Received: 31 August 2025 | Revised: 9 January 2026 | Accepted: 26 March 2026 | Published online: 10 April 2026

REVIEW

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–17
DOI: 10.47852/bonviewAIA62027539

Prediction of Drug–Drug Interactions Based on Artificial Intelligence: A Systematic Literature Review

Faisal Asad ur Rehman[1,*], Abdulkarim Kanaan Jebna[1], Touqeer Ahmad[2], Arif Ur Rahman[3] and Fasee Ullah[4]

[1] *Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, Malaysia*

[2] *Department of Mathematics, University of Oslo, Norway*

[3] *Department of Computer Science, Bahria University, Pakistan*

[4] *Department of Computing, Universiti Teknologi PETRONAS, Malaysia*

Abstract: The comprehensive knowledge about the simultaneous use of multiple drugs to treat a disease is essential for the medical community to determine the best decisions for patient health. The use of various drugs at the same time to treat a disease can result in drug–drug interaction, raising the possibility of serious side effects. This study conducted a systematic literature review that describes the declarative information about drug–drug interactions, including the research papers from 2019 to 2025. The study focused on significant areas that can enhance modern research in drug–drug interactions, which were not included in previous studies. It is composed of artificial intelligence techniques, particularly those based on machine learning and deep learning for predicting drug–drug interactions. The PRISMA-based flow chart concept is used in the literature review stage. After a thorough review of the research papers, 33 studies were chosen. This work presents four research questions that were addressed and answered by the obtained results. The study found that the drug–drug interaction trend increased starting from 2021. It also found that deep learning models and their hybrid frameworks are the most commonly used. It was also observed that most studies used DrugBank and TWOSIDE data repositories. The findings also reveal that the F1-score is the most frequent evaluation measure. It found that there is a bridge to validate the model's performance on a real-time clinical dataset, thereby assessing its true power. The research findings highlight significant study trends and knowledge gaps in the medical industry.

Keywords: artificial intelligence, deep learning, drug–drug interactions, machine learning

1. Introduction

The advancement of artificial intelligence (AI) approaches to predict drug–drug interactions (DDIs) has recently garnered interest in bioinformatics. The DDIs can develop when multiple drugs are taken together [1]. Such interactions might increase or decrease the effectiveness of medications, result in undesirable drug responses that could be life-disastrous in certain situations [2], and cause drug withdrawal from pharmacies [3]. According to the US Centers for Disease Control and Prevention, over 10% of people use five or more medicines concurrently. Even more concerning, 20% of senior citizens use more than 10 different drugs, which considerably increases the chance of harmful drug reactions [4]. The chance of DDIs increases in proportion to the expansion of the number of new permitted drugs [5]. Therefore, predicting DDIs in the early stages of drug development is both important and challenging. Figure 1 shows the DDIs network.

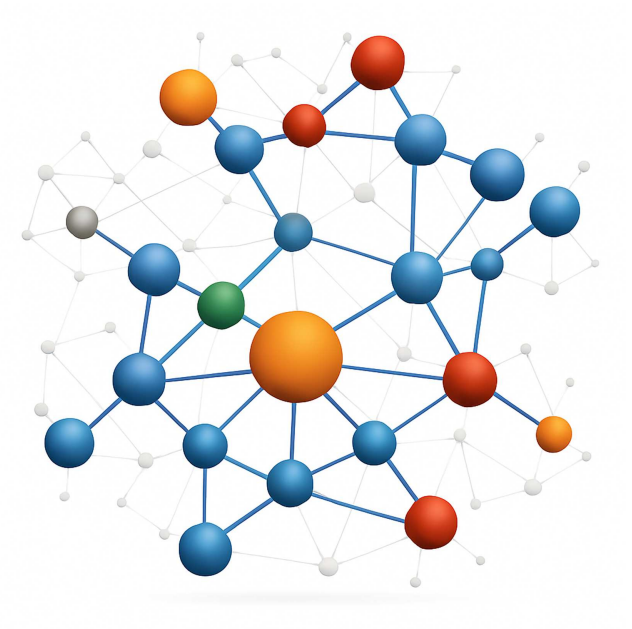
There are various traditional approaches used to identify DDIs, with in vitro and in vivo tests being the most common. However, they are sometimes impossible due to the lab's technical or financial constraints in identifying DDIs [6]. These traditional approaches to DDI prediction rely on manual expert knowledge and experimental studies to protect patient safety and improve treatment outcomes. Therefore, it becomes necessary to shift toward AI techniques to address issues with DDI identification. Two types of AI-based methods are now used to determine DDIs: (1) DDIs can be extracted from written sources, computerized medical records, and unprompted reports; and (2) DDIs can be predicted using known DDIs.

1.1. Extraction of DDIs

Unstructured research encompasses many DDIs; however, with the increasing volume of biomedical literature, it is currently

***Corresponding author:** Faisal Asad ur Rehman, Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, Malaysia. Email: faisal.asad.rehman92@1utar.my

Figure 1
Drug–drug interaction network



quite challenging to extract relevant data from the vast literature and synchronize it within drug databases [7, 8].

Two types of methods are used to extract DDIs: (1) pattern-based techniques and (2) characteristics-based machine learning (ML). The existing pattern-based techniques are outdated because they rely on manual categorization of DDIs, which requires domain expertise. Using the annotated dataset has increased the popularity of the ML approach to extracting DDIs [9]. Furthermore, the extraction of DDIs from textual data does not give an early indication or discover unknown DDIs, but ML can predict them successfully in advance [10–12].

1.2. Prediction of DDIs

To anticipate DDI and detect potential dangers related to the concurrent use of various drugs, AI approaches are used [13]. Due to the growing complexity of this research subject, there has been a significant increase in interest in drug regimens in recent years, highlighting the need to optimize patient safety and treatment outcomes [14]. ML approaches in drug prediction utilize algorithms that learn from historical data on drug interactions and patient profiles to develop predictive models. These models can analyze large datasets containing information about drug properties, molecular structures, pharmacokinetics, and clinical data to identify patterns and relationships that contribute to DDI [15].

The ML systems analyze and learn patterns from numerous data sources to anticipate DDIs. There is a need to first distinguish between traditional and nontraditional ML's methods for bioinformatics/network-based prediction problems. The two categories of traditional ML methods are classification-based and similarity-based. Nontraditional ML methods falls into four major types:

1) Network propagation technique. According to various techniques for network processing, the network propagation technique is separated into graph embedding and link prediction. The link prediction approach utilizes biological items as nodes and their complex relationships as edges to predict unknown connections and identify false or insufficient interactions. Embedding a graph involves transforming the high-dimensional structure of the known graph into a lower-dimensional space while preserving the information related to the graph. The network propagation technique is based on the guilt-by-association rule, where closely connected nodes share similar functions or properties within a network [16].
2) Factorization of a matrix. The matrix of known drug interactions is decomposed into N matrices of low dimension using various decomposition techniques. These matrices are then combined to create the matrix that predicts DDIs [17].
3) Ensemble approach. The ensemble approach integrates several methodologies to estimate drug interactions, yielding more accurate results [18].
4) Literature-based approach. This method uses natural language processing to get drug interactions on unstructured data, and unknown DDIs are predicted using extracted data [19].

Different ML models have frequently been used in DDI prediction, each with strengths and characteristics. Decision trees (DTs) are popular ML models used in DDI prediction [20]. They consist of a hierarchical structure where decisions are made based on the features of the data. The DTs can be well trained on the molecular features, different drug properties, and other relevant factors to predict the DDIs by recursively splitting the data based on specific criteria.

Another effective ML approach to predict DDIs is the support vector machine (SVM) [21, 22]. It separates the different data classes by identifying the optimal hyperplane. It maps the different drug features to high-dimensional spaces. SVM identifies the boundaries between non-interacting and interacting drug pairs, thereby increasing the margin between the classes and accurately predicting the DDIs.

Random forests (RFs) are one of the ensemble algorithms that integrate multiple DTs to enhance the predictive performance [23]. It uses a collection of the DTs, each trained on a different subset of data, and averages their results to obtain reliable results. Therefore, it reduces overfitting and captures complex interactions to achieve robust prediction [14]. Naïve Bayes (NB) is a widely used probabilistic classifier. NB assumes independence between features, but in reality, this assumption cannot hold for all drug interactions [24]. Despite this limitation, it can provide useful insights by calculating the likelihood of drug interactions using the conditional probabilities and the prior knowledge. However, NB can be useful in such scenarios where a limited dataset is available.

Another ensemble algorithm is gradient boosting. Basically, it combines multiple weak learners in an ensemble way to create a strong predictive model [25]. It iteratively built DTs, each of which attempted to address the errors in the previous tree [26, 27]. It improves the prediction accuracy by focusing on the misclassified instances. So, the healthcare professionals and researchers can effectively predict DDIs using these ML models. They facilitate better decision-making in reducing the risk of negative reactions and in prescribing the correct medications.

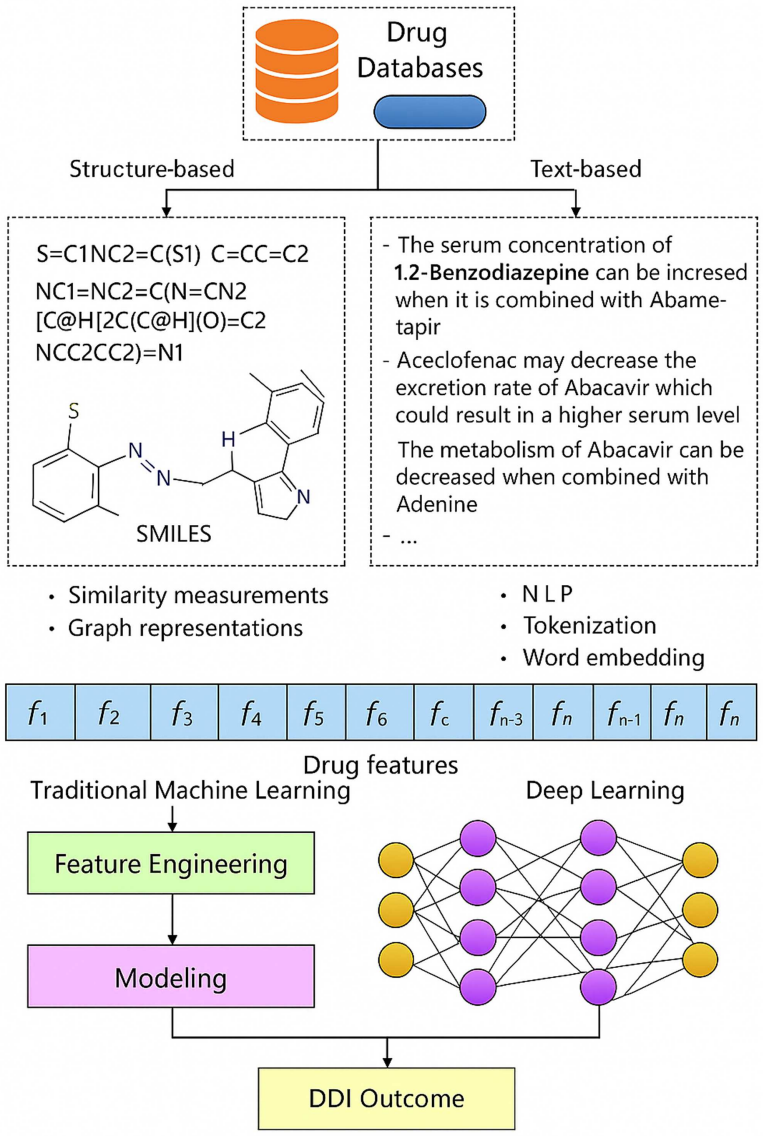
Different deep learning (DL) approaches are also used to predict the DDIs. They use multiple-layer artificial neural networks to extract information and automatically handle complex DDI patterns. In DL, deep neural networks (DNNs), convolutional neural networks (CNNs), graph neural networks (GNNs), deep belief networks, recurrent neural networks (RNNs), and autoencoders are commonly used to develop hierarchical representations that capture the dependencies between medications

and complex interactions [28–33]. These DL models can handle the scalability requirements and also enable the comprehensive analysis of DDIs at scale by providing powerful tools for processing large datasets comprising diverse drug properties, patient characteristics, and molecular structures [34]. Moreover, these techniques have the advantage of automatically performing feature engineering by extracting useful features from the raw data. However, it is useful for DDI prediction, as it identifies the relevant features from the complex nature of drug interactions.

Through pattern recognition and network analysis, these models can identify recurring patterns, associations, and structural characteristics within DDI networks. It can help identify potential negative or synergistic effects of drug combinations and provide valuable insights into the underlying processes of DDI [35, 36]. In summary, ML and DL techniques have the potential to revolutionize DDI prediction by offering scalability, feature extraction capabilities, adaptability, and the ability to uncover hidden patterns and relationships within DDI networks [37]. The network of the ML and DL models for predicting DDI is shown in Figure 2 [38].

By leveraging these advanced AI approaches, researchers and healthcare professionals can make more informed decisions about drug prescriptions, optimize treatment plans, and mitigate the risks of adverse drug interactions. Integrating AI-based drug-prediction algorithms into the medical domain holds significant promise for improving patient care and treatment outcomes [18]. Leveraging these algorithms can improve the precision of DDI predictions using ML and DL techniques. With the capacity to analyze vast amounts of data and identify intricate patterns, these algorithms provide healthcare professionals with valuable insights into potential interactions that may have been previously overlooked [23, 39]. Improved accuracy and understanding of DDI can significantly aid in making more informed decisions when prescribing medications, thereby reducing the risk of adverse reactions and ultimately enhancing patient safety [25, 40]. Additionally, these algorithms can detect interactions early, enabling proactive measures to mitigate potential risks before they

Figure 2
Workflow of the DL and ML models for predicting DDI



manifest in clinical practice. Drug prediction in the medical field can be revolutionized by AI techniques, leading to more personalized and effective treatment strategies for patients.

1.3. Evaluation metrics

The literature has employed various evaluation metrics to assess the performance of ML and DL-based models for predicting DDI that capture different aspects of classification quality. These include accuracy (ACC), F1-score, recall, precision, area under the precision-recall curve (AUPRC), area under the receiver operating characteristic curve (AUROC), and Matthews correlation coefficient (MCC). Each offers different insights into the classification purpose, based on the confusion matrix values, that is, true negatives (TN), true positives (TP), false negatives (FN), and false positives (FP) [41].

ACC measures the overall correctness of a model by quantifying the proportion of correctly classified negative and positive examples out of all predictions.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

A high value of ACC represents that the model correctly predicts most of the examples, but it may not provide good results for class imbalance cases.

Precision is based on the positive predictive values, and it quantifies the reliability of positive predictions by computing the portion of predictive positive examples that are TP.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall, also known as sensitivity, is based on the TP rate, and it quantifies the ability of the model to identify all actual positive examples.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

The F1-score is basically the harmonic average of recall and precision. The main aim of the F1-score is to balance the precision and recall aspects into a single evaluation metric.

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{4}$$

The F1-score is well-suited for situations where the class distribution is skewed, necessitating consideration of both FN and FP simultaneously.

AUROC precisely measures the trade-off between recall and FP rate among all possible classification thresholds.

$$AUROC = \int_0^1 (TP\ Rate)(FP\ Rate)\partial(FP\ Rate) \tag{5}$$

The value of AUROC is between 0 and 1. If the value is close to 1, it indicates strong discriminative performance across the threshold, and if the value is close to 0, it suggests poor discriminative performance across the threshold. Meanwhile, an average value of 0.5 indicates random discriminative performance.

AUPRC aggregates model performance over varying thresholds by plotting precision against recall. It generally performs well with class-imbalanced problems. It focuses on the performance of the positive class and overlooks the optimistic evaluation that may occur with the AUROC in cases of class imbalance.

$$AUPRC = \int_0^1 P(r)\partial r \tag{6}$$

Here, P(r) shows precision as a function of recall.

In MCC, all four categories of the confusion matrix are considered, providing a comprehensive and balanced measure for binary classification.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{7}$$

The value of MCC is between −1 and +1. If the value is −1, it indicates total disagreement, and if the value is +1, it means perfect classification. Meanwhile, zero suggests that performance is random, relying solely on chance. For binary classification, MCC is a robust and overall assessment metric, especially in class imbalance scenarios, compared to ACC or F1-score.

It is a good approach to use a set of metrics that consists of both threshold-independent measures (i.e., AUROC and AUPRC) and threshold-dependent measures (i.e., ACC, recall, precision, MCC, and F1-score) that enable a nuanced evaluation of ML and DL-based models' performance, especially in class-imbalanced cases such as DDI prediction.

1.4. Data source for DDIs

Direct DDIs analysis integrates a broad range of influencing factors derived from different drug-related data repositories. These data sources capture heterogeneous information, including detailed protein structural data, mechanisms of action, and pharmacogenomic effects. Therefore, the availability of a substantial amount of drug data has enabled researchers to classify potential interactions among drugs and model the underlying complexity of their interactions. It provides a deeper understanding of the molecular and biological mechanisms that drive DDIs. This study identifies several important chemical and bioinformatics databases, as shown in Table 1. It presents the information, including the publication year, and provides concise descriptions of the features and their uses. With the rapid advancement of drug discovery and analysis, Table 1 serves as a valuable reference, presenting the features of each dataset to help researchers select one that aligns with their specific needs.

1.5. Problem statement

DDIs present a serious clinical risk and challenges to patient safety, especially in the context of polypharmacy [41–43]. Traditional rule-based and experimental approaches are time-consuming, costly, and often insufficient for capturing the nonlinearity and high dimensionality of pharmacological interactions [44–47]. Consequently, AI-based approaches, particularly ML and DL, have been broadly used for predictive DDIs over the past decade due to their powerful model capability.

Several review studies have highlighted progress in AI-based DDI prediction. Current systematic literature reviews (SLRs) mainly focus on either selected data representation or ML models, particularly DL architectures, while representing a limited combination across DDIs data sources, modeling paradigms,

Table 1
Comprehensive overview of key data sources related to DDIs

Dataset (pub. year)	Drugs/items	Interactions/-pairs/entries	Context	URL
Protein Data Bank 1971	200,000 structures	Molecule complexes	Protein and ligand 3D structures	https://www.rcsb.org/
KEGG 1995	11,147	324,183	Biological pathways and interactions	https://www.kegg.jp/
Liverpool HIV DDI DB 1999	300 + HIV drugs + many co-administered drugs	1,400 clinically validated DDIs	This dataset focuses on antiretrovirals, providing management guidance and severity information	https://www.hiv-druginteractions.org/
UniProt 2003	Millions of proteins	Functional/interaction annotations	Central protein database	https://www.uniprot.org/
PubChem 2004	Millions of compounds	Assay/activity results	Broad chemical repository	https://pubchem.ncbi.nlm.nih.gov/
DrugBank 2006	1706	191,808 DDIs	Core DDI database with transporters, targets, and enzymes	https://go.drugbank.com/
SIDER 2008	1430	139,756 drug-side effect pairs	Drug side effects are linked to drugs	https://zenodo.org/records/7877720
Bio2RDF 2008	Bio-entities as RDF triples	Chemical/drug relationships	Cross-linked biomedical RDF dataset	https://bio2rdf.org/
TWOSIDES 2012	645	4,649,441 DDIs	Compiled adverse pairwise interactions	https://tatonettilab.org/offsides/
OFFSIDES 2012	1,332	18,842 drug-event associations	Off-label side effects mined from FDA reports	https://tatonettilab.org/offsides/
SemEval-2013/ DDIExtraction 2013	13029	4037 DDIs	Text corpus of abstracts and DrugBank excerpts with interaction labels	https://aclanthology.org/S13-2056/
ChEMBL22 2016	1.69 million distinct compounds and 11,224 targets	14.37 million compound-target activity values	It is a comprehensive bioactivity dataset with assay data, chemical structures, and target information	https://www.ebi.ac.uk/chembl/
ZhangDDI 2017	Derived from TWOSIDES / OFFSIDES	–	Used in computational DDI prediction	https://github.com/zw9977129/drug-drug-interaction/tree/master/dataset
BIOSNAP 2018	1332	41,520 DDIs	Drug interaction network for FDA-approved drugs	https://figshare.com/articles/dataset/BIOSNAP_DDI_and_genes_data/23600565/1
TAC DDI Extraction corpus 2018	–	–	NLP annotated corpus, training/test SPLs (8,000+ sentences, 128 test SPLs) with manually corrected interactions	https://bionlp.nlm.nih.gov/tac2018druginteractions/

(*Continued*)

Table 1
(*Continued*)

Dataset (pub. year)	Drugs/items	Interactions/-pairs/entries	Context	URL
DrugComb 2019	8397	21.5 million data points	Large drug combination activity database	https://drugcomb.org/download/
DDInter 2021	1833	236,834 interactions	Bioinformatics drug interaction database	https://ddinter2.scbdd.com/
HKG 2022	11,516 drug nodes (linked to DrugBank IDs)	20,988 total entity nodes	It is a heterogeneous knowledge graph constructed from the DrugBank dataset and linked to biomedical resources	https://github.com/tticoin/HKG-DDIE.git
DDI interaction dataset (Kaggle) 2024	–	–	Kaggle hosted dataset (not standardized)	https://www.kaggle.com/datasets/mghobashy/drug-drug-interactions
MeTDDI 2024	1700	191,000 predicted DDIs	Computationally predicted DDIs using ML. It is suitable for model training and research, but not for clinical validation	https://github.com/LabWeng/MeTDDI

Note: "–" represents that the data source does not provide the information.

evaluation strategies, and future directions in this field. These make it challenging to compare performance across studies, hindering meaningful progress. For instance, a recent review study highlighted ML-based models and drug-related features but does not thoroughly analyze the comparative evaluation of ML and DL approaches in terms of dataset usage trends and their evaluation criteria [48]. An existing SLR focused on a broad overview of AI applications in different biomedical interaction types and concentrates on AI approaches and challenges. However, it did not offer a DDI-based synthesis of AI models, evaluation metrics, and specific research gaps [46]. Moreover, there are other recent review studies [44, 45, 47, 49–51] that focus on the utilization of ML and DL models, along with their challenges and data sources. However, there is a rapid growth of AI-based DDI prediction studies, but still, there is a lack of a comprehensive and PRISMA-guided SLR that jointly examines:

1) The AI-based research trends of DDIs from 2019 to 2025.
2) The most commonly used datasets for DDI prediction.
3) The evaluation metrics used to assess the performance of ML and DL models for predicting DDI.
4) The research gaps and future research directions in the context of DDI.

To the best of our knowledge, there is no current SLR that presents an integrated and up-to-date synthesis of these aspects within a single study. This lack of integration limits model comparisons and remains a significant challenge, such as data bias, inconsistencies in evaluation metrics, explainability, and clinical applicability. However, there is a need for a precise SLR that follows PRISMA guidelines to combine recent research trends and advances and support the development of interpretable and reliable clinical AI-driven DDI prediction systems.

2. Methodology

The methodology of this study follows the SLR guidelines proposed by Kitchenham [52]. It has been continuously adopted across diverse domains [44, 50, 53]. The Kitchenham guidelines offer a transparent, reproducible, and rigorous framework for planning, conducting, and reporting SLR.

2.1. Search strategy and data sources

Several stages are included in the search review process, such as database identification, keyword selection, and defining search strings using specific criteria for the exclusion and inclusion of the articles. Initially, a comprehensive search was conducted in the Scopus database to identify the relevant articles on the use of AI approaches to predict the DDIs. Scopus was selected as the database for this SLR due to its extensive coverage of over 25,000 peer-reviewed journals. It also covers extensive multidisciplinary articles from high-impact-factor journals in AI, biomedical science, computer science, and pharmacology that are central to DDI prediction research. It also provides advanced search capabilities and standardized indexing that offer transparency and reproducibility in SLR.

The search terms and keywords were carefully selected to ensure the retrieval of articles specifically focused on the topic of interest. The search string for this SLR is as follows:

Table 2
Inclusion and exclusion criteria of this study

Collection criteria	Inclusion criteria	Exclusion criteria
Publication time period	Articles published between 2019 and 2025	Articles published before the year 2019
Publication coverage	All countries	NA
Publication language	English	Articles that were written in languages other than English
Method	Machine learning and deep learning techniques	Other than machine learning and deep learning techniques
Target outcome	Drug–drug interaction	Other than drug–drug interaction

("Drug-to-Drug Interaction" OR "DDI") AND ("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "ML" OR "Deep Learning" OR "DL")

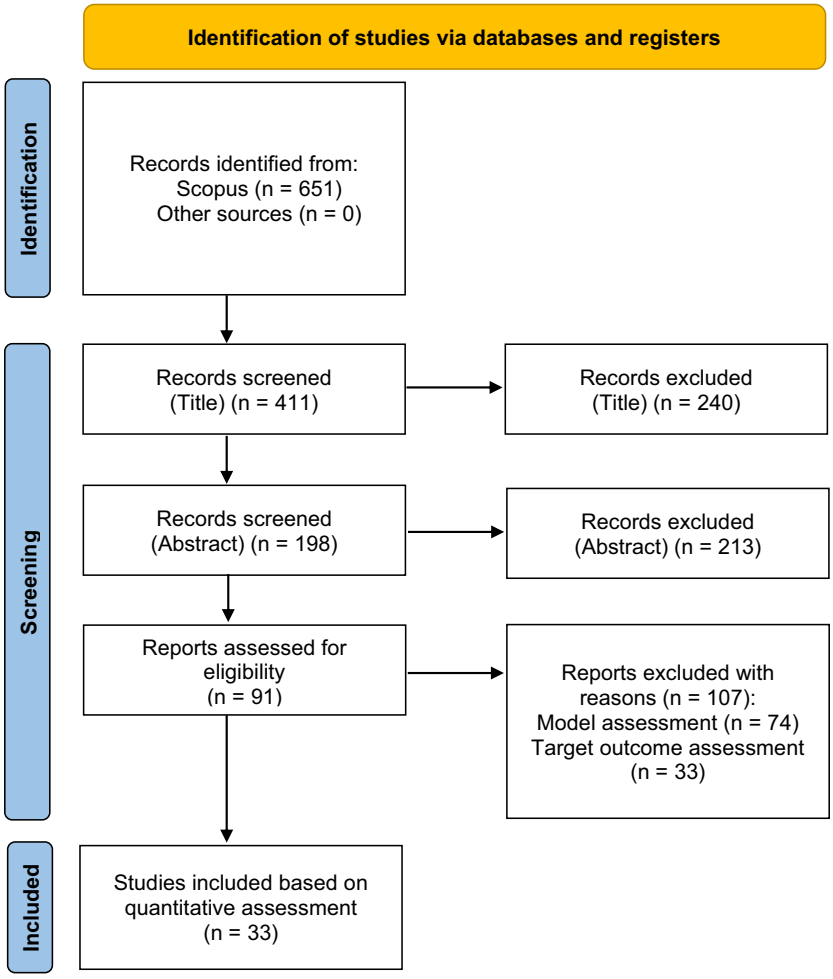
2.2. Eligibility criteria

The searches are not limited to a particular study design or country. Only studies published between 2019 and 2025 were selected. The articles were excluded if the ML or DL models were not used. Studies focusing on DDIs were selected. However, only those articles written in English were selected for consideration. Table 2 shows the inclusion and exclusion criteria of this study.

2.3. Study selection

The PRISMA flow chart in Figure 3 shows the results of the study's search string and selection process, which included the articles in this study. The search process yielded a total of 651 articles. These articles underwent a screening process based on predefined inclusion and exclusion criteria, ensuring that only the most pertinent information was included, and suitable studies were

Figure 3
PRISMA flow chart showing the review process with the study selection at the various stages



selected for the review. After the initial screening, 411 articles were selected for further evaluation.

Multiple stages were involved in finalizing the selected studies. First, both authors separately extracted the data using the same search string. To ensure consistency and accuracy in the data extraction, the discrepancies were resolved through mutual agreement. Second, retrieved records from the search string were screened based on titles, and then the abstracts were again reviewed against the inclusion and exclusion criteria of this study. This primary screening of the records was performed independently by both authors of this study. Lastly, the full texts of the shortlisted articles were read by the same authors to assess their eligibility for this SLR. Any disagreements regarding the inclusion or exclusion of articles were resolved through discussion among the authors.

The selected articles underwent a careful assessment to determine their relevance to the study objectives. Based on this assessment, the 33 articles were deemed highly relevant and included in a final review. The chosen articles cover the various methodologies and approaches used in the GNNs, CNNs, DNNs, ensembles, and attention-based models to predict the DDIs. These methodologies provide insights into the advancements, challenges, and recent progress in using DL and ML to predict the DDIs. Overall, the systematic review methodology employed in this research ensures a comprehensive and rigorous selection of the relevant articles and provides a solid foundation for analyzing the current state of research in DDI prediction.

2.4. Research question formulation

This SLR focuses on identifying the AI approaches for predicting the DDI. The study explores the effectiveness and potential of these techniques in accurately predicting the DDIs. This research aims to identify different approaches, data sources, and evaluation metrics. It also aims to identify research gaps and future directions in this domain by analyzing the relevant articles. This study will enhance our understanding of the capabilities of AI in the domain of DDIs. It ultimately improves patient care and drug safety. The focus of this study will be conducted using the following research questions (RQs):

1) RQ1: What are the AI-based research trends of the DDIs from the year 2019 to 2025?
2) RQ2: What are the most commonly used datasets for the DDI prediction?
3) RQ3: What are the most common evaluation metrics used to assess the performance of the ML and DL models for predicting the DDI?
4) RQ4: What are the research gaps and the future research directions in the context of the DDI?

3. Results

The information extraction phase of the study focused on identifying the relevant details for predicting the DDIs using AI methods. It involved identifying datasets and methodologies and assessing the evaluation metrics for selected articles. Table 3 presents a summary of all articles included in this SLR.

This study utilized descriptive statistics to evaluate the selected articles. The descriptive analysis shows different trends in DDI using AI approaches. The year-wise trend is shown in Figure 4.

The yearly distribution of the published articles on AI-based DDI prediction reports an upward trend from 2019 to 2025.

Table 3
Summary of the included articles of this study

S. No.	Author	Dataset	Method/technique	Evaluation metrics	Target outcome
1	Tahir et al. [54]	SemEval-2013	CNNs	ACC, precision, recall, and F1-score	DDI
2	Li et al. [55]	DrugBank	The article used attention-based (AB) with GNNs	ACC, precision, F1-score, and AUROC	DDI
3	Luo et al. [56]	DrugBank, ZhangDDI, and TWOSIDES	Hybrid (GNNs-AB)	F1-score and AUROC	DDI
4	Sun and Zheng [57]	DrugBank and TWOSIDES	Hybrid (GNNs-AB)	ACC, F1-score, AUROC, and AUPRC	DDI
5	Islam et al. [58]	DrugBank, Protein Data Bank, SIDER, and UniProt	DNNs	ACC, precision, recall, F1-score, and AUROC	DDI
6	Jia et al. [59]	DrugBank, DDI extraction 2013, and TAC 2018 DDI Extraction corpus	AB	Precision, recall, and F1-score	DDI
7	Halder et al. [60]	DrugBank and DDI interaction dataset (Kaggle)	Hybrid (RNNs-AB)	ACC, precision, recall, and F1-score	DDI

(*Continued*)

Table 3
(*Continued*)

S. No.	Author	Dataset	Method/technique	Evaluation metrics	Target outcome
8	Feng and Huang [61]	DrugBank, KEGG, and PubChem	DNNs	ACC, precision, recall, and F1-score	DDI
9	Im and Ko [62]	DrugBank	Hybrid (DNNs-AB)	ACC	DDI
10	Xiong et al. [63]	DrugBank, ZhangDDI, BIOSNAP, and MeTDDI	GNNs	ACC, F1-score, AUROC, and AUPRC	DDI
11	Zhao et al. [64]	DDInter database	Hybrid (GNNs-AB)	ACC, precision, recall, F1-score, AUROC, and MCC	DDI
12	Li et al. [65]	DrugBank	Hybrid (CNNs-GNNs-AB)	ACC, precision, recall, F1-score, and AUROC	DDI
13	Zhu et al. [66]	DrugBank and TWOSIDES	Hybrid (CNNs-GNNs-AB)	ACC, precision, recall, F1-score, and AUROC	DDI
14	Asfand-e-yar et al. [67]	DrugBank	Hybrid (CNNs-DNNs)	ACC, F1-score, AUROC, and AUPRC	DDI
15	Kha et al. [68]	DrugBank	XGBoost	Precision, recall, and F1-score	DDI
16	Han et al. [69]	TWOSIDES	The study used tensor factorization (TF) with DNNs	ACC, precision, recall, F1-score, and AUROC	DDI
17	Pham et al. [70]	Liverpool HIV DDI DB	Ensemble DNNs	ACC, precision, recall, and F1-score	DDI
18	Wang et al. [71]	DrugBank and TWOSIDES	GNNs	ACC, F1-score, AUROC, and AUPRC	DDI
19	Kpanou et al. [72]	DrugBank and ChEMBL22	Hybrid (CNNs-DNNs)	ACC, precision, recall, F1-score, AUROC, and AUPRC	DDI
20	Qiu et al. [73]	DrugBank and BIOSNAP	Hybrid (RNNs-AB-DNNs)	F1-score, AUROC, and AUPRC	DDI
21	Kumari et al. [74]	DrugBank, PubChem, and SIDER	The study used similarity network fusion (SNF) with CNNs	ACC, precision, recall, F1-score, and AUROC	DDI
22	Vo et al. [75]	DrugBank	Ensemble DNNs	ACC	DDI
23	Yang et al. [76]	DrugBank	The study used CNNs-Siamese with multimodal fusion	ACC, precision, recall, F1-score, AUROC, and AUPRC	DDI
24	Asada et al. [77]	HKG	Hybrid (KGE-AB)	F1-score	DDI
25	Wang et al. [18]	DrugComb	Ensemble DNNs	AUROC and AUPRC	DDI
26	Hao et al. [78]	DrugBank, and Bio2RDF	The study used knowledge graph embedding (KGE) with three-way decision (3WD)	ACC, F1-score, AUROC, and AUPRC	DDI
27	Feng and Zhang [15]	DrugBank	Hybrid (GNNs-AB)	ACC, precision, recall, F1-score, AUROC, and AUPRC	DDI
28	Al-Rabeah and Lakizadeh [79]	DrugBank and KEGG	Hybrid (GNNs-DNNs)	ACC, precision, recall, F1-score, AUROC, and AUPRC	DDI
29	Zhang et al. [80]	DrugBank	CNNs	ACC, precision, recall, F1-score, AUROC, and AUPRC	DDI

(*Continued*)

Table 3
(*Continued*)

S. No.	Author	Dataset	Method/technique	Evaluation metrics	Target outcome
30	Mei and Zhang [81]	DrugBank	Logistic regression (LoR)	ACC, F1-score, AUROC, AUPRC, and MCC	DDI
31	Schwar et al. [82]	TWOSIDES	Hybrid (DNNs-AB)	AUROC and AUPRC	DDI
32	Monteiro et al. [14]	DrugBank, UniProt, PubChem	Hybrid (CNNs-DNNs)	ACC, recall, F1-score, AUROC, and AUPRC	DDI
33	Zheng et al. [83]	DrugBank, and TWOSIDES	SVM	Precision, recall, and F1-score	DDI

Figure 4
Year-wise publication trend on DDI using AI-based techniques

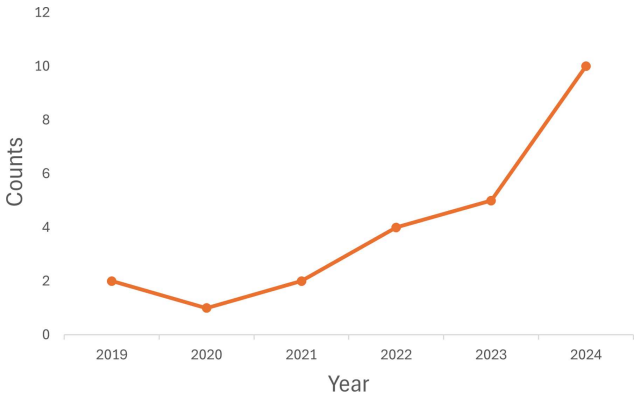
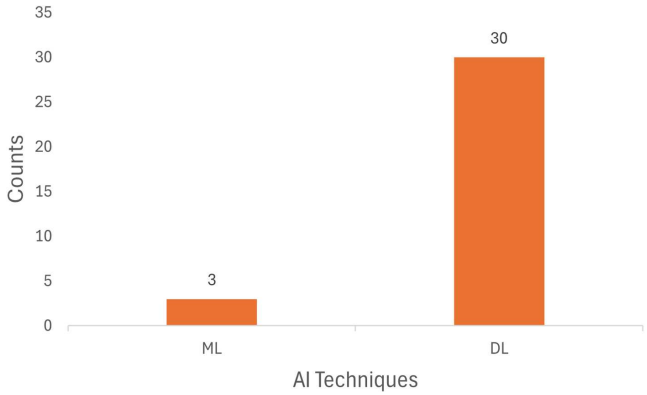


Figure 5
Comparison trend of the different AI techniques in the context of the DDI



Initially, in 2019 and 2020, only one study was published in both years that reflects the early adoption phase of ML and DL approaches within DDI research. However, from 2021 onward, there was an increasing publication count from year to year, and 2025 had the highest number of articles (i.e., 12). This growth aligns with the review studies that documented the expanding application of AI-based approaches in DDI prediction, particularly ML and DL models that have matured with the passage of time to address the complexity of DDI more reliably and effectively than traditional methods [44, 45, 50, 51]. So, these studies demonstrate advances in methodology and increasing research interest in using AI to predict the DDIs.

The distribution of AI approaches across the selected articles shows a preference for DL over ML in the domain of DDI prediction. Figure 5 shows that only three articles employed ML methods, whereas 30 articles used DL techniques. This discrepancy highlights the field's shift toward DL-based mechanisms, which are well-suited to capture the nonlinear, hidden, and complex patterns in high-dimensional drug data. In contrast to ML models such as the LR, DT, RF, or SVM, neural networks, including the RNNs, CNNs, DNNs, and GNNs, have shown superior capabilities for modeling the underlying relationships among drugs [54, 58, 63, 73]. This trend aligns with the recent studies indicating that DL not only enhances the DDI prediction but also facilitates the integration of heterogeneous datasets, such as the real-world healthcare records, drug interaction sets, and knowledge graphs [63, 73, 78]. The dominance of the DL in the selected articles for this SLR reveals both methodological advancements and the rising consensus that DL techniques offer superior scalability and flexibility for complex prediction tasks in pharmacoinformatics.

The analysis of the methodological approaches used in the selected articles reveals a clear preference for hybrid modeling strategies to predict the DDIs. Figure 6 shows that hybrid approaches combining two models yielded the highest numbers. A total of 16 articles used a hybrid approach that combines the two models, followed by the 13 articles using a single-model approach. Moreover, only four studies used a hybrid approach that integrates three models, which remain comparatively limited.

Figure 6
Comparison trend of the different AI approaches (architectures) in the context of the DDI

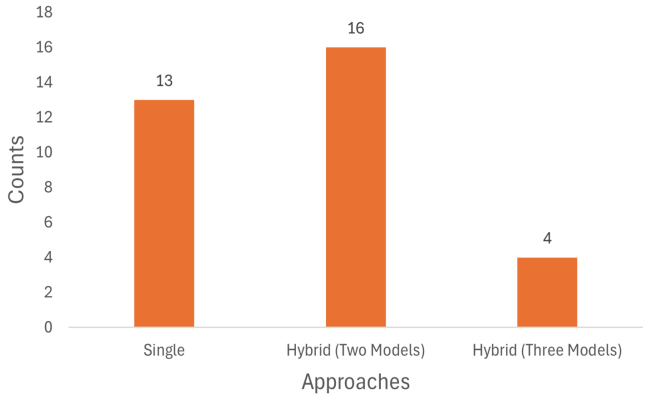
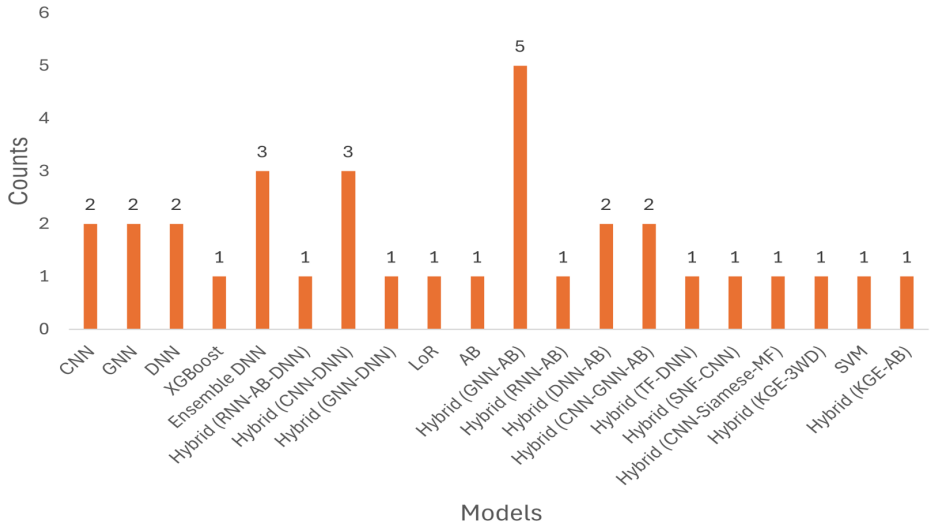


Figure 7
Comparison trend of the AI models in the context of the DDI



This distribution suggests that the researchers are increasingly favoring the combination of different modeling approaches to enhance the predictive robustness and capture the diverse and hidden pharmacological properties. The single-model approach is still used, but it is less dominant, which reflects the increasing recognition of its limitations in modeling the diverse nature and complex drug interactions data. Moreover, the relatively lower adoption of the three-model hybrid approach suggests that while it offers good performance, it may introduce challenges such as reduced interpretability, increased computational cost, and difficulties with model optimization due to its complex architecture. However, these findings reveal that most studies have adopted a balanced hybrid AI-based approach to predict the DDI.

The distribution of the AI-based models across the selected articles shows a clear preference for advanced DL approaches for the prediction of DDIs. Figure 7 shows the distribution of the different models for the reviewed studies in this SLR. The CNNs, GNNs, and DNN models, each reported with moderate frequency, indicate their extensive adoption for capturing the complex drug-related features and interaction patterns from the high-dimensional and diverse drug data. CNN models are frequently used for their ability to hierarchically extract features from the chemical representations, such as the SMILES strings. It achieves good predictive accuracy in the DDI task, as reported in recent studies [54, 80].

The GNN variants are also prominently represented in reported studies that align with the trend toward using topological networks and relational information about drugs to improve the interaction inference [63, 71]. The DNNs and ensemble DNNs further highlight the importance of diverse learning paradigms in capturing the nonlinear, hidden, and complex relationships in the drug pairs [18, 58, 61, 70, 75]. Hybrid models that combine neural architectures, such as GNN-AB and CNN-DNN, are the most common and indicate that multi-model frameworks are increasingly favored to leverage the strengths of different models and improve the overall performance [14, 15, 56, 57, 64, 67, 72]. This trend is consistent with recent review studies highlighting the growing application of hybrid and DL-based techniques in DDI-related research [45, 47].

Traditional ML models, including the SVM, LR, and XGBoost, remain widely used [68, 81, 83]. They are underrepresented compared to the DL models. It indicates a shift toward models capable of handling the diverse biomedical datasets and capturing the hidden interactions among drugs that traditional ML models may overlook. The dominance of DL and hybrid models in the reported studies indicates methodological advancement in the DDI field.

Figure 8 shows that the benchmark data sources used in the AI-based DDI prediction research exhibit substantial heterogeneity. DrugBank is the most frequently used data source among the different data sources considered in this SLR. It shows that it is the main repository for supporting binary and multiclass DDI prediction tasks. Its dominance is evident in the studies reported in this SLR, where it is extensively used to predict the DDIs using various ML and DL frameworks. The TWOSIDES dataset appears seven times; it is the most frequent after the DrugBank. Several mid-tier data repositories, such as SemEval-2013, ChEMBL22, and PubChem, are also used in tasks requiring the DDI prediction. Figure 8 shows that other datasets are also used but are not frequently employed. This pattern reveals broader trends in the DDI field, in which researchers combine biomedical and chemical knowledge repositories to improve the predictive performance. It often leverages knowledge graphs or diverse data representations to address the characteristic limitations of individual datasets.

Overall, the dominance of the DrugBank and TWOSIDES data sources in DDI prediction demonstrates their consistent relevance and makes them the benchmark datasets. Moreover, the proliferation of novel and niche data repositories underscores the ongoing efforts to enhance dataset reliability, coverage, and clinical relevance.

Figure 9 demonstrates the distribution of the evaluation metrics used in the reported studies of this SLR and shows distinct preferences for assessing performance. The F1-score was the most frequently used evaluation metric and appeared 29 times. It is followed by ACC, AUROC, recall, and precision.

The frequent use of the F1-score underscores its importance in handling the class imbalance. It is a common issue in DDI-related datasets. A recent study supports the statement that the F1-score is a robust metric for providing a more realistic and accurate evaluation of imbalanced binary classification tasks in the DDI prediction [80]. Furthermore, the frequent use of AUROC

Figure 8
Data sources analysis for the DDI prediction

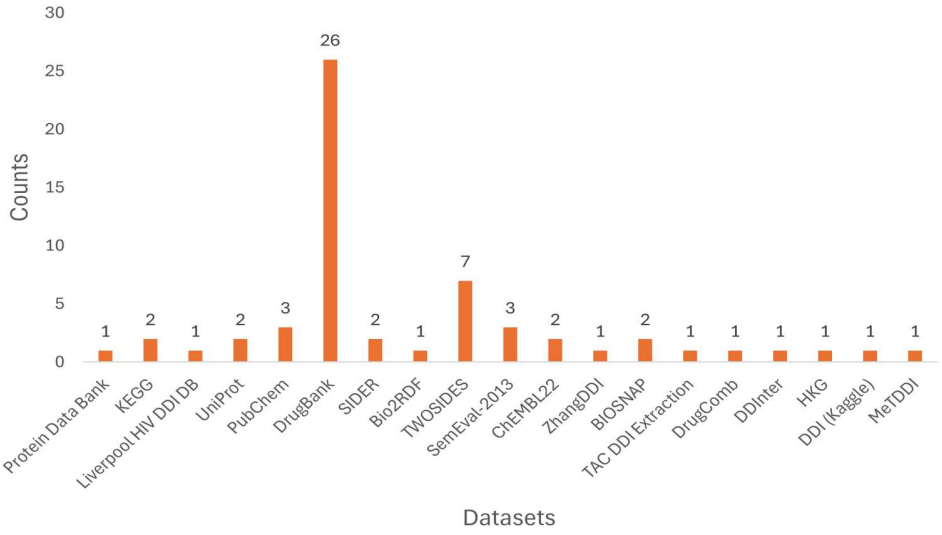
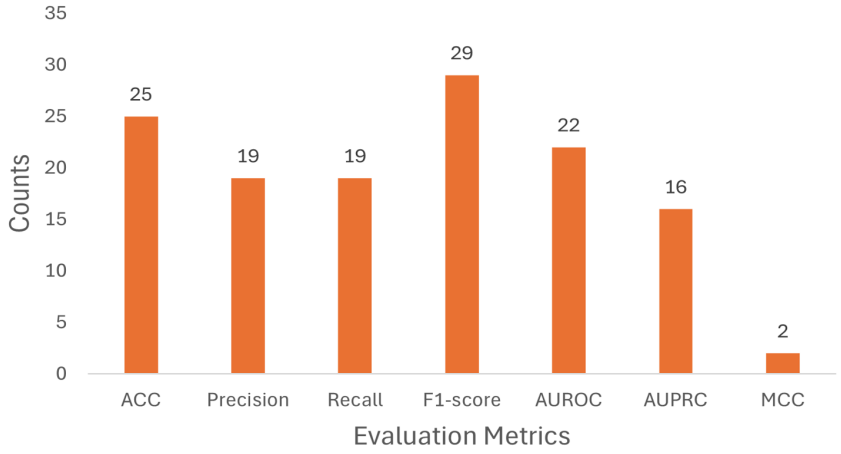


Figure 9
The evaluation metrics used in the AI-based models for predicting the DDI



in the selected studies underscores the assessment of classification thresholds' discriminative capability. It aligns with trends in DL-based research that prioritize threshold-agnostic metrics for evaluating the model performance and generalizability [18].

AUPRC is also commonly used because it provides improved insights into the model performance under severe imbalance. Moreover, MCC remains underutilized despite its theoretical advantages for binary classifiers, particularly in cases with skewed class distributions. However, these results indicate a strong preference for the evaluation metrics that balance specificity and sensitivity. It supports a detailed assessment of model performance in high-imbalance and complex DDI prediction tasks.

4. Discussion

This SLR provides a comprehensive and balanced synthesis of the latest advances in the AI-based DDI prediction. It highlights various methodological limitations, critical research gaps, and emerging challenges that permit further investigation. The increasing number of published articles since 2021 shows a growing research momentum and the maturity of the AI models. However, the findings also reveal a structural imbalance in the use of data repositories, model design, evaluation metrics, and clinical relevance. It hinders the translational impact of current AI-based DDI prediction methods.

A primary research gap found in this study is the overreliance on a few benchmark datasets, such as the DrugBank and TWOSIDES. Meanwhile, these data sources provide high-quality drug interaction data and facilitate methodological benchmarking. However, their frequent use raises concerns about the data homogeneity and limited clinical and external validity. The underrepresentation of the real-world clinical datasets (i.e., pharmacovigilance reports, electronic patient health records, and post-marketing surveillance data sources) indicates a disconnect between model performance and real-world applicability. This research gap is very critical, as clinically relevant DDIs often come from complex patient-oriented factors that are not adequately captured in these available data repositories.

Another critical gap is identified in the evaluation metrics used in the reported articles of this study. To assess the clinical utility or decision reliability, the frequent use of measures such as F1-score, ACC, and AUROC remains largely unexplored. The infrequent use of the MCC and the absence of clinical

utility-based measures highlight an evaluation gap that may lead to presenting an optimistic interpretation of the model performance and generalization in highly imbalanced drug-related data. This gap limits the model comparisons and the assessment of their readiness for clinical decision support systems.

Moreover, the DL and hybrid models dominate the AI-based DDI prediction literature. However, their interpretability and explainability remain insufficiently addressed [54, 59, 62, 65, 68, 73, 77, 79, 81, 82]. This gap creates a significant hurdle to clinical adoption, as trust, reliability, and transparency are essential for integrating it into healthcare.

While many of the studies acknowledge this issue, few adopt systematic mitigation strategies such as resampling, cost-sensitive learning, or uncertainty-aware modeling. As a result, reported performance gains may not generalize to rare or clinically critical interaction types.

The reported articles of this study have several limitations that affect methodological generalizability. First, imbalanced classes in the DDI datasets remain a pervasive issue, with most data sources having a smaller number of confirmed DDI as compared to noninteractive drug pairs. Some studies acknowledge this challenge, and a few of them use data balancing techniques to overcome this issue [61]. However, their model's performance is not generalizable to clinically critical interaction types. It is also noticed that external validation is rarely performed in the selected studies. Most studies rely solely on train–test splits within the same dataset. However, for the deployment of the model in industry, it is important to test the model's performance on a cross-dataset that may enhance the reliability and generalizability of the model.

The dominance of DL and their hybrid models introduces methodological trade-offs; they also show strong predictive performance. However, it may increase architectural complexity, especially in multi-component hybrid models, which often lead to higher computational costs, reduced interpretability, and difficulty in hyperparameter optimization. The limited utilization of the three-model hybrid system in the selected studies reveals that these critical challenges may restrict researchers.

Compared with earlier reviews on DDI prediction, this study extends existing knowledge in several important ways. Previous SLRs largely emphasized traditional ML methods, similarity-based approaches, and rule-based systems. However, this review highlights a clear shift toward the DL, graph-based, and hybrid AI models. In contrast to earlier reviews that focused primarily on algorithmic categories, this work provides a quantitative synthesis of datasets, evaluation metrics, and modeling strategies that enable a more granular assessment of methodological trends and gaps.

This SLR extends existing knowledge in several important ways compared to recent review studies on the DDI prediction. As the existing reviews did not provide a DDI-based synthesis of recent ML and DL models, details of the data repositories, evaluation measures, and the associated research gaps were lacking [44–47, 49–51]. Additionally, the previous review studies have often discussed the performance of ML and DL models separately. However, this study systematically provides details regarding recent trends, trending models, frequently used data repositories, and commonly adopted evaluation measures. It offers a more comprehensive understanding of the AI-based DDI research within a single study. By doing so, this study provides explainability, differentiates future directions, and highlights the areas where future research efforts may be concentrated.

The findings of this discussion suggest that future research on the AI-based DDI prediction should focus on:

1) Broder integration of the multimodal and the real-world clinical datasets
2) Incorporation of the latest DL models and specifically utilizing them in a hybrid framework
3) Standardized with a clinical point of view evaluation framework
4) Accurate external validation strategies

5. Conclusion

This study provides a structured and comprehensive overview of the recent advances in the AI-based DDI predictions, with an emphasis on key trends, the latest methodological modifications, dominant datasets, evaluation metrics, and emerging challenges in the field. The analysis shows growth in research since 2021, along with the latest methodological shifts toward the DL frameworks that can capture nonlinear, hidden, and complex relationships in pharmacological datasets. The frequent use of the benchmark data repositories, such as the DrugBank and TWOSIDES, along with the evaluation metrics like FI-score and AUROC, indicates a growing consensus on consistent evaluation practices for the DDI prediction tasks, particularly those involving imbalanced data.

Despite these advances, this study identifies several limitations, including heavy reliance on two benchmark datasets, limited use of real-world clinical data sources, insufficient validation of model performance across datasets, and inadequate evaluation metrics. Furthermore, the explainability, interpretability, and clinical relevance of existing AI-based DDI prediction models remain underexplored, which potentially limits their translational impact. The complex architecture of hybrid DL models also raises concerns about their generalizability and feasibility in real-world clinical settings.

Moreover, the SLR is constrained by the articles reported in the Scopus database and the defined inclusion and exclusion criteria. It may have excluded some relevant literature. The rapidly advancing nature of the AI field implies that very recent enhancements may not be adequately captured at analysis time. Finally, this study does not conduct a quantitative meta-analysis to predict the real-world performance due to inconsistencies in the evaluation settings across the selected studies.

In conclusion, this study highlights the growing importance of AI approaches for DDI prediction, while underscoring the need for future research to shift from performance-centric modeling toward clinically meaningful, interpretable, and generalizable solutions. Addressing evaluation standardization, data diversity, explainability, and the real-world clinical validation will be essential to advancing AI-based DDI prediction, with practitioners and researchers serving as the guides for future development.

Ethical Statement

This research does not include any studies with human or animal subjects conducted by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Faisal Asad ur Rehman: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Abdulkarim Kanaan Jebna:** Validation, Investigation, Visualization, Supervision, Project administration. **Touqeer Ahmad:** Methodology, Resources. **Arif Ur Rahman:** Data curation, Writing – review & editing, Project administration. **Fasee Ullah:** Formal analysis, Writing – review & editing, Visualization.

References

[1] McCluggage, L. K., & Davis, L. E. (2010). Book review: Stockley's drug interactions 2010: Pocket companion. *Annals of Pharmacotherapy*, *44*(7-8), 1355. https://doi.org/10.1345/aph.1P048

[2] Agarwal, S., Agarwal, V., Agarwal, M., & Singh, M. (2020). Exosomes: Structure, biogenesis, types and application in diagnosis and gene and drug delivery. *Current Gene Therapy*, *20*(3), 195–206. https://doi.org/10.2174/1566523220999200731011702

[3] Lazarou, J., Pomeranz, B. H., & Corey, P. N. (1998). Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA*, *279*(15), 1200–1205. https://doi.org/10.1001/jama.279.15.1200

[4] Hohl, C. M., Dankoff, J., Colacone, A., & Afilalo, M. (2001). Polypharmacy, adverse drug-related events, and potential adverse drug interactions in elderly patients presenting to an emergency department. *Annals of Emergency Medicine*, *38*(6), 666–671. https://doi.org/10.1067/mem.2001.119456

[5] Khori, V., Semnani, S. H., & Roshandel, G. H. (2006). Frequency distribution of drug interactions and some of related factors in prescriptions. *Medical Journal of Tabriz University of Medical Sciences*, *27*(4), 29–32. https://mj.tbzmed.ac.ir/Article/6993

[6] Safdari, R., Ferdousi, R., Aziziheris, K., Niakan-Kalhori, S. R., & Omidi, Y. (2016). Computerized techniques pave the way for drug-drug interaction prediction and interpretation. *BioImpacts*, *6*(2), 71–78. https://doi.org/10.15171/bi.2016.10

[7] Kontsioti, E., Maskell, S., Bensalem, A., Dutta, B., & Pirmohamed, M. (2022). Similarity and consistency assessment of three major online drug-drug interaction resources. *British Journal of Clinical Pharmacology*, *88*(9), 4067–4079. https://doi.org/10.1111/bcp.15341

[8] Rivas, A., & Vidal, M. E. (2021). Capturing knowledge about drug-drug interactions to enhance treatment effectiveness. In *Proceedings of the 11th Knowledge Capture Conference*, 33–40. https://doi.org/10.1145/3460210.3493560

[9] Spanakis, M., Tzamali, E., Tzedakis, G., Koumpouzi, C., Pediaditis, M., Tsatsakis, A., & Sakkalis, V. (2025). Artificial intelligence models and tools for the assessment of drug-herb interactions. *Pharmaceuticals*, *18*(3), 282. https://doi.org/10.3390/ph18030282

[10] Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Research*, *49*(D1), D545–D551. https://doi.org/10.1093/nar/gkaa970

[11] Han, K., Cao, P., Wang, Y., Xie, F., Ma, J., Yu, M., . . . , & Wan, J. (2022). A review of approaches for predicting drug–drug interactions based on machine learning. *Frontiers in Pharmacology*, *12*, 814858. https://doi.org/10.3389/fphar.2021.814858

[12] Song, B., Li, F., Liu, Y., & Zeng, X. (2021). Deep learning methods for biomedical named entity recognition: A survey and qualitative comparison. *Briefings in Bioinformatics*, *22*(6), bbab282. https://doi.org/10.1093/bib/bbab282

[13] Allesøe, R. L., Lundgaard, A. T., Hernández Medina, R., Aguayo-Orozco, A., Johansen, J., Nissen, J. N., . . . , & Consortium, IMI DIRECT. (2023). Discovery of drug–omics associations in type 2 diabetes with generative deep-learning models. *Nature Biotechnology*, *41*(3), 399–408. https://doi.org/10.1038/s41587-022-01520-x

[14] Monteiro, N. R., Ribeiro, B., & Arrais, J. P. (2021). Drug-target interaction prediction: End-to-end deep learning approach. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *18*(6), 2364–2374. https://doi.org/10.1109/TCBB.2020.2977335

[15] Feng, Y.-H., & Zhang, S.-W. (2022). Prediction of drug-drug interaction using an attention-based graph neural network on drug molecular graphs. *Molecules*, *27*(9), 3004. https://doi.org/10.3390/molecules27093004

[16] Picart-Armada, S., Thompson, W. K., Buil, A., & Perera-Lluna, A. (2021). The effect of statistical normalization on network propagation scores. *Bioinformatics*, *37*(6), 845–852. https://doi.org/10.1093/bioinformatics/btaa896

[17] Ling, C., Zeng, T., Dang, Q., Liang, Y., Liu, X., & Xie, S. (2024). Predicting drug-target interactions using matrix factorization with self-paced learning and dual similarity information. *Technology and Health Care*, *32*, 49–64. https://doi.org/10.3233/THC-248005

[18] Wang, Z., Dong, J., Wu, L., Dai, C., Wang, J., Wen, Y., . . . , & Bo, X. (2023). DEML: Drug synergy and interaction prediction using ensemble-based multi-task learning. *Molecules*, *28*(2), 844. https://doi.org/10.3390/molecules28020844

[19] Qi, H., Li, X., Zhang, C., & Zhao, T. (2025). Improving drug-drug interaction prediction via in-context learning and judging with large language models. *Frontiers in Pharmacology*, *16*, 1589788. https://doi.org/10.3389/fphar.2025.1589788

[20] Datta, A., Flynn, N. R., Barnette, D. A., Woeltje, K. F., Miller, G. P., & Swamidass, S. J. (2021). Machine learning liver-injuring drug interactions with non-steroidal anti-inflammatory drugs (NSAIDs) from a retrospective electronic health record (EHR) cohort. *PLOS Computational Biology*, *17*(7), e1009053. https://doi.org/10.1371/journal.pcbi.1009053

[21] Hauben, M. (2023). Artificial intelligence and data mining for the pharmacovigilance of drug-drug interactions. *Clinical Therapeutics*, *45*(2), 117–133. https://doi.org/10.1016/j.clinthera.2023.01.002

[22] Nyamabo, A. K., Yu, H., Liu, Z., & Shi, J.-Y. (2022). Drug–drug interaction prediction with learnable size-adaptive molecular substructures. *Briefings in Bioinformatics*, *23*(1), bbab441. https://doi.org/10.1093/bib/bbab441

[23] Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Artificial intelligence to deep learning: Machine intelligence approach for drug discovery.

Molecular Diversity, *25*(3), 1315–1360. https://doi.org/10.1007/s11030-021-10217-3

[24] Ren, Z.-H., Yu, C.-Q., Li, L.-P., You, Z.-H., Guan, Y.-J., Wang, X.-F., & Pan, J. (2022). BioDKG–DDI: Predicting drug–drug interactions based on drug knowledge graph fusing biochemical information. *Briefings in Functional Genomics*, *21*(3), 216–229. https://doi.org/10.1093/bfgp/elac004

[25] Qiu, Y., Zhang, Y., Deng, Y., Liu, S., & Zhang, W. (2022). A comprehensive review of computational methods for drug-drug interaction detection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *19*(4), 1968–1985. https://doi.org/10.1109/TCBB.2021.3081268

[26] Kumar Shukla, P., Kumar Shukla, P., Sharma, P., Rawat, P., Samar, J., Moriwal, R., & Kaur, M. (2020). Efficient prediction of drug–drug interaction using deep learning models. *IET Systems Biology*, *14*(4), 211–216. https://doi.org/10.1049/iet-syb.2019.0116

[27] Lee, G., Park, C., & Ahn, J. (2019). Novel deep learning model for more accurate prediction of drug-drug interaction effects. *BMC Bioinformatics*, *20*(1), 415. https://doi.org/10.1186/s12859-019-3013-0

[28] Ali, S., Alam, W., Tyara, H., & Chong, K. T. (2025). An accurate prediction of drug–drug interactions and side effects by using integrated convolutional and BiLSTM networks. *Chemometrics and Intelligent Laboratory Systems*, *257*, 105304. https://doi.org/10.1016/j.chemolab.2024.105304

[29] Zhang, S., Huang, W., & Zhang, Y. (2021). Drug-drug interaction extraction based on bidirectional gated recurrent unit networks and capsule networks. *Open Access Journal of Biomedical Science*, *3*, 805–811.

[30] Patro, P., Kumar, K., Swain, D., & Sama, H. R. (2022). An analysis of drug-drug interactions (DDIs) using machine learning techniques in the drug development process. In R. Raut, S. Krit, & P. Chatterjee (Eds.), *Machine vision for Industry 4.0* (pp. 195–206). CRC Press.

[31] Ahmed, R., Darwish, A., & Hassanien, A. E. (2024). Efficient prediction adverse drug-drug interactions with deep neural networks. In A. E. Hassanien, A. Darwish, & S. M. Elghamrawy (Eds.), *Artificial intelligence for environmental sustainability and green initiatives* (pp. 183–199). Springer. https://doi.org/10.1007/978-3-031-63451-2_11

[32] Zhang, Y., Qiu, Y., Cui, Y., Liu, S., & Zhang, W. (2020). Predicting drug-drug interactions using multimodal deep auto-encoders based network embedding and positive-unlabeled learning. *Methods*, *179*, 37–46. https://doi.org/10.1016/j.ymeth.2020.05.007

[33] Ma, M., & Lei, X. (2023). A dual graph neural network for drug–drug interactions prediction based on molecular structure and interactions. *PLOS Computational Biology*, *19*(1), e1010812. https://doi.org/10.1371/journal.pcbi.1010812

[34] Saad, A. I., Omar, Y. M., & Maghraby, F. A. (2019). Predicting drug interaction with adenosine receptors using machine learning and SMOTE techniques. *IEEE Access*, *7*, 146953–146963. https://doi.org/10.1109/ACCESS.2019.2946314

[35] Shi, J.-Y., Mao, K.-T., Yu, H., & Yiu, S.-M. (2019). Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi-nonnegative matrix factorization. *Journal of Cheminformatics*, *11*(1), 28. https://doi.org/10.1186/s13321-019-0352-9

[36] Shtar, G., Greenstein-Messica, A., Mazuz, E., Rokach, L., & Shapira, B. (2022). Predicting drug characteristics using biomedical text embedding. *BMC Bioinformatics*, *23*(1), 526. https://doi.org/10.1186/s12859-022-05083-1

[37] Szulc, N. A., Mackiewicz, Z., Bujnicki, J. M., & Stefaniak, F. (2023). Structural interaction fingerprints and machine learning for predicting and explaining binding of small molecule ligands to RNA. *Briefings in Bioinformatics*, *24*(4), bbad187. https://doi.org/10.1093/bib/bbad187

[38] Vo, T. H., Nguyen, N. T. K., Kha, Q. H., & Le, N. Q. K. (2022). On the road to explainable AI in drug-drug interactions prediction: A systematic review. *Computational and Structural Biotechnology Journal*, *20*, 2112–2123. https://doi.org/10.1016/j.csbj.2022.04.021

[39] Song, T., Zhang, X., Ding, M., Rodriguez-Paton, A., Wang, S., & Wang, G. (2022). DeepFusion: A deep learning based multi-scale feature fusion method for predicting drug-target interactions. *Methods*, *204*, 269–277. https://doi.org/10.1016/j.ymeth.2022.02.007

[40] Ran, B., Chen, L., Li, M., Han, Y., & Dai, Q. (2022). Drug-drug interactions prediction using fingerprint only. *Computational and Mathematical Methods in Medicine*, *2022*(1), 7818480. https://doi.org/10.1155/2022/7818480

[41] Alhumaidi, R. M., Bamagous, G. A., Alsanosi, S. M., Alqashqari, H. S., Qadhi, R. S., Alhindi, Y. Z., . . . , & Falemban, A. H. (2023). Risk of polypharmacy and its outcome in terms of drug interaction in an elderly population: A retrospective cross-sectional study. *Journal of Clinical Medicine*, *12*(12), 3960. https://doi.org/10.3390/jcm12123960

[42] Wolff, J., Hefner, G., Normann, C., Kaier, K., Binder, H., Hiemke, C., . . . , & Klimke, A. (2021). Polypharmacy and the risk of drug–drug interactions and potentially inappropriate medications in hospital psychiatry. *Pharmacoepidemiology and Drug Safety*, *30*(9), 1258–1268. https://doi.org/10.1002/pds.5310

[43] Khan, E., Ijaz, A., Jan, F., Rashid, S., & Mercado, G. M. (2025). Investigating specific drug-drug interactions and their clinical implications in polypharmacy, particularly in elderly patients. *Journal of Medical & Health Sciences Review*, *2*(2), 3745–3758. https://doi.org/10.62019/cn4rwa95

[44] Gheorghita, F. I., Bocanet, V. I., & Iantovics, L. B. (2025). Machine learning-based drug-drug interaction prediction: A critical review of models, limitations, and data challenges. *Frontiers in Pharmacology*, *16*, 1632775. https://doi.org/10.3389/fphar.2025.1632775

[45] Luo, H., Yin, W., Wang, J., Zhang, G., Liang, W., Luo, J., & Yan, C. (2024). Drug-drug interactions prediction based on deep learning and knowledge graph: A review. *Iscience*, *27*(3), 109148. https://doi.org/10.1016/j.isci.2024.109148

[46] Marzouk, N. H., Selim, S., Elattar, M., Mabrouk, M. S., & Mysara, M. (2025). A comprehensive landscape of AI applications in broad-spectrum drug interaction prediction: A systematic review. *Journal of Cheminformatics*, *17*(1), 141. https://doi.org/10.1186/s13321-025-01093-2

[47] Xia, Y., Xiong, A., Zhang, Z., Zou, Q., & Cui, F. (2025). A comprehensive review of deep learning-based approaches for drug–drug interaction prediction. *Briefings in Functional Genomics*, *24*, elae052. https://doi.org/10.1093/bfgp/elae052

[48] Wang, N.-N., Zhu, B., Li, X.-L., Liu, S., Shi, J.-Y., & Cao, D.-S. (2024). Comprehensive review of drug–drug interaction prediction based on machine learning: Current status, challenges, and opportunities. *Journal of Chemical Information and Modeling*, *64*(1), 96–109. https://doi.org/10.1021/acs.jcim.3c01304

[49] Tuama, M. H., Najm, S. M., & Lakizadeh, A. (2025). Challenges in computational drug-drug interaction prediction: A survey. *Ingénierie des Systèmes d'Information*, *30*(6), 1653–1664. https://doi.org/10.18280/isi.300623

[50] Behrooznia, S., & Hooshmand, M. (2025). A systematic review of drug-related interactions–Utilizing deep learning and LLMs for prediction and mitigation. *ACS Omega*, *10*(50), 61048–61075. https://doi.org/10.1021/acsomega.5c04997

[51] Ru, X., Li, Z., Wei, L., Liu, Y., & Zou, Q. (2025). Drug–drug interaction prediction: Paradigm shifts, key bottlenecks, and future directions. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *15*(6), e70056. https://doi.org/10.1002/wcms.70056

[52] Kitchenham, B. (2004). *Procedures for performing systematic reviews*. *NICTA Technical Report 0400011T.1*. https://www.elizabete.com.br/site/Outros/Entradas/2012/11/19_Revisao_Sistematica_files/Conceitos_RevisaoSistematica_kitchenham_2004.pdf

[53] Rehman, F. A. U., Jebna, A. K., Ayyasamy, R. K., & Senadjki, A. (2025). Project performance prediction based on machine learning: A systematic literature. *Journal of Information Systems and Technology Management*, *10*(41), 405–427. https://doi.org/10.35631/JISTM.1041026

[54] Tahir, M. T., Ibrahim, M., Sarwar, N., Irshad, A., & Atteia, G. (2025). Enhanced drug-drug interaction extraction from biomedical text using deep learning-based sentence representations. *Scientific Reports*, *15*(1), 37842. https://doi.org/10.1038/s41598-025-21782-0

[55] Li, W., Zhou, Y., Ma, W., Liang, W., & Tang, X. (2025). MDG-DDI: Multi-feature drug graph for drug-drug interaction prediction. *BMC Bioinformatics*, *26*(1), 268. https://doi.org/10.1186/s12859-025-06288-w

[56] Luo, Y., Deng, L., & Huang, Z. (2025). HLN-DDI: Hierarchical molecular representation learning with co-attention mechanism for drug-drug interaction prediction. *BMC Bioinformatics*, *26*(1), 152. https://doi.org/10.1186/s12859-025-06157-6

[57] Sun, J., & Zheng, H. (2025). HDN-DDI: A novel framework for predicting drug-drug interactions using hierarchical molecular graphs and enhanced dual-view representation learning. *BMC Bioinformatics*, *26*(1), 28. https://doi.org/10.1186/s12859-025-06052-0

[58] Islam, S., Abbasi, A., Agarwal, N., Zheng, W., Doretto, G., & Adjeroh, D. A. (2025). PS3N: Leveraging protein sequence-structure similarity for novel drug-drug interaction discovery. *Scientific Reports*, *15*(1), 37191. https://doi.org/10.1038/s41598-025-14690-w

[59] Jia, Y., Yuan, Z., Zhu, L., & Xiang, Z.-L. (2025). A meta-contrastive learning approach for clinical drug-drug interaction extraction from biomedical literature. *PLOS Computational Biology*, *21*(12), e1013722. https://doi.org/10.1371/journal.pcbi.1013722

[60] Halder, A., Saha, B., Roy, M., & Majumder, S. (2025). A novel deep sequential learning architecture for drug drug interaction prediction using DDINet. *Scientific Reports*, *15*(1), 9337. https://doi.org/10.1038/s41598-025-93952-z

[61] Feng, Q., & Huang, X. (2025). Multi-feature machine learning for enhanced drug–drug interaction prediction. *Journal of Biomedical Informatics*, *171*, 104923. https://doi.org/10.1016/j.jbi.2025.104923

[62] Im, S., & Ko, Y. (2025). LLM-enhanced multimodal framework for drug–drug interaction prediction. *Biomedicines*, *13*(10), 2355. https://doi.org/10.3390/biomedicines13102355

[63] Xiong, P., Chen, H., Zhou, J., Zeng, Y., & Dai, Q. (2025). MGRL-DDI: Multiview graph representation learning for accurate drug–drug interaction prediction. *Journal of Chemical Information and Modeling*, *65*(18), 9800–9814. https://doi.org/10.1021/acs.jcim.5c01773

[64] Zhao, L., Shang, J., Tang, X., Kong, X., Sun, Y., & Liu, J.-X. (2025). A mutual-guided co-attention mechanism and heterogeneous attribute graph-based framework for drug-drug interaction event prediction. *Chemometrics and Intelligent Laboratory Systems*, *264*, 105440. https://doi.org/10.1016/j.chemolab.2025.105440

[65] Li, J., Zhao, Y., Wang, Z., Lei, C., Wu, L., Zhang, Y., . . . , & Xiao, J. (2025). Identify drug-drug interactions via deep learning: A real world study. *Journal of Pharmaceutical Analysis*, *15*(6), 101194. https://doi.org/10.1016/j.jpha.2025.101194

[66] Zhu, J., Che, C., Jiang, H., Xu, J., Yin, J., & Zhong, Z. (2024). SSF-DDI: A deep learning method utilizing drug sequence and substructure features for drug–drug interaction prediction. *BMC Bioinformatics*, *25*(1), 39. https://doi.org/10.1186/s12859-024-05654-4

[67] Asfand-e-yar, M., Hashir, Q., Shah, A. A., Malik, H. A. M., Alourani, A., & Khalil, W. (2024). Multimodal CNN-DDI: Using multimodal CNN for drug to drug interaction associated events. *Scientific Reports*, *14*(1), 4076. https://doi.org/10.1038/s41598-024-54409-x

[68] Kha, Q.-H., Nguyen, N. T. K., Le, N. Q. K., & Kang, J.-H. (2024). Development and validation of a machine learning model for predicting drug-drug interactions with oral diabetes medications. *Methods*, *232*, 81–88. https://doi.org/10.1016/j.ymeth.2024.10.012

[69] Han, G., Peng, L., Ding, A., Zhang, Y., & Lin, X. (2024). CTF-DDI: Constrained tensor factorization for drug–drug interactions prediction. *Future Generation Computer Systems*, *161*, 26–34. https://doi.org/10.1016/j.future.2024.06.060

[70] Pham, T., Ghafoor, M., Grañana-Castillo, S., Marzolini, C., Gibbons, S., Khoo, S., . . . , & Siccardi, M. (2024). DeepARV: Ensemble deep learning to predict drug-drug interaction of clinical relevance with antiretroviral therapy. *npj Systems Biology and Applications*, *10*(1), 48. https://doi.org/10.1038/s41540-024-00374-0

[71] Wang, Y., Yang, Z., & Yao, Q. (2024). Accurate and interpretable drug-drug interaction prediction enabled by knowledge subgraph learning. *Communications Medicine*, *4*(1), 59. https://doi.org/10.1038/s43856-024-00486-y

[72] Kpanou, R., Dallaire, P., Rousseau, E., & Corbeil, J. (2024). Learning self-supervised molecular representations for drug–drug interaction prediction. *BMC Bioinformatics*, *25*(1), 47. https://doi.org/10.1186/s12859-024-05643-7

[73] Qiu, J., Yan, X., Tian, Y., Li, Q., Liu, X., Yang, Y., . . . , & Liu, H. (2024). PTB-DDI: An accurate and simple framework for drug–drug interaction prediction based on pre-trained tokenizer and BiLSTM model. *International Journal of Molecular Sciences*, *25*(21), 11385. https://doi.org/10.3390/ijms252111385

[74] Kumari, D., Agrawal, D., Nema, A., Raj, N., Panda, S., Christopher, J., . . . , & Behera, S. (2024). A study on improving drug–drug interactions prediction using convolutional neural networks. *Applied Soft Computing*, *166*, 112242. https://doi.org/10.1016/j.asoc.2024.112242

[75] Vo, T. H., Nguyen, N. T. K., & Le, N. Q. K. (2023). Improved prediction of drug-drug interactions using ensemble deep neural networks. *Medicine in Drug Discovery*, *17*, 100149. https://doi.org/10.1016/j.medidd.2022.100149

[76] Yang, Z., Tong, K., Jin, S., Wang, S., Yang, C., & Jiang, F. (2023). CNN-Siam: Multimodal Siamese CNN-based deep learning approach for drug–drug interaction prediction. *BMC Bioinformatics*, *24*(1), 110. https://doi.org/10.1186/s12859-023-05242-y

[77] Asada, M., Miwa, M., & Sasaki, Y. (2023). Integrating heterogeneous knowledge graphs into drug–drug interaction extraction from the literature. *Bioinformatics*, *39*(1), btac754. https://doi.org/10.1093/bioinformatics/btac754

[78] Hao, X., Chen, Q., Pan, H., Qiu, J., Zhang, Y., Yu, Q., . . . , & Du, X. (2023). Enhancing drug–drug interaction prediction by three-way decision and knowledge graph embedding. *Granular Computing*, *8*(1), 67–76. https://doi.org/10.1007/s41066-022-00315-4

[79] Al-Rabeah, M. H., & Lakizadeh, A. (2022). Prediction of drug-drug interaction events using graph neural networks based feature extraction. *Scientific Reports*, *12*(1), 15590. https://doi.org/10.1038/s41598-022-19999-4

[80] Zhang, C., Lu, Y., & Zang, T. (2022). CNN-DDI: A learning-based method for predicting drug–drug interactions using convolution neural networks. *BMC Bioinformatics*, *23*(1), 88. https://doi.org/10.1186/s12859-022-04612-2

[81] Mei, S., & Zhang, K. (2021). A machine learning framework for predicting drug–drug interactions. *Scientific Reports*, *11*(1), 17619. https://doi.org/10.1038/s41598-021-97193-8

[82] Schwarz, K., Allam, A., Perez Gonzalez, N. A., & Krauthammer, M. (2021). AttentionDDI: Siamese attention-based deep learning method for drug–drug interaction predictions. *BMC Bioinformatics*, *22*(1), 412. https://doi.org/10.1186/s12859-021-04325-y

[83] Zheng, Y., Peng, H., Zhang, X., Zhao, Z., Gao, X., & Li, J. (2019). DDI-PULearn: A positive-unlabeled learning method for large-scale prediction of drug-drug interactions. *BMC Bioinformatics*, *20*(19), 661. https://doi.org/10.1186/s12859-019-3214-6

How to Cite: Rehman, F. A. ur, Jebna, A. K., Ahmad, T., Rahman, A. U., & Ullah, F. (2026). Prediction of Drug–Drug Interactions Based on Artificial Intelligence: A Systematic Literature Review. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62027539