion-based attention mechanism. *J Biomed Inform*. 2021;115:103707. <https://doi.org/10.1016/j.jbi.2021.103707> PMID: 33571676
52. Chen J, Sun X, Jin X, Sutcliffe R. Extracting drug-drug interactions from no-blinding texts using key semantic sentences and GHM loss. *J Biomed Inform*. 2022;135:104192. <https://doi.org/10.1016/j.jbi.2022.104192> PMID: 36064114
53. Deng H, Li Q, Liu Y, Zhu J. MTMG: a multi-task model with multi-granularity information for drug-drug interaction extraction. *Heliyon*. 2023;9(6):e16819. <https://doi.org/10.1016/j.heliyon.2023.e16819> PMID: 37484258
54. Zhang T, Yu C, Zhang S. CA-SQBG: cross-attention guided Siamese quantum BiGRU for drug-drug interaction extraction. *Comput Biol Med*. 2025;186:109655. <https://doi.org/10.1016/j.compbiomed.2025.109655> PMID: 39864333
55. Jia Y, Wang Z, Zhang H, Li P, Xie P, Yuan Z. Hierarchical feature modeling with data augmentation and focal loss for drug–drug interaction extraction. *Biomedical Signal Processing and Control*. 2025;110:108199. <https://doi.org/10.1016/j.bspc.2025.108199>
56. Tang S, Zhang Q, Zheng T. Two step joint model for drug drug interaction extraction. *arXiv preprint* 2020. <https://doi.org/arXiv:2008.12704>
57. Yang J, Ding Y, Long S, Poon J, Han SC. DDI-MuG: multi-aspect graphs for drug-drug interaction extraction. *Front Digit Health*. 2023;5:1154133. <https://doi.org/10.3389/fgdth.2023.1154133> PMID: 37168529
58. KafiKang M, Hendawi A. Drug-drug interaction extraction from biomedical text using relation BioBERT with BLSTM. *MAKE*. 2023;5(2):669–83. <https://doi.org/10.3390/make5020036>
59. Hu H, Yang A, Deng S. Drug-drug interaction extraction from biomedical text using relation BioBERT with BLSTM. *Machine Learning and Knowledge Extraction*. 2025;273:126953.
60. Yang Y, Hu L, Li G, Li D, Hu P, Luo X. Link-based attributed graph clustering via approximate generative bayesian learning. *IEEE Trans Syst Man Cybern, Syst*. 2025;55(8):5730–43. <https://doi.org/10.1109/tsmc.2025.3572738>
61. Su X, Hu P, Li D, Zhao B, Niu Z, Herget T, et al. Interpretable identification of cancer genes across biological networks via transformer-powered graph representation learning. *Nat Biomed Eng*. 2025;9(3):371–89. <https://doi.org/10.1038/s41551-024-01312-5> PMID: 39789329