

Article

LLM-Enhanced Multimodal Framework for Drug–Drug Interaction Prediction

Song Im  and Younhee Ko * 

Department of Biomedical Engineering, Hankuk University of Foreign Studies, Yongin 17035, Gyeonggi-do, Republic of Korea; song.im@hufs.ac.kr

* Correspondence: younko@hufs.ac.kr

Abstract

Background: Drug–drug interactions (DDIs) involve pharmacokinetic or pharmacodynamic changes that occur when multiple drugs are co-administered, potentially leading to reduced efficacy or adverse effects. As polypharmacy becomes more prevalent, especially among patients with chronic diseases, scalable and accurate DDI prediction has become increasingly important. Although numerous computational approaches have been proposed to predict DDIs using various modalities such as chemical structure and biological networks, the intrinsic heterogeneity of these data complicates unified modeling; **Methods:** We address this challenge with a multimodal deep learning framework that integrates three complementary, heterogeneous modalities: (i) chemical structure, (ii) BioBERT-derived semantic embeddings (a domain-specific large language model, LLM), and (iii) pharmacological mechanisms through the CTET proteins. To incorporate indirect biological pathways within the PPI network, we apply a random walk with restart (RWR) algorithm. **Results:** Across features combinations, fusing structural feature with BioBERT embedding achieved the highest classification accuracy (0.9655), highlighting the value of readily available data and the capacity of domain-specific language models to encode pharmacological semantics from unstructured texts. **Conclusions:** BioBERT embeddings were particularly informative, capturing subtle pharmacological relationships between drugs and improving prediction of potential DDIs. Beyond predictive performance, the framework is readily applicable to real-world clinical workflows, providing rapid DDI references to support the polypharmacy decision-making.

Keywords: drug–drug interaction; DDI prediction; deep learning; CTET; large language model (LLM) embeddings



Academic Editors: Amirata Saei Dibavar, Emily G. Werth and Lin-Zhi Chen

Received: 21 August 2025

Revised: 20 September 2025

Accepted: 23 September 2025

Published: 26 September 2025

Citation: Im, S.; Ko, Y.

LLM-Enhanced Multimodal Framework for Drug–Drug Interaction Prediction. *Biomedicines* **2025**, *13*, 2355.

<https://doi.org/10.3390/biomedicines13102355>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The concurrent use of multiple medications—commonly referred to as polypharmacy—has become increasingly common, particularly among older adults and patients with multiple chronic conditions [1]. While polypharmacy is often necessary for comprehensive disease management, it also elevates the risk of drug–drug interactions (DDIs), which can lead to unexpected adverse effects or reduced therapeutic efficacy [2,3]. Identifying potential DDIs therefore remains a critical challenge in clinical pharmacology and drug development.

Traditional methods for detecting DDIs—including in vitro assays, animal studies, clinical trials, and post-marketing surveillance—continue to provide critical insights [4–7]. However, these approaches are time- and resource-intensive and infeasible for evaluating the vast number of possible drug combinations. This challenge is particularly pronounced

for newly approved or investigational drugs, which often lack sufficient interaction data to support empirical assessment [8].

To address these limitations, a range of computational approaches have been proposed for DDI prediction [9–13]. Early methods focused on structural similarity or molecular fingerprints, based on the assumption that chemically similar drugs exhibit similar interaction profiles [14,15]. Another model also incorporated biological features, such as protein–protein interaction (PPI) networks, to infer DDIs through indirect functional associations [16]. More recently, natural language processing (NLP) techniques have been applied to extract semantic features from biomedical studies, drug labels, and curated knowledge bases [17,18]. These approaches aim to capture multi-level semantics information, including drug-specific context, to improve the accuracy of DDI prediction.

Furthermore, recent deep learning frameworks have integrated multiple modalities—structural, biological, and semantic—into unified models. Several studies have also approached DDI prediction as a multi-class classification problem, targeting fine-grained interaction types (e.g., 65 or more categories) [15,19–24]. Despite these advances, many existing models still rely on fixed input formats or curated interaction graphs, which may hinder generalization to newly approved or less-characterized compounds.

To overcome these limitations, we present a multimodal deep learning framework that combines three complementary sources of information: (i) molecular structural similarity, (ii) protein similarity derived from PPI networks, and (iii) semantic embeddings derived from BioBERT, a biomedical language model pre-trained on large-scale corpora [25]. Large language models (LLMs) have transformed biomedical research by enabling the extraction of complex semantic knowledge from unstructured texts. In particular, they have shown promising applications in drug discovery tasks, including drug–drug interaction and drug–target prediction [18,26]. As a domain-specific large language model trained on PubMed abstracts and full-text articles, BioBERT extracts pharmacologically meaningful representations from drug descriptions that are well-suited for DDI prediction. By integrating the three modalities, the model captures both chemical structure and semantic/biological contexts for each drug.

The proposed model is implemented as a flexible MLP architecture that accommodates varying input combinations. While the full framework supports all three modalities, empirical evaluation demonstrated that the combination of structural similarity and semantic embeddings—without protein-based features—yielded the best classification accuracy. This modular design enables the model to perform robustly even when certain data types are unavailable. The findings indicate that structural and semantic features alone can achieve high accuracy, while reducing dependence on complex biological inputs.

Importantly, the required inputs are broadly accessible. The semantic embeddings are derived from general drug descriptions without the need for curated DDI annotations or complex graph structures and structural features can be computed from widely available molecular representations (e.g., SMILES). In contrast, protein similarity features derived from PPI networks require more extensive biological information, which is often incomplete or unavailable, particularly for investigational compounds. Our results showed that the combination of structural and semantic features (SSP + BioBERT) outperformed the full model that also included protein-based features (PSP). This suggests that the best-performing configuration depends on readily available data, improving both scalability and applicability across a wide range of drug compounds, including newly approved or less-characterized drugs.

2. Materials and Methods

2.1. Drug–Drug Interaction Data

We obtained drug information from DrugBank (version 5.1), including chemical structure such as SMILES strings, drug names, and textual descriptions. DDI type (class) labels were obtained from DrugBank via DeepDDI [13]. Of the 86 distinct interaction types, we excluded 7 types due to insufficient data, resulting in 79 DDI types used for training and evaluation (Supplementary Table S1). In total, 1705 drugs were included in the final dataset, yielding 178,849 drug pairs used for model training and evaluation.

2.2. BioBERT Embedding

To extract semantic features of each drug, we used BioBERT, a pre-trained biomedical language model that produces 768-dimensional embeddings [24]. Drug names and general descriptions were obtained from DrugBank. Among the 9377 drugs with available descriptions, 9294 drugs had input within BioBERT's maximum sequence length (e.g., 512-token) and were processed without truncation. If a drug's name was not included in its description, we prepended the name. Otherwise, the description was used as provided. Each input was encoded once, and hidden state of the [CLS] token was taken as drug's semantic representation, resulting in a 768-dimension BioBERT embedding per drug.

2.3. SSP: Structural Similarity Profile

SSP was calculated from molecular fingerprints derived from SMILES representations. To maximize coverage, we collected SMILES for 13,082 drugs from DrugBank [27], PubChem [28] and ChEMBL [29]. Each SMILES was converted into an extended-connectivity fingerprint (ECFP) using RDKit, yielding binary vectors that indicate the presence or absence of specific chemical substructures. We evaluated both ECFP4 and ECFP6 to examine the effect of fingerprint resolution. While ECFP6 generally resulted in higher average accuracy, the ECFP4-based SSP yielded the single best performance in one setting (Supplementary Table S4), suggesting that optimal fingerprint resolution may depend on the embedding combination. Overall, however, performance differences due to fingerprint resolution were minimal.

The pairwise structural similarity between drugs was computed using the Tanimoto coefficient. For each drug, we then constructed an SSP feature vector consisting of its Tanimoto similarities to all other drugs, resulting in an initial dimension of 13,082. To reduce dimensionality while preserving most of the variance, principal component analysis (PCA) was applied, and 200 components were retained for model training.

2.4. PSP: Protein Similarity Profile

To capture protein-level similarities between drugs, we constructed Protein Similarity Profile (PSP) features from drug–protein associations and a protein–protein interaction (PPI) network. The PPI network was compiled from STRING [30], a comprehensive resource of known protein–protein interactions that include direct (physical) and indirect (functional) associations derived from experimental, co-expression, computational inference, and curated databases.

We adopted the definition of CTET proteins from PRID [16] and collected a total of 3955 CTET proteins associated with 6747 drugs in DrugBank. These proteins serve as molecular mediators of pharmacokinetics and pharmacodynamics, and drug–drug interactions (DDIs) often arise from their functional overlap or perturbation of shared biological pathways.

Following the PSP procedure in PRID, we applied the random walk with restart (RWR) algorithm [31] to the PPI network to incorporate indirect functional connections

among CTET proteins associated with different drugs. Since most drugs are linked to only a few CTET proteins and often do not share direct overlap, RWR allows consideration of indirect functional relationships by propagating signals from each drug's seed proteins to their neighbors [32]. The resulting diffusion scores were then used to construct the PSP vector for each drug, representing the network-informed protein similarity profile. This method enables the model to reflect pharmacologically relevant similarities even when two drugs do not share CTET proteins directly by considering the topology of the underlying PPI network.

2.5. Model Architecture: Feature Processing and DDI Prediction

The overall framework of our proposed DDI prediction model is shown in Figure 1, and its step-by-step workflow is detailed in Supplementary Algorithm S1. We integrate three modalities to represent each drug: text-based embeddings, structural similarity features, and protein interaction-based features. For each drug pair, different combinations of SSP, PSP, and BioBERT embeddings were evaluated depending on the experimental setting. Using PCA, we reduced SSP to 200 dimensions and PSP to 300, while BioBERT embeddings were retained at their native 768 dimensions per drug.

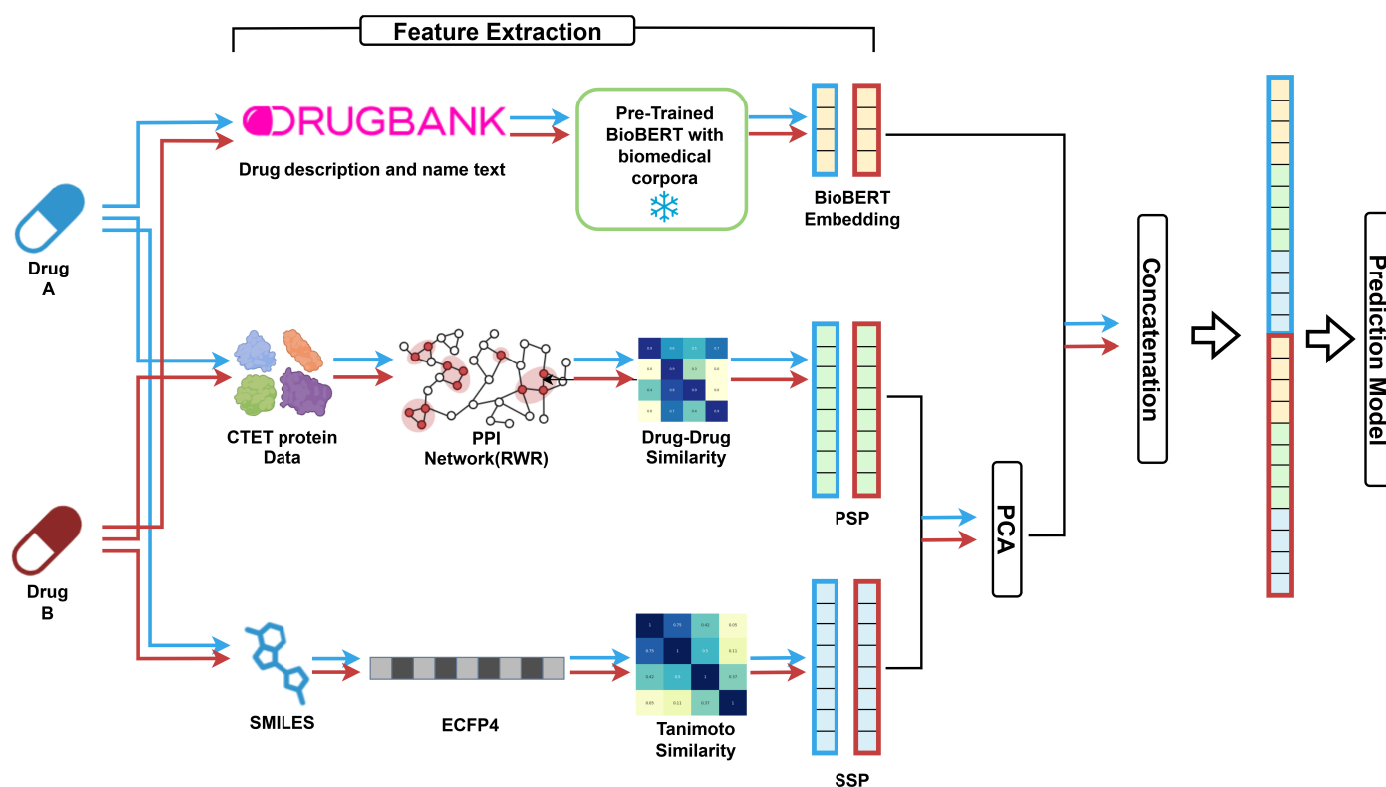


Figure 1. Overall framework of the proposed DDI prediction model. Features are extracted from three modalities: BioBERT embeddings (drug description and name), protein similarity profiles (PSP) computed via random walk with restart (RWR) on CTET (Carrier, Transporter, Enzyme, and Target) over PPI networks, and structure similarity profiles (SSP) based on SMILES-derived ECFP4 fingerprints with Tanimoto coefficient. In the heatmaps, more intense colors represent higher similarity between two drugs in terms of their targets and chemical structures, respectively. Extracted features are concatenated and fed into multi-layer perceptron classifiers for final prediction. The source code and data are available on GitHub (<https://github.com/SongIm/ddi-prediction>, accessed on 21 August 2025).

- PCA was applied to SSP and PSP to reduce dimensionality while preserving $\geq 95\%$ of the variance. In contrast, applying PCA to BioBERT embeddings generally de-

graded performance in preliminary experiments and we therefore used the original 768-dimensional vectors, suggesting that their semantic information was critical for prediction.

- After feature extraction, the selected representations were concatenated into a single feature vector (Figure 1). Before any dimensionality reduction, the input vector was passed through a projection layer—a learnable linear map with the same input and output size as the concatenated feature. The projection layer re-weights and linearly mixes feature dimensions to mitigate cross-modal scale/variance mismatches and map them into a better-aligned joint space before compression. The projection output was then fed to a hidden layer, followed by an output layer that mapped the hidden dimension to logits for the 79 DDI classes.
- The network architecture was determined using a two-stage hyperparameter optimization process. First, grid search was used to identify the number of hidden layers (1–5). We then employed Optuna [33] to tune the hidden dimension, dropout rate, and learning rate for each feature combination. The final architecture consisted of a projection layer for feature re-weighting/alignment, followed by a hidden layer for dimensionality reduction, and an output layer for classification. ReLU activation and dropout were applied after each linear transformation to improve generalization.
- After fixing the best hyperparameters, we trained from scratch for up to 300 epochs with Adam and cross-entropy loss. To ensure sufficient convergence, early stopping (patience = 20) was activated only after epoch 200, and the best validation-accuracy checkpoint was saved. Next, we resumed from this checkpoint and performed fine-tuning for up to 100 additional epochs with ReduceLROnPlateau (mode = max, factor = 0.1, patience = 5, min_lr = 1×10^{-6}); early stopping (patience = 20) was also applied in this stage. Test metrics were reported from the best validation checkpoint of the fine-tuning stage. The overall MLP configuration is summarized in Supplementary Table S2, and detailed hyperparameters for each feature combination are provided in Supplementary Table S3. The dataset was split into training, validation, and test sets with a ratio of 0.64:0.16:0.20, using stratified sampling to preserve the distribution of all 79 interaction classes.

3. Results and Discussion

3.1. Performance Evaluation

To evaluate the contribution of each modality, we conducted an ablation study across different feature combinations (Supplementary Tables S4 and S5). Models using single modality features, such as BioBERT only or SSP only, achieved moderate accuracy with 0.9584 and 0.9495, respectively. Incorporating additional modalities into the model consistently improved performance. For example, combining SSP and BioBERT yielded the highest accuracy of 0.9655, outperforming all single-modality models.

These results highlight the complementary nature of structural similarity and semantic embedding. While protein similarity profiles (PSP) alone showed relatively lower performance (0.9292), integrating them with BioBERT or SSP features (PSP + SSP or PSP + BioBERT) improved performance relative to each single-modality baseline, demonstrating the benefit of multi-modal feature fusion for DDI prediction; however, adding PSP on top of the best two-modality setting (SSP + BioBERT) did not yield additional gains. The performance gain from the projection layer was consistent across both shallow and deeper MLP architectures, indicating its general applicability.

Additionally, introducing a projection layer consistently improved performance across most feature combinations. For example, in the SSP + BioBERT setting, the model with a projection layer achieved an accuracy of 0.9655, compared to 0.9611 without it. This

suggests that the projection layer helps re-weight and align heterogeneous feature spaces before compression. Studies in multimodal and contrastive learning—such as CLIP [34] and SimCLR [35]—have demonstrated that applying a learnable projection head prior to task-specific loss functions improves alignment in shared latent spaces and enhances downstream performance.

We also observed that applying PCA to BioBERT embeddings generally led to slightly reduced performance compared to using the full 768-dimensional embeddings. As shown in Figure 2, this pattern was consistent across all evaluated feature combinations, including when BioBERT was combined with SSP or PSPs. In every case, PCA-reduced embeddings led to lower prediction accuracy. This suggests that BioBERT embeddings already provide a compact yet semantically dense representation, and further compression may compromise their predictivity.

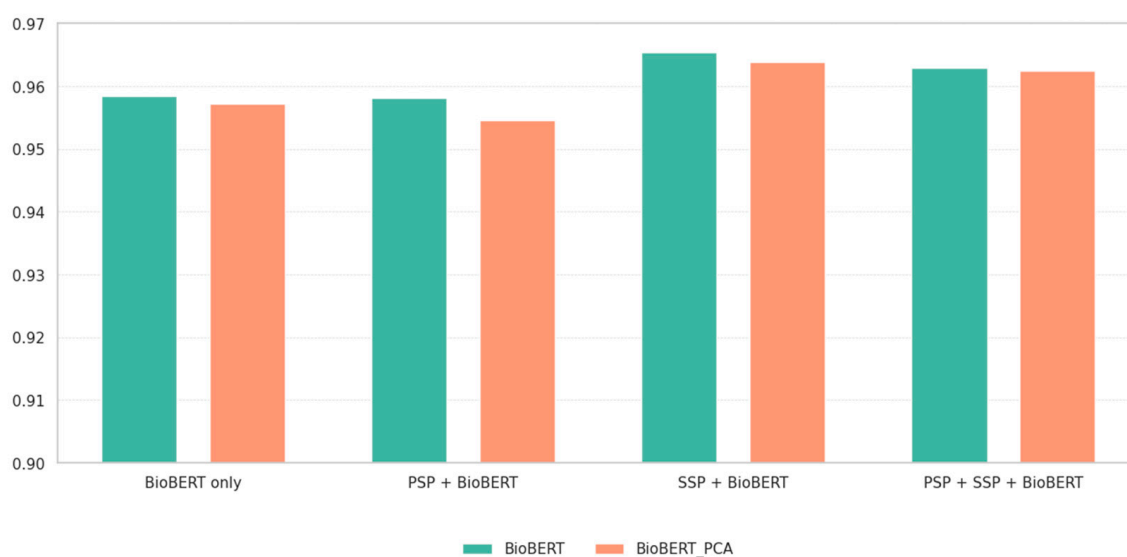


Figure 2. Accuracy comparison of BioBERT vs. BioBERT_PCA embeddings across feature combinations. PCA consistently reduced performance compared to using full BioBERT embeddings.

Beyond accuracy, our best-performing model (SSP + BioBERT) achieved a macro-average AUC of 0.990 and a macro-averaged AUPR of 0.9688 on the test set, confirming its strong discriminative capability across all interaction types.

To further assess whether the learned embeddings capture meaningful pharmacological information, we visualized the raw and learned features using t-SNE (Figure 3). While the raw features showed little separation across interaction classes, the learned representations exhibited clearer clustering patterns, indicating that the model encodes clinically relevant relationships between drugs.

3.2. Comparison with Existing Methods

We evaluated the class-wise prediction accuracy of our proposed model (SSP + BioBERT) and compared it with PRID, a state-of-the-art DDI prediction model. Figure 4 summarizes global performance: our model (SSP + BioBERT) shows higher accuracy and macro-F1 than PRID, with similar macro-precision and higher macro-recall. As shown in Figure 5, the radar chart illustrates the per-class performance across the 79 interaction types. Overall, our model demonstrated comparable or superior performance to PRID for most classes. Notably, our model outperformed PRID in 38 classes, while PRID performed better in 15 classes.

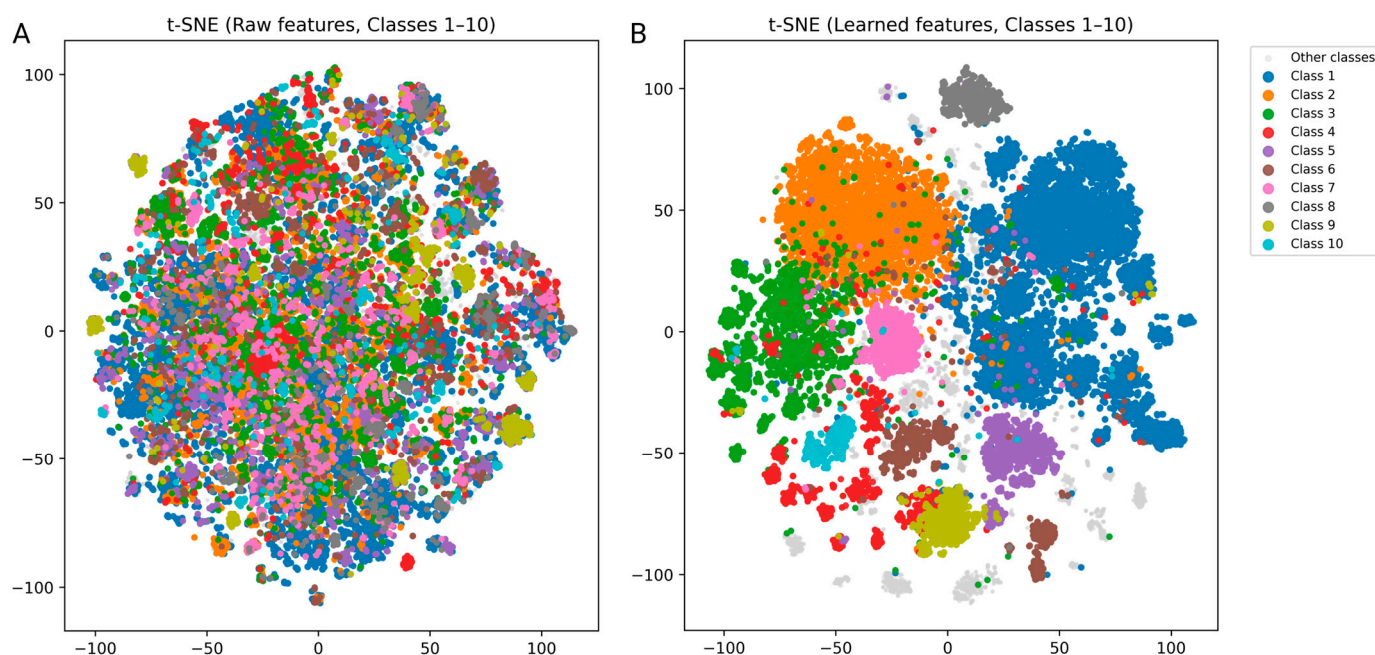


Figure 3. (A) Raw input features show no clear separation among interaction classes. (B) Learned representations form distinct clusters corresponding to different DDI classes, suggesting that the model captures pharmacologically meaningful structure.

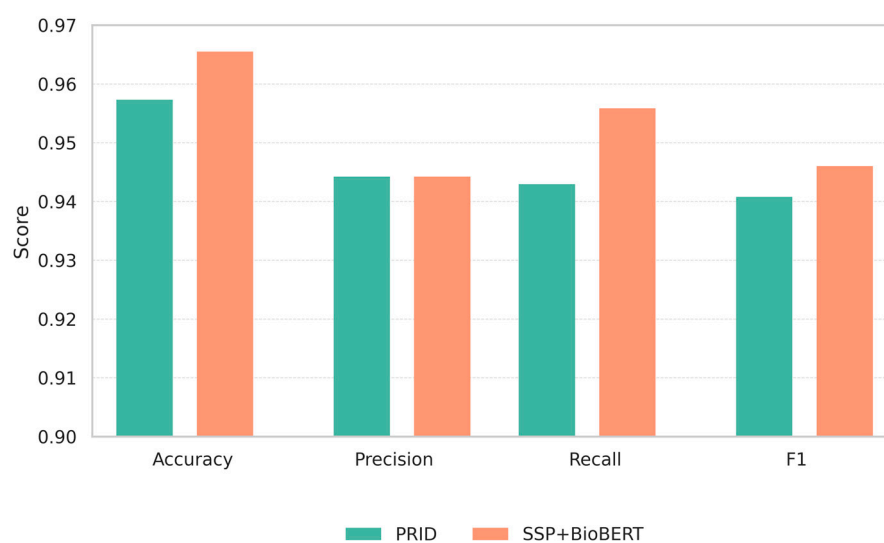


Figure 4. Comparison of overall accuracy and macro-averaged precision, recall and F1 between PRID and our proposed model on their respective test sets.

For example, in type 68 (“the risk of severity of hypotension can be increased when Drug A is combined with Drug B”), our model achieved substantially higher accuracy than PRID. Similar improvements were observed in type 63 (“the absorption of Drug B can be decreased when combined with Drug A”). These types tend to have relatively few positive samples in the training data, suggesting that our model generalizes better under data scarcity. We attribute this robustness to the effective fusion of structural similarity and semantic embedding, which provides more informative and complementary features than the combination of structural and protein similarity used in PRID.

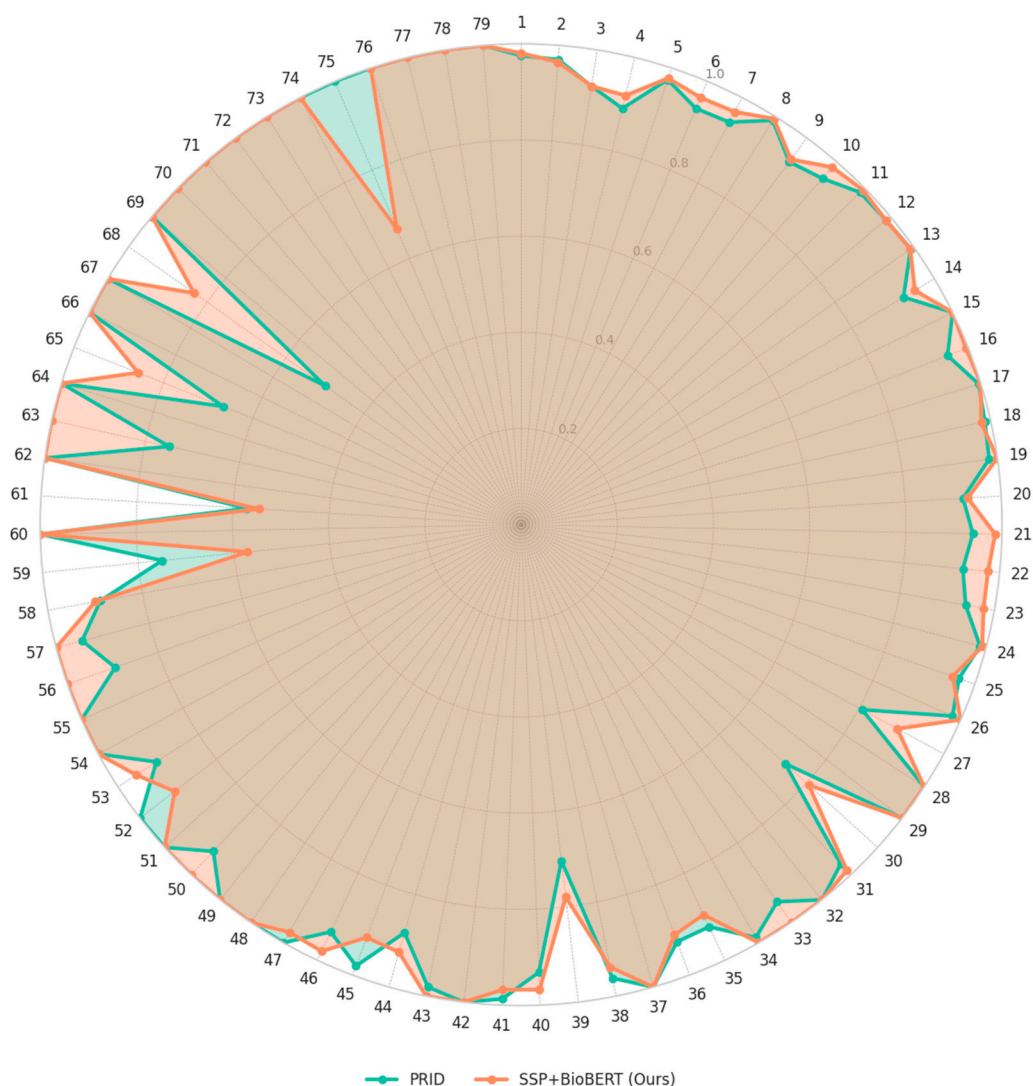


Figure 5. Class-wise comparison of prediction accuracy between PRID and our proposed SSP + BioBERT model. The radar plot shows that our model generally outperforms PRID across most classes.

3.3. Clinical Implication of False Positive Cases

We further investigated the misclassification patterns of our model to identify potential limitations and areas for improvement. Figure 6 shows the heatmap of the top 10 most misclassified interaction types, revealing that interaction types 1, 2, 3 and 4 accounted for the majority of errors. DDI type 3 showed the highest misclassification ratio, accounting for 21.72% of all misclassified samples, despite comprising only 11.52% of the test set (Table 1). DDI type 2 similarly exhibited a high misclassification ratio of 21.07%, with a test sample proportion of 17.29%. DDI type 1, while the most frequent DDI type in the dataset (32.38% of test samples), accounted for 12.32% of misclassification. These three DDI types are pharmacologically related. Inhibition of drug metabolism (DDI type 2) can lead to increased serum concentration (DDI type 3), which may result in heightened toxicity (DDI type 1). This mechanistic relationship likely contributes to model confusion among these classes.

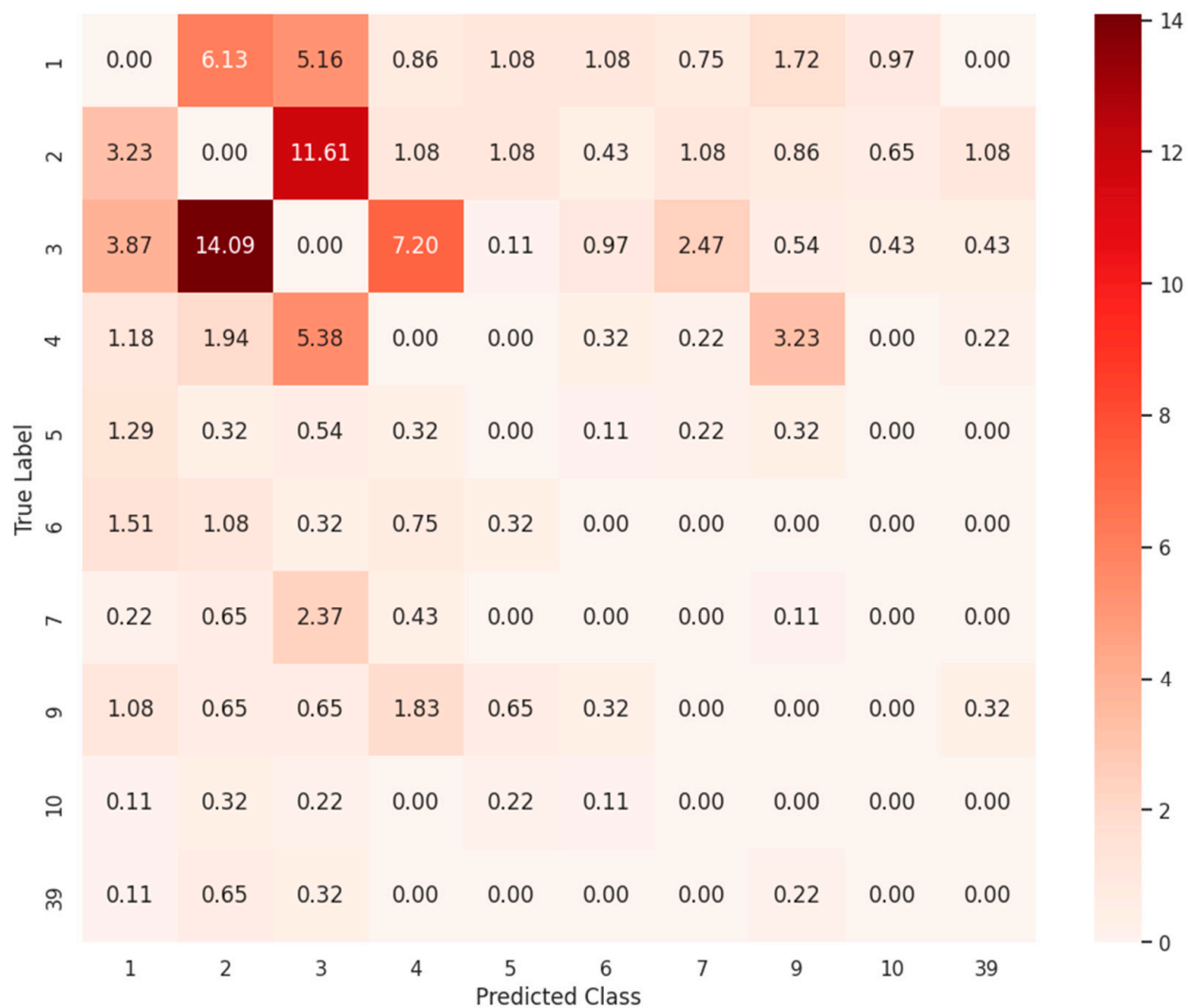


Figure 6. Misclassification heatmap for the top 10 most frequently misclassified DDI types. The values in the heatmap represent the normalized proportion of misclassified samples between the corresponding true and predicted interaction types, where higher values indicate greater confusion.

To determine whether some false positives might reflect plausible pharmacological reasoning rather than model error, we first examined cases labeled as Class 2 but predicted as Class 3. As summarized in Table 2, these pairs involve inhibition of CYP enzymes, particularly CYP3A4 affecting Drug B, which is expected to increase systemic exposure. This consistent pharmacokinetic (PK) pattern is well documented in the literature and drug labels and suggests that the model is capturing mechanistic nuances that are not explicitly reflected in the current DDI class labels. In some cases, the predicted Class 3 outcome may actually offer a more accurate pharmacological interpretation than the original Class 2 label.

We next examined cases labeled as Class 3 but predicted as Class 7. A representative example is mifepristone and granisetron shown in Figure 7. Mifepristone inhibits CYP3A4, thereby increasing the serum concentration of granisetron (Class 3). The higher exposure produces hERG block, which reduces IKr, prolongs APD, and consequently increases the risk of QTc prolongation (Class 7). Hence, this type of misclassification follows a plausible PK → PD cascade rather than a random error; the upstream PK effect (Class 3) can manifest as a downstream PD risk (Class 7).

Table 1. Top 10 most misclassified DDI types ranked by misclassification rate.

Class	Description	Test Samples	Test Sample Ratio (%)	Misclassified Count	Misclassified Ratio (%)
3	The serum concentration of Drug b can be increased when it is combined with Drug a.	4122	11.52	268	21.72
2	The metabolism of Drug b can be decreased when combined with Drug a.	6186	17.29	260	21.07
1	The risk or severity of adverse effects can be increased when Drug a is combined with Drug b.	11,583	32.38	152	12.32
4	The serum concentration of Drug b can be decreased when it is combined with Drug a.	1683	4.71	128	10.37
9	The metabolism of Drug b can be increased when combined with Drug a.	935	2.61	68	5.51
7	Drug a may increase the QTc-prolonging activities of Drug b.	1180	3.3	51	4.13
6	The therapeutic efficacy of Drug b can be decreased when used in combination with Drug a.	1488	4.16	37	3
5	Drug a may increase the hypotensive activities of Drug b.	1647	4.6	36	2.92
10	Drug a may increase the anticoagulant activities of Drug b.	631	1.76	20	1.62
39	The serum concentration of the active metabolites of Drug b can be reduced when Drug b is used in combination with Drug a resulting in a loss in efficacy.	65	0.18	19	1.54

Table 2. Representative misclassified cases where the predicted Class may better reflect the true DDI mechanism than the true labels.

True Class	Predicted Class	Drug A	Drug B	Comment
2	3	Ketoconazole	Delavirdine	CYP3A4 inhibition by ketoconazole increases delavirdine serum levels [36].
2	3	Voriconazole	Amiodarone	CYP3A4 inhibition by voriconazole increases amiodarone serum levels [37].
2	3	Cobicistat	Ketamine	CYP3A4 inhibition by cobicistat increases oral ketamine serum levels [38].
3	7	Mifepristone	Granisetron	CYP3A4 inhibition by mifepristone may increase the risk of QTc prolongation with granisetron [39,40].
3	7	Mifepristone	Ziprasidone	CYP3A4 inhibition by mifepristone may increase the risk of QTc prolongation with ziprasidone [40,41].
3	7	Mifepristone	Iloperidone	CYP3A4 inhibition by mifepristone may increase the risk of QTc prolongation with iloperidone [41,42].
9	4	Rifampicin	Lamotrigine	UGT induction by rifampicin decreases lamotrigine serum levels [43].
9	4	Rifapentine	Dabrafenib	CYP3A4/2C8 induction by rifapentine decreases dabrafenib serum levels [44,45].
41	5	Silodosin	Lofexidine	Coadministration of the α 1-blocker silodosin and the central α 2-agonist lofexidine may increase the risk of hypotension through additive effects [46,47].

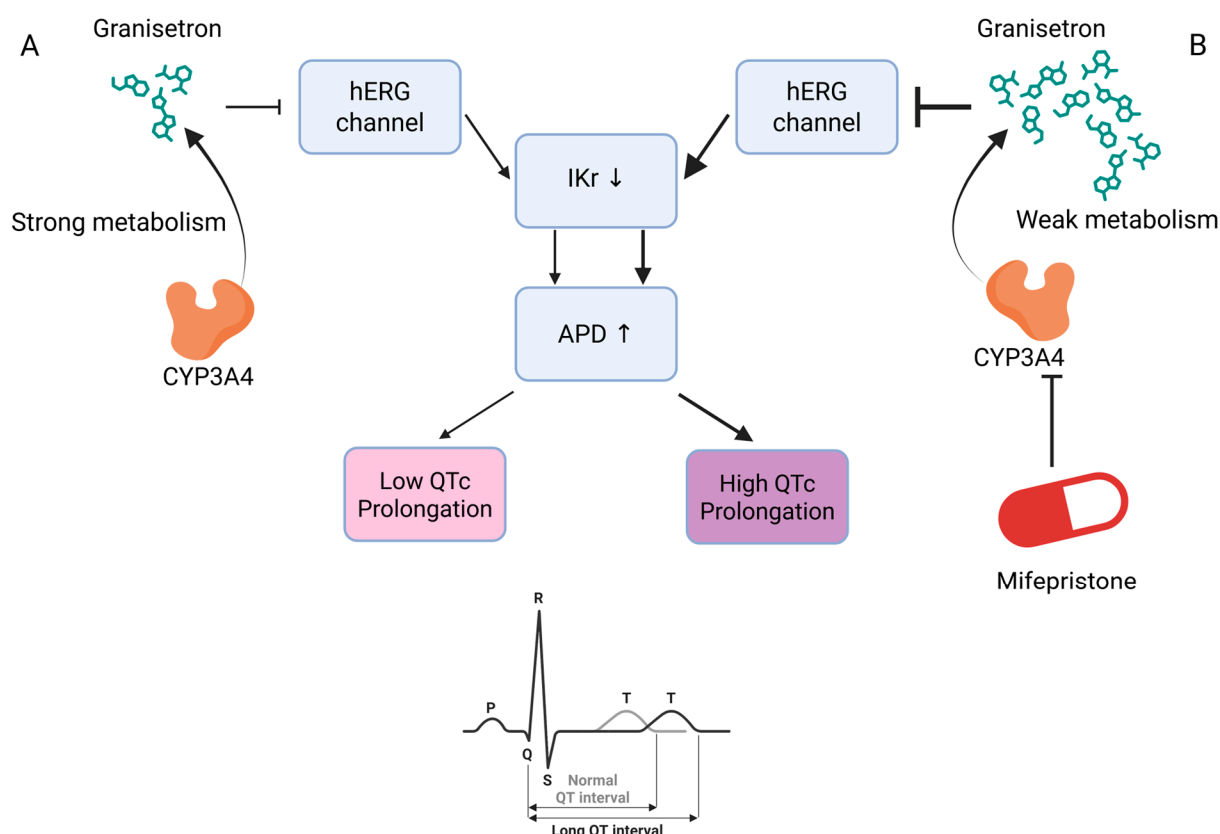


Figure 7. Granisetron–mifepristone pathway. (A) Baseline (CYP3A4 active): low granisetron exposure results in low QTc prolongation. (B) Granisetron + mifepristone (CYP3A4 inhibition): high granisetron exposure resulted in high QTc prolongation. $\dashv$ inhibition; $\rightarrow$ causal flow; IKr: rapid delayed rectifier K^+ current; APD: action potential duration. Created in BioRender. Im, S. (2025) <https://BioRender.com/i1bqtg8>.

Third, several pairs labeled Class 9 but predicted Class 4 show the mirror-image cause–effect of the first pattern. Enzyme induction (e.g., UGT or CYP3A4/2C8) increases clearance, which lowers exposure—a mechanistically coherent path from induction (Class 9) to reduced concentration (Class 4).

Lastly, some pairs labeled Class 41 (reduced vasoconstrictor activity) but predicted as Class 5 (increased hypotensive effect) are more naturally interpreted as additive pharmacodynamic effects when two BP-lowering agents are co-administered. In such cases, the model’s Class 5 prediction more closely matches the combined PD liability than the antagonistic framing implied by Class 41.

Collectively these findings indicate that a subset of false positives arise from mechanistically consistent cascade or additive effects that fall between or across our current label definitions. This highlights both the model’s sensitivity to pharmacological cues and an opportunity to refine label granularity.

4. Conclusions

We presented a multimodal deep learning framework that integrates structural, semantic, and protein-based features to predict drug–drug interaction (DDI) types. By leveraging BioBERT-derived embeddings alongside chemical and biological information, our model achieved high predictive performance across 79 interaction classes. These results highlight the potential of language models to encode latent pharmacological knowledge, while also revealing limitations in capturing fine-grained mechanistic distinctions.

Several limitations of the proposed model warrant discussion. First, it appears to rely more heavily on underlying mechanisms (e.g., enzyme inhibition) than on observable pharmacological outcomes. While this tendency may benefit mechanistic interpretation, it also suggests the model may not fully capture whether a drug increases or decreases an effect, as described in drug labels.

Second, the current model lacks explicit knowledge of pharmacological pathways or ontologies. As a result, it cannot reason effectively over complex or multi-step interactions that extend beyond surface-level textual patterns. This limitation likely affects its performance in cases where drug effects depend on indirect or context-specific mechanisms.

Third, while the semantic embeddings derived from BioBERT contributed meaningfully to model performance, they are primarily based on surface-level language patterns and pharmacological descriptions, rather than mechanistic interaction evidence. As a result, two drugs with similar indications or side effect profiles may yield similar embeddings—even if their roles in interactions differ substantially. This can lead to confusion between interaction classes, especially when mechanistic distinctions are underrepresented in the textual input. In such cases, semantic similarity may not align with interaction-specific differences, highlighting a limitation of language-based representations.

Future work can proceed in several directions to further improve DDI prediction. First, while BioBERT embeddings demonstrated substantial utility in capturing semantic features, these representations were derived from a general biomedical corpus without explicit DDI supervision. Fine-tuning BioBERT or other biomedical language models (e.g., PubMedBERT [26]) on DDI-specific corpora—such as curated interaction reports, clinical notes, or pharmacological texts—may yield more informative embeddings that capture finer-grained interaction semantics. Furthermore, generative LLMs could be leveraged not only for extracting but also for generating enriched textual representations, thereby providing deeper pharmacological semantics for DDI prediction. This approach could enhance the model's ability to distinguish mechanistic nuances and resolve ambiguities among similar DDI classes. Second, the interaction type labels in this study were based on those defined by DeepDDI [14], which reflect DrugBank at the time of its publication in 2018. Given the substantial updates to DrugBank since then, future work could benefit from re-extracting and reorganizing the interaction labels using the most recent version, potentially capturing newly discovered interactions and improving label completeness.

Overall, our findings provide a foundation for future developments through domain-specific fine-tuning and integration of updated interaction annotations.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines13102355/s1>, Algorithm S1: Workflow of the proposed DDI prediction framework. Table S1: Full list of Drug–Drug Interaction (DDI) types and their definitions; Table S2: Summary of MLP architecture and training setting; Table S3: Hyperparameter settings for the best performing feature combinations; Table S4: Accuracy by feature combination and MLP layer configuration (ECFP4); Table S5: Accuracy by feature combination and MLP layer configuration (ECFP6).

Author Contributions: Conceptualization and methodology, Y.K.; software, S.I.; validation and formal analysis, Y.K. and S.I.; investigation, Y.K. and S.I.; data curation, S.I.; writing—original draft preparation, Y.K. and S.I.; writing—review and editing, Y.K. and S.I.; project administration, Y.K.; funding acquisition, Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (2023R1A2C1007756) and the Hankuk University of Foreign Studies Research Fund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Code and data are available at GitHub: <https://github.com/SongIm/ddi-prediction> (accessed on 21 August 2025).

Acknowledgments: The authors would like to thank the anonymous reviewers for their insightful suggestions and careful review of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LLM	Large Language Model
CTET	Carrier, Transporter, Enzyme, and Target
DDI	Drug–drug Interaction
RWR	Random Walk with Restart
NLP	Natural Language Processing
PPI	Protein–protein Interaction
SMILES	Simplified Molecular Input Line Entry System
ECFP	Extended-Connectivity Fingerprint
PD	Pharmacodynamics
PK	Pharmacokinetics

References

- Nicholson, K.; Liu, W.; Fitzpatrick, D.; Hardacre, K.A.; Roberts, S.; Salerno, J.; Stranges, S.; Fortin, M.; Mangin, D. Prevalence of multimorbidity and polypharmacy among adults and older adults: A systematic review. *Lancet Healthy Longev.* **2024**, *5*, e287–e296. [\[CrossRef\]](#)
- Tsoukalas, N.; Brito-Dellán, N.; Font, C.; Butler, T.; Rojas-Hernandez, C.M.; Butler, T.; Escalante, C.; MASCC Hemostasis Study Group. Complexity and clinical significance of drug–drug interactions (DDIs) in oncology: Challenging issues in the care of patients regarding cancer-associated thrombosis (CAT). *Support. Care Cancer* **2022**, *30*, 8559–8573. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xu, S.; Song, Z.; Bai, J.; Wang, J. Prevalence and clinical significance of potential drug–drug interactions of antimicrobials in Intensive Care Unit patients: A retrospective study. *BMC Pharmacol. Toxicol.* **2025**, *26*, 104. [\[CrossRef\]](#)
- Tornio, A.; Filppula, A.M.; Niemi, M.; Backman, J.T. Clinical studies on drug–drug interactions involving metabolism and transport: Methodology, pitfalls, and interpretation. *Clin. Pharmacol. Ther.* **2019**, *105*, 1345–1361. [\[CrossRef\]](#)
- Cai, R.; Liu, M.; Hu, Y.; Melton, B.L.; Matheny, M.E.; Xu, H.; Duan, L.; Waitman, L.R. Identification of adverse drug–drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artif. Intell. Med.* **2017**, *76*, 7–15. [\[CrossRef\]](#)
- Lu, C.; Di, L. In vitro and in vivo methods to assess pharmacokinetic drug–drug interactions in drug discovery and development. *Biopharm. Drug Dispos.* **2020**, *41*, 3–31. [\[CrossRef\]](#)
- Tang, C.; Prueksaritanont, T. Use of in vivo animal models to assess pharmacokinetic drug–drug interactions. *Pharm. Res.* **2010**, *27*, 1772–1787. [\[CrossRef\]](#) [\[PubMed\]](#)
- Huang, S.M.; Temple, R.; Throckmorton, D.; Lesko, L. Drug interaction studies: Study design, data analysis, and implications for dosing and labeling. *Clin. Pharmacol. Ther.* **2007**, *81*, 298–304. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, Z.; Zhu, S.; Shao, B.; Zeng, X.; Wang, T.; Liu, T.-Y. DSN-DDI: An accurate and generalized framework for drug–drug interaction prediction by dual-view representation learning. *Brief. Bioinform.* **2023**, *24*, 1–12. [\[CrossRef\]](#)
- Yan, C.; Duan, G.; Zhang, Y.; Wu, F.-X.; Pan, Y.; Wang, J. Predicting drug–drug interactions based on integrated similarity and semi-supervised learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *19*, 168–179. [\[CrossRef\]](#)
- Yan, X.-Y.; Yin, P.-W.; Wu, X.-M.; Han, J.-X. Prediction of the drug–drug interaction types with the unified embedding features from drug similarity networks. *Front. Pharmacol.* **2021**, *12*, 794205. [\[CrossRef\]](#)
- Zhu, J.; Che, C.; Jiang, H.; Xu, J.; Yin, J.; Zhong, Z. SSF-DDI: A deep learning method utilizing drug sequence and substructure features for drug–drug interaction prediction. *BMC Bioinform.* **2024**, *25*, 39. [\[CrossRef\]](#)
- Rifaioğlu, A.S.; Atas, H.; Martin, M.J.; Cetin-Atalay, R.; Atalay, V.; Dogan, T. Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools and databases. *Brief. Bioinform.* **2019**, *20*, 1878–1912. [\[CrossRef\]](#)
- Ryu, J.Y.; Kim, H.U.; Lee, S.Y. Deep learning improves prediction of drug–drug and drug–food interactions. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4304–E4311. [\[CrossRef\]](#)

15. Wang, H.; Lian, D.; Zhang, Y.; Qin, L.; Lin, X. Gognn: Graph of graphs neural network for predicting structured entity interactions. *arXiv* **2020**, arXiv:2005.05537. [[CrossRef](#)]
16. Seo, J.; Jung, H.; Ko, Y. PRID: Prediction model using RWR for interactions between drugs. *Pharmaceutics* **2023**, *15*, 2469. [[CrossRef](#)] [[PubMed](#)]
17. Yang, J.; Ding, Y.; Long, S.; Poon, J.; Han, S.C. Ddi-mug: Multi-aspect graphs for drug–drug interaction extraction. *Front. Digit. Health* **2023**, *5*, 1154133. [[CrossRef](#)]
18. Qi, H.; Li, X.; Zhang, C.; Zhao, T. Improving drug–drug interaction prediction via in-context learning and judging with large language models. *Front. Pharmacol.* **2025**, *16*, 1589788. [[CrossRef](#)]
19. Deng, Y.; Xu, X.; Qiu, Y.; Xia, J.; Zhang, W.; Liu, S. A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics* **2020**, *36*, 4316–4322. [[CrossRef](#)]
20. Jin, Q.; Xie, J.; Huang, D.; Zhao, C.; He, H. Msff-ma-ddi: Multi-source feature fusion with multiple attention blocks for predicting drug–drug interaction events. *Comput. Biol. Chem.* **2024**, *108*, 108001. [[CrossRef](#)] [[PubMed](#)]
21. Lin, S.; Wang, Y.; Zhang, L.; Chu, Y.; Liu, Y.; Fang, Y.; Jiang, M.; Wang, Q.; Zhao, B.; Xiong, Y. MDF-SA-DDI: Predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism. *Brief. Bioinform.* **2022**, *23*, bbab421. [[CrossRef](#)] [[PubMed](#)]
22. Yang, Z.; Tong, K.; Jin, S.; Wang, S.; Yang, C.; Jiang, F. CNN-Siam: Multimodal siamese CNN-based deep learning approach for drug–drug interaction prediction. *BMC Bioinform.* **2023**, *24*, 110. [[CrossRef](#)] [[PubMed](#)]
23. Yu, Y.; Huang, K.; Zhang, C.; Glass, L.M.; Sun, J.; Xiao, C. SumGNN: Multi-typed drug interaction prediction via efficient knowledge graph summarization. *Bioinformatics* **2021**, *37*, 2988–2995. [[CrossRef](#)]
24. Zhang, C.; Lu, Y.; Zang, T. CNN-DDI: A learning-based method for predicting drug–drug interactions using convolution neural networks. *BMC Bioinform.* **2022**, *23*, 88. [[CrossRef](#)]
25. 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* **2020**, *36*, 1234–1240. [[CrossRef](#)]
26. Gu, Y.; Tinn, R.; Cheng, H.; Lucas, M.; Usuyama, N.; Liu, X.; Naumann, T.; Gao, J.; Poon, H. Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans. Comput. Healthc. (HEALTH)* **2021**, *3*, 1–23. [[CrossRef](#)]
27. 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. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [[CrossRef](#)]
28. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B. PubChem 2023 update. *Nucleic Acids Res.* **2023**, *51*, D1373–D1380. [[CrossRef](#)]
29. Zdrazil, B.; Felix, E.; Hunter, F.; Manners, E.J.; Blackshaw, J.; Corbett, S.; De Veij, M.; Ioannidis, H.; Lopez, D.M.; Mosquera, J.F. The ChEMBL Database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **2024**, *52*, D1180–D1192. [[CrossRef](#)]
30. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P. STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [[CrossRef](#)] [[PubMed](#)]
31. Valdeolivas, A.; Tichit, L.; Navarro, C.; Perrin, S.; Odelin, G.; Levy, N.; Cau, P.; Remy, E.; Baudot, A. Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics* **2019**, *35*, 497–505. [[CrossRef](#)]
32. Ko, Y.; Cho, M.; Lee, J.-S.; Kim, J. Identification of disease comorbidity through hidden molecular mechanisms. *Sci. Rep.* **2016**, *6*, 39433. [[CrossRef](#)]
33. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631.
34. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J. Learning transferable visual models from natural language supervision. In Proceedings of the International Conference on Machine Learning, Online, 18–24 July 2021; pp. 8748–8763.
35. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning, Online, 13–18 July 2020; pp. 1597–1607.
36. Vadlapatla, R.K.; Patel, M.; Paturi, D.K.; Pal, D.; Mitra, A.K. Clinically relevant drug–drug interactions between antiretrovirals and antifungals. *Expert Opin. Drug Metab. Toxicol.* **2014**, *10*, 561–580. [[CrossRef](#)] [[PubMed](#)]
37. Mourad, A.; Stiber, J.A.; Perfect, J.R.; Johnson, M.D. Real-world implications of QT prolongation in patients receiving voriconazole and amiodarone. *J. Antimicrob. Chemother.* **2019**, *74*, 228–233. [[CrossRef](#)]
38. Amariles, P.; Rivera-Cadavid, M.; Ceballos, M. Clinical Relevance of Drug Interactions in People Living with Human Immunodeficiency Virus on Antiretroviral Therapy—Update 2022: Systematic Review. *Pharmaceutics* **2023**, *15*, 2488. [[CrossRef](#)] [[PubMed](#)]

39. Bloomer, J.; Baldwin, S.; Smith, G.; Ayrton, A.; Clarke, S.; Chenery, R. Characterisation of the cytochrome P450 enzymes involved in the in vitro metabolism of granisetron. *Br. J. Clin. Pharmacol.* **1994**, *38*, 557–566. [[CrossRef](#)]
40. He, K.; Woolf, T.F.; Hollenberg, P.F. Mechanism-based inactivation of cytochrome P-450-3A4 by mifepristone (RU486). *J. Pharmacol. Exp. Ther.* **1999**, *288*, 791–797. [[CrossRef](#)]
41. Beedham, C.; Miceli, J.J.; Obach, R.S. Ziprasidone metabolism, aldehyde oxidase, and clinical implications. *J. Clin. Psychopharmacol.* **2003**, *23*, 229–232. [[CrossRef](#)] [[PubMed](#)]
42. Potkin, S.G.; Preskorn, S.; Hochfeld, M.; Meng, X. A thorough QTc study of 3 doses of iloperidone including metabolic inhibition via CYP2D6 and/or CYP3A4 and a comparison to quetiapine and ziprasidone. *J. Clin. Psychopharmacol.* **2013**, *33*, 3–10. [[CrossRef](#)] [[PubMed](#)]
43. Ebert, U.; Thong, N.; Oertel, R.; Kirch, W. Effects of rifampicin and cimetidine on pharmacokinetics and pharmacodynamics of lamotrigine in healthy subjects. *Eur. J. Clin. Pharmacol.* **2000**, *56*, 299–304. [[CrossRef](#)]
44. Zheng, C.; Hu, X.; Zhao, L.; Hu, M.; Gao, F. Clinical and pharmacological hallmarks of rifapentine's use in diabetes patients with active and latent tuberculosis: Do we know enough? *Drug Des. Dev. Ther.* **2017**, *11*, 2957–2968. [[CrossRef](#)] [[PubMed](#)]
45. Lawrence, S.K.; Nguyen, D.; Richards-Peterson, L.; Skordos, K.W. The metabolic drug-drug interaction profile of Dabrafenib: In vitro investigations and quantitative extrapolation of the P450-mediated DDI risk. *Drug Metab. Dispos.* **2014**, *42*, 1180–1190. [[CrossRef](#)]
46. Rivasi, G.; Rafanelli, M.; Mossello, E.; Brignole, M.; Ungar, A. Drug-related orthostatic hypotension: Beyond anti-hypertensive medications. *Drugs Aging* **2020**, *37*, 725–738. [[CrossRef](#)] [[PubMed](#)]
47. Gish, E.C.; Miller, J.L.; Honey, B.L.; Johnson, P.N. Lofexidine, an α_2 -receptor agonist for opioid detoxification. *Ann. Pharmacother.* **2010**, *44*, 343–351. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.