

Multimodal Deep Learning for Drug–Drug Interaction Prediction using Structural Biological and Semantic Information

Mahdi Nouroozi

Institute for Advanced Studies in Basic Sciences

Mohsen Hooshmand

`mohsen.hooshmand@iasbs.ac.ir`

Institute for Advanced Studies in Basic Sciences

Fatemeh Nasiri

Institute for Advanced Studies in Basic Sciences

Research Article

Keywords: Drug–Drug Interaction Prediction, Multimodal Learning, BioGPT, Drug–Target Interactions, Squeeze-and-Excitation Networks, Multi-Head Self-Attention, DrugBank, Computational Pharmacology

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10948072/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Multimodal Deep Learning for Drug–Drug Interaction Prediction using Structural Biological and Semantic Information

Mahdi Nouroozi¹, Mohsen Hooshmand^{1*}, Fatemeh Nasiri¹

¹Department of Computer Science and Information Technology,
Institute for Advanced Studies in Basic Sciences (IASBS) , 444 Prof
Yousef Sobouti Blvd, , Zanzan, 45137-66731, Zanzan, Iran.

*Corresponding author(s). E-mail(s): mohsen.hooshmand@iasbs.ac.ir;

Contributing authors: m.nouroozi@iasbs.ac.ir;

fatemeh.nasiri@iasbs.ac.ir;

Abstract

Drug–drug interactions are a major source of adverse drug events and represent a significant challenge for medication safety and clinical decision-making. Although experimental methods provide reliable evidence for DDI identification, they are often costly, time-consuming, and unable to keep pace with the rapidly increasing number of approved drugs and potential drug combinations. Recent advances in machine learning have enabled large-scale computational prediction of DDIs; however, effectively integrating heterogeneous drug information remains a key challenge. To address this limitation, we propose HySEAtt-DDI, a multimodal deep learning framework for DDI prediction that integrates chemical, biological, and semantic drug information. The proposed framework combines structural similarity derived from molecular fingerprints, biological similarity obtained from drug–protein interaction profiles, and semantic similarity extracted from drug descriptions using BioGPT. These complementary representations are fused into a unified feature space and compressed through an autoencoder to learn compact latent representations. A hybrid architecture consisting of convolutional neural networks (CNNs), Squeeze-and-Excitation blocks, and multi-head self-attention is then employed to capture local patterns, channel-wise feature importance, and long-range dependencies associated with drug interactions. In addition, we construct a curated DrugBank-derived dataset containing 2,148 approved small-molecule drugs, 2,056 protein targets, and 592,840 drug–drug interaction instances. Extensive experiments and ablation studies demonstrate the effectiveness of the proposed framework and highlight the contribution of each modality

and architectural component. The results show that HySEAtt-DDI consistently outperforms conventional machine learning methods and competitive deep learning baselines, demonstrating the value of multimodal representation learning for DDI prediction. Overall, the proposed framework provides an effective and scalable approach for identifying potential drug-drug interactions and supporting computational pharmacology research.

Keywords: Drug-Drug Interaction Prediction, Multimodal Learning, BioGPT, Drug-Target Interactions, Squeeze-and-Excitation Networks, Multi-Head Self-Attention, DrugBank, Computational Pharmacology

1 Introduction

Drug-drug interactions (DDIs) are a major challenge in modern healthcare and can cause adverse drug events, reduced therapeutic efficacy, increased healthcare costs, and, in severe cases, patient mortality [1–3]. A DDI occurs when one drug alters the pharmacological effect of another, potentially resulting in toxicity, treatment failure, or unexpected clinical outcomes. The risk is particularly high among older adults and patients with complex chronic conditions, who are frequently exposed to polypharmacy, long-term treatment regimens, and care from multiple healthcare providers [4]. As the prevalence of diseases such as cancer, diabetes, and cardiovascular disorders continues to rise, the number of patients receiving concurrent medications has increased, further elevating the likelihood of clinically significant DDIs.

Although some drug combinations provide synergistic therapeutic benefits, many interactions compromise patient safety and disease management. Polypharmacy is a major contributor to adverse drug events and hospitalization among older adults, emphasizing the need to identify harmful interactions before they occur [5]. Conventional DDI detection relies on in vitro experiments, clinical studies, spontaneous adverse-event reporting systems, curated pharmacological databases, and rule-based clinical decision-support systems. While these approaches have improved medication safety, they are often costly, time-consuming, and limited to previously documented interactions. Consequently, rare, emerging, and unobserved DDIs may remain undetected [2]. Computational approaches can complement pharmacovigilance systems by efficiently screening large-scale biomedical and pharmacological data to predict potential novel interactions and support safer clinical decision-making.

Artificial intelligence (AI) and machine learning (ML) have become increasingly important in drug discovery and safety assessment, where conventional computational techniques such as molecular docking, pharmacophore modeling, and molecular dynamics simulations may be limited by simplified assumptions, restricted experimental conditions, and imperfect translation to clinical settings [6]. AI-driven methods can analyze large-scale chemical, biological, structural, and clinical datasets and have shown potential across the drug development pipeline, including target identification, candidate discovery, preclinical assessment, clinical studies, and post-market surveillance [7, 8].

Small-molecule drugs are especially relevant to computational DDI prediction because they are widely used in clinical practice and can be represented efficiently through well-defined chemical structures [9, 10]. Their molecular structures can be encoded using SMILES strings, enabling computational analysis, simulation, and molecular redesign [11]. Structural properties, including functional groups, atomic composition, and molecular geometry, influence pharmacokinetic and pharmacodynamic behavior, such as target binding, metabolism, and interactions with other compounds [4]. Accordingly, molecular structure has become a fundamental information source for DDI prediction.

Recent ML approaches have enabled the extraction of predictive patterns from large-scale molecular datasets and can estimate drug properties such as bioactivity, toxicity, solubility, target associations, and potential drug interactions [12]. However, molecular structure alone may not fully characterize the mechanisms underlying DDIs. Complementary information from drug-protein associations and textual drug descriptions can provide additional insights into drug mechanisms, therapeutic roles, and interaction patterns. Integrating these heterogeneous sources can therefore produce richer drug representations and improve the prediction of clinically relevant DDIs.

Despite significant advances in publicly available datasets and the development of AI and deep learning methods for DDI prediction, several challenges remain. Existing datasets are often limited in terms of completeness, quality, coverage, and consistency, which can negatively affect the reliability and generalizability of predictive models. Furthermore, although recent studies have introduced a variety of molecular, biological, and semantic features for DDI prediction, there is still a gap in effectively exploiting multimodal information, particularly through large language model (LLM)-based representations and their integration with other drug-related modalities [3].

To address these limitations, this study proposes a novel Hybrid Squeeze-and-Excitation Attention-based framework for DDI prediction, termed HySEAtt-DDI. The proposed approach integrates complementary information derived from molecular structures, drug-protein interactions, and textual drug descriptions to construct rich multimodal representations of drugs. HySEAtt-DDI combines autoencoder-based feature compression, convolutional layers for local pattern extraction, squeeze-and-excitation blocks for channel recalibration, and multi-head self-attention for modeling global dependencies across drug representations.

In addition, we introduce a new curated DDI dataset derived from DrugBank. The dataset was collected from validated sources and carefully processed through cleaning, integration, and quality-control procedures to ensure the inclusion of approved drugs with reliable information. Specifically, the dataset incorporates multimodal representations generated from SMILES strings, drug-target interaction profiles, and textual descriptions, providing a comprehensive basis for DDI prediction tasks.

To evaluate the proposed framework, extensive experiments are conducted using a set of baseline methods, including conventional ML models, deep learning approaches, and state-of-the-art graph-based architectures such as 3DGT [13]. Furthermore, an ablation study is performed to investigate the contribution of each data modality and architectural component to the overall predictive performance of HySEAtt-DDI.

Pharmaceutical therapeutics can generally be categorized into small-molecule drugs and biotechnology-based therapeutics. While interactions involving biologics and their combinations with small-molecule drugs constitute an active and important research area [14, 15], the scope of this study is limited to DDI prediction among small-molecule drugs.

The main contributions of this study are summarized as follows:

- We construct and publicly introduce a curated DrugBank-derived DDI dataset containing approved small-molecule drugs together with structural, biological, and semantic representations for reliable multimodal DDI prediction.
- We propose HySEAtt-DDI, a novel attention-based DDI prediction framework that combines autoencoder-based representation learning, convolutional feature extraction, Squeeze-and-Excitation blocks, and multi-head self-attention mechanisms.
- Extensive experiments validate the effectiveness of HySEAtt-DDI through comparisons with conventional ML, deep learning, and state-of-the-art graph-based DDI prediction methods.
- We conduct extensive ablation studies to quantify the contribution of individual data modalities and architectural components within the proposed framework.

The remainder of this paper is organized as follows. Section 2 reviews the related work on DDI prediction. Section 3 presents the proposed dataset (Section 3.1) and the HySEAtt-DDI framework (Section 3.2). Section 4 reports the experimental results and compares the proposed method with competing approaches. Finally, Section 5 concludes the paper and discusses future research directions.

2 Related work

DDI prediction has attracted considerable attention due to its importance in drug safety and clinical decision-making. Early research in this area primarily focused on extracting DDIs from biomedical literature. One of the seminal works was conducted by Segura-Bedmar et al. [16], who introduced a framework for identifying drug entities and extracting DDIs from biomedical texts. The authors defined a DDI as any modification in the effect of one drug caused by the presence of another, emphasizing its significance for patient safety. Since a large number of DDIs are reported in scientific publications, automated extraction systems can substantially support pharmacovigilance and clinical decision-making processes. The study was conducted within the DDIExtraction shared task, which expanded the scope of DDI research beyond interaction extraction to include both *drug named-entity recognition (NER)* and *DDI classification*. Compared with earlier versions of the challenge, the updated task incorporated documents from DrugBank and abstracts from MEDLINE, providing a richer and more diverse corpus for model development and evaluation.

The advancement of deep learning methods for molecular representation further accelerated research on DDI prediction. Duvenaud et al. [17] proposed one of the earliest graph neural network (GNN)-based approaches for learning molecular fingerprints. Their objective was to automatically generate vector representations of molecular structures without relying on handcrafted descriptors. The proposed architecture

consists of three stages. First, atom-level embeddings are generated using a neural network. Second, information from all atoms is aggregated to produce a fixed-length molecular fingerprint. Finally, permutation invariance with respect to atom ordering is achieved through a summation-based aggregation function. This work demonstrated the effectiveness of graph-based learning for capturing molecular structure and laid the foundation for subsequent GNN-based approaches in DDI prediction.

Liu et al. [18] proposed a CNN-based approach for extracting DDIs from biomedical literature. Their study was evaluated on the DDIExtraction 2013 benchmark dataset [16], which classifies interactions into four positive categories, namely *Advice*, *Effect*, *Mechanism*, and *Interaction*, together with a negative class representing the absence of interaction.

The proposed framework consisted of two main stages: data preprocessing and model training. During preprocessing, sentences containing candidate drug pairs were normalized by masking drug mentions and identifying unrelated drugs with separate entity markers. The text was subsequently tokenized and converted to lowercase. To reduce noise and improve the quality of negative samples, the authors incorporated additional filtering criteria based on synonymous drug names and chemically related compounds.

Unlike the best-performing systems in the DDIExtraction challenge, which relied on support vector machines with carefully engineered linguistic features, Liu et al. employed a CNN architecture to automatically learn informative representations directly from text. Their results demonstrated that deep neural networks can effectively capture semantic patterns associated with DDIs without extensive manual feature engineering, highlighting the potential of representation learning for biomedical relation extraction.

Zitnik et al. [19] proposed *Decagon*, a graph convolutional neural network framework for modeling polypharmacy side effects caused by DDIs. Unlike conventional DDI prediction methods that mainly determine whether two drugs interact, Decagon was designed to predict the specific side effect through which a drug combination may manifest clinically. The study formulated DDI prediction as a multi-relational link prediction problem over a large multimodal biomedical graph.

The constructed graph integrates several types of biomedical information, including protein-protein interactions, drug-protein target associations, and DDIs associated with polypharmacy side effects. In this graph, drugs and proteins are represented as nodes, while the relationships between them are modeled as edges. DDI edges are further categorized by side-effect type, meaning that each side effect corresponds to a distinct relation type between a pair of drugs.

To learn representations from this heterogeneous network, the authors developed a graph convolutional architecture capable of handling multiple node and edge types. The encoder learns low-dimensional embeddings for drugs and proteins by aggregating information from their graph neighborhoods, while relation-specific transformations are used to capture the semantics of different interaction types. A decoder is then applied to predict whether a given pair of drugs is associated with a particular side effect.

The experimental results showed that Decagon achieved strong performance in predicting polypharmacy side effects and outperformed several baseline methods. The model was particularly effective for side effects with a strong molecular basis, demonstrating the benefit of incorporating drug targets and protein-protein interaction information.

Asada et al. [20] proposed a hybrid deep learning framework for DDI prediction that integrates both molecular structure and textual interaction information. Their study was evaluated on the DDIEExtraction benchmark dataset [16], which categorizes interactions into four positive classes (*Advice*, *Effect*, *Mechanism*, and *Interaction*) together with a negative class representing the absence of interaction.

To represent drug structures, the authors utilized SMILES strings and molecular graphs, where atoms and chemical bonds were modeled as graph components. The proposed framework combined CNNs and recurrent neural networks (RNNs) to learn structural representations of drugs. CNN layers were employed to extract local structural patterns and generate molecular fingerprints, whereas recurrent networks were used to capture sequential dependencies within molecular representations. Drug embeddings were subsequently obtained through the aggregation of atom-level representations.

In addition to structural information, textual descriptions associated with drug interactions were processed using a convolutional architecture. The textual representation was constructed from word embeddings together with position embeddings that encode the relative locations of target drug entities within the sentence. The resulting structural and textual embeddings were fused and provided as input to a multi-class classifier with a Softmax output layer for interaction prediction.

The results demonstrated that integrating molecular structure with textual information improves DDI prediction performance compared with approaches relying on a single modality. This study highlighted the importance of multimodal representation learning and motivated subsequent research on combining complementary sources of drug-related information for interaction prediction.

In 2021, Asada et al. [21] extended their previous DDI prediction framework by introducing richer drug representations derived from multiple complementary information sources. Their approach incorporated three types of information for each drug pair: molecular structure, textual descriptions of drug interactions, and independent textual descriptions of individual drugs. Separate embedding vectors were generated for each modality and subsequently integrated into a unified representation for interaction prediction.

To capture semantic information from biomedical text, the authors employed SciBERT, a transformer-based language model pretrained on scientific literature. Drug-specific textual descriptions were transformed into contextual embeddings, while molecular information was represented using structural features extracted from Drug-Bank. The resulting structural and semantic representations were fused and provided as input to a neural classification network, which predicted interaction categories through a Softmax output layer.

The study demonstrated that integrating drug-specific semantic information with molecular representations can improve DDI prediction performance. By leveraging

contextualized language representations in addition to structural information, the proposed framework provided a more comprehensive view of drug properties and their potential interactions. This work further emphasized the importance of multimodal representation learning in computational pharmacology and inspired subsequent efforts to combine chemical and biomedical-textual information for DDI prediction.

In a subsequent study, Asada et al. [22] proposed a heterogeneous biomedical knowledge graph to enrich drug representations for DDI prediction. The constructed knowledge graph integrated multiple biomedical entities, including drugs, proteins, pathways, and other biological concepts, together with diverse relationships such as drug–drug, drug–protein, and pathway-associated interactions. This design enabled the model to capture complex biological dependencies beyond direct molecular similarity.

To generate drug representations, the authors combined knowledge graph embedding techniques with contextual language representations derived from transformer-based models. Specifically, BERT [23, 24] was used to obtain semantic embeddings from textual drug descriptions, while graph-based learning captured structural and relational information from the heterogeneous knowledge graph. The resulting representations were further refined through data augmentation and alignment strategies to improve robustness and representation quality.

Experimental results demonstrated that incorporating heterogeneous biomedical knowledge substantially improves the characterization of drugs and their interactions. By integrating semantic information with biological relationships, the proposed framework was able to model complex pharmacological mechanisms more effectively than approaches relying solely on molecular or textual features. This work highlighted the potential of knowledge-enhanced representation learning for advancing DDI prediction and related biomedical informatics tasks.

He et al. [13] proposed 3DGT-DDI, a multimodal framework for DDI prediction that combines three-dimensional molecular graph representations with contextual biomedical text information. The model integrates two complementary components: a three-dimensional GNN for structural representation learning and a pretrained language model for semantic feature extraction.

For molecular representation, the authors utilized three-dimensional molecular graphs together with positional and geometric information to capture spatial relationships between atoms. To incorporate textual knowledge, SciBERT was employed to generate contextualized embeddings from biomedical literature describing drug interactions. The structural and semantic representations were subsequently fused and used for interaction prediction.

Their findings indicated that combining three-dimensional molecular information with biomedical text improves DDI prediction performance compared with approaches relying on a single modality. By integrating geometric and semantic information, 3DGT-DDI [13] highlighted the importance of multimodal representation learning and established a strong baseline for subsequent DDI prediction models.

Yu et al. [25] proposed an attention-based framework for predicting drug relationships using heterogeneous biomedical information. Their model simultaneously considers intrinsic drug characteristics and structural information derived from a

heterogeneous graph containing drugs, genes, and diseases. This design enables the framework to capture complex biological dependencies associated with drug interactions.

To represent intrinsic drug properties, chemical similarity-based structural fingerprints were extracted and compressed using principal component analysis (PCA). For gene and disease entities, feature representations were generated using the TransE embedding model. After constructing a heterogeneous graph, the authors employed a graph attention network (GAT) to learn structural representations while adaptively weighting neighboring nodes according to their relevance. Multi-layer attention propagation was further applied to capture higher-order neighborhood information and alleviate sparsity issues.

The intrinsic and graph-derived representations were subsequently fused to generate a unified embedding for each drug. These embeddings were then used to predict drug relationships. The study demonstrated that attention-based aggregation over heterogeneous biomedical networks can effectively capture both local and global biological information, leading to improved interaction prediction performance.

Shi et al. [26] proposed SubGE-DDI, a multimodal framework that combines textual information and heterogeneous knowledge graph representations for DDI prediction. The model leverages both unstructured biomedical text and structured biological knowledge to construct comprehensive drug representations.

For textual representation learning, the authors employed a pretrained BERT model to extract semantic information from drug descriptions and biomedical documents. In parallel, a heterogeneous biomedical knowledge graph was constructed by integrating relationships among drugs, proteins, genes, pathways, and other biological entities. To improve representation quality, the model focused on local subgraphs and utilized attention mechanisms to filter irrelevant edges and preserve informative biological relationships.

The semantic representations derived from BERT and the structural representations obtained from the knowledge graph were subsequently fused to generate a unified drug embedding. The resulting multimodal representation was used to classify drug pairs into one of four interaction categories (*Advice*, *Effect*, *Mechanism*, and *Interaction*) or a negative class indicating the absence of interaction. The results demonstrated that combining textual semantics with knowledge graph information substantially improves DDI prediction performance and provides a richer understanding of pharmacological relationships.

More recently, Li et al. [27] proposed MDG-DDI, a multi-feature drug graph framework for DDI prediction. The model jointly exploits molecular graph structures and substructure-level semantic information to enhance predictive performance. Rather than relying on a single representation, MDG-DDI integrates multiple molecular features to construct a more expressive drug graph, enabling the capture of both local chemical patterns and global molecular relationships. The study highlights the importance of rich molecular representations and suggests that single-source drug features may be insufficient for modeling complex pharmacological interactions. However, MDG-DDI primarily focuses on molecular graph information. In contrast, the proposed HySEAtt-DDI incorporates structural similarity, drug-protein

interaction profiles, and semantic representations derived from biomedical text. Consequently, HySEAtt-DDI extends the multimodal perspective by integrating biological and textual information alongside chemical structure.

Chen et al. [28] introduced ACaps-DDI, a DDI prediction model that combines capsule neural networks with self-attention mechanisms. The model aims to improve feature fusion by leveraging multi-source drug information and attention-based representation learning. Specifically, the self-attention module identifies important feature relationships, while the capsule network preserves hierarchical feature information relevant to DDI prediction. Similar to HySEAtt-DDI, ACaps-DDI employs attention mechanisms to enhance predictive performance. However, HySEAtt-DDI differs by integrating multimodal similarity-based representations with an autoencoder, a Squeeze-and-Excitation block, and multi-head self-attention. Whereas ACaps-DDI emphasizes capsule-based feature modeling, HySEAtt-DDI focuses on learning complementary representations from structural, biological, and semantic drug information.

Wang et al. [29] proposed DDI-AttendNet, an attention-based framework designed to capture both intra-drug atomic relationships and inter-drug interaction patterns. The model employs graph representations and cross-attention mechanisms to identify important molecular substructures and their interactions across drug pairs. The study demonstrates the effectiveness of attention mechanisms for modeling complex drug relationships. Unlike DDI-AttendNet, which primarily relies on molecular graph structures, HySEAtt-DDI integrates multiple complementary modalities, including structural similarity, biological target similarity, and semantic similarity derived from biomedical literature. Therefore, while DDI-AttendNet validates the utility of attention-based learning, HySEAtt-DDI extends this paradigm through a multimodal framework that combines chemical, biological, and textual sources of drug information.

Chianumba et al. [4] studied the effect of multimodal feature integration on DDI prediction. Their work showed that combining diverse drug information can improve prediction performance across different ML and graph-based models. In particular, their results indicated that graph attention networks benefited strongly from multimodal integration, suggesting that attention-based graph models can effectively exploit complementary drug features.

This work is important because it supports a central assumption of HySEAtt-DDI: drug interactions cannot be fully explained by a single modality. Chemical structure, biological activity, and pharmacological information may each describe different aspects of a drug. Therefore, integrating multiple sources of information can provide a more complete representation for DDI prediction.

Compared with their study, HySEAtt-DDI proposes a specific multimodal deep learning architecture that combines similarity-based feature construction, autoencoder-based dimensionality reduction, Squeeze-and-Excitation recalibration, and multi-head self-attention. Thus, while Chianumba et al. demonstrate the general value of multimodal integration, HySEAtt-DDI provides a dedicated architecture for exploiting such multimodal drug representations.

Collectively, recent advances in DDI prediction have progressively moved from single-modality approaches toward multimodal representation learning frameworks that combine chemical, biological, and textual information. Nevertheless, most existing

Table 1: Summary of representative DDI prediction methods.

Study	Modality	Method	Key Idea
Segura-Bedmar et al. [16]	Text	SVM / CNN features	First benchmark for DDI extraction from biomedical text
Duvenaud et al. [17]	Molecular graph	GNN fingerprints	Learns molecular structure automatically from graphs
Liu et al. [18]	Text	CNN model	Learns DDI patterns without handcrafted features
Zitnik et al. [19]	Graph network	GCN (Decagon)	Predicts polypharmacy side effects using biomedical graph
Asada et al. [20]	Multimodal	CNN + RNN	Combines molecular + text representations
Asada et al. (2021a) [21]	Multimodal	SciBERT fusion	Adds contextual biomedical semantics
Asada et al. (2021b) [22]	Knowledge graph	KG + BERT	Uses biomedical knowledge graph + text embeddings
He et al. [13]	3D + text	3D GNN + SciBERT	Combines spatial molecular info with text
Yu et al. [25]	Heterogeneous graph	GAT + TransE	Uses attention over biomedical graph
Shi et al. [26]	KG + text	BERT + subgraph KG	Combines text + local graph structures
Li et al. [27]	Molecular graph	Multi-feature GNN	Uses rich molecular substructure features
Chen et al. [28]	Multi-source	Capsule + attention	Improves feature fusion with capsule networks
Wang et al. [29]	Molecular graph	Attention GNN	Models atomic and inter-drug interactions
Chianumba et al. [4]	Multimodal	Empirical study	Shows benefit of multimodal integration

methods rely on molecular graphs, heterogeneous knowledge graphs, or textual representations independently, while limited attention has been given to jointly exploiting structural similarity, drug-protein interaction profiles, and contextual biomedical semantics within a unified attention-based architecture. To address this limitation, we propose HySEAtt-DDI, a multimodal framework that integrates structural, biological, and semantic representations through autoencoder-based feature compression, Squeeze-and-Excitation recalibration, and multi-head self-attention learning for accurate DDI prediction.

As summarized in Table 1, existing DDI prediction studies have evolved from text-based extraction and handcrafted feature-based methods toward deep learning, GNNs, knowledge graphs, and multimodal representation learning. However, many existing approaches still emphasize a single dominant source of information, such as molecular graphs, biomedical text, or heterogeneous knowledge graphs. Although recent

multimodal methods have improved DDI prediction by combining complementary drug features, limited attention has been given to jointly exploiting structural similarity, drug-protein interaction profiles, and contextual biomedical semantic similarity within a unified attention-based framework. To address this limitation, we propose HySEAtt-DDI, a multimodal framework that integrates structural, biological, and semantic representations through autoencoder-based feature compression, Squeeze-and-Excitation recalibration, and multi-head self-attention learning for accurate DDI prediction.

3 Materials and Methods

3.1 Dataset Construction and Quality Control

A well-defined and comprehensive dataset is essential for the development and evaluation of reliable DDI prediction models. Therefore, we conducted a comprehensive study of the available dataset on DDI prediction. He et al. [13] collected their data from Drugbank [30] and MEDLINE [31]. Although the authors made a considerable effort to construct the dataset, our further analysis discovered some structural and conceptual inconsistencies in the dataset. The most important inconsistency is associating incorrect or incomplete SMILES strings with corresponding drugs, e.g., non-consistent with the syntactic rules of SMILES Generation, recording incomplete sequences, and assigning to an irrelevant drug. Besides that, there is no proper standard for naming the drugs in the He dataset. For example, some drugs have been recorded with their generic name, some others with their commercial name, which caused some of the drugs to be recorded more than once in the dataset. There are other inconsistencies, such as inconsistencies in letter capitalization in saving drug names and properties; consequently, another source of redundancy and repetition of drugs in the dataset. Last but not least, there are rows or drugs in the dataset with incomplete properties, e.g., a lack of sequence drug name or DB ID. Such inconsistencies may introduce artifacts and biases that can adversely affect model training, evaluation, and downstream inference. Therefore, revising and refining the chosen dataset is critical for proper downstream tasks and analysis. Shi et al. [26], while the method is innovative and novel, used a dataset where the same drugs have been repeated several times with different labels. Segura et al. [16] used a dataset with the majority of negative classes and a negligible ratio of positive classes.

These investigations led us to create a new dataset from scratch. To do this, all the information has been gathered from DrugBank. This work covers two main entities – small molecular drugs and those proteins that have an established interaction with these types of drugs. It is worth mentioning that in the pharmacology domain, there are different types of drug interactions; among them, this work concentrates on the small molecule-small molecule drug interactions, which we call drug-drug interactions (DDI) in this work. For each small molecule drug, its information, i.e., its generic name, its corresponding SMILES string, its description, its interactions with other drugs, and its protein target have been collected. In more detail, each drug has been identified not only with its chemical structure properties, but also by its functional description based on molecular targets and its relation with other drugs and proteins.

A combination of drug-drug interactions and drug-protein interactions ensures that the ML approach, in addition to identifying and utilizing the structural similarities, becomes sensitive to functional similarities and dissimilarities. Using these types of information simultaneously can lead to enhanced performance in learning complicated patterns of drug interactions and play a crucial role in the final prediction.

The final dataset contains information on 2,148 approved small-molecule drugs and 2,056 protein targets from DrugBank. The labels of DDIs are categorized into four positive classes (types of known interactions) and one zero class (representative of unknown or negative interactions). As mentioned above, the aim of this study is the prediction of the interaction type between each pair of small-molecule drugs based on their extracted structural and descriptive data. Structural properties are based on the SMILES strings of drugs, and descriptive properties are the formal text recorded in DrugBank. Overall, the new dataset contains 592840 drug-drug interactions and 7405 drug-protein interactions. These direct and indirect interactions among drugs help to improve the prediction performance. Table 2 shows the statistical properties of the proposed dataset briefly.

Table 2: Dataset Statistics

Entity	Count
Small-molecule drugs	2148
Protein targets	2056
Drug-target interactions	7405
Drug-Drug interactions	592840

Table 3: Distribution of DDI labels in the dataset.

Label	Description	Count	Percentage (%)
0	No interaction or unknown interaction	296,420	50.00
1	Known interaction type I	118,028	19.91
2	Known interaction type II	99,692	16.81
3	Known interaction type III	39,915	6.73
4	Known interaction type IV	38,785	6.54
Total		592,840	100.00

Table 3 reports the label distribution of DDIs. Label 0 shows the unknown or lack of interaction between corresponding drug pair and labels 1 through 4 are assigned to types of known interactions between corresponding drug pairs. Due to a high number of available zero classes, the dataset could be highly imbalanced; therefore, the number of zero classes has been chosen in order to be equal to the total number of positive classes. In other words, label zero has 296420 samples (50% of the whole dataset, label

z1 has 118028 samples, and classes 2, 3, and 4 have 99692, 39915, and 38785 samples, respectively. Totally, there are 296420 positive samples in the proposed dataset.

drug1 may decrease the excretion rate of drug2 which could result in a higher serum level.
The metabolism of drug1 can be decreased when combined with drug2.
The serum concentration of drug2 can be increased when it is combined with drug1.
The risk or severity of CNS depression can be increased when drug1 is combined with drug2.

Fig. 1: Positive classes present in the proposed dataset.

Figure 1 presents representative examples of positive DDI instances in the dataset. To ensure data consistency and prevent entity-specific bias, all drug names were systematically anonymized using generic placeholders such as *drug1* and *drug2*. This preprocessing step standardizes input representations across the corpus and eliminates variability arising from heterogeneous drug nomenclature.

Such anonymization is crucial in mitigating the risk of spurious feature learning, where models may inadvertently exploit drug identity tokens or frequency patterns rather than learning the underlying semantic and contextual cues that characterize true pharmacological interactions [16, 18]. In particular, the presence of explicit drug names may introduce dataset bias, allowing the model to associate frequently occurring entities with specific interaction labels without capturing meaningful relational semantics.

By masking drug entities, the learning process is constrained to rely on contextual linguistic information, including interaction-triggering verbs, pharmacodynamic and pharmacokinetic descriptions, dosage-related expressions, and other surrounding biomedical cues. This encourages the model to focus on relational semantics rather than lexical memorization.

Furthermore, this normalization strategy enhances the generalization capability of the model by ensuring invariance to lexical variations and repeated occurrences of the same drug entities across different samples. Consequently, the model becomes more robust and better suited for real-world biomedical text, where unseen drug combinations frequently occur.

3.2 HySEAtt-DDI Framework

Figure 2 shows the general view of the proposed framework. It contains five modules: dataset preparation, preprocessing phase, dimensionality reduction, model learning, and prediction phase.

3.2.1 Multimodal Drug Representation Learning

To construct a comprehensive representation of each drug, HySEAtt-DDI integrates three complementary sources of drug information: chemical structure, drug-protein interactions, and textual descriptions. Each source captures a distinct aspect of drug knowledge. Chemical structure provides a molecular basis for assessing similarity

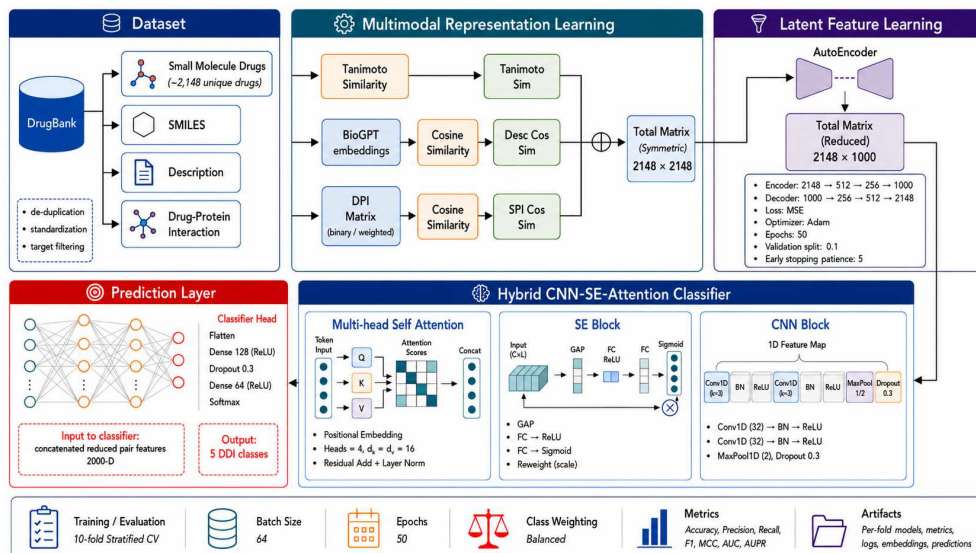


Fig. 2: Overview of the proposed HySEAtt-DDI framework. Structural, biological, and textual drug information are first integrated to construct multimodal drug representations. An autoencoder is then employed to learn compact latent features, which are subsequently processed by CNN, Squeeze-and-Excitation (SE), and multi-head self-attention modules to capture local, channel-wise, and global feature dependencies. Finally, an MLP classifier predicts the DDI type.

among compounds, as structurally similar drugs often exhibit comparable biological properties and pharmacological activities. Drug-protein interaction information offers insights into biological mechanisms, target engagement, and functional relationships between drugs and biological systems [32, 33]. Furthermore, textual descriptions capture conceptual, pharmacological, and therapeutic information that may not be directly inferred from chemical structures or protein interactions. The preprocessing stage aims to transform these heterogeneous data sources into a unified representation with compatible dimensions, enabling their effective integration into a single feature matrix.

Using chemical information, HySEAtt-DDI extracts the Morgan fingerprint $\mathbf{fp}_i$ of each drug i from its corresponding SMILES representation. Morgan fingerprints are employed because of their high sensitivity to molecular substructures and their ability to capture subtle structural similarities among compounds. Subsequently, the Tanimoto similarity coefficient is computed for every pair of drugs. This process results in a $2,148 \times 2,148$ structural similarity matrix, denoted by S_{str} , whose values lie within the interval $[0, 1]$, with diagonal entries equal to one.

$$S_{str}(i, j) = \text{Tanimoto}(\mathbf{fp}_i, \mathbf{fp}_j) \quad (1)$$

Biological information is derived from drug–target interactions and represented in a binary format. Specifically, each drug is associated with a binary vector $\mathbf{p}_i$, $\forall i \in \mathcal{D}$, where each entry indicates either the presence (1) or absence (0) of an interaction with a corresponding protein. The binary representation is intentionally adopted to minimize uncertainties arising from experimental heterogeneity and varying interaction confidence levels. HySEAtt-DDI then computes the cosine similarity between the interaction vectors of every pair of drugs. The resulting similarities are assembled into a biological similarity matrix, $S_{spi} \in \mathbb{R}^{2148 \times 2148}$, which captures the functional similarity of drugs based on their protein interaction profiles.

$$S_{spi}(i, j) = \cos(\mathbf{p}_i, \mathbf{p}_j), \forall i, j \in 1, \dots, N \quad (2)$$

To extract semantic information from drug descriptions, HySEAtt-DDI employs BioGPT [34], a transformer-based language model pretrained on large-scale biomedical corpora. Unlike traditional approaches such as TF-IDF [35] and Word2Vec [36], BioGPT captures contextual semantics and long-range dependencies at both sentence and document levels. Moreover, it demonstrates superior generalization capabilities compared with several alternative biomedical language models, such as BioBERT [37]. Consequently, BioGPT is particularly effective in processing medical and pharmacological descriptions that contain specialized terminology and domain-specific concepts.

For each drug, the corresponding textual description obtained from DrugBank is tokenized using the BioGPT tokenizer with a maximum sequence length of 256 tokens. Shorter sequences are padded, whereas longer sequences are truncated in a controlled manner. The hidden representations from the final transformer layer are aggregated through attention-mask-weighted averaging to generate a fixed-length embedding while excluding the influence of padding tokens. In addition, HySEAtt-DDI utilizes the special classification token (e.g., [CLS]) instead of the output pooler representation, as it provides a more robust semantic representation [24].

Formally, for each drug i , HySEAtt-DDI processes its textual description through BioGPT and generates a semantic embedding vector $\mathbf{b}_i$ of dimension 256. A semantic similarity matrix is then constructed by applying cosine similarity to the embedding vectors of every pair of drugs.

$$S_{text}(i, j) = \cos(\mathbf{b}_i, \mathbf{b}_j) \quad (3)$$

The final feature representation is obtained by integrating the three similarity matrices derived from chemical structures, protein interaction profiles, and textual descriptions. Specifically, the matrices are combined through element-wise summation:

$$F = S_{str} + S_{spi} + S_{text} \quad (4)$$

The resulting feature matrix captures complementary structural, biological, pharmacological, and semantic characteristics of drugs, providing a comprehensive representation for downstream learning tasks.

3.2.2 Latent Feature Learning

To reduce redundancy and noise in the fused feature space, perform nonlinear dimensionality reduction, and learn compact latent representations, HySEAtt-DDI employs an autoencoder (AE) [38] on the final feature vectors $\mathbf{f}_i \in F$, where $i \in 1, \dots, N$ and N denotes the total number of drugs. Let $\mathbf{f}_i \in \mathbb{R}^d$ represent the input feature vector and $\hat{\mathbf{f}}_i \in \mathbb{R}^d$ its reconstruction generated by the autoencoder, where d is the input dimensionality. The encoder consists of a multi-layer fully connected network that progressively compresses the input into a low-dimensional bottleneck representation, while the decoder reconstructs the original input from this latent space. The resulting latent representation, denoted by $\mathbf{z}_i$, preserves the most informative characteristics of the original feature vector while reducing noise and redundancy.

The autoencoder is trained by minimizing the mean squared error (MSE) between the original and reconstructed feature vectors. The reconstruction loss is defined as follows:

$$\mathcal{L}_{\text{AE}} = \frac{1}{N} \sum_{i \in \mathcal{D}} \left\| \mathbf{f}_i - \hat{\mathbf{f}}_i \right\|_2^2, \quad (5)$$

where $\hat{\mathbf{f}}_i = De(En(\mathbf{f}_i))$, $En(\cdot)$ denotes the encoder, and $De(\cdot)$ denotes the decoder of the autoencoder. After training, the latent representation of each drug is obtained as

$$\mathbf{z}_i = En(\mathbf{f}_i). \quad (6)$$

The latent representations $\mathbf{z}_i$ are subsequently used as inputs to the classification network. To effectively capture local patterns, channel-wise feature importance, and long-range dependencies, HySEAtt-DDI employs a hybrid CNN-SE-Attention architecture.

3.2.3 Hybrid CNN-SE-Attention Classifier

This section aims to learn informative representations from the input feature vectors and subsequently perform the downstream classification task. To achieve this objective, HySEAtt-DDI comprises three main components: a CNN block, a Squeeze-and-Excitation (SE) block, and a multi-head self-attention block. These components are designed to capture complementary feature dependencies at different levels, ranging from local patterns to long-range interactions, thereby enabling a more comprehensive representation of the input features.

Figure 3 illustrates the CNN block employed in HySEAtt-DDI. The block comprises two consecutive one-dimensional convolutional layers, each followed by a ReLU activation function and batch normalization. Both convolutional layers use three kernels and identical padding to preserve the input dimensionality. This design enables the extraction of local feature patterns while promoting stable and efficient training through batch normalization. In short, the CNN acts as follows.

$$\mathbf{h}^{(l)} = \text{ReLU} \left(\text{BN} \left(\mathbf{W}^{(l)} * \mathbf{h}^{(l-1)} + \mathbf{b}^{(l)} \right) \right), \quad (7)$$



Diagram illustrating the proposed channel-wise reweighting module:

- Input feature map:** $X \in \mathbb{R}^{C \times H \times W}$ (dimensions H, W, C).
- Excitation Path:**
 - Squeeze:** Global Average Pooling operation.
 - Intermediate vector:** $z \in \mathbb{R}^C$.
 - Excitation Block (dashed box):**
 - FC, ReLU C/r** (where r is the reduction ratio).
 - FC, Sigmoid C** .
 - Output vector:** $s \in \mathbb{R}^C$.
- Scale Operation:**
 - Combines the input X and the output vector s .
 - Operation: $X' = s \otimes X$ (channel-wise reweighting).
- Output feature map:** $X' \in \mathbb{R}^{C \times H \times W}$ (dimensions H, W, C).

Fig. 4: Architecture of the Squeeze-and-Excitation block.

17

In the squeeze stage, global average pooling is applied to aggregate the information of each channel into a single descriptor. Specifically, the channel descriptor $\mathbf{z}_i^c$ is computed as

$$\mathbf{z}_i^c = \frac{1}{L} \sum_{i=1}^L U_c(i), \quad c = 1, \dots, C, \quad (8)$$

where $U_c(i)$ denotes the i -th element of the c -th channel. This operation generates a compact channel-wise representation $\mathbf{z} \in \mathbb{R}^C$ that summarizes the global information of each feature channel.

In the excitation stage, the channel descriptors are passed through a lightweight gating mechanism composed of two fully connected layers (FC). The channel attention weights are obtained as

$$\mathbf{s} = \sigma(W_2 \delta(W_1 \mathbf{z})), \quad (9)$$

where W_1 and W_2 are trainable weight matrices corresponding to the dimensionality reduction and expansion operations, respectively, $\delta(\cdot)$ denotes the ReLU activation function, and $\sigma(\cdot)$ is the sigmoid activation function. The resulting vector $\mathbf{s} \in [0, 1]^C$ contains channel-wise attention weights that quantify the relative importance of each channel.

The recalibrated feature map is then obtained through channel-wise multiplication:

$$\tilde{\mathbf{U}}_c = s_c \cdot \mathbf{U}_c, \quad c = 1, \dots, C, \quad (10)$$

where s_c denotes the attention weight associated with the c -th channel. This operation amplifies informative channels while suppressing less relevant or noisy ones, thereby enhancing the discriminative capability of the learned representations.

After channel recalibration, a one-dimensional max-pooling layer with a pool size of 2 is applied to reduce the effective sequence length and computational complexity. Finally, a dropout layer with a rate of 0.3 is employed to improve generalization and reduce the risk of overfitting.

Although the CNN and SE blocks effectively capture local patterns and channel-wise feature importance, their receptive fields remain limited in modeling long-range dependencies and global interactions among features. To address this limitation, HySEAtt-DDI incorporates a multi-head self-attention (MHSA) layer after the CNN-SE module. Figure 5 illustrates the multi-head self-attention module employed in HySEAtt-DDI. Self-attention enables each feature position to dynamically attend to all other positions in the sequence, thereby capturing complex and non-local relationships that may not be adequately represented by convolutional operations alone.

Let $\mathbf{X} \in \mathbb{R}^{L \times d}$ denote the input feature sequence, where L is the sequence length and d is the feature dimension. The self-attention mechanism first projects the input into query, key, and value representations:

$$\mathbf{Q} = \mathbf{X}W_Q, \quad \mathbf{K} = \mathbf{X}W_K, \quad \mathbf{V} = \mathbf{X}W_V, \quad (11)$$

where W_Q , W_K , and W_V are learnable projection matrices. The attention weights are then computed using the scaled dot-product attention mechanism:

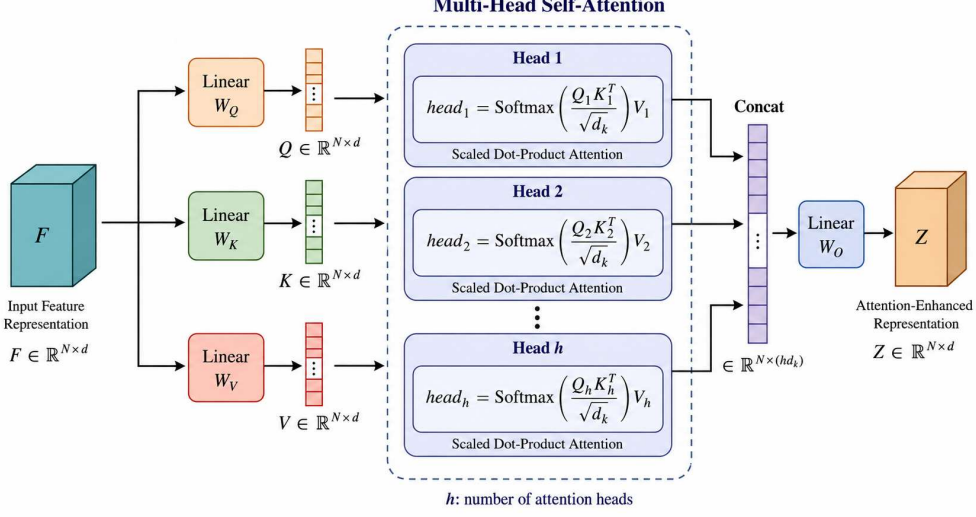


Fig. 5: Architecture of the multi-head self-attention module.

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax} \left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}} \right) \mathbf{V}, \quad (12)$$

where d_k denotes the dimensionality of the key vectors. This formulation enables the model to assign higher importance to relevant features regardless of their distance in the sequence.

To enhance representation learning, multiple attention heads are employed in parallel. Specifically, HySEAtt-DDI utilizes four attention heads with a key dimension of 16. Each head learns complementary feature relationships in different representation subspaces. The outputs of all heads are concatenated and linearly transformed to obtain the final attention representation:

$$\text{MHSA}(\mathbf{X}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W_O, \quad (13)$$

where

$$\text{head}_i = \text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i), \quad i = 1, \dots, h, \quad (14)$$

$h = 4$ is the number of attention heads, and W_O is a learnable output projection matrix.

A residual connection is employed around the attention layer to facilitate information propagation and mitigate gradient degradation. Subsequently, layer normalization is applied to stabilize training and improve convergence:

$$\mathbf{Y} = \text{LayerNorm}(\mathbf{X} + \text{MHSA}(\mathbf{X})). \quad (15)$$

Finally, a dropout layer with a rate of 0.3 is applied to the attention outputs to improve generalization and reduce overfitting.

The attention-enhanced feature representations are subsequently flattened and passed to a multilayer perceptron (MLP) classifier consisting of a dense layer with 128 units and ReLU activation, dropout with a rate of 0.3, a second dense layer with 64 units and ReLU activation, and a final dense output layer with Softmax activation to produce a probability distribution over five interaction classes.

Let C denote the total number of classes. For each input sample, the corresponding ground-truth label is defined as

$$y \in \{0, 1, \dots, C - 1\}. \quad (16)$$

The classifier outputs a probability distribution over the C classes. Let $\hat{p}_k$ denote the predicted probability of the sample belonging to class k , where $k \in \{0, 1, \dots, C-1\}$. The predicted probabilities are computed using the softmax function applied to the output logits z_k :

$$\hat{p}_k = \frac{\exp(z_k)}{\sum_{j=0}^{C-1} \exp(z_j)}. \quad (17)$$

The hybrid classification network is trained using sparse categorical cross-entropy loss. For a single training instance, the loss is defined as $\mathcal{L}_{\text{CE}} = -\log(\hat{p}_y)$. To mitigate class imbalance, a class-weighted loss function is employed. Let w_y denote the weight associated with class y . The weighted cross-entropy loss for a single instance is given by $\mathcal{L}_{\text{WCE}} = -w_y \log(\hat{p}_y)$.

For a dataset consisting of N samples, the empirical risk minimized during training is

$$\mathcal{L}_{\text{WCE}} = -\frac{1}{N} \sum_{n=1}^N w_{y_n} \log(\hat{p}_{y_n}). \quad (18)$$

3.2.4 Optimization Objective

The model is optimized in two sequential stages. In the first stage, the autoencoder parameters are learned by minimizing $\mathcal{L}_{\text{AE}}$. In the second stage, the classifier parameters are learned by minimizing the class-weighted cross-entropy loss $\mathcal{L}_{\text{WCE}}$. The overall optimization objective is therefore

$$\mathcal{L} = \begin{cases} \mathcal{L}_{\text{AE}}, & \text{autoencoder training,} \\ \mathcal{L}_{\text{WCE}}, & \text{classifier training.} \end{cases} \quad (19)$$

Algorithm 1 summarizes the complete workflow of HySEAtt-DDI, including multimodal similarity fusion, latent representation learning, CNN-SE-attention-based classification, model optimization, and performance evaluation. Lines 1–2 define the inputs and outputs of the framework. The inputs consist of a drug set $\mathcal{D}$ with corresponding SMILES representations, drug–target interaction profiles, textual descriptions, and multi-class DDI labels. The outputs are the trained HySEAtt-DDI model and the associated evaluation metrics reported as the mean and standard deviation over K cross-validation folds.

Lines 3–6 describe the multimodal similarity fusion stage. First, structural, biological, and semantic similarity matrices are computed according to Eqs. (1), (2), and (3), respectively (Line 2). These matrices capture complementary structural, functional, and semantic relationships among drugs. Subsequently, the similarity matrices are integrated using Eq. (4) to generate the final feature vectors $\mathbf{f}_i$ for all drugs (Line 3).

Lines 4–6 present the latent representation learning stage. An autoencoder is trained using the fused feature vectors in order to learn compact and informative latent representations while reducing noise and redundancy (Line 5). After optimization of the reconstruction objective, the encoder produces the latent representation $\mathbf{z}_i = \text{En}(\mathbf{f}_i)$ for each drug (Line 6), which serves as the basis for subsequent drug-pair construction.

Lines 7–11 describe the drug-pair construction procedure. For each interacting drug pair (i, j) , the latent representations of the two drugs are concatenated to form the pairwise feature vector $\mathbf{x}_{ij} = [\mathbf{z}_i | \mathbf{z}_j]$ (Lines 8–10). The resulting feature vectors are then assembled into the final dataset X used for classification (Line 11).

Lines 12–32 describe the model training and evaluation process using stratified 10-fold cross-validation. The dataset is partitioned into $K = 10$ folds while preserving the class distribution (Lines 13–14). For each fold, class weights are computed to alleviate class imbalance (Line 16), and the HySEAtt-DDI classifier is trained using the training partition.

The architecture of HySEAtt-DDI is presented in Lines 17–26. First, the input feature vectors are reshaped and processed by two one-dimensional convolutional layers with batch normalization and ReLU activation to extract local feature patterns (Lines 18–19). Max-pooling and dropout are subsequently applied to reduce feature dimensionality and improve generalization (Line 20). An SE block is then employed to recalibrate channel-wise feature responses and emphasize informative channels while suppressing less relevant ones (Line 21). Positional embeddings are added to preserve sequential information prior to attention modeling (Line 22). Next, a multi-head self-attention module captures long-range dependencies and global feature interactions according to Eq. (13) (Line 23). Residual connections and layer normalization are subsequently applied to stabilize training and facilitate information flow (Line 24). The resulting attention-enhanced representations are flattened and passed through fully connected layers (Line 25), followed by a Softmax output layer that predicts the DDI class probabilities (Line 26).

Lines 27–29 describe the optimization stage. The classifier parameters are learned by minimizing the weighted cross-entropy loss defined in Eq. (18) using the Adam optimizer. Early stopping is employed to prevent overfitting and improve generalization performance.

Finally, Lines 30–32 present the evaluation procedure. For each cross-validation fold, the predictive performance of HySEAtt-DDI is assessed using Accuracy, Precision, Recall, F1-score, Matthews Correlation Coefficient (MCC), Area Under the Receiver Operating Characteristic Curve (AUROC), and Area Under the Precision–Recall Curve (AUPR). The final reported results correspond to the mean and standard deviation of these metrics across all 10 folds.

Algorithm 1 HySEAtt-DDI: Multimodal similarity fusion with CNN-SE-Attention network for DDI prediction

Require: Drug set $\mathcal{D}$ with corresponding SMILES, target interactions, and descriptions; labels $\mathbf{y} \in \{0, 1, 2, 3, 4\}$; latent dim $d_z = 1000$; folds no $K = 10$

Ensure: Trained model and evaluation metrics (mean $\pm$ std over K folds)

- 1: **Step 1: Multimodal similarity fusion**
- 2: Compute structural, biological, and semantic similarity matrices using Eqs. (1), (2), (3)
- 3: Generate feature vectors $\mathbf{f}_i$ using Eq. (4)
- 4: **Step 2: Autoencoder-based representation learning**
- 5: Train autoencoder on $\mathbf{f} \in F$ using loss function (5)
- 6: Obtain latent representation: $\mathbf{z}_i = En_\theta(\mathbf{f}_i)$
- 7: **Step 3: Drug-pair construction**
- 8: **for** each $i, j \in \mathcal{D}$ **do**
- 9: Construct pair feature: $\mathbf{x}_{ij} = [\mathbf{z}_i \parallel \mathbf{z}_j]$
- 10: **end for**
- 11: Form dataset X
- 12: **Step 4: 10-fold stratified training**
- 13: Split $X, \mathbf{y}$ into $K = 10$ folds
- 14: **for** $k = 1$ to K **do**
- 15: Train/test split
- 16: Compute class weights
- 17: **Step 4.1: HySEAtt-DDI model**
- 18: Reshape input $\mathbf{x}$
- 19: Apply Conv1D layers + BatchNorm + ReLU
- 20: Apply MaxPooling + Dropout
- 21: Apply Squeeze-and-Excitation block
- 22: Add positional embedding
- 23: Apply multi-head attention using Eq. (13)
- 24: Residual connection + normalization using Eq. (15)
- 25: Flatten and apply Dense layers
- 26: Output prediction using Eq. (17):
- 27: **Step 4.2: Optimization**
- 28: Minimize weighted loss using Eq. (18)
- 29: Adam optimizer + early stopping
- 30: **Step 4.3: Evaluation**
- 31: Accuracy, Precision, Recall, F1, MCC, AUROC, AUPR
- 32: **end for**

4 Results

This section presents the experimental results obtained using the proposed HySEAtt-DDI framework. All experiments were conducted using Python 3 and evaluated under a stratified 10-fold cross-validation protocol. The results are organized into several subsections to provide a comprehensive analysis of the proposed approach and its comparison with competing methods.

4.1 Experimental Setup and Evaluation Protocol

To ensure a reliable and unbiased evaluation, a stratified 10-fold cross-validation strategy was employed. This protocol preserves the original class distribution in each fold, which is particularly important for handling imbalanced datasets. Consequently, both the training and testing partitions contain representative proportions of samples from each class.

The dataset consists of five labels, including four positive interaction classes and one negative class. To mitigate class imbalance and reduce model bias, the negative class was sampled to match the combined size of the positive classes. Furthermore, class-weighted training was employed during model optimization.

Model training was performed using the Adam optimizer with a learning rate of 10^{-4} and the sparse categorical cross-entropy loss function. To improve model generalization and mitigate overfitting, several regularization strategies were incorporated, including dropout, weight decay, validation-based monitoring, and early stopping.

Since the task involves multi-class classification, performance is reported using three averaging schemes: *micro*, *macro*, and *weighted*. These perspectives provide complementary insights into classifier performance, particularly in the presence of class imbalance.

4.2 Performance Metrics

The effectiveness of HySEAtt-DDI was assessed using several widely adopted evaluation metrics and compared against state-of-the-art approaches. These measures are computed from the numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (20)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (21)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (22)$$

$$\text{AUROC} = \sum_{i=1}^n \text{TPR}_i \Delta \text{FPR}_i, \quad (23)$$

$$\text{AUPR} = \sum_{i=1}^n \text{Precision}_i \Delta \text{Recall}_i. \quad (24)$$

These metrics collectively evaluate classification performance from different perspectives, including predictive capability, class discrimination, and robustness to class imbalance.

4.3 Computational Environment

All experiments were conducted on a centralized computing platform running Ubuntu 20.04.6 LTS. The system was equipped with a 64-bit Intel(R) Core(TM) i5-8400 processor operating at 2.80 GHz, 64 GB DDR4 RAM at 2133 MHz, a 1 TB internal hard disk drive, and an NVIDIA GeForce GTX 1080 GPU. BioGPT-based feature extraction was performed using GPU acceleration, while similarity matrices for the autoencoder framework were represented using `float32` precision to improve computational efficiency.

4.4 Baseline Models and Hyperparameter Tuning

To demonstrate the effectiveness of HySEAtt-DDI, we compared its performance against three classical ML methods, namely Random Forest (RF), Support Vector Machine (SVM), and XGBoost. Hyperparameter optimization for these models was performed using grid search. The optimal hyperparameter configurations are summarized in Table 4.

Table 4: Best hyperparameter settings for classical ML models

XGBoost	Random Forest	SVM
Number of estimators: 100	Number of estimators: 10	Kernel: Linear
Maximum depth: 3	Criterion: Log loss	Regularization parameter (C): 10
Learning rate: 0.1	Max features: $\log_2$	
Subsample: 0.5		
Colsample by tree: 0.5		

In addition to the classical ML models, HySEAtt-DDI was compared with representative deep learning baselines, including an MLP-based model [14] and a CNN-based model [39]. For all baseline methods, the feature representations of each drug pair were concatenated and subsequently used as input to the corresponding classifier.

The following subsections present a detailed comparison of the proposed framework with these baseline methods and provide an in-depth analysis of the obtained results.

4.5 Micro-Averaged Results

Table 5 summarizes the micro-averaged performance of the evaluated machine learning and deep learning methods. Among the classical ML methods, random forest has the highest performance and achieved 97% and 92% for AUCROC and AUPR, respectively. XGBoost, on the other hand, achieves the lowest overall performance. To evaluate the proposed method, we have used several configuration for the proposed HySEAtt-DDI. HySEAtt-DDI with MLP outperforms all other methods in all metrics.

Table 5: Micro-averaged evaluation of the methods (Best results in bold).

Method	F1	AUROC	AUPR
SVM	0.7297 ± 0.0023	0.9365 ± 0.0005	0.8115 ± 0.0013
RF	0.8517 ± 0.0031	0.9776 ± 0.0005	0.9244 ± 0.0017
XGBoost	0.6076 ± 0.0014	0.8953 ± 0.0005	0.6985 ± 0.0008
MLP	0.7019 ± 0.0046	0.9319 ± 0.0016	0.7812 ± 0.0054
CNN	0.7405 ± 0.0300	0.9449 ± 0.0122	0.8317 ± 0.0313
HySEAtt-DDI (no CNN)	0.8577 ± 0.0079	0.9815 ± 0.0016	0.9368 ± 0.0053
HySEAtt-DDI (with CNN)	0.8562 ± 0.0052	0.9815 ± 0.0011	0.9367 ± 0.0037
HySEAtt-DDI (MLP→no CNN)	0.8698 ± 0.0043	0.9843 ± 0.0009	0.9458 ± 0.0029
HySEAtt-DDI (MLP→CNN)	0.8562 ± 0.0052	0.9815 ± 0.0011	0.9367 ± 0.0037

4.6 Macro-Averaged Results

Table 6 shows the results based on macro-average evaluation of the methods. XGBoost has the lowest performance among all methods. in contrast, random forest has the highest performance among ML models. HySEAtt-DDI especially with MLP and CNN has the highest performance among all methods. This outperforming is not only in AUCROC but in AUPR, which shows that for imbalanced datasets, the proposed method can learn the patterns effectively for correct prediction of DDIs. Additionally, CNN shows acceptable performance, it exhibits a relatively larger standard deviation and sensitivity in comparison to other methods.

Table 6: Macro-averaged evaluation of the methods (Best results in bold).

Method	F1	AUROC	AUPR
SVM	0.6265 ± 0.0037	0.9244 ± 0.0009	0.7559 ± 0.0030
RF	0.8033 ± 0.0038	0.9704 ± 0.0007	0.8695 ± 0.0024
XGBoost	0.4496 ± 0.0036	0.8784 ± 0.0011	0.6424 ± 0.0029
MLP	0.6959 ± 0.0042	0.9360 ± 0.0010	0.7786 ± 0.0030
CNN	0.7289 ± 0.0251	0.9538 ± 0.0077	0.8244 ± 0.0188
HySEAtt-DDI (no CNN)	0.8348 ± 0.0069	0.9778 ± 0.0008	0.8931 ± 0.0028
HySEAtt-DDI (with CNN)	0.8364 ± 0.0042	0.9800 ± 0.0006	0.8937 ± 0.0011
HySEAtt-DDI (MLP→no CNN)	0.8456 ± 0.0038	0.9800 ± 0.0005	0.9005 ± 0.0017
HySEAtt-DDI (MLP→CNN)	0.8360 ± 0.0047	0.9795 ± 0.0006	0.8987 ± 0.0024

4.7 Weighted-Averaged Results

Table 7 presents the weighted-averaged evaluation results of the methods. In this way, the ratio of each class has been considered in the evaluations and therefore, better reflects performance differences among the evaluated methods more clearly. Still, XGBoost has the lowest performance and random forest gains the highest performance among ML methods, even that MLP and CNN. HySEAtt-DDI (with CNN) and HySEAtt-DDI (with MLP) outperform all competing methods. while the former

has higher values in AUCROC and AUPR, the latter has the highest performance according to F1-score.

Table 7: Weighted-averaged evaluation of the methods (Best results in bold).

Method	F1-score	AUROC	AUPR
SVM	0.6986 ± 0.0027	0.9081 ± 0.0009	0.8181 ± 0.0019
RF	0.8471 ± 0.0032	0.9657 ± 0.0009	0.9118 ± 0.0019
XGBoost	0.5501 ± 0.0021	0.8505 ± 0.0009	0.7245 ± 0.0018
MLP	0.7068 ± 0.0052	0.9154 ± 0.0017	0.8284 ± 0.0029
CNN	0.7462 ± 0.0312	0.9443 ± 0.0123	0.8753 ± 0.0185
HySEAtt-DDI (no CNN)	0.8631 ± 0.0068	0.9735 ± 0.0010	0.9301 ± 0.0021
HySEAtt-DDI (with CNN)	0.8618 ± 0.0043	0.9774 ± 0.0006	0.9438 ± 0.0023
HySEAtt-DDI (MLP→no CNN)	0.8745 ± 0.0034	0.9763 ± 0.0007	0.9359 ± 0.0015
HySEAtt-DDI (MLP→CNN)	0.8615 ± 0.0045	0.9767 ± 0.0007	0.9352 ± 0.0016

The whole three sets of metric orchestrations shows the proposed HySEAtt-DDI improves the prediction performance and generalizability. To have a visual comparison of the method, Figure 6 contains visual comparison of methods regarding AUCROC and AUPR for micro-averaged metrics’ computation. As the figure shows the HySEAtt-DDI with MLP have the highest performance. An interesting observation is the strong performance of the Random Forest baseline in comparison to the other methods. This observation suggests that, it is necessary in many cases to check ML methods due to their high performance.

Table 8 compares the performance of HySEAtt-DDI and 3DGT [13] based on weighted evaluation metrics. The results indicate that the proposed HySEAtt-DDI model achieves superior performance across all considered metrics, including Precision, Recall, F1-score, AUROC, and AUPR. This consistent improvement demonstrates the effectiveness of the proposed architecture in capturing complex DDI patterns and producing more reliable predictions than the state-of-the-art 3DGT method.

Table 8: Performance comparison between HySEAtt-DDI and 3DGT [13] for small-molecule DDI prediction (weighted metrics; best results in bold).

Method	Precision	Recall	F1-score	AUROC	AUPR
3DGT[13]	0.8697 ± 0.0051	0.8445 ± 0.0090	0.8507 ± 0.0083	0.9702 ± 0.0020	0.9208 ± 0.0036
HySEAtt-DDI	0.8833 ± 0.0025	0.8561 ± 0.0046	0.8618 ± 0.0043	0.9774 ± 0.0006	0.9438 ± 0.0023

4.8 Ablation study

A comprehensive ablation study was conducted to evaluate the contribution of the individual components of HySEAtt-DDI. The results presented in Tables 6–7 illustrate

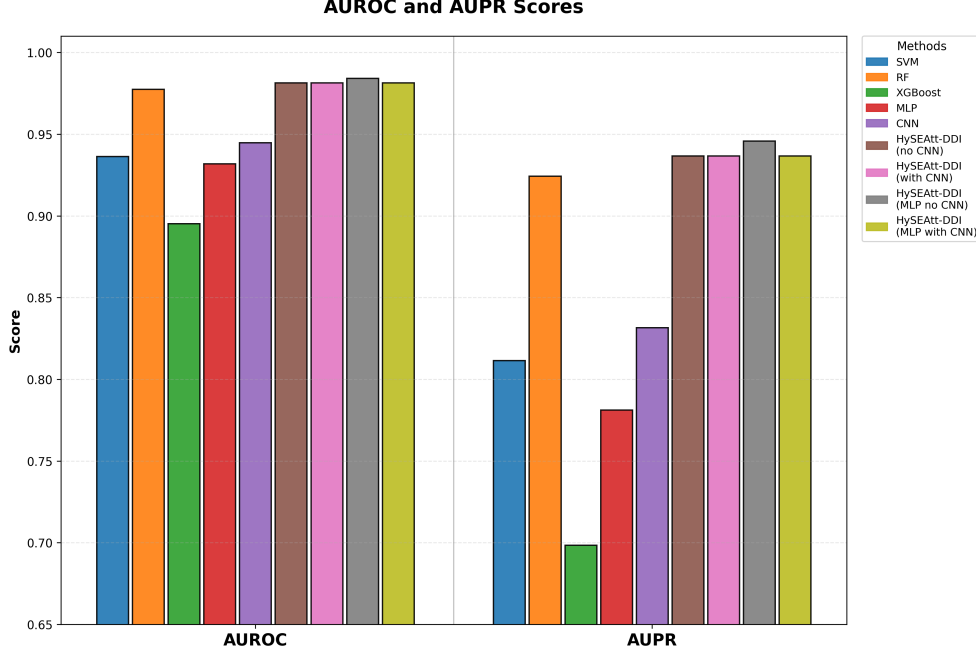


Fig. 6: Comparison of AUROC and AUPR scores across baseline ML models (SVM, RF, XGBoost, MLP, CNN) and the proposed HySEAtt-DDI variants.

the impact of each module on the overall performance of the proposed framework. Based on the weighted evaluation metrics reported in Table 7, the inclusion of the CNN component leads to a noticeable improvement in predictive performance, highlighting its effectiveness in extracting discriminative feature representations.

Furthermore, the contribution of each input feature type was investigated through an additional ablation analysis, the results of which are reported in Table 9. The best performance was achieved when all input representations were jointly utilized, including the features derived from Tanimoto similarity, SPI, and BioGPT embeddings. This finding demonstrates that each feature source provides complementary information and confirms the importance of integrating all three feature modalities to enhance the generalization capability and predictive accuracy of HySEAtt-DDI.

Table 9: Ablation study evaluating the contribution of structural similarity (Tanimoto) and SPI features to the proposed HySEAtt-DDI model (Best results in bold).

Method	Precision	Recall	F1-score	AUROC	AUPR
Tanimoto	0.8200 $\pm$ 0.0024	0.7676 $\pm$ 0.0046	0.7736 $\pm$ 0.0045	0.9485 $\pm$ 0.0014	0.8825 $\pm$ 0.0025
Tanimoto + SPI	0.8309 $\pm$ 0.0020	0.7881 $\pm$ 0.0057	0.7935 $\pm$ 0.0054	0.9551 $\pm$ 0.0008	0.8966 $\pm$ 0.0015
HySEAtt-DDI	0.8833 $\pm$ 0.0025	0.8561 $\pm$ 0.0046	0.8618 $\pm$ 0.0043	0.9774 $\pm$ 0.0006	0.9438 $\pm$ 0.0023

5 Conclusion

In this study, we have proposed HySEAtt-DDI, a multimodal deep learning framework for multiclass DDI prediction. To construct comprehensive drug representations, structural information from SMILES, biological information from drug-protein interactions, and semantic information from textual drug descriptions were integrated through similarity-based feature fusion. The resulting representations were processed by a hybrid architecture combining autoencoder-based latent feature learning, convolutional feature extraction, channel-wise attention, and multi-head self-attention mechanisms.

Experimental results demonstrated that HySEAtt-DDI consistently outperformed conventional ML and deep learning baselines across multiple evaluation metrics, including F1-score, AUROC, and AUPR. Notably, the strong performance achieved by the Random Forest baseline suggests that the proposed multimodal feature representation is highly informative and effectively captures relevant characteristics of drug pairs. Nevertheless, HySEAtt-DDI further improved predictive performance by leveraging convolutional and attention-based mechanisms to model complex nonlinear relationships and long-range dependencies within the fused feature space. These findings indicate that both the quality of the multimodal representation and the proposed deep learning architecture contribute to the overall effectiveness of the framework.

The results further revealed that textual drug descriptions provide complementary information beyond structural and biological features, highlighting the value of incorporating biomedical language models into DDI prediction systems. Moreover, the ablation studies confirmed that each feature modality contributes to the final prediction performance and that their integration yields superior results compared with any individual feature source.

Despite its effectiveness, the proposed framework is currently limited to small-molecule drugs and relies on a dataset derived from a single source. Future work will investigate the integration of additional biomedical data resources, graph-based drug representations, transfer learning strategies, and large-scale foundation models to improve the prediction of rare and complex drug interactions. The development of standardized benchmark datasets would also facilitate more comprehensive, reproducible, and fair evaluations across competing methods.

Overall, the findings demonstrate that multimodal integration of structural, biological, and semantic drug information provides an effective strategy for DDI prediction. By combining similarity-based feature fusion with hybrid CNN and attention mechanisms, HySEAtt-DDI captures complementary molecular, functional, and semantic characteristics while jointly modeling local and global feature dependencies. Consequently, the proposed framework offers a robust and scalable approach for learning informative drug representations and represents a promising direction for computational pharmacoinformatics and drug safety research.

Declarations

Clinical Trial Number

Clinical Trial Number: not applicable.

Consent to Publish

Consent to Publish declaration: not applicable.

Consent to Participate

Consent to Participate declaration: not applicable.

Ethics Declaration

Ethics declaration: not applicable.

Code and Data Availability

The source code used to construct the dataset, generate multimodal drug representations, train HySEAtt-DDI, and reproduce the experiments is available at <https://github.com/BioinformaticsIASBS/HySEAtt-DDI>.

Funding

The authors declared that no financial support was received for this work.

References

- [1] Li, L., Baker, J., Quirk, R., Deidun, D., Moran, M., Salem, A.A., Aryal, N., Van Dort, B.A., Zheng, W.Y., Hargreaves, A., *et al.*: Drug–drug interactions and actual harm to hospitalized patients: a multicentre study examining the prevalence pre-and post-electronic medication system implementation. *Drug Safety* **47**(6), 557 (2024)
- [2] Hashir, Q., Yar, M.A.E., Ullah, A., Kamal, S., Ullah, F., Abdullah, Z.H.: Computational approaches for drug–drug interaction prediction: a systematic review of data sources, modeling strategies, and evaluation frameworks. *Frontiers in Pharmacology* **17**, 1816394 (2026)
- [3] Behrooznia, S., Hooshmand, M.: A systematic review of drug-related interactions-utilizing deep learning and llms for prediction and mitigation. *ACS omega* **10**(50), 61048–61075 (2025)
- [4] Chianumba, E.C., Liu, H., Varde, A.S., Zhuang, Y.: Advancing drug-drug interaction prediction using multi-modal feature integration with graph neural networks. PhD thesis, Montclair State University (2025)

- [5] Weng, Y.-A., Deng, C.-Y., Pu, C.: Targeting continuity of care and polypharmacy to reduce drug–drug interaction. *Scientific reports* **10**(1), 21279 (2020)
- [6] Kamboj, S., Kumar, M.: Integration of artificial intelligence and machine learning technologies in antiviral drug discovery: strategies and challenges, 261–278 (2026)
- [7] Sun, D., Gao, W., Hu, H., Zhou, S.: Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B* **12**(7), 3049–3062 (2022)
- [8] Zhang, K., Yang, X., Wang, Y., Yu, Y., Huang, N., Li, G., Li, X., Wu, J.C., Yang, S.: Artificial intelligence in drug development. *Nature medicine* **31**(1), 45–59 (2025)
- [9] Liu, X., Yu, X., Dai, Q.: Graph-based deep learning for drug-drug interaction prediction: a systematic review. *Frontiers in Chemistry* **14**, 1806549 (2026)
- [10] Southey, M.W., Brunavs, M.: Introduction to small molecule drug discovery and preclinical development. *Frontiers in Drug Discovery* **3**, 1314077 (2023)
- [11] Ma, Y.-S., Xin, R., Yang, X.-L., Shi, Y., Zhang, D.-D., Wang, H.-M., Wang, P.-Y., Liu, J.-B., Chu, K.-J., Fu, D.: Paving the way for small-molecule drug discovery. *American journal of translational research* **13**(3), 853 (2021)
- [12] Li, K., Xiong, Y., Zhang, H., Cai, X., Du, B., Hu, W.: Small molecule drug discovery through deep learning: Progress, challenges, and opportunities. *arXiv preprint arXiv:2502.08975* (2025)
- [13] He, H., Chen, G., Yu-Chian Chen, C.: 3dgt-ddi: 3d graph and text based neural network for drug–drug interaction prediction. *Briefings in Bioinformatics* **23**(3), 134 (2022)
- [14] Nasiri, F., Hooshmand, M.: A graph-attention-based deep learning network for predicting biotech–small-molecule drug interactions. *bioRxiv*, 2025–05 (2025)
- [15] Nasiri, F., Hooshmand, M.: Machine and deep learning models for prediction of small molecule-biotech drug pair’s interactions. In: 2024 14th International Conference on Computer and Knowledge Engineering (ICCKE), pp. 055–058 (2024). IEEE
- [16] Segura-Bedmar, I., Martínez, P., Herrero-Zazo, M.: Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). In: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pp. 341–350 (2013)
- [17] Duvenaud, D.K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., Adams, R.P.: Convolutional networks on graphs for learning

- molecular fingerprints. *Advances in neural information processing systems* **28** (2015)
- [18] Liu, S., Tang, B., Chen, Q., Wang, X.: Drug-drug interaction extraction via convolutional neural networks. *Computational and mathematical methods in medicine* **2016**(1), 6918381 (2016)
 - [19] Zitnik, M., Agrawal, M., Leskovec, J.: Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* **34**(13), 457–466 (2018)
 - [20] Asada, M., Miwa, M., Sasaki, Y.: Enhancing drug-drug interaction extraction from texts by molecular structure information. *arXiv preprint arXiv:1805.05593* (2018)
 - [21] Asada, M., Miwa, M., Sasaki, Y.: Using drug descriptions and molecular structures for drug–drug interaction extraction from literature. *Bioinformatics* **37**(12), 1739–1746 (2021)
 - [22] Asada, M., Gunasekaran, N., Miwa, M., Sasaki, Y.: Representing a heterogeneous pharmaceutical knowledge-graph with textual information. *Frontiers in Research Metrics and Analytics* **6**, 670206 (2021)
 - [23] Koroteev, M.V.: Bert: a review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943* (2021)
 - [24] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. *NAACL-HLT* (2019)
 - [25] Yu, H., Li, K., Dong, W., Song, S., Gao, C., Shi, J.: Attention-based cross domain graph neural network for prediction of drug–drug interactions. *Briefings in Bioinformatics* **24**(4), 155 (2023)
 - [26] Shi, Y., He, M., Chen, J., Han, F., Cai, Y.: Subge-ddi: a new prediction model for drug-drug interaction established through biomedical texts and drug-pairs knowledge subgraph enhancement. *PLOS Computational Biology* **20**(4), 1011989 (2024)
 - [27] Li, W., Zhou, Y., Ma, W., Liang, W., Tang, X.: Mdg-ddi: multi-feature drug graph for drug-drug interaction prediction. *BMC bioinformatics* **26**(1), 268 (2025)
 - [28] Chen, X., Wang, Z., Miao, Z., Nie, B.: Research on drug-drug interaction prediction using capsule neural network based on self-attention mechanism. *BMC bioinformatics* **26**(1), 293 (2025)
 - [29] Wang, J., Du, H., Li, Y.: Ddi-attendnet: cross attention with structured graph learning for inter-drug connectivity analysis. *Frontiers in Pharmacology* **16**, 1680655 (2025)

- [30] Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., *et al.*: Drugbank 5.0: a major update to the drugbank database for 2018. *Nucleic acids research* **46**(D1), 1074–1082 (2018)
- [31] Greenhalgh, T.: How to read a paper: the medline database. *Bmj* **315**(7101), 180–183 (1997)
- [32] Willett, P., Barnard, J.M., Downs, G.M.: Chemical similarity searching. *Journal of chemical information and computer sciences* **38**(6), 983–996 (1998)
- [33] Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., Kanehisa, M.: Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* **24**(13), 232–240 (2008)
- [34] Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., Liu, T.-Y.: Biogpt: generative pre-trained transformer for biomedical text generation and mining. *Briefings in bioinformatics* **23**(6), 409 (2022)
- [35] Qaiser, S., Ali, R.: Text mining: use of tf-idf to examine the relevance of words to documents. *International journal of computer applications* **181**(1), 25–29 (2018)
- [36] Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013)
- [37] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H., Kang, J.: Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **36**(4), 1234–1240 (2020)
- [38] Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *Science* **313**(5786), 504–507 (2006) <https://doi.org/10.1126/science.1127647>
- [39] Zabihian, A., Asghari, J., Hooshmand, M., Gharaghani, S.: A comparative analysis of computational drug repurposing approaches: proposing a novel tensor-matrix-tensor factorization method. *Molecular Diversity* **28**(4), 2177–2196 (2024)