Computers in Biology and Medicine 171 (2024) 108104

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Enhancing drug–food interaction prediction with precision representations through multilevel self-supervised learning

Jinhang Wei [a,1], Zhen Li [b,1], Linlin Zhuo [a], Xiangzheng Fu [c], Mingjing Wang [a], Keqin Li [c,d], Chengshui Chen [e,f,*]

[a] Wenzhou University of Technology, Wenzhou, 325000, China
[b] Institute of Computational Science and Technology, Guangzhou University, Guangzhou, 510006, China
[c] College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410006, China
[d] Department of Computer Science, State University of New York, New York, 12561, USA
[e] Department of Pulmonary and Critical Care Medicine, Quzhou People's Hospital, The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou, 324000, China
[f] Key Laboratory of Interventional Pulmonology of Zhejiang Province, Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, 325000, China

ARTICLE INFO

Keywords:
Drug–food interaction
Self-supervised learning
Enhanced representations quality
Feature alignment
Domain separation

ABSTRACT

Drug–food interactions (DFIs) crucially impact patient safety and drug efficacy by modifying absorption, distribution, metabolism, and excretion. The application of deep learning for predicting DFIs is promising, yet the development of computational models remains in its early stages. This is mainly due to the complexity of food compounds, challenging dataset developers in acquiring comprehensive ingredient data, often resulting in incomplete or vague food component descriptions. DFI-MS tackles this issue by employing an accurate feature representation method alongside a refined computational model. It innovatively achieves a more precise characterization of food features, a previously daunting task in DFI research. This is accomplished through modules designed for perturbation interactions, feature alignment and domain separation, and inference feedback. These modules extract essential information from features, using a perturbation module and a feature interaction encoder to establish robust representations. The feature alignment and domain separation modules are particularly effective in managing data with diverse frequencies and characteristics. DFI-MS stands out as the first in its field to combine data augmentation, feature alignment, domain separation, and contrastive learning. The flexibility of the inference feedback module allows its application in various downstream tasks. Demonstrating exceptional performance across multiple datasets, DFI-MS represents a significant advancement in food presentations technology. Our code and data are available at https://github.com/kkkayle/DFI-MS.

1. Introduction

Over recent decades, the study of drug–food interactions (DFIs) has gained significant attention. DFIs, encompassing both pharmacokinetic and pharmacodynamic interactions, can significantly alter the effectiveness and safety of drugs by affecting their absorption, distribution, metabolism, and excretion [1]. Understanding these interactions is crucial for patient safety and improving treatment outcomes. Clinical physicians can customize drug treatment plans based on patients' dietary habits and nutritional needs, integrating this knowledge into patient care. This approach is especially important for the treatment of medications with a narrow therapeutic index and chronic diseases with a significant diet-related aspect [2]. Patient education on DFIs is also a key aspect, guiding them on when to take medications relative to meals and which foods to avoid. This comprehensive understanding of DFIs helps in minimizing adverse effects, improving medication adherence, and overall, enhances treatment outcomes.

In recent years, as scientific research has deepened, people have gained a more detailed understanding of drug–food interactions. For example, resveratrol is a common component of common foods, found in grapes, blueberries, and some nuts and vegetables. Resveratrol inhibits the activity of various drug-metabolizing enzymes, such as CYP1A1 [3]. Grapefruit and certain cholesterol-lowering drugs (such as statins):

* Corresponding author.
E-mail addresses: zhuoninnin@163.com (L. Zhuo), fxz326@hnu.edu.cn (X. Fu), wangmingjing.style@gmail.com (M. Wang), chenchengshui@wmu.edu.cn (C. Chen).

[1] Jinhang Wei and Zhen Li contributed equally to this work.

https://doi.org/10.1016/j.compbiomed.2024.108104
Received 2 December 2023; Received in revised form 27 January 2024; Accepted 4 February 2024
Available online 6 February 2024

grapefruit juice may increase the plasma concentration of statins, thereby increasing the risk of muscle damage and liver damage [4]. High-potassium foods (such as bananas, oranges, and tomatoes) and potassium-sparing diuretics (such as amiloride, spironolactone) may cause hyperkalemia, leading to arrhythmia [5].

Our current understanding of DFIs relies on costly and less efficient clinical studies [6]. These studies use traditional experimental design methods, often requiring extensive time to ensure the reliability of the results [7]. Additionally, they often struggle to fully capture the complexity of biomarkers and the diversity of patient characteristics, limiting their application in precision medicine and the design of personalized treatment plans. This is particularly significant in clinical settings where the understanding of DFIs is essential for optimizing patient care. Healthcare professionals rely on these studies to guide prescription practices, considering how certain foods can alter the effectiveness and metabolism of medications. In particular, new drugs often require extensive testing to determine their safety before clinical use [8], including DFIs. Drug–food interactions have prompted the medical industry to pay attention to drug–food interactions. For the treatment of diseases, it is imperative for physicians to factor in the interactions between medications and food, thereby mitigating potential adverse reactions. However, with the onset of the big data era, we have been presented with unprecedented opportunities [9]. Methods grounded in molecular simulations [10] and machine learning have now been implemented across various sectors of food science [11–14], including chemical composition analysis [15] and biological transformation processes [16]. The advent of these methodologies addresses the inefficiencies of traditional techniques, hence offering invaluable insights for the evolution of the food science discipline [17].

While significant advancements have been made in computational methods across various fields, the area of DFI remains largely underexplored. Consequently, there is an urgent need to develop an effective and accurate computational approach to expedite the identification of DFIs. Such advancements would not only enhance the precision of biochemical experiments but also reduce associated costs [18]. To our knowledge, the application of deep learning models for predicting DFI relationships in this field is exceedingly limited. For instance, DFinder [19] employs a graph neural network based on drug structural similarity scoring to predict DFI relationships, demonstrating commendable performance in recommendation tasks. However, this method has room for improvement in exploring DFIs on a large scale. Other DFI computational models primarily originate from supplementary experiments on drug–drug interactions [20]. The challenge of acquiring comprehensive information about food composition not only results in high costs but also leads to the scarcity and inadequacy of DFI datasets. The majority of existing datasets lack detailed and clear descriptions of food components, presenting significant obstacles for the development and refinement of computational models, particularly in the aspect of feature engineering. This limitation underscores the need for more robust and innovative approaches in the DFI research field.

The essence of the DFI problem is still that two or more compounds undergo chemical reactions or produce certain effects on the human body through complex regulatory mechanisms. DDI research has many similarities with DFI in many respects. Compared with DFIs, DDI research has certain advantages in data availability, quality, and feature engineering. Drugs usually have relatively complete molecular structure information and pharmacokinetic parameters, which helps construct feature vectors for model use. This has great reference value for our research on DFI problems. For example, DeepDDI [20] uses the names and structural information of drug–drug or drug–food ingredient pairs as input to predict DDIs. DeepDrug [21] uses residual graph convolutional networks (RGCNs) and convolutional networks (CNNs) to learn the integrated structure and sequence representation of drugs and proteins to improve the accuracy of DDI prediction. SSI-DDI [22] proposes substructure-substructure interaction and drug–drug interaction, achieving richer feature extraction. Although these models are developed based on DDI tasks, they can also be migrated to DFI tasks to some extent. We will use their performance on DFIs as a baseline for DFI-MS. To mitigate the negative impact of insufficient feature information on model performance, we have been inspired by some methods [23–26] and have tried to mine more effective information from known DFI relationships to find multi-level feature drugs Interaction relationships with food, using these known DFI relationships to guide the embedding layer of the model, in order to obtain accurate and effective drug (food) feature representation. We divide this process into three sub-modules for training feature representations: perturbation interaction module, feature alignment and domain separation module, and inference feedback module.

The Perturbation Interaction module, underpinned by data augmentation principles, employs a masking layer to perturb Drug (Food) features, thereby constructing two sets of analogous feature representations. These similar sets are forwarded to the feature interaction layer with the objective of minimizing differences in features output from the interaction layer.

The Feature Alignment module is designed to mitigate the impact of long-tail distribution and heterogeneous datasets on feature representation and model performance. As various datasets often exhibit long-tail distributions [27], models' fitting results tend to favor high-frequency samples. Feature alignment achieves global balance among different Drug (Food) features within the same dimension, thereby alleviating these issues' impact on model performance.

The Inference Feedback module establishes an inference network by calculating Drug (Food) features within neural networks. Through loss calculation, it updates network parameters to obtain higher-quality Drug (Food) feature representations. Our primary contributions are summarized below:

1. We propose a novel multi-level feature optimization method that can extract effective information from higher-order features and be used to train high-quality drug (food) feature representations.
2. We propose DFI-MS, a contrastive learning DFI prediction model based on multi-level self-supervised feature optimization, which achieves the best performance on multiple datasets.
3. We leverage effective information extracted from higher-order features to guide the training of high-quality features. This approach offers a fresh perspective for DFI and datasets with limited data availability.

2. Materials and methods

2.1. Dataset

The data used in this study comes from the work of DFinder [19], involving two datasets: DrugBank-DFI and PubMed-DFI.

First, researchers collected DFI data from the DrugBank (v5.1.7) database, which contains specific information about each drug, such as drug interactions, pharmacological properties, chemical structures, target actions, and metabolic pathways [19,28,29]. During the process of parsing the database, researchers filtered out mutually exclusive relationships, invalid relationships, and other unclear relationships between food components and drugs and manually extracted DFI information from the text. In this process, only small-molecule drugs and food components were retained. Eventually, the DrugBank-DFI dataset contains 1784 interactions, covering 143 drugs and 213 food components [19].

In addition, researchers realized that a large amount of valuable DFI information was still hidden in biomedical literature, so they attempted to extract DFI information from PubMed publications and established the following rules:

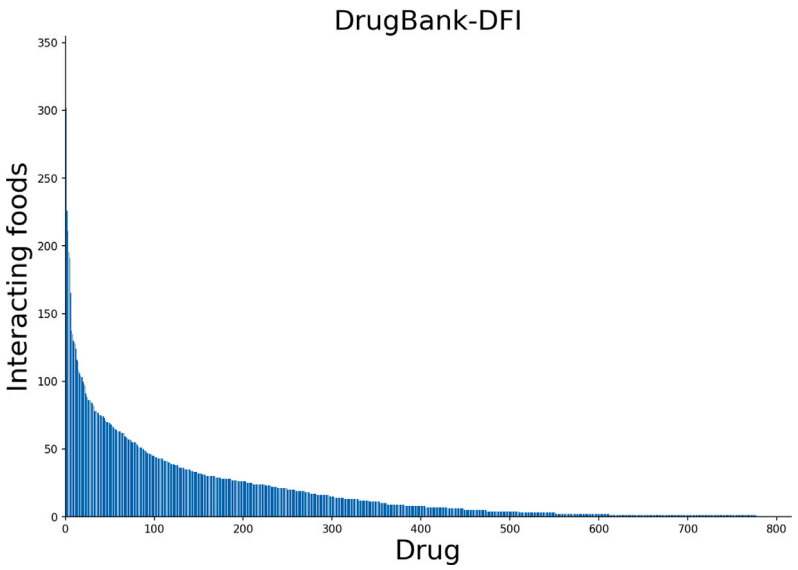

Fig. 1. The interaction distribution plot of the DrugBank dataset, in which more than half of the nodes have fewer than 20 interactions.

1. In the retrieved literature, each DFI pair should appear at least 20 times.
2. DFIs with the same food component and drug component are removed.
3. Semantic similarity is calculated between food and drug components in the same literature; if the similarity is high, it is discarded.

Under these three rules, researchers obtained the PubMed-DFI dataset, which contains 15,890 drug–food component interactions, representing 779 drugs and 818 food components.

2.2. Feature alignment and domain separation module

2.2.1. Long-tail distribution

The long-tail distribution refers to a phenomenon in a dataset where certain categories contain substantially more samples than others, resulting in most data points residing in the "tail", thereby creating a "long tail." This distribution pattern is characteristic of a majority of natural datasets.

Within the realm of deep learning, the long-tail distribution poses unique challenges. Given that some categories within the dataset contain a relatively low number of samples, the model may inadequately learn from these sparse categories during training. This often results in the model's tendency to be biased towards high-frequency categories. Despite the under-representation of these sparse categories, their cumulative total is substantial. If not properly addressed, the long-tail distribution can significantly impact overall performance [27,30,31].

Notably, the DrugBank-DFI dataset used in this experiment exhibits a pronounced long-tail distribution, as illustrated in Fig. 1.

Current methodologies in the mainstream predominantly address the issue of data imbalance [32], using techniques such as data augmentation for underrepresented categories and undersampling for overrepresented ones. While these techniques can indeed mitigate the effects of long-tail distribution, they come with their own set of challenges, including the introduction of redundant and noisy data. Specifically, the risk of overfitting looms with data augmentation, whereas undersampling could potentially lead to underfitting. Therefore, finding the sweet spot between these strategies presents a significant challenge.

2.2.2. Heterogeneous data

Heterogeneous data is defined as data exhibiting substantial variations in distribution, features, and labels. For instance, within a single dataset, there exists a degree of heterogeneity between the training set and the test set, with even greater heterogeneity present between different datasets. The presence of such heterogeneous data can adversely affect the model's performance, compromising its ability to generalize from the training set to the test set, which subsequently results in diminished generalization performance. Macroscopically, the heterogeneity can manifest in the feature representation of Drug (Food) entities. In the DFI model, the occurrence of heterogeneous data is often attributed to biased data distributions. This situation typically arises at the tail nodes of long-tail distributions and in cases where there is an extreme imbalance between negative and positive data in some samples. Due to this imbalance, the model may face difficulties in processing these data. A major problem with heterogeneous data is that they are prone to causing overfitting in the model. Overfitting arises when the model becomes excessively attuned to specific patterns and noise present in the training data, leading to the capture of idiosyncratic characteristics that are not universally applicable. This leads to the model performing well on the training set but struggling to maintain the same performance on test data because it has not effectively learned underlying principles that can be generalized to new data. A brief representation of this concept through data visualization is provided in Fig. 2. These feature values diverge significantly from the majority of the dataset, presenting a challenge for the model to adequately accommodate during the fitting process. To navigate the issue of heterogeneous data, researchers have employed strategies primarily encompassing transfer learning. Here, the model is pre-trained on the source domain and fine-tuned on the target domain (which houses the test data), facilitating the transfer of knowledge from the source domain to the target domain [33]. Multi-task learning, another popular strategy, involves the simultaneous training of multiple related tasks, thereby enabling the model to learn more generalized feature representations [34]. These methods augment the model's generalization ability with respect to heterogeneous data. However, these methods typically require a substantial amount of training data to optimize model performance. Therefore, they may not perform well in the DFI domain, which is relatively data-deficient.

2.2.3. Feature alignment

To address the challenges delineated above, we propose the deployment of a technique known as feature alignment. This technique

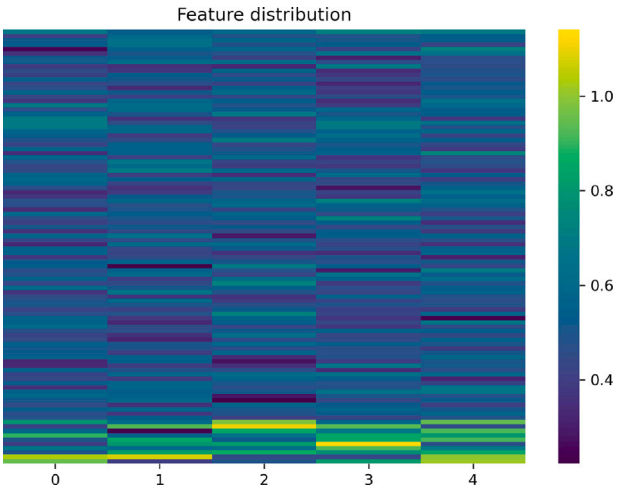

Fig. 2. The heat map representation of heterogeneous data in a low-dimensional feature space. In this heat map, the data at the bottom represent data with unique feature values, the significant differences of which indicate that these data are far apart in the feature space, highlighting their heterogeneity compared to other data.

identifies meaningful correspondences across heterogeneous datasets or disparate data representations and minimizes the distance among respective features. Consequently, feature alignment can attenuate data variance and stabilize the feature data, thereby leading to a more robust data representation [35]. In a more comprehensive context, feature alignment encourages features of identical dimensions to cluster more compactly within the feature space. This configuration facilitates the learning process of feature representation and mapping relationships, thereby enhancing the accuracy and robustness of the model. While feature alignment is a prevalent technique in fields such as computer vision and natural language processing, its application in bioinformatics is limited.

In the context of the long-tail distribution of pharmaceutical and food categories, feature alignment identifies common characteristics between head and tail samples. This approach significantly enhances the recognition of minority class samples, which are often overshadowed by high-frequency categories. Traditional deep learning models typically perform better on these high-frequency categories because the models have more opportunities to learn the features and patterns of high-frequency samples during training, optimizing their internal representations in the process. To reduce model bias towards high-frequency categories while maintaining data representativeness, this study introduces a feature alignment method. In a shared feature space, knowledge encapsulated in high-frequency nodes is used to guide the learning of low-frequency nodes. Through this method, knowledge from high-frequency samples can be transferred to low-frequency ones. This transfer of knowledge is not only reflected in the feature space but also encompasses higher levels of learning in the model, including pattern learning and relational inference. The application of feature alignment technology has effectively enhanced the model's performance in handling complex and diverse data.

In the context of heterogeneous data, feature alignment enhances the knowledge transfer process across such datasets, aiding in the discovery of shared features and mitigating the negative impacts of data inconsistency. This methodology facilitates the learning of heterogeneous data under the guidance of homogeneous data knowledge, thereby bolstering the model's ability to generalize. The learned features are, therefore, predicated on shared information rather than specific data sources, resulting in more effective decision boundaries. A schematic representation of the feature alignment process is provided in Fig. 3. In DFI-MS, feature alignment is implemented through the following loss function:

$$L_a = \sum_{i=1}^{batchsize} \left| E_i - E_O \right|_{L_2} \tag{1}$$

Where E denotes the embedding of Drug(Food), O denotes the Drug(Food) in a batch except i, and $|\cdot|_{L2}$ denotes the $L2$ distance.

2.2.4. Domain separation

Although feature alignment alleviates the impact of long-tailed distributions and heterogeneous data to some extent, it also introduces new problems. For example, feature alignment may overfit between the source and target domains, leading to decreased generalization performance in new domains or tasks. In feature space, this can be manifested as multiple nodes being tightly distributed within the same dimension of features. To address this issue and further improve the quality of Drug (Food) feature representation, we introduce the concept of domain separation. The feature vectors for drugs and food are defined as distinct domains, and our objective is to maximize the difference between these domains. This allows the model to better distinguish the patterns of different domains. A simple schematic of this domain separation is provided in Fig. 3. In DFI-MS, domain separation is implemented through the following loss function:

$$L_s = \sum_{i=1}^{batchsize} \cos\mathrm{sim}\left(F_i d, F_i f\right) \tag{2}$$

Where F represents a certain dimension of the embedding of Drug(Food), d represents the Drug, f represents the Food, and $\cos\mathrm{sim}$ represents cosine similarity.

2.3. Perturbation interaction module

2.3.1. Feature perturbation

Feature perturbation, as a data augmentation method, has been widely applied in various deep learning tasks. Although in the past feature perturbation was mainly used for supervised learning tasks, recent research results indicate that the application potential of feature perturbation in unsupervised learning environments cannot be ignored, and is even more competitive in some scenarios than supervised learning [36]. By introducing minor modifications to the original data, feature perturbation generates new data points, thereby enhancing the diversity of the dataset and aiding the model in learning more generalized features. Compared to other data augmentation methods, such as interpolation and adversarial training, the advantage of feature perturbation lies in its independence from specific data structures, enabling its application across a broader spectrum of scenarios [37].

However, to ensure the effectiveness of feature perturbation, it is imperative to accurately determine the optimal degree of perturbation. Excessive perturbation may lead to the loss of crucial features, while insufficient perturbation might not significantly enhance the model's generalization capabilities. Identifying this balance requires precise experimentation. In summary, although feature perturbation presents potential advantages in unsupervised learning, its ultimate efficacy depends on meticulous adjustment and experimental validation of the perturbation level.

In this paper, we apply feature perturbation to unsupervised learning to help the model mine effective information from feature data. We adopt a random feature perturbation algorithm, with the formula as follows:

$$\hat{E} = E \cdot I, I \sim \mathrm{Bernoulli}(p) \tag{3}$$

Where $\mathrm{Bernoulli}(\cdot)$ is a Bernoulli distribution and I is a matrix of Bernoulli random variables where each variable has p with probability 1, $1-p$ with probability 0.

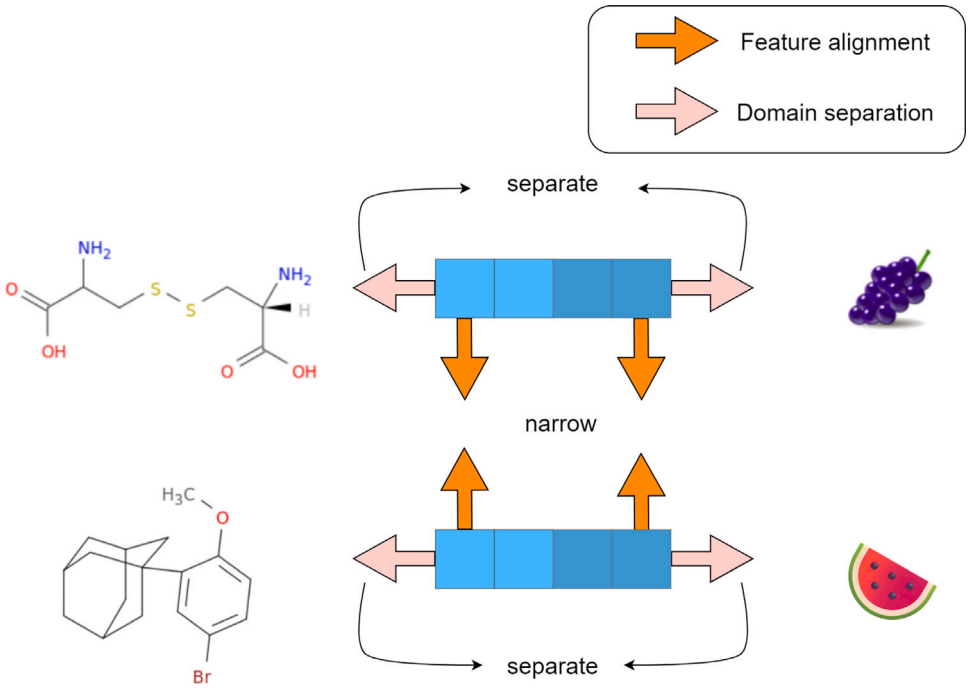

Fig. 3. A simple schematic of the feature alignment and domain separation modules.

2.3.2. Feature interaction

Feature perturbation can hardly guarantee the quality of the perturbed feature embeddings, and it is easy to introduce irrelevant noise into the unsupervised learning process, thus affecting the accuracy of the final feature representation. To address this issue, we introduce a Transformer-based [38] feature interaction encoder, which is widely used to extract vector-level relationships of features, thereby increasing the robustness of the self-supervised model to perturbed features. We deploy a shared Transformer feature interaction encoder in the model, which, after inputting two pairs of perturbed feature interaction pairs (Drug–food embedding pair) at a set ratio, outputs two pairs of feature interaction representations extracted from the features. We then use a contrastive loss function to reduce the distance between the two pairs of feature interaction representations, thus training the accuracy of the feature interaction encoder and features. By processing through the feature interaction encoder, we can more effectively capture the potential associations between features, mine long-distance dependencies between features, and further enhance the expression of feature interaction, thereby improving the accuracy of feature representation. The reduction of the distance between perturbed feature interaction pairs is achieved through the following loss function:

$$L_{pi} = \frac{1}{\text{batchsize}} \sum_{i=1}^{\text{batchsize}} \left| I\left(\hat{E}_1\right) - I\left(\hat{E}_2\right) \right|_{L_2} \tag{4}$$

where $|\cdot|_{L2}$ denotes the $L2$ distance, $\hat{E}$ represents the perturbed Drug(Food) embedding, and I represents the feature interaction layer.

2.4. Inference feedback module

The inference feedback module's purpose is to estimate the existence of an interaction relationship between food and drugs based on Drug (Food) features. We experimented with several modules, for the DFI inference task. After a performance comparison, we selected FM [39] as the DFI-MS's inference feedback module. The FM used in this module has the advantage of capturing the interplay among features, and representing Drug (Food) features with latent vectors enables effective expression of higher-order interaction terms. Further, FM exhibits robust adaptability to sparse data, such as DFI. By feeding the corresponding features of Drug (Food) into the model, we yield the probability value of a Drug–Food Pair interaction. The model parameters are then trained using the binary cross-entropy loss function. In this paper's experiments, we categorize Drug–Food Pairs with a predicted probability value exceeding 0.5 as positive and those with a value below 0.5 as negative.

2.5. Model overview

The DFI-MS model, illustrated in Fig. 4, is designed for deriving high-quality drug and food features from DFI relationships through self-supervised learning. The model integrates four key sub-modules in its framework.

The Embedding Module initially transforms the raw data into a continuous feature representation. Initially, embedding might not accurately depict the true properties of food and drugs, but through the model's training process, it is refined for better accuracy. Following this, the Inference Feedback Module, which processes Drug–Food Pairs, employs Factorization Machines (FM) to efficiently calculate interaction terms between feature vectors. It does so by computing the difference between the sum of squares and the square of sums, thus significantly enhancing computational efficiency. Simultaneously, the Perturbation Interaction Module introduces random perturbations to the input embeddings. This results in two distinct sets of perturbed embeddings that are further processed using a shared transformer encoding layer. The process is constrained within the feature space using an L2 loss function, which contributes to the model's robustness. Additionally, the Feature Alignment and Domain Separation Module plays a pivotal role. It tunes a batch of input feature pairs using L2 norm and cosine similarity loss functions, focusing on both batch and domain dimensions. This module ensures the alignment of features and separation of domains effectively.

The training of the model adopts a multi-task strategy, optimizing these modules collectively in an end-to-end manner. The feature alignment and domain separation module, along with the perturbation interaction module, are set for interval training to optimize their performance, while the inference feedback module is involved in every

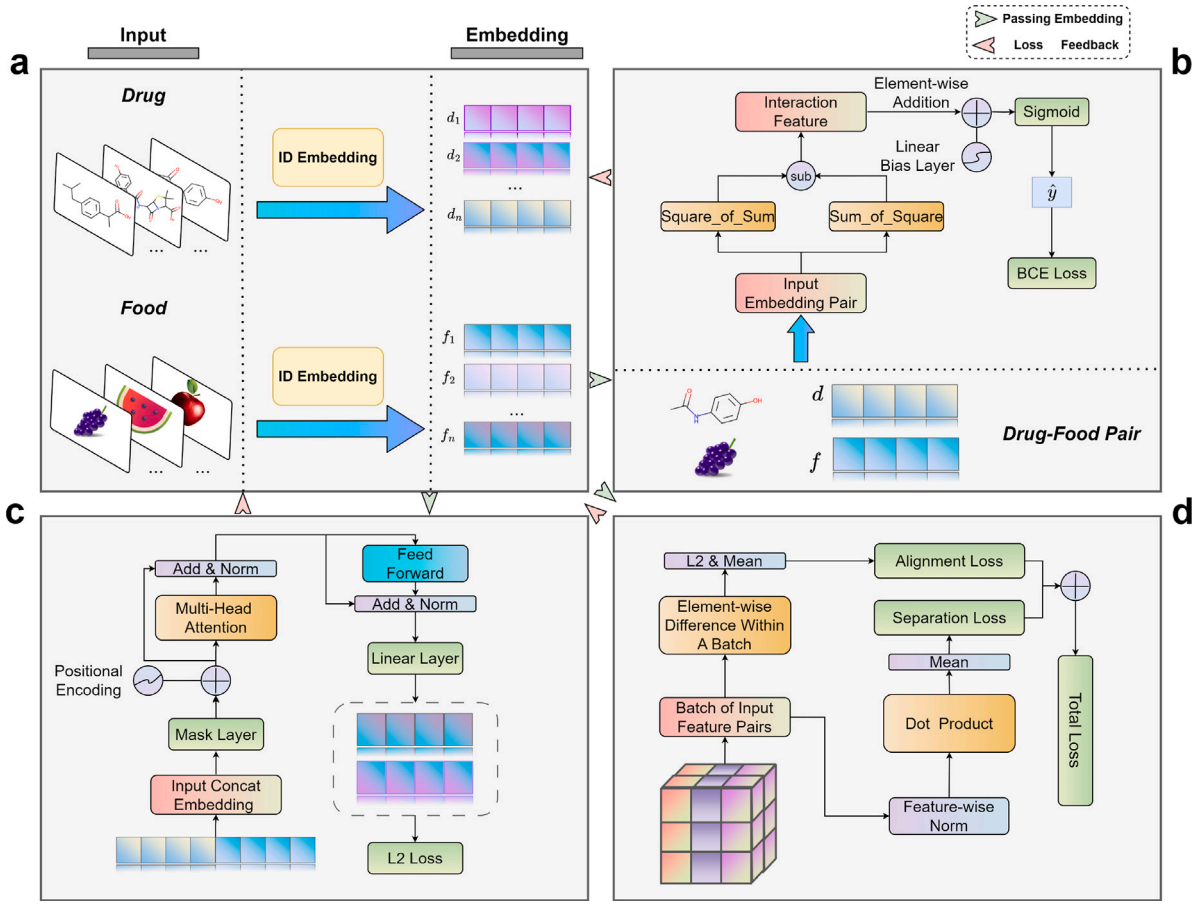

Fig. 4. The DFI-MS model consists of four key sub-modules:
a. Embedding Module: Converts raw input data into continuous feature representations.
b. Inference Feedback Module: Utilizes Drug–Food Pairs and employs Factorization Machines (FM) for efficient computation of interaction terms.
c. Perturbation Interaction Module: Applies random perturbations to input embeddings and uses a shared transformer encoding layer for robustness.
d. Feature Alignment and Domain Separation Module: Adjusts input feature pairs using L2 norm and cosine similarity loss functions for feature alignment and domain separation.

training cycle. The final objective function of the model is formulated as:

$$L = L_{BCE} + \alpha \cdot L_{pi} + \beta \cdot \left(L_a + L_s\right) \tag{5}$$

where L_{BCE} represents the BCELOSS, and α and β are adjustable hyper-parameters.

In the model's testing phase, only the input of trained Drug–Food pair features into the inference feedback module is required for prediction. This module's design allows for adaptability to various downstream tasks by modifying its network structure and loss functions. Similarly, the feature alignment and domain separation module, as well as the perturbation interaction module, can be easily adjusted for new tasks with simple changes in parameter settings.

3. Result

3.1. Evaluation metrics

In all experiments conducted in this paper, we are more concerned with the discovery of positive samples, so we chose AUC, AUPR, F1-score, Recall, and Precision to evaluate the performance of DFI-MS and other comparison models. The specific metric explanations are as follows:

$$Precision = \frac{TP}{FP + TP}, \quad Recall = \frac{TP}{TP + FN}, \quad F1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{6}$$

where TP and TN denote the number of positive and negative samples correctly predicted by the model; FP and FN denote the number of positive and negative samples incorrectly predicted by the model, respectively.

AUC (Area Under the Curve): Refers to the area under the ROC curve (Receiver Operating Characteristic Curve). The ROC curve is a curve plotted with the True Positive Rate (TPR) as the vertical axis and the False Positive Rate (FPR) as the horizontal axis. The closer the AUC value is to 1, the more effective the classifier is in distinguishing between positive and negative samples.

AUPR (Area Under the Precision–Recall Curve): Refers to the area under the Precision–Recall Curve. The Precision–Recall Curve is a curve plotted with Recall (also known as the True Positive Rate) as the vertical axis and Precision as the horizontal axis. AUPR focuses more on the recognition effect of positive cases, especially in cases of class imbalance.

3.2. Comparison models

To evaluate the performance of DFI-MS on the two datasets, we compared DFI-MS with seven DDI prediction models and one DFI prediction model. Here are some brief introductions to these comparison models:

1. CASTER [40]: A DDI prediction model that effectively characterizes drug functional substructures based on DDI mechanism sequential pattern mining modules while using an autoencoder module to improve generalization ability and interpretability.

Table 1
Comparison of the performance of DFI-MS with eight state of the art methods on the DrugBank-DFI dataset.

Method	AUC	AUPR	F1-score	Recall	Precision
CASTER	60.72%	8.47%	8.37%	7.67%	9.22%
MR-GNN	63.24%	10.32%	17.83%	17.4%	18.31%
GCN-BMP	74.88%	19.45%	24.14%	22.99%	25.42%
EPGCN-DS	77.89%	21.55%	3.50%	2.41%	6.45%
SSI-DDI	92.84%	57.46%	*42.15%*	30.20%	69.76%
DeepDrug	89.85%	37.75%	40.93%	*39.20%*	42.84%
DeepDDI	92.51%	64.84%	44.80%	31.67%	*76.56%*
DFinder	**97.53%**	*81.58%*	20.20%	11.24%	**99.58%**
DFI-MS	*96.92%*	**87.10%**	**81.27%**	**70.34%**	*96.22%*

Bold indicates the optimal value among the compared methods, and italic indicates the suboptimal value.

Table 2
Comparison of the performance of DFI-MS with eight state of the art methods on the PubMed-DFI dataset.

Method	AUC	AUPR	F1-score	Recall	Precision
CASTER	63.88%	6.24%	12.59%	11.41%	14.05%
MR-GNN	68.15%	10.73%	15.39%	14.24%	16.76%
GCN-BMP	76.64%	12.65%	19.13%	17.81%	20.67%
EPGCN-DS	75.52%	14.88%	14.90%	8.48%	61.43%
SSI-DDI	84.28%	23.37%	10.66%	5.91%	54.57%
DeepDrug	89.22%	23.78%	30.06%	28.77%	31.48%
DeepDDI	90.04%	*46.83%*	*47.49%*	**46.99%**	48.02%
DFinder	*90.08%*	40.22%	20.20%	4.81%	**98.79%**
DFI-MS	**93.78%**	**61.35%**	**55.59%**	*43.38%*	*77.39%*

Bold indicates the optimal value among the compared methods, and italic indicates the suboptimal value.

2. MR-GNN [41]: An end-to-end graph neural network DDI model that uses a multi-resolution architecture to extract node features and employs a dual-graph state long short-term memory network to capture interaction features between entities.
3. GCN-BMP [42]: A DDI prediction method that adopts end-to-end graph representation learning and key-aware information propagation mechanism, achieving high-performance prediction and interpretability provided by the built-in attention mechanism.
4. EPGCN-DS [43]: A molecular structure DDI detection method based on graph convolutional networks and deep ensembles, which uses more discriminative convolutional layers and maintains permutation invariance of input predictions while capturing complex interactions.
5. SSI-DDI [22]: A DDI prediction method that decomposes and transforms the drug–drug interaction prediction task into recognizing pairwise interactions between structures by extracting rich features through manipulating the original molecular graph representations of drugs.
6. DeepDrug [21]: A DDI prediction model that leverages residual graph convolutional networks (RGCN) and convolutional networks (CNN) to learn comprehensive structural and sequential representations of drugs and proteins.
7. DeepDDI [20]: A DDI prediction model that uses drug pair and drug–food component pair names and structural information, through structural similarity analysis and multi-label classification model.
8. DFinder [19]: An end-to-end DFI prediction method based on graph embedding combined with LightGCN [44] to aggregate node attribute features and topological structure features for learning drug and food component representations.

3.3. Experimental setup

We set the embedding size of Drug(Food) to 64, the batch size to 256, the learning rate to 0.015, and the weight decay to 1e−6. In the perturbation interaction module, the perturbation ratio is set to 0.2, and the training intervals of the feature alignment and domain separation modules and the perturbation interaction module are set to 30 epochs in DrugBank and 10 epochs in PubMed. Additionally, in the DrugBank dataset, we set alpha in the multi-task loss function to 0.05 and beta to 0.1. In the PubMed dataset, alpha is set to 0.1 and beta is set to 0.01. Furthermore, we randomly divided the positive and negative samples, with 20% used as the test set and the remaining 80% for five-fold cross-validation in the training set.

3.4. Experimental results

The performance of DFI-MS was compared to eight alternative models, with the results presented in Tables 1 and 2. The table enumerates the performances of DFI-MS on two distinct datasets.

On the DrugBank dataset, DFI-MS demonstrated a substantial improvement in comprehensive metrics, such as AUPR and F1-Score, compared to the alternative models. Due to the relative sparsity of the DFI dataset, there was an imbalance in the ratio of positive to negative samples in the test set, which reduced the reliability of AUC in the experiment. However, AUPR, being indicative of the model's performance on such imbalanced datasets, can more comprehensively demonstrate the model's discriminative ability across categories, as it focuses on precision and recall across the board. Please refer to Fig. 5 for comparisons of AUPR performance. The seven DDI prediction models' performance on the DFI task was slightly lacking when compared to the dedicated DFI model. As compared to the highest performing DDI model, DFI-MS's AUPR index rose by 21.92%, and when compared to the latest DFI model, the AUPR increased by 5.18%. In addition, DFI-MS achieved the highest F1-score at the default threshold, reflecting the robustness and generalizability of the features trained by the DFI-MS model, along with its resistance to the influence of imbalanced training samples.

On the PubMed dataset, DFI-MS outperformed the alternatives in both AUPR and F1-Score. The model's ability to recognize positive samples improved noticeably, as indicated by a significant enhancement in the AUPR metric. In addition, it achieved the highest F1-Score at the default threshold on both datasets. We attribute this to the feature alignment and domain separation module and the perturbation interaction module that trained high-quality features. These effectively modulated the range between various feature values during the training process, pushing dissimilar types of features apart in the feature space, while bringing similar features closer together. This strategy not only adeptly addressed the challenges posed by the incomplete nature of the DFI dataset, but also had beneficial effects on a variety of downstream tasks.

In comparison with other baseline models, DFI-MS demonstrates balanced characteristics across various performance metrics, which is particularly noteworthy. In the context of DDI models, most models are not specifically optimized for DFI tasks. Precise predictions of DDI models depend on accurate drug (or food) characterization extraction, which is especially challenging in DFI tasks, leading to relatively weaker performance of DDI models in these tasks. Moreover, the DFinder model, which uses graph neural networks to predict DFI relationships, shows higher performance on the densely linked DrugBank dataset. However, its performance significantly decreases on the more sparsely connected PubMed dataset, particularly in terms of the AUPR metric, which reflects the model's ability to predict positive samples. Also, when using the default threshold (0.5), DFinder performs poorly in terms of recall rate. In contrast, DFI-MS leverages multi-layer self-supervised learning to develop more versatile embedding representations, significantly boosting its recall rate in environments with sparse connections. This enhancement is particularly relevant in the context of sparse interactions with low-frequency samples, indicating that DFI-MS is capable of learning more universally applicable feature

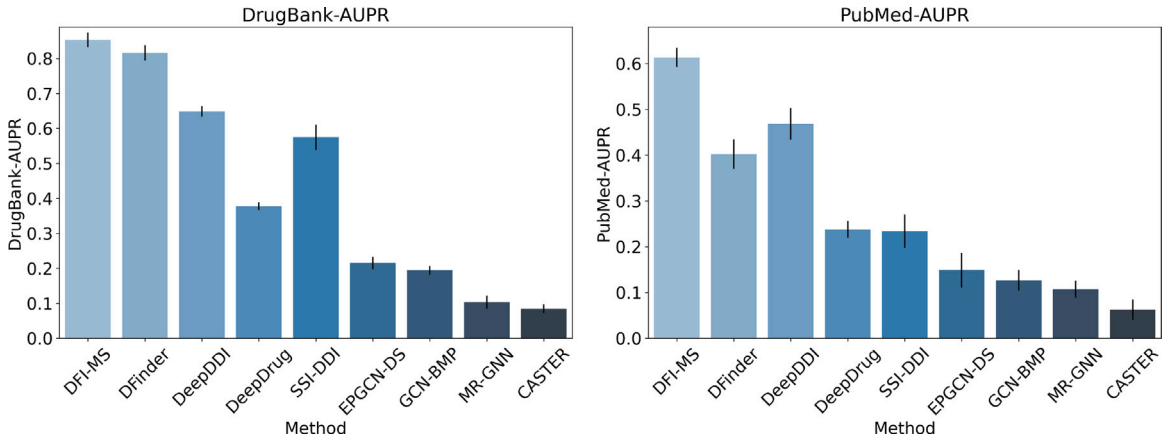

Fig. 5. Comparison of AUPR metrics between DFI-MS and eight methods.

representations. It does so without excessively depending on sparse or random data points. This method proves extremely effective in environments with sparse data, providing robust support for accurately identifying DFI relationships.

3.5. Parameter experiment

To verify the role of the feature alignment and domain separation module and the perturbation interaction module in DFI-MS, we conducted parameter experiments. We achieved this goal by changing the loss function ratio (α and β). In the experiment, we selected the DrugBank dataset and the PubMed dataset, and varied the values of α and β proportionally. Heatmaps were generated based on the experimental results, where Fig. 6 represents the experimental results for DrugBank and PubMed. These experimental findings corroborate our initial hypothesis, indicating that the self-supervised learning enabled module enhances the quality of Drug(Food) feature sets. This, in turn, augments the feature distinctiveness among different categories and sharpens the model's decision boundary. Similarly, the perturbation interaction module contributes to an ongoing update of features and optimizes model performance, drawing on a training approach reminiscent of generative adversarial training.

The specific relationship between the parameters of the feature alignment and domain separation modules and the experimental results in AUPR shows a significant correlation. In the two datasets studied, we observed that higher β parameters tend to decrease model performance. This phenomenon can be attributed to the fact that excessively high β values might lead to an overemphasis on the distribution of features of drugs (or food), thereby impairing the model's ability to capture true semantic information. However, a moderate β parameter, by promoting knowledge sharing among high-frequency samples, significantly enhances the model's generalization ability for low-frequency samples. Furthermore, the perturbation interaction module also plays a key role in enhancing model performance. Although it does not show a significant linear correlation with the overall performance of the model, this is mainly because the design of the module involves the introduction of random noise data. An appropriate level of noise can not only cultivate more generalizable embedding layers but also, in combination with the feature alignment and domain separation modules, further enhance the overall efficacy of the model. Despite its marginal contribution possibly being less directly observable, the introduction of the perturbation interaction module undoubtedly provides an indispensable perspective for improving model performance.

Additionally, the batch size and embedding size have an impact on the self-supervised loss function. We conducted further experiments on these two parameters on two datasets, with the results shown in Fig. 7. Relatively, the PubMed dataset requires a smaller embedding size and a larger batch size to achieve better performance. We believe this is due to the large number of elements in PubMed but a lower proportion of positive samples. An excessively large embedding size is detrimental to achieving good generalization of positive samples in larger datasets; meanwhile, a smaller batch size might lead to unavoidable errors in feature alignment due to limited receptive fields. The DrugBank dataset exhibits better robustness to these two parameters, but an excessively large embedding size still significantly harms model performance, which is also due to high generalization errors caused by overfitting. Overall, reasonably adjusting the batch size and embedding size according to the different characteristics of datasets is crucial for improving the performance and generalization ability of the DFI-MS.

4. Feature distribution analysis

To verify the effectiveness of the Feature Alignment and Domain Separation Module (FD) and the Perturbation Interaction Module (PI), we conducted an in-depth analysis of the embedding layer of our model. Firstly, we reduced the dimensionality of the embedding using the t-SNE algorithm, and the visualization results are shown in Figs. 8–11. Secondly, we quantitatively evaluated the performance of the embedding layer using three metrics: Silhouette Coefficient, Davies–Bouldin Index, and Calinski–Harabasz Index, with the results presented in Table 3.

The Silhouette coefficient is a valuable metric for evaluating the similarity among samples, taking into account both the cohesion within a category and the separation from the nearest neighboring category. A higher Silhouette coefficient suggests superior feature separation in the embedding layer, indicating that samples within the same category are similar to each other and distinct from those in different categories [45]. The Davies–Bouldin index is defined based on the ratio of intra-cluster similarity to inter-cluster dissimilarity. A lower Davies–Bouldin index value indicates that the samples within a cluster are closely packed and there is clear separation between different clusters, which is important for the performance of feature embedding [46]. Similarly, the Calinski–Harabasz index is based on the ratio of between-cluster dispersion to within-cluster dispersion. A higher value of this index indicates that the clusters are not only well-separated but also compact, demonstrating the effectiveness of the embedding layer in organizing and distinguishing data [47]. These metrics comprehensively reflect the performance of the embedding layer in effectively separating and grouping data, thereby indicating the quality of features in the embedding layer.

Experiments demonstrate that FD and PI significantly enhance feature quality. In the DrugBank dataset, the visualization of the embedding using the complete model (Figs. 8–11) displays clear clustering effects and inter-class separation, indicating that different class features

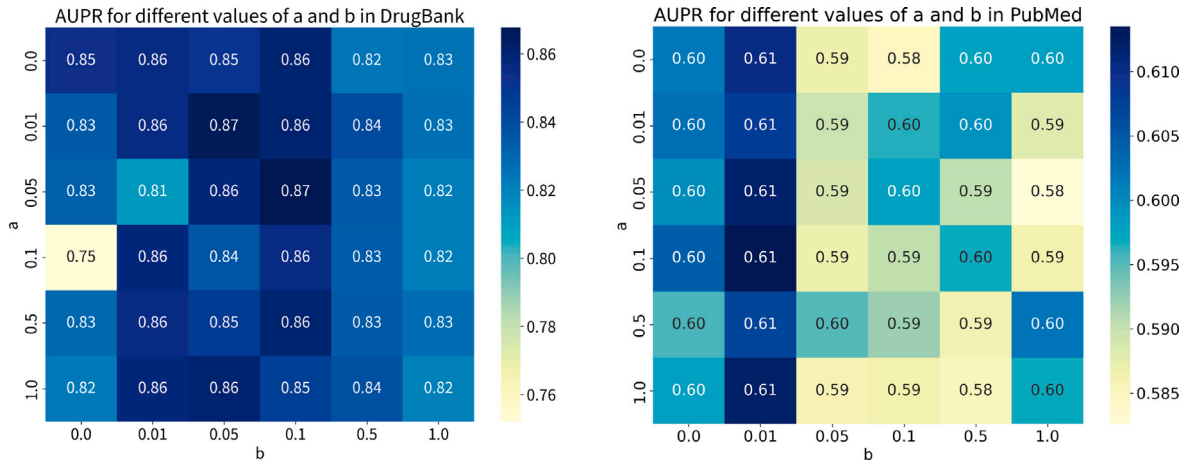

Fig. 6. Parameter experiments of perturbation interaction module and feature alignment and domain separation module.

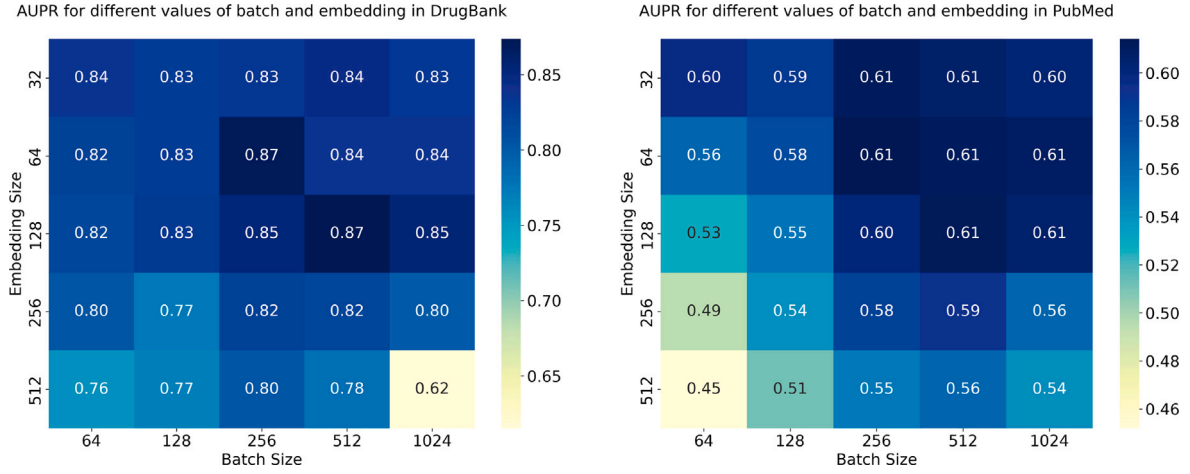

Fig. 7. Parameter experiments of batch size and embedding size.

are clearly and effectively separated in space. In contrast, models without these two modules show chaotic visualization in embedding, with blurred inter-class boundaries and low feature distinction. Additionally, we observed that FD has a more significant impact on the quality of embedding, showing clear inter-class separation in feature distribution, which is crucial for model performance. On the PubMed dataset, although there is more data in the embedding layer and reduced dimensionality leads to less obvious distinctions, the comparative visualization of contour maps still allows us to clearly see the advantages and disadvantages between different methods. Models including FD demonstrate more significant inter-class separation, indicating the semantic space separation of Drugs and Food, which aligns with their inherent differences in biological and chemical properties. This distinction is vital for the application of deep learning models in these fields, directly affecting the accuracy and reliability of classification, prediction, or recommendation systems.

Further validation of the effectiveness of each module is provided by comparing the results of the Silhouette Coefficient, Davies–Bouldin Index, and Calinski–Harabasz Index (see Table 3). The complete model has a higher Silhouette Coefficient value, indicating high intra-class cohesion and clear inter-class separation. The relatively lower Davies–Bouldin Index and higher Calinski–Harabasz Index underscore the excellent clustering quality and inter-class separation effect. Conversely, the baseline model without these modules performs poorly on these metrics, reflecting its inadequacy in feature representation.

Table 3
Feature quality evaluation with t-SNE based clustering metrics.

Dataset	Methods	Silhouette coefficient ↑	Davies–Bouldin index ↓	Calinski–Harabasz index ↑
DrugBank	Complete model	**0.387**	**1.039**	**231.474**
	w/o FD	0.073	3.238	22.669
	w/o PI	0.297	1.371	116.592
	w/o All	0.024	7.922	3.889
PubMed	Complete model	**0.065**	**3.778**	**92.844**
	w/o FD	0.027	7.122	25.669
	w/o PI	0.056	4.045	81.171
	w/o All	0.012	14.202	6.562

FD:Feature Alignment and Domain Separation Module PI:Perturbation Interaction Module.
Bold indicates the optimal value among the compared methods.

5. Discuss and conclusion

In this paper, we propose a computational model for Drug–food interaction (DFI) named DFI-MS, which is based on feature alignment and domain separation modules, perturbation interaction modules, and inference feedback modules to obtain high-quality Drug(Food) embeddings. By using a simple inference model, we achieved the best results on the Drugbank-DFI and PubMed-DFI datasets. To further evaluate the effectiveness of the proposed modules, we conducted a series of Parameter experiments, and the results demonstrated that our proposed modules significantly improved the model's performance. Moreover,

Feature Distribution in the Embedding Layer of the DFI-MS

DrugBank Dataset:

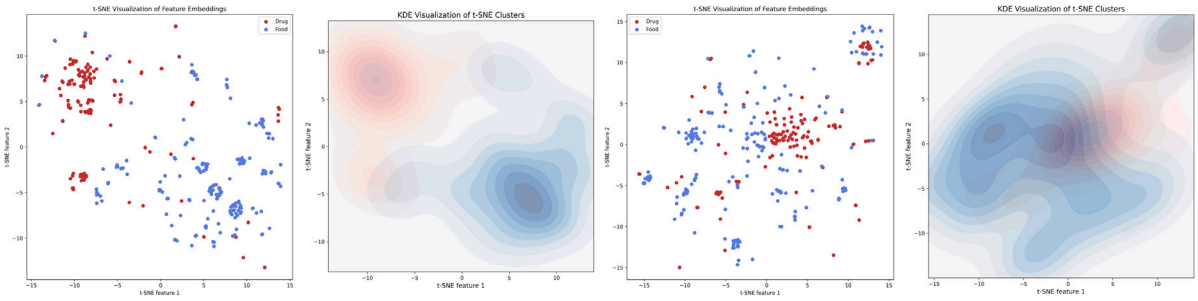

Fig. 8. Complete Model(left) and Model without the Feature Alignment and Domain Separation Module(right).

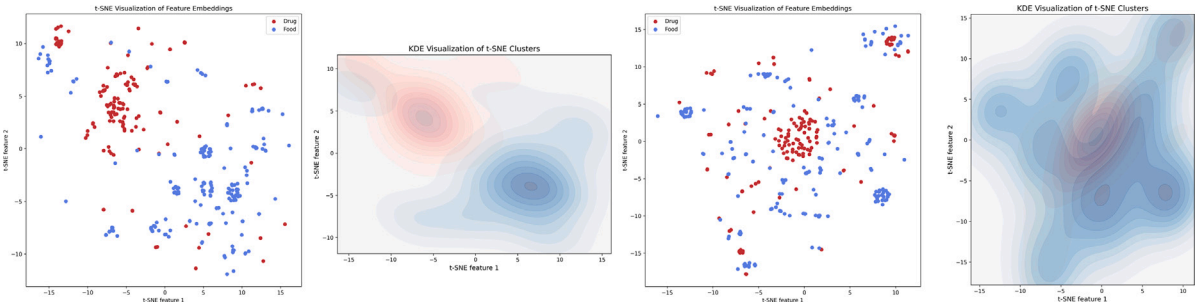

Fig. 9. Model without the Perturbation Interaction Module(left) and Model without any module(right).

PubMed Dataset:

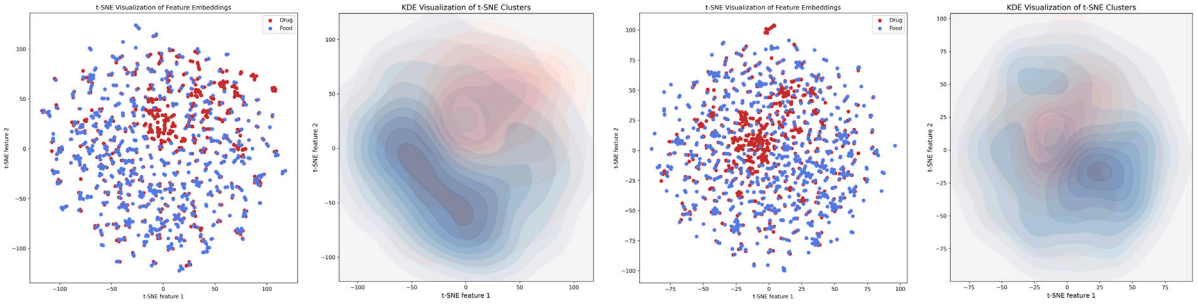

Fig. 10. Complete Model(left) and Model without the Feature Alignment and Domain Separation Module(right).

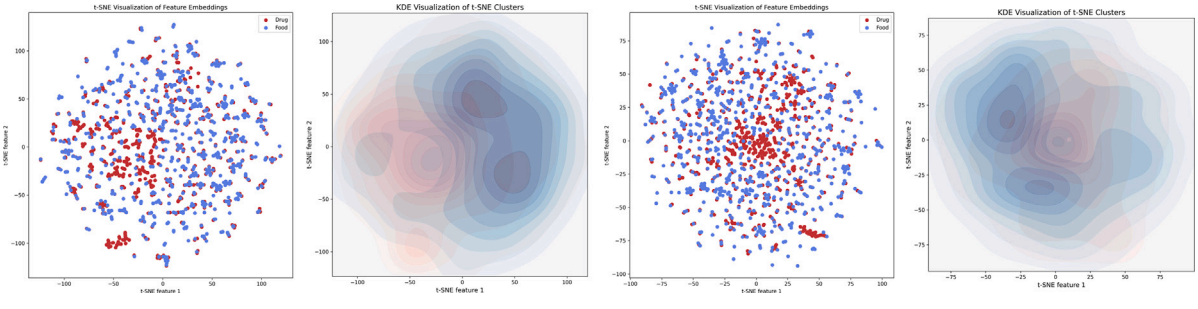

Fig. 11. Model without the Perturbation Interaction Module(left) and Model without any module(right).

DFI-MS can be applied to different downstream tasks by replacing the inference feedback modules, thus providing high practical application value.

Due to the complexity of food ingredients, DFI research faces many limitations, and most DFI studies are based on clinical findings or derived from DDI (drug–drug interaction) studies. Our computational model offers a novel direction for addressing this issue. By using computational methods, we can assess whether there are interactions between drugs and food in a shorter period, which can guide clinical practice to some extent and improve the safety and effectiveness of medication for patients. Furthermore, the feature alignment and domain separation modules provide robustness for the model to handle

heterogeneous data, allowing researchers to integrate data from various sources into the model, thereby improving the prediction accuracy and reliability. In addition, we believe there is room for improvement in the model. The model needs to rely on known DFI relationships to train accurate embeddings for drugs (foods). Although the model employs multi-layer self-supervised learning to mitigate biases introduced by different data distributions, the parameter setting of self-supervised learning requires extensive experimentation to achieve ideal results. Future work may focus on how to improve the parameter setting of self-supervised learning and feature engineering. Furthermore, considering the limitations of data availability, future research might explore the use of transfer learning from the DDI domain and generative model techniques to create data, addressing the scarcity of real-world data. This synthetic data can not only be used for training and improving the model but also for validating the model's generalizability and robustness. We also hope that such computational models can promote the development of DFI research, provide directional guidance for wet experiments, reduce experimental costs and time, and offer more effective references for clinical practice. In the field of personalized medicine, this information can assist doctors in predicting the effects and side effects of medications, taking into account the patient's dietary habits and lifestyle. This enables the customization of safer and more effective treatment plans for each patient. Additionally, knowledge of DFIs can guide the formulation of public health policies. For example, educating communities about drug–food interactions can raise public awareness and reduce adverse drug reactions. Our computational model fundamentally offers a novel approach in the field of DFI. This method not only enhances the safety and efficacy of medications in practical applications but also holds potential to make significant contributions to the fields of personalized medicine and public health.

CRediT authorship contribution statement

Jinhang Wei: Writing – original draft, Validation, Methodology. **Zhen Li:** Writing – original draft, Validation, Methodology. **Linlin Zhuo:** Writing – original draft, Validation, Methodology. **Xiangzheng Fu:** Writing – review & editing, Investigation. **Mingjing Wang:** Writing – review & editing, Formal analysis. **Keqin Li:** Writing – review & editing, Validation. **Chengshui Chen:** Writing – review & editing.

Declaration of competing interest

We declare that we have no conflict of interest.

Acknowledgments

Linlin Zhuo, Jinhang Wei and Zhen Li are mainly responsible for the tasks of experimental design and paper writing and revision. Xiangzheng Fu, Mingjing Wang, Keqin Li and Chengshui Chen are responsible for the tasks of experiment guidance, thesis guidance and revision. This study was supported by the National Key Research and Development Program of China grants 2016YFC1304000 (C Chen); The National Natural Scientific Foundation of China 82170017 (C Chen); Zhejiang Provincial Key Research and Development Program 2020C03067 (C Chen).

References

[1] Mirko Koziolek, Stefano Alcaro, Patrick Augustijns, Abdul W. Basit, Michael Grimm, Bart Hens, Caroline L. Hoad, Philipp Jedamzik, Christine M. Madla, Marc Maliepaard, et al., The mechanisms of pharmacokinetic food-drug interactions–A perspective from the UNGAP group, Eur. J. Pharm. Sci. 134 (2019) 31–59.

[2] Sarah Downer, Seth A. Berkowitz, Timothy S. Harlan, Dana Lee Olstad, Dariush Mozaffarian, Food is medicine: actions to integrate food and nutrition into healthcare, bmj 369 (2020).

[3] Lucie Hyrsova, Alena Vanduchova, Jan Dusek, Tomas Smutny, Alejandro Carazo, Veronika Maresova, Frantisek Trejtnar, Pavel Barta, Pavel Anzenbacher, Zdenek Dvorak, et al., Trans-resveratrol, but not other natural stilbenes occurring in food, carries the risk of drug-food interaction via inhibition of cytochrome P450 enzymes or interaction with xenosensor receptors, Toxicol. Lett. 300 (2019) 81–91.

[4] David G. Bailey, J. Malcolm, O. Arnold, J. David Spence, Grapefruit juice-drug interactions, Br. J. Clin. Pharmacol. 58 (7) (2004) S831.

[5] Anthony J. Viera, Noah Wouk, Potassium disorders: hypokalemia and hyperkalemia, Am. Fam. Physician 92 (6) (2015) 487–495.

[6] Andrea Ricci, Giovanni N. Roviello, Exploring the protective effect of food drugs against viral diseases: Interaction of functional food ingredients and SARS-CoV-2, Influenza Virus, and HSV, Life 13 (2) (2023) 402.

[7] Daan A.C. Lanser, Simon P. de Leeuw, Esther Oomen-de Hoop, Peter de Bruijn, Marthe S Paats, Daphne W Dumoulin, Stijn L.W. Koolen, Anne-Marie C. Dingemans, Ron H.J. Mathijssen, G.D. Marijn Veerman, Influence of food with different fat concentrations on Alectinib exposure: A randomized crossover pharmacokinetic trial, J. Natl. Compr. Cancer Netw. 21 (6) (2023) 645–651.

[8] Jingjing Yu, Ichiko D. Petrie, René H. Levy, Isabelle Ragueneau-Majlessi, Mechanisms and clinical significance of pharmacokinetic-based drug-drug interactions with drugs approved by the US food and drug administration in 2017, Drug Metab. Dispos. 47 (2) (2019) 135–144.

[9] Ge-Fei Hao, Wei Zhao, Bao-An Song, Big Data Platform: An Emerging Opportunity for Precision Pesticides, ACS Publications, 2020.

[10] Yuandong Yu, Shiqi Xu, Ran He, Guizhao Liang, Application of molecular simulation methods in food science: Status and prospects, J. Agricult. Food Chem. 71 (6) (2023) 2684–2703.

[11] Zhiyong Cui, Zhiwei Zhang, Tianxing Zhou, Xueke Zhou, Yin Zhang, Hengli Meng, Wenli Wang, Yuan Liu, A TastePeptides-meta system including an umami/bitter classification model Umami_YYDS, a TastePeptidesDB database and an open-source package Auto_Taste_ML, Food Chem. 405 (2023) 134812.

[12] Junho Lee, Seon Bin Song, You Kyoung Chung, Jee Hwan Jang, Joonsuk Huh, BoostSweet: Learning molecular perceptual representations of sweeteners, Food Chem. 383 (2022) 132435.

[13] Meetali Sinha, Deepak Kumar Sachan, Roshni Bhattacharya, Prakrity Singh, Ramakrishnan Parthasarathi, ToxDP2 database: Toxicity prediction of dietary polyphenols, Food Chem. 370 (2022) 131350.

[14] Zheng-Fei Yang, Ran Xiao, Guo-Li Xiong, Qin-Lu Lin, Ying Liang, Wen-Bin Zeng, Jie Dong, Dong-sheng Cao, A novel multi-layer prediction approach for sweetness evaluation based on systematic machine learning modeling, Food Chem. 372 (2022) 131249.

[15] Xingran Kou, Peiqin Shi, Chukun Gao, Peihua Ma, Huadong Xing, Qinfei Ke, Dachuan Zhang, Data-driven elucidation of flavor chemistry, J. Agricult. Food Chem. 71 (18) (2023) 6789–6802.

[16] Dachuan Zhang, Cancan Jia, Dandan Sun, Chukun Gao, Dongheng Fu, Pengli Cai, Qian-Nan Hu, Data-driven prediction of molecular Biotransformations in food fermentation, J. Agricult. Food Chem. (2023).

[17] Xia Hu, Zhen Zeng, Jing Zhang, Di Wu, Hui Li, Fang Geng, Molecular dynamics simulation of the interaction of food proteins with small molecules, Food Chem. 405 (2023) 134824.

[18] Yizheng Wang, Yixiao Zhai, Yijie Ding, Quan Zou, SBSM-pro: support bio-sequence machine for proteins, 2023, arXiv preprint arXiv:2308.10275.

[19] Tao Wang, Jinjin Yang, Yifu Xiao, Jingru Wang, Yuxian Wang, Xi Zeng, Yongtian Wang, Jiajie Peng, DFinder: a novel end-to-end graph embedding-based method to identify drug–food interactions, Bioinformatics 39 (1) (2023) btac837.

[20] Mark Lennox, Neil Robertson, Barry Devereux, Modelling drug-target binding affinity using a BERT based graph neural network, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 4348–4353.

[21] Qijin Yin, Rui Fan, Xusheng Cao, Qiao Liu, Rui Jiang, Wanwen Zeng, DeepDrug: A general graph-based deep learning framework for drug-drug interactions and drug-target interactions prediction, Quant. Biol. 11 (3) (2021) 260–274.

[22] Arnold K. Nyamabo, Hui Yu, Jian-Yu Shi, SSI–DDI: substructure–substructure interactions for drug–drug interaction prediction, Brief. Bioinform. 22 (6) (2021) bbab133.

[23] Zhecheng Zhou, Zhenya Du, Jinhang Wei, Linlin Zhuo, Shiyao Pan, Xiangzheng Fu, Xinze Lian, MHAM-NPI: Predicting ncRNA-protein interactions based on multi-head attention mechanism, Comput. Biol. Med. (2023) 107143.

[24] Katie Ovens, Farhad Maleki, B. Frank Eames, Ian McQuillan, Juxtapose: a gene-embedding approach for comparing co-expression networks, BMC Bioinformatics 22 (2021) 1–26.

[25] Jinhang Wei, Linlin Zhuo, Zhecheng Zhou, Xinze Lian, Xiangzheng Fu, Xiaojun Yao, GCFMCL: predicting miRNA-drug sensitivity using graph collaborative filtering and multi-view contrastive learning, Brief. Bioinform. (2023) bbad247.

[26] Yizheng Wang, Yixiao Zhai, Yijie Ding, Quan Zou, SBSM-pro: Support bio-sequence machine for proteins, 2023, arXiv preprint arXiv:2308.10275.

[27] Jinhang Wei, Linlin Zhuo, Shiyao Pan, Xinze Lian, Xiaojun Yao, Xiangzheng Fu, HeadTailTransfer: An efficient sampling method to improve the performance of graph neural network method in predicting sparse ncRNA–protein interactions, Comput. Biol. Med. 157 (2023) 106783.

[28] David S. Wishart, Craig Knox, An Chi Guo, Savita Shrivastava, Murtaza Hassanali, Paul Stothard, Zhan Chang, Jennifer Woolsey, DrugBank: a comprehensive resource for in silico drug discovery and exploration, Nucleic Acids Res. 34 (suppl_1) (2006) D668–D672.
[29] David S. Wishart, Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, Murtaza Hassanali, DrugBank: a knowledgebase for drugs, drug actions and drug targets, Nucleic Acids Res. 36 (suppl_1) (2008) D901–D906.
[30] Vitaly Feldman, Chiyuan Zhang, What neural networks memorize and why: Discovering the long tail via influence estimation, Adv. Neural Inf. Process. Syst. 33 (2020) 2881–2891.
[31] Songyang Zhang, Zeming Li, Shipeng Yan, Xuming He, Jian Sun, Distribution alignment: A unified framework for long-tail visual recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2361–2370.
[32] Sukwon Yun, Kibum Kim, Kanghoon Yoon, Chanyoung Park, LTE4G: Long-tail experts for graph neural networks, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 2434–2443.
[33] Weihua Li, Ruyi Huang, Jipu Li, Yixiao Liao, Zhuyun Chen, Guolin He, Ruqiang Yan, Konstantinos Gryllias, A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges, Mech. Syst. Signal Process. 167 (2022) 108487.
[34] Fuhao Zhang, Bi Zhao, Wenbo Shi, Min Li, Lukasz Kurgan, DeepDISOBind: accurate prediction of RNA-, DNA-and protein-binding intrinsically disordered residues with deep multi-task learning, Brief. Bioinform. 23 (1) (2022) bbab521.
[35] Zhenguang Liu, Runyang Feng, Haoming Chen, Shuang Wu, Yixing Gao, Yunjun Gao, Xiang Wang, Temporal feature alignment and mutual information maximization for video-based human pose estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11006–11016.
[36] Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton, A simple framework for contrastive learning of visual representations, in: International Conference on Machine Learning, PMLR, 2020, pp. 1597–1607.
[37] Fangye Wang, Yingxu Wang, Dongsheng Li, Hansu Gu, Tun Lu, Peng Zhang, Ning Gu, CL4CTR: A contrastive learning framework for CTR prediction, in: Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, 2023, pp. 805–813.
[38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[39] Steffen Rendle, Factorization machines, in: 2010 IEEE International Conference on Data Mining, IEEE, 2010, pp. 995–1000.
[40] Kexin Huang, Cao Xiao, Trong Hoang, Lucas Glass, Jimeng Sun, Caster: Predicting drug interactions with chemical substructure representation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 01, 2020, pp. 702–709.
[41] Nuo Xu, Pinghui Wang, Long Chen, Jing Tao, Junzhou Zhao, Mr-gnn: Multi-resolution and dual graph neural network for predicting structured entity interactions, 2019, arXiv preprint arXiv:1905.09558.
[42] Xin Chen, Xien Liu, Ji Wu, GCN-BMP: investigating graph representation learning for DDI prediction task, Methods 179 (2020) 47–54.
[43] Mengying Sun, Fei Wang, Olivier Elemento, Jiayu Zhou, Structure-based drug-drug interaction detection via expressive graph convolutional networks and deep sets (student abstract), in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 10, 2020, pp. 13927–13928.
[44] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, Meng Wang, Lightgcn: Simplifying and powering graph convolution network for recommendation, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020, pp. 639–648.
[45] Ketan Rajshekhar Shahapure, Charles Nicholas, Cluster quality analysis using silhouette score, in: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA, IEEE, 2020, pp. 747–748.
[46] Slobodan Petrovic, A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters, in: Proceedings of the 11th Nordic Workshop of Secure IT Systems, Vol. 2006, Citeseer, 2006, pp. 53–64.
[47] Ujjwal Maulik, Sanghamitra Bandyopadhyay, Performance evaluation of some clustering algorithms and validity indices, IEEE Trans. Pattern Anal. Mach. Intell. 24 (12) (2002) 1650–1654.