
Ziyan Wang †, Zhankun Xiong †, Feng Huang, Xuan Liu and Wen Zhang ∗
College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
{wangziyan, xiongzk, fhuang233, lx666}@webmail.hzau.edu.cn, zhangwen@mail.hzau.edu.cn

Abstract

Drug-drug interactions (DDIs) can result in various pharmacological changes, which can be categorized into different classes known as DDI events (DDIEs). In recent years, previously unobserved/unseen DDIEs have been emerging, posing a new classification task when unseen classes have no labelled instances in the training stage, which is formulated as a zero-shot DDIE prediction (ZS-DDIE) task. However, existing computational methods are not directly applicable to ZS-DDIE, which has two primary challenges: obtaining suitable DDIE representations and handling the class imbalance issue. To overcome these challenges, we propose a novel method named ZeroDDI for the ZS-DDIE task. Specifically, we design a biological semantic enhanced DDIE representation learning module, which emphasizes the key biological semantics and distills discriminative molecular substructure-related semantics for DDIE representation learning. Furthermore, we propose a dual-modal uniform alignment strategy to distribute drug pair representations and DDIE semantic representations uniformly in a unit sphere and align the matched ones, which can mitigate the issue of class imbalance. Extensive experiments showed that ZeroDDI surpasses the baselines and indicate that it is a promising tool for detecting unseen DDIEs. Our code has been released in https://github.com/wzy-Sarah/ZeroDDI.

1 Introduction

Taking multiple drugs concurrently can cause drug-drug interactions (DDIs), which may lead to various pharmacological changes [Vilar et al., 2014]. According to the different specific consequences, the huge number of DDIs can be categorized into hundreds of classes [Ryu et al., 2018], named DDI events (DDIEs), as shown in Figure 1(a). In recent years, researchers have paid much attention to DDIE prediction, i.e., classifying a DDI into a specific DDIE, which is helpful for researchers and clinicians to investigate the mechanism behind the consequences of polypharmacy. Note that to avoid confusion, we will use the term “drug pairs” instead of DDIs in this paper.

With further clinical observation and the development of therapeutic drugs, the number of DDIEs is increasing, as shown in Figure 1 (b). Moreover, inspired by [Deng et al., 2020; Feng et al., 2023], we notice that the DDIE contains three key concepts: the effect (such as “bradycardia”), the “increased” or “decreased” pharmacological changes, and the directional interaction between two drugs (which is reflected in the sentence pattern). To simplify, we term them as three kinds of attributes, i.e., Effect, Sign, and Pattern, respectively. We observed that annotated DDIEs (seen DDIEs) are less than all compositions of three kinds of attributes, as shown in Figure 1 (c), thus we can infer the existence of unseen DDIEs without labelled drug pairs. Predicting unseen DDIEs has a high practical value because novel drug pairs may come from unseen DDIEs. We formulate classifying drug pairs of unseen DDIEs as a zero-shot DDIE prediction (ZS-DDIE) task.

Although many DDIE prediction methods have been proposed, they all neglect the ZS-DDIE task. When no trainable...
instances are used to transfer knowledge from seen to unseen classes, the semantic information (or auxiliary information) is needed to be attached to every class for constructing semantic relatedness between seen and unseen classes. However, most DDIE prediction methods view DDIEs as either labels [Zhu et al., 2023], random initialized vectors [Lin et al., 2021] or one-hot vectors [Zhu et al., 2022a], which do not have realistic semantic significance. Therefore, they cannot be applied to the ZS-DDIE task. An intuitive solution for ZS-DDIE is to learn the experience from zero-shot learning (ZSL) classification methods, which are developed by computer vision. There is a commonly used compatibility framework in ZSL [Xian et al., 2018], which transfers knowledge from seen to unseen classes by projecting all classes into a common semantic space, and realizes the unseen class prediction by aligning the representations of instances and matched classes. Nevertheless, applying this framework to the ZS-DDIE task needs to resolve two major challenges. The first challenge is to obtain suitable DDIE representations, which are required to construct reasonable relatedness between different classes. For example, based on DDIE textual descriptions, the DDIE representations with the same Pattern may be closer than that with the same Effect in semantic space (as shown in Appendix A), which is inconsistent with the biological significance and will hinder the prediction performance. The second challenge is the class imbalance. Researchers have revealed that the top three most frequent DDIEs contain more than half of all instances [Deng et al., 2021], while a great number of DDIEs have less than 10 instances. The class imbalance problem may lead to unclear decision boundaries of classification, causing a decline of discriminative ability for the prediction of classes with few or even no instances [Ji et al., 2022].

In this work, we propose a novel method, called ZeroDDI, for zero-shot DDIE prediction and resolve the challenges above. For the first challenge, we design a biological semantic enhanced DDIE representation learning module (BRL) to obtain suitable DDIE representation. Since Effect attributes are the key biological semantics, BRL first extracts attribute-level semantics of Effect and together with class-level semantics of DDIE textual descriptions. Then, considering that molecular substructure plays an important role in drug properties [Jin et al., 2023], BRL establishes the fine-gained interaction between substructures and semantic tokens of text to guide the fusion of bi-level semantics, which can distill discriminative semantics for learning DDIE representations. For the second challenge, we design a dual-modal uniform alignment strategy (DUA) to enable structure and text modal representations (i.e., drug pair and DDIE representations) to be uniformly distributed in unit sphere, aiming to handle the class imbalance issue and thus improve the discrimination of the model.

The extensive experiments show that ZeroDDI can achieve superior performances in the ZS-DDIE task than baselines and reveal that ZeroDDI is a promising tool for detecting unseen DDIEs.

2 Related Work

2.1 Drug-Drug Interaction Event Prediction

Existing DDIE prediction methods can be classified into three categories: DNN-based methods [Deng et al., 2021; Lin et al., 2021], Tensor factorization-based methods [Jin et al., 2017; Yu et al., 2022], and GNN-based methods [Zhu et al., 2022a; Lv et al., 2023; Xiong et al., 2023]. However, none of the DDIE prediction methods are specifically tailored to tackle the ZS-DDIE task. Most of them cannot be applied to the ZS-DDIE task in that they neglect to encode DDIE or lack semantic significance in their DDIE embeddings [Zhu et al., 2023; Zhu et al., 2022a; Lin et al., 2022]. We find that 3DGT-DDI [Lv et al., 2023], which utilized a pre-trained language model (SCIBERT) to learn DDIE representations from textual descriptions, shows potential adaptability for handling the ZS-DDIE task. Despite this, zero-shot DDIE prediction remains uncharted territory and needs to be explored. More discussions of related work are in Appendix B.

2.2 Zero-Shot Learning

Zero-Shot learning (ZSL) methods, which were introduced in computer vision, achieve transferability by projecting all classes into a common semantic space, and facilitate the classification of instances from unseen classes through a compatibility function [Nawaz et al., 2022]. For learning semantic representation, current ZSL methods rely on three primary information sources: class names, class attributes, and class textual descriptions. Word-based methods [Frome et al., 2013; Li et al., 2015] represent the classes by simple words of the class name. Attribute-based methods represent the classes by more accurate and specific shared characteristics among classes [Lampert et al., 2013; Al-Halah and Stiefelhagen, 2017]. Textual description-based methods represent the classes by textual descriptions of classes [Reed et al., 2016; Zhu et al., 2018], which can provide more context information about classes. For constructing compatibility functions, ZSL methods utilize/developed different metric functions to pull an instance close to its matched class and push it away with a ZS-DDIE dataset.
3 Methodology
3.1 Preliminaries
Given the training data $S = \{(g_i, g_i', y_i)\}_{i=1}^I$, where each element includes a pair of drugs $(g_i, g_i')$ as an instance and a matched DDIE $y_i \in \mathcal{Y}^s$ as the label, and $\mathcal{Y}^s$ denotes the set of seen DDIE labels. Let $\mathcal{Y}^u$ denote the set of unseen DDIE labels, which is disjoint from $\mathcal{Y}^s$, i.e. $\mathcal{Y}^s \cap \mathcal{Y}^u = \emptyset$. We denote $\mathcal{T} = \{(g_j, g_j', y_j)\}_{j=1}^J$ as the testing data and $\mathcal{Y}^t$ as the set of test DDIE labels, $y_j \in \mathcal{Y}^t$. Herein we test the model in two scenarios: one is that $\mathcal{Y}^t = \mathcal{Y}^u$, i.e., all test classes come from the unseen class set, named conventional ZSL (CZSL); the other is that $\mathcal{Y}^t = \mathcal{Y}^s \cup \mathcal{Y}^u$, $\mathcal{Y}^s \subseteq \mathcal{Y}^u$, i.e., the test classes come from both seen class set and unseen class set, named generalized ZSL (GZSL).

3.3 Biological Semantic Enhanced DDIE Representation Learning (BRL)
To obtain DDIE representations, we design a BRL module, which is composed of two steps: class-level and attribute-level semantic extraction, and substructure-guided bi-level semantic fusion (SSF). The former extracts the bi-level semantics of DDIEs from the attribute-level and class-level text. The latter adaptive fuses the bi-level semantic to obtain DDIE representations, preserving discriminative information as much as possible.
Class-level and attribute-level semantic extraction. For the class-level texts, we extract the DDIE textual descriptions from the DrugBank database [Wishart et al., 2018]. For the attribute-level texts, we utilize a StanfordNLP tool [Qi et al., 2019] to locate the Effect words in every DDIE textual description, and we extract the textual description of the Effect from the MeSH database\(^1\). After that, we employ a language model BioBERT [Lee et al., 2020] pre-trained on large-scale biomedical corpora, to get the token features \( \{ t^c_m \in \mathbb{R}^{d_c} \}_{m=1}^{M} \) and \( \{ t^a_i \in \mathbb{R}^{d_a} \}_{i=1}^{L} \) from class-text and attribute-text, respectively, where \( M \) and \( L \) are the numbers of the tokens and \( d_{c} \) denote the token feature dimension. Since the attribute is part of a DDIE, we take the average over all the rows of \( R \) related token semantics, where \( R \) is used to normalize the features that come from two-level semantic sources.

Substructure-guided bi-level semantic fusion (SSF). Different from averaging the tokens \( T \) directly, we utilize a cross-modality attention mechanism to fuse the tokens discriminatively, i.e., \( \mathbb{R}^{(M+1) \times d_t} \rightarrow \mathbb{R}^{1 \times d_t} \). In particular, we take the molecular substructure embeddings of drug pair \( P \in \mathbb{R}^{N \times d_n} \) (learned from Drug Pair Encoder) to establish the fine-grained interaction with \( T \), and distill the substructure-related token semantics, where \( N \) is the number of substructures. More specifically, \( P \) and \( T \) are transferred to:

\[
Q = PW_Q, K = TW_K, V = TW_V
\]

where \( Q \in \mathbb{R}^{N \times d_r} \) denotes the queries, \( K \in \mathbb{R}^{(M+1) \times d_r} \), and \( V \in \mathbb{R}^{(M+1) \times d_r} \) denote keys and values. \( W_Q, W_K \) and \( W_V \) are learnable linear transformations. Then the cross-modality attention can be calculated by:

\[
A = \text{softmax} \left( \frac{QK^T}{\sqrt{d_r}} \right), O = AV,
\]

where \( A \in \mathbb{R}^{N \times (M+1)}, O \in \mathbb{R}^{N \times d_r} \). Finally, averaging over all the rows of \( O \), we can obtain the DDIE representation \( z \in \mathbb{R}^{d_r} \).

3.4 Dual-Modal Uniform Alignment (DUA)

We introduce a DUA as the compatibility function, which includes an alignment function to align the drug pair representations and its matched DDIE representations for classification, and a dual-modal uniformity loss to constraint both modal representations uniformly distributed in unit sphere.

To realize the classification of drug pairs, we use the dot-product to measure the similarity of the drug pair representation with all DDIE semantic representations and then employ contrastive loss [Han et al., 2021] to enforce the drug pair to have the highest similarity score with its matched DDIE:

\[
L_{\text{ALIGN}} = - \sum_{i=1}^{t} \log \frac{\exp(h_i \cdot z_{i}^{y_i}/\tau)}{\sum_{j \in Y_i} \exp(h_i \cdot z_{i}^{y_i}/\tau)}
\]

where \( h_i \) denotes the current drug pair representation, \( z_{i}^{y_i} \) denotes the matched DDIE representation of \( h_i \), \( z_{i}^{y_i} \in \{z_{i}^{1}, z_{i}^{2}, ..., z_{i}^{|Y_i|}\} \) denotes the \( j \)-th DDIE representations of all DDIE representations corresponding to \( h_i \). \( \tau \) is a temperature coefficient.

To realize the uniform distribution, i.e., maximize the representation distance between classes in unit sphere space, inspired by [Lu et al., 2022], we take the center of all the DDIE representations as the center of the sphere. Then we utilize the class uniformity loss \( L_{\text{cla}} \) to push every DDIE representation to unit sphere:

\[
L_{\text{cla}} = \frac{1}{|Y|} \sum_{i=1}^{t} \left[ \sum_{j \in Y_i} \left( \frac{\|z_{i}^{j} - c_i\|}{\|z_{i}^{j}\|} \right \| \frac{\|z_{i}^{k} - c_i\|}{\|z_{i}^{k}\|} \right) \right]
\]

where \( c_i = 1/|Y_i| \cdot \sum_{j=1}^{|Y_i|} z_{i}^{j} \) and \( z_{i}^{j} \) denotes the center of all DDIE representations of drug pair \( i \). \( k \in Y \setminus \{j\} \) means sampling all DDIEs except the \( j \)-th DDIE. This loss function uses the cosine similarity to pull every distance between the DDIE representations and the center representation approaching to same. Thus all DDIE representations are unified and distributed on a unit sphere.

To make the drug pair representation get consistency constraints with DDIE representations. We introduce an instance uniformity loss \( L_{\text{ins}} \) to unify every drug pair \( i \) to the sphere with the center of \( c_i \):

\[
L_{\text{ins}} = \frac{1}{|Y|} \sum_{i=1}^{t} \left[ \frac{1}{|S\{i\}|} \sum_{k \in S\{i\}} \left( \frac{\|h_i - c_i\|}{\|h_k - c_i\|} \right \| h_k - c_i \right) \right]
\]

where \( k \in S\{i\} \) is the sample randomly drawn from a batch set of training data excluding \( i \). According to Eq. (5) and Eq. (6), the dual-modal uniformity loss \( L_{\text{UNI}} \) can be expressed as:

\[
L_{\text{UNI}} = L_{\text{cla}} + L_{\text{ins}}
\]

3.5 Training and Prediction

For training our model ZeroDDI, we optimize the total loss that combines Eq. (4) and Eq. (7):

\[
L = L_{\text{ALIGN}} + \lambda L_{\text{UNI}}
\]

where \( \lambda \) is hyper-parameters.

For the zero-shot DDIE prediction, we use the trained Drug Pair Encoder to obtain drug pair representation \( h \). Its DDIE \( \hat{y} \)
can be predicted by researching the highest dot-product similarity between $h$ and all DDIE representations in test set:

$$\hat{y} = \arg \max_{j \in \mathcal{Y}^u} h \cdot z^j,$$

(9)

where $\mathcal{Y}^u$ is the set of test DDIEs, $z^j$ denotes the $j$-th DDIE representation learned by the trained BRL.

4 Experiments

4.1 Experimental Setup

Datasets

DrugBank [Wishart et al., 2018] is a widely used data source for DDIE prediction, which contains drugs, DDIs, textual descriptions of DDIEs, and the molecular structures of drugs. Based on raw data from DrugBank v5.1.9, we construct a ZS-DDIE dataset by annotating attributes for every DDI manually, which contains a total of 2,004 approved drugs, 394,118 DDIs, 175 DDIEs with unique textual descriptions, 2 Signs, 3 Patterns and 114 Effects with corresponding attribute-level text. We rank the classes of DDIEs in the order of the number of instances from more to less and split DDIEs into seen classes and unseen classes, in which we take the less number of DDIEs as unseen classes to simulate the real situation of DDIE that are newly observed. 107 DDIEs are viewed as seen classes $\mathcal{Y}^s$, which contains all attributes, and the rest 68 DDIEs are viewed as unseen classes $\mathcal{Y}^u$. Additionally, every DDI is associated with a DDIE in this dataset. Drug pairs in the dataset are asymmetric, in that the roles of drugs are different in one DDIE [Feng et al., 2023]. More details of the dataset can be found in Appendix D.1.

Baselines

As mentioned in the related work, only one DDIE method (3DGT-DDI) has the potential to be applied to the ZS-DDIE task. In addition, under the compatibility framework, we also compare several popular compatibility functions in computer vision and equip them with two kinds of DDIE semantic representations. One learned from a binary vector of attributes (namely attribute-based representation) and the other learned from a PLM embedding of class textual description (namely class-based representation), respectively. These baselines have the same Drug Pair Encoder as our method.

- **3DGT-DDI** [Lv et al., 2023]. The DDIE semantic encoder of 3DGT-DDI is composed of a PLM and a CNN. To enable this model to deal with the ZS-DDIE task, we replace its binary classifier with a compatibility function incorporating cross-entropy loss.

- **ZSLHinge**. ZSLHinge utilizes a hinge rank loss from the classic ZSL method DeViSE [Frome et al., 2013], which can produce a higher dot-product similarity between an instance and the matched class than between the instance and other randomly chosen classes by a margin.

- **ZSLCE**. ZSLCE utilizes the cross-entropy loss from [Chen et al., 2023] and [Sumbul et al., 2017], which maximizes the similarity score of the instance and its matching class.

- **ZSLTriplet**. ZSLTriplet utilizes a ZSL triplet loss from [Cacheux et al., 2019], which improves the standard triplet loss by adding a flexible semantic margin, partial normalization, and relevance weighting.

More details of baselines can be found in Appendix D.2.

Evaluation Protocols

The experiments are performed on two common zero-shot learning scenarios: CZSL and GZSL, which are described in Preliminaries. We train the model on data from seen classes. In the CZSL scenario, we validate and test the model in 3-fold of the unseen classes, i.e., in each fold, one-third of unseen classes are used for the validation, and the rest are used for the testing. In the GZSL scenario, to evaluate the performance of seen and unseen prediction simultaneously, a part of the seen classes and instances are added into the GZSL validation and testing set. Note that all instances from seen classes in the validation and testing set are not included in the training set.

Metrics

In the CZSL scenario, the performance is measured with top-k accuracy (k=1, 3, 5 and denoted as $\text{acc}^s_{k=1}$, $\text{acc}^s_{k=3}$ and $\text{acc}^s_{k=5}$). In addition, we calculate the average of top-1 accuracy across all unseen classes, denoted as $\text{acc}^u$. In the GZSL scenario, we calculate the $\text{acc}^u_{s}$ or $\text{acc}^u_{u}$ on seen classes (denoted as $\text{acc}^u_{s,\text{ave}}$ or $\text{acc}^u_{u,\text{ave}}$) and unseen classes, and calculate the harmonic mean of them: $\text{acc}^u = \frac{2 \cdot (\text{acc}^u_{s,\text{ave}} \cdot \text{acc}^u_{u,\text{ave}})}{\text{acc}^u_{s,\text{ave}} + \text{acc}^u_{u,\text{ave}}}$ [Xian et al., 2018]. In addition, $\text{acc}^u_{s,\text{ave}}$ or $\text{acc}^u_{u,\text{ave}}$ is used to evaluate the seen-unseen binary classification results. We also utilize the proportion $P^u = \text{acc}^u_{s,\text{ave}}/\text{acc}^u_{u,\text{ave}}$ to evaluate the multi-class classification performance under the correct binary classification.

The implementation details, hyper-parameter sensitivity analysis, and model configuration can be found in Appendix D due to space constraints.

4.2 Comparison with Baselines

We compare ZeroDDI with baselines in both CZSL and GZSL scenarios, and the results are shown in Table 1. Overall, ZeroDDI achieves the best performances on all unseen metrics (i.e., $\text{acc}^u$, $P^u$ and $\text{acc}^u$), which indicates that our method has a superior ability to deal with unseen DDIE prediction problem. Specifically, we have the following observations: (1) Compared with 3DGT-DDI, our method has salient advantages in the ZS-DDIE task, as evidenced by the 9.32% performance gain on the unseen metrics $\text{acc}^u_{u,\text{ave}}$ of CZSL and 7.92% performance gain on $\text{acc}^u_{u,\text{ave}}$ of GZSL. This not only shows the promising unseen prediction performance of our method but also underscores the necessity of specifically tailoring methods for ZS-DDIE tasks. (2) Comparing the performances of class-based and attribute-based models, we find that, in most cases (i.e., ZSLHinge, ZSLCE, and ZeroDDI), the class-based representations perform better. This demonstrates that, in the ZS-DDIE task, the transferability of class-based representations learned by a PLM may be greater than that of attribute-based representations encoded by the shared attributes among classes. Thus our BRL module is constructed based on the PLM. (3) ZeroDDI
outperforms ZeroDDI in the CZSL and GZSL scenarios. Figure 3: Performance (in %) comparisons of ZeroDDI with baselines in the CZSL and GZSL scenario. The best and suboptimal results are highlighted in bold and underline, respectively. Note that Model1 (or Model2) denotes using attribute-based (or class-based) representations learned from a binary vector of attributes (or the PLM of class textual description where the semantic encoder is the same as that in ZeroDDI).

4.3 Ablation Study

We conduct an ablation study to validate the contribution of important components in ZeroDDI, i.e., attribute-level semantics, the SSF component, and dual-modal uniformity loss. The results are shown in Figure 3. Specifically, (1) after removing the input of attribute-level semantics (w/o Attri.), the performance drops, which indicates that the attribute-level semantics extracted by our method can improve the performance of the model for the ZS-DDIE task. We will further discuss this in Section 4.4. (2) The model shows a significant decline after replacing the fusion component SSF with the average operation of bi-level token-wise features (w/o SSF), which proves the necessity of selective bi-level fusion and the effectiveness of molecular substructure for distilling discriminative semantic information. (3) Removing the dual-modal uniformity loss (w/o $L_{UNI}$) or any one part of this loss function (w/o $L_{cla}$ or w/o $L_{ins}$) undermines the accuracy, demonstrating that the uniformity loss assist in improve the discriminative ability, and $L_{cla}$ and $L_{ins}$ are all useful for ZS-DDI task. We will further discuss this in Section 4.5.

PLM | Attri. Level | Semantic | CZSL | GZSL
--- | --- | --- | --- | ---
| | $\text{acc^u_{ave}}$ | $\text{acc^u_{@1}}$ | $\text{acc^u_{@3}}$ | $\text{acc^u_{@5}}$ | $\text{acc^H_{ave}}$ | $\text{acc^H_{@1}}$ | $\text{P^u}$ | $\text{P^H}$
| SCIBERT ✓ | 13.79 | 12.47 | 21.31 | - | 10.84 | 8.58 | 14.65
| BioBERT ✓ | 17.57 | 16.13 | 26.48 | - | 14.66 | 11.72 | 20.21

Table 2: Performance (in %) comparisons of ZeroDDI with its variants that replace the BioBERT by different PLM models in CZSL and GZSL scenarios.

4.4 Effectiveness of Bio-Enhanced Representation Learning

In this section, we further discuss the effectiveness of two components (attribute-level semantics and SSF) of BRL in unseen DDIE prediction.

For attribute-level semantics, we compare the variants of ZeroDDI (by replacing the BioBERT with SCIBERT [Beltagy et al., 2019] or PubMedBERT [Gu et al., 2022] whose versions are shown in Appendix D) equipped with/without the attribute-level semantic information. Table 2 shows that our attribute-level semantics can provide additional stable gains of transferability, leading to better performance in unseen DDIE prediction no matter which PLM is based. For SSF, we visualize an example to verify that SSF can con-
struct a reasonable relationship between molecular substructures and the bi-level text tokens. We randomly select two drug pairs from the hits of a DDIE in the test set, then visualize them and highlight one word "hypertensive" with its most relevant substructures in Figure 4. As the website \(^2\) reports that "coadministration with adrenergic agonists with hypertensive potential may lead to increased cardiac output and blood pressure due to the vasoconstricting effects of ergot derivatives". In the example, Cabergoline and Dihydroergotamine are ergot derivatives, and the SSF can highlight the part of their parent structure of ergot derivatives, which proves the effectiveness of the constructed relationship between tokens and substructures and indicate the potential interpretability of SSF.

### 4.5 Effectiveness of Dual-Modal Uniform Distribution

We further discuss the effectiveness of dual-modal uniformity loss for the class imbalance issue in seen and unseen DDIE prediction by comparing ZeroDDI with ZeroDDI (w/o \(\mathcal{L}_{\text{UNI}}\)). To simulate the relatively balanced scenario, we reconstruct the training set with the imbalance ratio \(\rho\) is 1:100, where \(\rho\) is defined as the number of samples in the most frequent class divided by that of the least frequent class [Li et al., 2022] (here we denote least frequent class has ten instances at least). We randomly select five unseen and seen DDIE classes from the results of the test set, respectively, and visualize them in Figure 5. From the results, we have the following observations: (1) For unseen DDIE prediction, the drop of accuracy from the scenario \(\rho = 1:100\) to \(1:10000\) shows that the class imbalance has an impact on unseen DDIE prediction to some extent, which demonstrates that the necessity of tackle class unbalance challenge in unseen DDIE prediction. (2) The class imbalance can cause a drop of performance in both seen and unseen classes and lead to unclear and inseparable decision boundaries as shown in the right column of Figure 5(a) and (b), while our method can decelerate this problem. This indicates that \(\mathcal{L}_{\text{UNI}}\) can improve the discriminability of the model and mitigate the adverse effects of class imbalance.

\(^2\) https://go.drugbank.com/drug-interaction-checker

### 4.6 Zero-Shot DDIE Application Analysis

In this section, we conduct an application analysis to validate the factuality of the zero-shot DDIE setting in this work and evaluate the practical capabilities of ZeroDDI. Firstly, we choose the latest dataset (i.e., DrugBank v5.1.11) as the Novel dataset and take the dataset used in this work as the Existing Dataset. Through the data process for the Novel dataset, we find that, in the Novel Dataset, 6 DDIEs are composed of the attributes from the Existing Dataset but are not included in the Existing Dataset. It not only proves that novel DDIEs are increasing but also proves that novel DDIEs can be composed of existing attributes and further indicates the practical significance of our evaluation setting. Then, we use the Existing Dataset to train the ZeroDDI and use the trained model to predict 6 novel DDIEs in the Novel Dataset. When there are no labelled instances in the training set, ZeroDDI can achieve 61.11% in \(\text{acc}_{\text{ave}}\), which shows the application ability of our method. More details and experiments are shown in Appendix E.

### 5 Conclusion

This is the first work that pays attention to the zero-shot DDIE prediction and proposes a novel method, called ZeroDDI, to predict zero-shot DDIEs. Specifically, we designed a biological semantic enhanced DDIE representation learning module to learn suitable DDIE representations containing enhanced key biological semantic and substructure-guided discriminative semantics, leading to better knowledge transferability from seen DDIEs to unseen DDIEs. Moreover, we design a dual-modal uniformity alignment strategy to uniform the distribution of drug pair representations and DDIE representations in unit sphere and thus mitigate the class imbalance issue. The extensive experiments show that ZeroDDI can produce superior performance in zero-shot DDIE prediction, our designed module can effectively promote the performances, and ZeroDDI is a promising tool for practice application.
Acknowledgments
This work was supported by the National Natural Science Foundation of China (62372204, 62072206, 61772381, 62102158); Huazhong Agricultural University Scientific & Technological Self-innovation Foundation; Fundamental Research Funds for the Central Universities (2662021JC008, 2662022JC004). The funders have no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Contribution Statement
Wen Zhang is the corresponding author of this study. Ziyan Wang is responsible for the paper writing, data annotating, experimental code, figure drawing, and experimental design. Zhankun Xiong helped with paper writing, figure drawing, and experimental design, which contributed equally to this paper. Feng Huang provided help in paper writing and Xuan Liu provided help in data annotating.

References
[18] Li et al., 2015 Xiaoyong Du, Gang Yang. Zero-shot image tagging by


