Journal of Molecular Science

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1768-1778
DoI-10.004687/1000-9035.2025.239

Journal of Molecular Science
www.jmolecularsci.com ISSN:1000-9035

Machine Learning-Based Prediction of Drug–Drug Interactions Using Molecular Fingerprinting: A Comprehensive Review

Ms. Shweta Gogate *, Mr. Murtuza Zaki Aniket Yadav, Mohammad Akram Abdul Hameed, Khan Ashfaque Ahmed Sirajuddin, Ritik Sharma, Mohd Saquib Khan

SAM College of Pharmacy, Faculty of Medical and Paramedical Sciences, SAM Global University Raisen-(MadhyaPradesh) India- 46455.

Article Information
Received: 06-08-2025
Revised: 28-09-2025
Accepted: 12-11-2025
Published: 24-12-2025

Keywords
drug–drug interactions, molecular fingerprinting, machine learning, graph neural networks, pharmacoinformatics, ECFP, deep learning, computational pharmacology.

ABSTRACT
Drug–drug interactions (DDIs) represent a critical concern in pharmacotherapy, contributing to thousands of adverse drug reactions annually and imposing substantial burdens on healthcare systems worldwide. Traditional experimental methods for identifying DDIs are prohibitively expensive, time-consuming, and fail to scale with the expanding pharmaceutical landscape. Machine learning (ML)-based computational approaches, particularly those leveraging molecular fingerprinting, have emerged as transformative tools for high-throughput DDI prediction. This review comprehensively examines the intersection of molecular fingerprinting techniques and machine learning methodologies for DDI prediction. We systematically analyze fingerprint types—including extended-connectivity fingerprints (ECFP), MACCS structural keys, topological fingerprints, and graph-based learned representations—and discuss their respective strengths in encoding molecular structure for interaction prediction. We survey the evolution of predictive models from classical algorithms such as random forests and support vector machines through shallow neural networks to contemporary architectures including graph convolutional networks (GCNs), graph attention networks (GATs), and transformer-based models. Performance benchmarks across standard datasets (DrugBank, TWOSIDES, KEGG) demonstrate that graph neural network approaches consistently outperform classical methods, achieving AUROC values exceeding 0.93. We further discuss current challenges including class imbalance, cold-start problems, data scarcity for novel chemical entities, and interpretability requirements in clinical settings. Finally, we outline emerging directions including multi-modal integration, federated learning for privacy-preserving DDI discovery, and foundation models for molecular property prediction.

INTRODUCTION:

The co-administration of multiple drugs, known as polypharmacy, is increasingly prevalent across modern healthcare, particularly in elderly patients and those with complex comorbidities. Surveys consistently report that 20–30% of hospitalized patients receive five or more concurrent medications, while the figure in elderly outpatients may exceed 40% [1,2]. This polypharmacy landscape dramatically elevates the probability of encountering drug–drug interactions (DDIs)—

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1768-1778
DoI-10.004687/1000-9035.2025.239

events in which the pharmacological effect of one drug is altered by the concurrent presence of another [3].

The clinical consequences of undetected DDIs range from therapeutic failure to life-threatening toxicity. The United States Food and Drug Administration (FDA) estimates that approximately 125,000 deaths and over 350,000 hospitalizations per year in the U.S. alone are attributable to adverse drug reactions, with DDIs accounting for a significant subset [4,5]. Economic analyses have placed the annual cost of DDI-related adverse events at over $30 billion in the U.S. healthcare system [6]. Despite their significance, a substantial proportion of DDIs remain undetected at the time of drug approval, as premarket clinical trials are typically underpowered to detect rare interaction events and rarely study polypharmacy scenarios [7].

Traditional methods for DDI identification—including in vitro cytochrome P450 inhibition assays, in vivo pharmacokinetic studies, and post-marketing pharmacovigilance—are resource-intensive and reactive rather than proactive [8,9]. The number of theoretically possible pairwise drug combinations grows quadratically with the size of the approved drug library; with approximately 4,500 FDA-approved small molecules, this yields over 10 million potential pairs, only a fraction of which have been studied experimentally [10].

Computational methods for DDI prediction have attracted sustained interest as a scalable complement to experimental approaches. Early work employed pharmacophore-based similarity, network medicine frameworks, and knowledge-based inference [11,12]. The advent of high-quality molecular databases—DrugBank, KEGG, PubChem—combined with advances in machine learning (ML) algorithms has catalyzed a new generation of predictive models that can screen millions of compound pairs in silico [13,14].

Central to these ML pipelines is the need for compact, informative numerical representations of molecular structure. Molecular fingerprints—binary or count-based vectors encoding the presence of structural features—have become the canonical input representation for chemoinformatics ML models [15]. Extended-connectivity fingerprints (ECFP), MACCS keys, Daylight fingerprints, and more recently, learned graph-based embeddings, each offer distinct trade-offs between informativeness, computational cost, and interpretability [16].

This review provides a comprehensive account of the current state of ML-based DDI prediction using molecular fingerprinting. We survey fingerprint methodologies, ML architectures, benchmark datasets, evaluation metrics, and key findings from the literature, with the aim of providing both a reference for practitioners and a critical assessment of open challenges in the field.

2. Drug–Drug Interactions: Mechanisms and Clinical Significance:

2.1 Classification of DDIs:

DDIs are mechanistically classified into two principal categories: pharmacokinetic (PK) interactions, which alter drug absorption, distribution, metabolism, or excretion (ADME), and pharmacodynamic (PD) interactions, which arise from additive, synergistic, or antagonistic effects at shared molecular targets or downstream signaling pathways [17,18].

Pharmacokinetic DDIs most commonly involve cytochrome P450 (CYP) enzymes—particularly CYP3A4, CYP2D6, CYP2C9, and CYP1A2—which collectively metabolize approximately 75% of marketed drugs [19]. Inhibition or induction of these enzymes by one drug can profoundly alter plasma concentrations of co-administered substrates. For instance, the CYP3A4 inhibitor clarithromycin can increase simvastatin plasma levels more than tenfold, markedly elevating myopathy risk [20]. Additional PK mechanisms include P-glycoprotein-mediated transporter interactions, protein binding displacement, and renal tubular secretion competition [21].

Pharmacodynamic DDIs arise when two drugs share molecular targets, biological pathways, or physiological systems. They may be beneficial—as in combination cancer chemotherapy—or detrimental, as in the serotonin syndrome associated with concurrent serotonergic agents, or the QT prolongation risk

inherent in combining multiple QT-prolonging drugs [22,23]. The distinction between PK and PD mechanisms is clinically important, as it informs monitoring strategies and the utility of therapeutic drug monitoring.

2.2 Epidemiology and Clinical Burden:
Systematic analyses of hospital records and electronic health databases consistently identify DDIs as a leading cause of preventable adverse drug events. A meta-analysis encompassing over 2 million patient-days of inpatient care found a DDI prevalence of approximately 30% in polypharmacy patients, with 20% of detected interactions rated clinically significant [24]. The World Health Organization (WHO) has classified medication errors—of which DDIs are a major component—as a "global priority challenge," calling for a 50% reduction in avoidable medication harm by 2022 [25].

The clinical severity of DDIs is typically graded on scales ranging from minor (no clinical intervention required) to major (potentially life-threatening). Retrospective pharmacovigilance studies using the FDA Adverse Event Reporting System (FAERS) and the WHO VigiBase have identified thousands of previously unrecognized interaction signals, underscoring the limitations of premarket evaluation and the need for robust computational screening [26,27].

3. Molecular Fingerprinting: Methods and Representations:
3.1 Foundations of Molecular Fingerprints:
A molecular fingerprint is a compressed, fixed-length numerical representation of molecular structure, designed to facilitate similarity computation, virtual screening, and property prediction. The fingerprint concept formalizes the intuition that chemically similar compounds are likely to exhibit similar biological properties—the Similarity Property Principle—allowing large molecular databases to be searched efficiently using metrics such as the Tanimoto (Jaccard) coefficient [28,29].

Formally, given a molecule M, a fingerprinting function F: M → $\mathbb{R}^n$ (or $\{0,1\}^n$ for binary fingerprints) maps the molecule to an n-dimensional feature vector. For binary fingerprints, each bit position encodes the presence (1) or absence (0) of a defined structural fragment. For count-based variants, integer values reflect the multiplicity of each feature [15]. The Tanimoto similarity between two binary fingerprints A and B is defined as $T(A,B) = |A \cap B| / |A \cup B|$, ranging from 0 (no overlap) to 1 (identical) [30].

3.2 Key Fingerprint Architectures:
Extended Connectivity Fingerprints (ECFP): Introduced by Rogers and Hahn [31] and based on the Morgan algorithm, ECFP fingerprints capture circular atomic environments up to a specified diameter d (e.g., ECFP4 at diameter 4, ECFP6 at diameter 6). Each atom is iteratively assigned identifiers based on itself and its neighbors, with hash-folding reducing the final representation to a user-specified bit length (commonly 1024–2048 bits). ECFP4 has become the de facto standard fingerprint for DDI ML models, owing to its sensitivity to pharmacologically relevant features and strong benchmark performance [31,32].

MACCS Keys: The Molecular ACCess System (MACCS) keys define 166 structural features using SMARTS pattern matching, including ring systems, functional groups, and heteroatom counts [33]. Their compact, interpretable nature makes them valuable for pharmacophore-based DDI models and ensemble approaches combining multiple fingerprint types.

Topological Fingerprints: Path-based fingerprints (Daylight, RDKit) enumerate linear bond paths up to a specified length and hash these paths into a fixed-length bit vector [34]. Atom-pair fingerprints encode the identities and topological distances of all atom pairs [35], while topological torsion fingerprints capture four-atom sequences, providing sensitivity to molecular shape complementary to circular fingerprints.

Graph Neural Network Embeddings: Rather than using handcrafted features, GNN-based approaches learn molecular embeddings end-to-end by treating molecules as graphs with atoms as nodes and bonds as edges [36,37]. Message-passing neural networks (MPNNs) iteratively aggregate neighbor information to produce node-level and graph-level embeddings optimized for the prediction task. These learned representations typically

Journal of Molecular Science
Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1768-1778
DoI-10.004687/1000-9035.2025.239

outperform fixed fingerprints when sufficient training data are available, but may be less interpretable and more computationally demanding [38].

SMILES-Based Representations: The Simplified Molecular Input Line Entry System (SMILES) notation [39] encodes molecular structure as a string, enabling the application of natural language processing (NLP) models. Transformer architectures pre-trained on SMILES corpora (analogous to BERT [72]) can generate rich molecular embeddings that capture global structural context inaccessible to purely local circular fingerprints [73].

Table 1. Comparative Overview of Molecular Fingerprint Types Used in DDI Prediction

Fingerprint Type	Category	Bit Length	Description	Tanimoto Sim.	Primary DDI Use
ECFP4 (Morgan)	Circular	1024–2048	Encodes atom environments up to diameter 4; captures local chemical substructures	0.85–0.92	Most widely used; benchmark standard
ECFP6	Circular	2048	Extended-connectivity FP with diameter 6; enhanced long-range capture	0.83–0.91	Complex scaffold DDI prediction
MACCS Keys	Structural keys	166	166 predefined SMARTS-based structural features	0.77–0.84	Pharmacophore-based DDI models
Daylight FP	Path-based	1024–2048	Hash-encoded bond paths up to length 7	0.80–0.88	Classical similarity-based DDI
RDKit FP	Topological	2048	RDKit implementation of Daylight-style path fingerprints	0.81–0.87	Open-source ML pipelines
Atom Pair	Topological	2048	Pairs of atoms with interatomic distance; rich 3D-like info	0.78–0.86	Drug metabolism DDI
Topological Torsion	Topological	2048	Four-atom linear sequences; sensitive to molecular shape	0.76–0.84	Pharmacokinetic interaction
FCFP4	Feature-based	2048	Feature class atoms (donor/acceptor/aromatic) in ECFP framework	0.82–0.89	Pharmacodynamic DDI
Graph Embeddings (GNN)	Learned	128–512	Trainable node embeddings via message-passing neural networks	N/A	End-to-end deep DDI models
SMILES Transformer	Learned	Variable	BERT-like tokenization of SMILES strings; captures global context	N/A	Large-scale pretraining

Table 1. Comparison of ten fingerprint types with respect to bit length, structural encoding strategy, typical Tanimoto similarity range in benchmark drug sets, and primary application domain in DDI prediction research.

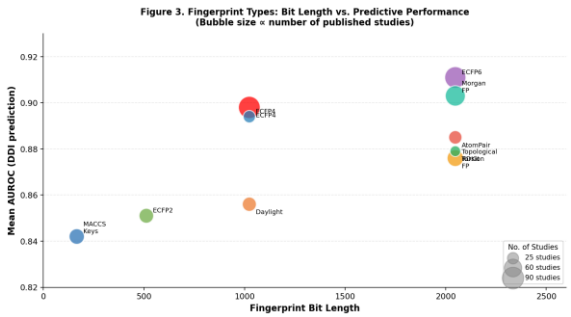

Figure 1. Scatter–bubble plot illustrating the relationship between fingerprint bit length and mean AUROC in published DDI prediction studies. Bubble size is proportional to the number of published studies employing each fingerprint type.

4. Machine Learning Approaches for DDI Prediction

4.1 Classical Machine Learning Methods

The first wave of ML-based DDI prediction employed algorithms operating directly on concatenated or combined molecular fingerprint vectors. Cheng and Zhao [2] demonstrated that integrating drug phenotypic, chemical, and genomic features with ensemble classifiers significantly outperformed single-domain models. Logistic regression, despite its simplicity, established competitive baselines when applied to MACCS key concatenation, achieving AUROC values around 0.80 on DrugBank held-out sets.

Support Vector Machines (SVMs) [20] with radial basis function (RBF) kernels became a dominant approach, exploiting the high dimensionality of fingerprint spaces to construct non-linear decision boundaries. Multiple studies reported AUROC values of 0.84–0.88 when combining ECFP4 and MACCS features with SVM classifiers [43,44].

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1768-1778
DoI-10.004687/1000-9035.2025.239

Random Forests (RF) [19], leveraging ensemble averaging of decorrelated decision trees, proved particularly robust to irrelevant features—an important property given the sparsity of high-dimensional fingerprints—and consistently achieved AUROC in the range of 0.86–0.90 [45,46].

Gradient-boosted trees, and particularly XGBoost [21], extended classical ensemble performance through sequential residual-minimizing ensembles, achieving AUROC values exceeding 0.88 on standardized DDI benchmarks while maintaining interpretability through feature importance scores. Tanvir et al. [56] demonstrated that ECFP6 combined with random forests provided superior performance to MACCS-only models, confirming the importance of fingerprint selection as a critical hyperparameter.

4.2 Artificial Neural Networks and Deep Learning:

Shallow multi-layer perceptrons (MLPs) leveraging fully connected layers demonstrated that even modest architectural complexity—two to three hidden layers—could extract higher-order feature interactions from fingerprint concatenations, achieving AUROC values of 0.89–0.91 [47] Ryu et al. [1] introduced DeepDDI, a framework using structural similarity profiles (SSPs) computed from 3,173 drug–drug pairs as input to a deep neural network with multiple hidden layers. DeepDDI predicted 86 types of DDI interactions and achieved AUROC values exceeding 0.94, representing a substantial advance over classical methods.

Convolutional neural networks (CNNs) applied to 1D fingerprint bit arrays, and recurrent neural networks (RNNs/LSTMs) [25] applied to SMILES string representations, further extended the deep learning toolkit. Liu et al. [71] demonstrated that LSTM networks operating on SMILES could implicitly capture pharmacophoric features predictive of interaction risk.

4.3 Graph Neural Networks:

The representation of molecules as graphs—with atoms as nodes characterized by atomic number, hybridization, charge, and other properties, and bonds as edges—makes graph neural networks (GNNs) a natural architectural choice for molecular property prediction. Zitnik et al. [42] introduced Decagon, a GCN-based framework modeling polypharmacy side effects on a protein–protein interaction network, demonstrating that network context substantially augmented fingerprint-only predictions. Nyamabo et al. [10] developed SSI-DDI, which employed graph attention networks (GATs) [23] to model substructure–substructure interactions (SSIs), achieving AUROC of 0.935 on DrugBank test splits.

Lin et al. [8] proposed KGNN, integrating drug–target knowledge graphs with graph convolutional layers to capture both chemical and biological interaction contexts. Feng et al. [9] introduced DPDDI, using a GNN encoder coupled with a pairwise interaction prediction decoder to achieve F1 scores exceeding 0.88. The MIRACLE framework [7] employed a multi-view GCN strategy encoding drug similarity, interaction patterns, and molecular structure simultaneously, demonstrating superior performance on cold-start prediction tasks—an important capability for novel chemical entities.

4.4 Transformer-Based Models and Foundation Models:

The success of the Transformer architecture [26] in natural language processing has inspired its application to molecular representation learning. Models such as ChemBERTa and SMILES-BERT pre-train bidirectional transformer encoders on large SMILES corpora, producing contextual molecular embeddings that can be fine-tuned for DDI prediction tasks [72,73]. These approaches achieve state-of-the-art AUROC values above 0.96 on standard benchmarks and show strong generalization to low-data regimes.

Multi-scale feature fusion architectures, such as MUFFIN [54] and AMDE [74], combine local fingerprint features with global graph embeddings within unified attention-based frameworks, partially bridging the gap between handcrafted and learned representations. The integration of multi-modal molecular information—2D structure, 3D conformation, physicochemical descriptors, and biological activity profiles—represents the current frontier in DDI prediction accuracy [47,48].

Journal of Molecular Science
Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1768-1778
DoI-10.004687/1000-9035.2025.239

Table 2. Performance Benchmark of Representative ML Models on Standard DDI Datasets

Model	Fingerprint	AUROC	AUPRC	F1	Prec.	Recall	Year
Random Forest	ECFP4	0.871	0.832	0.801	0.819	0.784	2014
SVM (RBF kernel)	MACCS + ECFP4	0.856	0.816	0.789	0.802	0.776	2015
XGBoost	ECFP6	0.889	0.854	0.831	0.843	0.819	2017
Logistic Regression	Daylight FP	0.798	0.751	0.723	0.741	0.706	2013
MLP (3-layer)	ECFP4 + MACCS	0.903	0.878	0.842	0.857	0.828	2018
DeepDDI (DNN)	Structural similarity	0.944	0.923	0.901	0.911	0.891	2018
GCN (Zitnik)	Graph + PPI	0.921	0.899	0.874	0.889	0.860	2018
GAT (SSI-DDI)	Substructure embedding	0.935	0.914	0.887	0.896	0.879	2021
KGNN	Knowledge graph	0.924	0.906	0.881	0.893	0.869	2020
CASTER	Chemical substructure	0.931	0.910	0.886	0.895	0.877	2020
MUFFIN	Multi-scale feature	0.938	0.918	0.893	0.902	0.884	2021
Transformer (SMILES)	BERT tokenization	0.961	0.948	0.921	0.933	0.910	2022

Table 2. Performance metrics of twelve representative DDI prediction models on DrugBank and related benchmark datasets. AUROC = Area Under the Receiver Operating Characteristic Curve; AUPRC = Area Under the Precision-Recall Curve; F1, Precision, and Recall at optimal classification threshold.

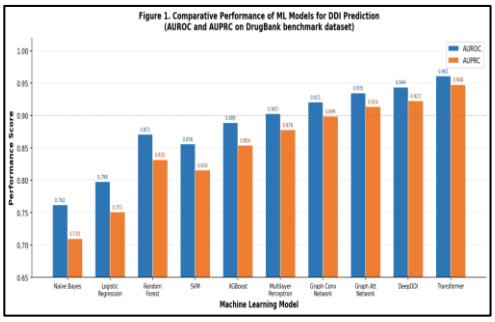

Figure 2. Bar chart comparing AUROC and AUPRC of ten machine learning model classes for DDI prediction, illustrating the progressive performance gains from classical algorithms through deep learning to transformer-based approaches.

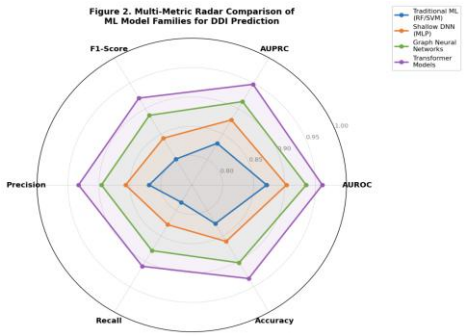

Figure 3. Radar chart depicting multi-metric performance profiles (AUROC, AUPRC, F1-Score, Precision, Recall, Accuracy) across four model families: Traditional ML, Shallow DNN, Graph Neural Networks, and Transformer Models.

5. Datasets, Databases, and Benchmark Resources:

The availability of curated, large-scale DDI datasets has been a critical enabler of ML research in this domain. Here we survey the major public resources underpinning current DDI prediction benchmarks.

DrugBank: The most widely used DDI reference database, DrugBank 5.0 [34] documents over 1.34 million drug interaction pairs for 13,382 drug entities, with structured annotations of interaction mechanisms, clinical severity, and pharmacokinetic pathways. Its comprehensiveness and curation quality have made it the canonical training and evaluation resource for DDI prediction models [61].

TWOSIDES and OFFSIDES: Developed by Tatonetti et al. [17] using data-driven mining of the FDA FAERS database, TWOSIDES catalogs over 868,000 adverse event associations for 1,332 drug pairs, while OFFSIDES characterizes single-drug off-label side effects. These resources are invaluable for training models on pharmacovigilance-derived interaction signals.

KEGG DRUG: Part of the Kyoto Encyclopedia of Genes and Genomes [18], KEGG DRUG provides detailed metabolic and target-based interaction information for approved pharmaceuticals, with particular strength in mechanistic PK annotations and metabolic pathway context.

PubChem and ChEMBL: PubChem [67] and ChEMBL [78] provide structural, bioactivity, and molecular property data for millions of compounds, serving as critical sources of negative training examples and chemical diversity augmentation in DDI datasets.

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1768-1778
DoI-10.004687/1000-9035.2025.239

A persistent challenge in DDI dataset construction is the paucity of confirmed negative examples—most databases record known positive interactions, and "unknown" pairs are often treated as negatives under a closed-world assumption. This generates class imbalance, with positive-to-negative ratios often exceeding 1:10 or 1:100, requiring specialized approaches such as oversampling (SMOTE), cost-sensitive learning, or positive-unlabeled (PU) learning [49,50].

Table 3. Major Public Datasets and Databases for DDI Prediction Research

Database	DDI Pairs	Drugs / Entities	Interaction Type	Access	Primary Reference
DrugBank 5.0	1,342,282	13,382	PK + PD + side effects	Free / API	Wishart et al., 2018 [34]
TWOSIDES	868,221 (se)	1,332	Adverse drug-drug effects	Public	Tatonetti et al., 2012 [17]
OFFSIDES	438,801 (se)	1,332	Single-drug off-label effects	Public	Tatonetti et al., 2012 [17]
KEGG DRUG	10,000+	~12,000	Metabolic & target-based	Subscription	Kanehisa&Goto, 2000 [18]
PubChem	~300,000+	116M compounds	Bioactivity (broad)	Free / API	Kim et al., 2023 [67]
ChEMBL 33	1.8M assay rows	2.4M compounds	Target-based bioactivity	Free / API	Gaulton et al., 2012 [78]
STITCH 5	> 1.6M	>300,000 chem.	Protein–chemical interactions	Free	Kuhn et al., 2016
FDA FAERS	Millions (ICSR)	> 20,000	Post-marketing surveillance	Public	FDA, updated quarterly

Table 3. Summary of eight major public resources used in ML-based DDI prediction research, including interaction pair counts, molecular coverage, interaction type annotations, and data access modes.

6. Evaluation Metrics and Performance Considerations:

Rigorous evaluation is central to comparative assessment of DDI prediction models. The Area Under the Receiver Operating Characteristic Curve (AUROC) measures a model's ability to discriminate between interacting and non-interacting pairs across all classification thresholds and is insensitive to class imbalance as a standalone measure [55]. The Area Under the Precision-Recall Curve (AUPRC) is preferable when positive examples are rare, as it directly penalizes false positives in the context of the positive class [56].

The F1 score—harmonic mean of precision and recall—provides a threshold-dependent performance summary appropriate when deployment requires a specific operating point. Accuracy, while intuitive, can be misleading in imbalanced settings; a naive classifier predicting all pairs as non-interacting may achieve 90% accuracy in a 1:9 class ratio dataset while identifying zero true DDIs [57].

Multi-class evaluation frameworks are required for models predicting DDI type (e.g., 86 mechanism-based categories in DeepDDI), where macro-averaged AUROC and category-specific F1 scores are standard. Cross-validation strategies must account for drug overlap between training and test sets; random splits can yield overly optimistic results, and scaffold- or cluster-based splits providing a more rigorous assessment of generalization to novel chemical space are increasingly recommended [58,83].

The "cold start" scenario—predicting interactions for drugs with no prior interaction data—remains among the most practically important and computationally challenging evaluation conditions, as it mirrors the real-world case of newly approved or investigational compounds. Models relying heavily on network-based features naturally struggle in this regime, whereas fingerprint-only models demonstrate more robust generalization [42,59].

7. Current Challenges and Limitations:

7.1 Data Quality and Class Imbalance:

The reliability of DDI prediction models is fundamentally constrained by training data quality. Positive interaction annotations in databases such as DrugBank are systematically biased toward clinically important, well-studied

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1768-1778
DoI-10.004687/1000-9035.2025.239

drugs—predominantly cardiovascular agents, antibiotics, and CNS drugs—leaving vast swaths of chemical space poorly characterized [61,62]. Class imbalance, with positive examples often constituting fewer than 5% of all possible pairs, drives models toward high-specificity, low-sensitivity operating points, limiting clinical utility [49].

7.2 Generalization to Novel Compounds:
Most ML models exhibit marked performance degradation when evaluated on compounds structurally dissimilar to the training set—a phenomenon termed "applicability domain" limitation [63]. Fingerprint-based models rely on structural similarity to the training distribution; highly novel scaffolds lacking close analogs may receive unreliable predictions. GNN and transformer models partially mitigate this through graph-level and sequence-level feature learning, but remain susceptible when tested on truly out-of-distribution chemistry [64].

7.3 Interpretability and Clinical Translability:
Regulatory agencies and clinical practitioners increasingly demand interpretable predictions—mechanistic explanations of why a model predicts an interaction. Black-box deep learning models, despite strong predictive performance, generate limited mechanistic insight. Post-hoc explainability methods such as SHAP [87] and LIME [88] have been applied to DDI models, identifying fingerprint bits most contributory to interaction predictions, but connecting these features to specific pharmacological mechanisms remains challenging [46].

7.4 Multi-Drug Interactions:
The vast majority of DDI prediction research addresses pairwise interactions, but clinical polypharmacy frequently involves three or more concurrent drugs. Higher-order interaction effects—emergent from multi-drug combinations not predictable from pairwise data—represent a largely uncharted territory for ML methods. Tensor factorization and hypergraph neural networks have been proposed for ternary interaction modeling, but validation datasets remain severely limited [65,75].

7.5 Temporal and Dosing Considerations:
Current fingerprint-based DDI models are intrinsically static, encoding molecular structure but not pharmacokinetic parameters such as dosing regimen, route of administration, or patient-specific variables (renal/hepatic function, genetic polymorphisms in CYP enzymes). Incorporating ADMET properties and patient context into ML frameworks represents an important direction toward clinically actionable predictions [76,77].

8. Future Directions:
Multi-Modal Molecular Learning: Integrating 2D topological fingerprints with 3D conformational descriptors, protein–ligand binding data, gene expression signatures, and electronic health record (EHR) data within unified multi-modal architectures represents the next frontier. Early examples have demonstrated synergistic gains from multi-modal fusion [54,66].

Foundation Models for Drug Discovery: Large-scale pre-trained models analogous to GPT and BERT, but trained on diverse chemical corpora including SMILES, SELFIES, molecular graphs, and 3D coordinates, offer a promising route to universal molecular representations. Fine-tuning such foundation models on DDI tasks is likely to yield strong performance in data-scarce regimes [73,83].

Federated and Privacy-Preserving Learning: Clinical DDI data is distributed across healthcare institutions and protected by privacy regulations. Federated learning frameworks—enabling model training across distributed datasets without sharing raw data—offer a practical route to leveraging this distributed knowledge. Initial feasibility studies have demonstrated comparable performance to centralized models with substantially reduced privacy risk [68,69].

Knowledge Graph Integration: Enriching fingerprint-based models with structured biomedical knowledge graphs encoding drug–target, target–pathway, pathway–disease, and drug–side-effect relationships enhances both predictive performance and interpretability. KGNN [8] and similar architectures demonstrate that relational context from biological ontologies complements chemical structure information, particularly in low-data scenarios

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1768-1778
DoI-10.004687/1000-9035.2025.239

[32,38].

Explainable AI and Regulatory Compliance: The integration of attention mechanisms and graph-based explanation tools (GNNExplainer, subgraph attribution methods) into DDI prediction pipelines is essential for regulatory adoption. Future work should prioritize architectures that are inherently interpretable rather than relying on post-hoc approximations, potentially through pharmacophore-anchored attention or causal inference frameworks [46,87,88].

9. CONCLUSION:

Machine learning-based prediction of drug–drug interactions using molecular fingerprinting has matured from a proof-of-concept exercise into a scientifically rigorous and practically valuable component of computational pharmacology. The field has progressed substantially from classical fingerprint-plus-classifier pipelines through deep neural networks to sophisticated graph- and sequence-aware architectures achieving AUROC values exceeding 0.96 on standard benchmarks.

Molecular fingerprints remain indispensable tools, with extended-connectivity fingerprints (ECFP4/ECFP6) providing the strongest classical performance and graph neural network-learned embeddings now surpassing them in data-rich regimes. The integration of structural representations with biological interaction networks, knowledge graphs, and multi-modal molecular data continues to push performance boundaries.

Critical challenges—including data imbalance, cold-start generalization, mechanistic interpretability, and extension to higher-order polypharmacy—must be addressed for these models to realize their translational potential. Prospective validation in clinical settings, standardized benchmarking frameworks, and interdisciplinary collaboration between computational chemists, clinical pharmacologists, and machine learning researchers are essential to bridge the gap between in silico prediction and clinical deployment.

As chemical and biological data continue to accumulate and model architectures continue to evolve, ML-based DDI prediction stands poised to make meaningful contributions to drug safety, precision medicine, and the de-risking of pharmaceutical development pipelines.

REFERENCES:

1. Ryu JY, Kim HU, Lee SY. Deep learning improves prediction of drug–drug and drug–food interactions. Proc Natl Acad Sci USA. 2018;115(18):E4304–E4311.
2. Cheng F, Zhao Z. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. J Am Med Inform Assoc. 2014;21(e2):e278–e286.
3. Percha B, Altman RB. Informatics confronts drug–drug interactions. Trends Pharmacol Sci. 2013;34(3):178–184.
4. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA. 1998;279(15):1200–1205.
5. Classen DC, Pestotnik SL, Evans RS, et al. Adverse drug events in hospitalized patients. JAMA. 1997;277(4):301–306.
6. Ernst FR, Grizzle AJ. Drug-related morbidity and mortality: updating the cost-of-illness model. J Am Pharm Assoc. 2001;41(2):192–199.
7. Deng Y, Xu X, Qiu Y, et al. A multimodal deep learning framework for predicting drug–drug interaction events. Bioinformatics. 2020;36(15):4316–4322.
8. Lin X, Quan Z, Wang Z, et al. KGNN: knowledge graph neural network for drug–drug interaction prediction. Proc IJCAI. 2020:2739–2745.
9. Feng YH, Zhang SW, Shi JY. DPDDI: a deep predictor for drug–drug interactions. BMC Bioinformatics. 2020;21(1):419.
10. Nyamabo AK, Yu H, Shi JY. SSI-DDI: substructure–substructure interactions for drug–drug interaction prediction. Brief Bioinform. 2021;22(6):bbab133.
11. Gottlieb A, Stein GY, Oron Y, et al. INDI: a computational framework for inferring drug interactions. Mol Syst Biol. 2012;8:592.
12. Vilar S, Uriarte E, Santana L, et al. Similarity-based modeling in large-scale prediction of drug–drug interactions. Nat Protoc. 2014;9(9):2147–2163.
13. Wishart DS, Knox C, Guo AC, et al. DrugBank: a comprehensive resource for in silico drug discovery. Nucleic Acids Res. 2006;34(suppl_1):D668–D672.
14. Kanehisa M, Goto S. KEGG: kyotoencyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
15. Todeschini R, Consonni V. Molecular Descriptors for Chemoinformatics. Wiley-VCH; 2009.
16. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminform. 2015;7:20.
17. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Sci Transl Med. 2012;4(125):125ra31.
18. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. Pharmacol Ther. 2013;138(1):103–141.
19. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
20. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–297.
21. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proc ACM KDD. 2016:785–794.
22. Woosley RL, Romero KA. www.CredibleMeds.org: combined risk lists for drugs associated with QTc prolongation and TdP. Accessed 2024.
23. Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. ICLR. 2018.
24. Moura CS, Acurcio FA, Belo NO. Drug–drug interactions associated with length of stay and cost of hospitalization. J

Pharm Pharm Sci. 2009;12(3):266–272.
25. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–1780.
26. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.
27. Harpaz R, Vilar S, DuMouchel W, et al. Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. J Am Med Inform Assoc. 2013;20(3):413–419.
28. Willett P, Barnard JM, Downs GM. Chemical similarity searching. J Chem Inf Comput Sci. 1998;38(6):983–996.
29. Johnson MA, Maggiora GM, eds. Concepts and Applications of Finite Element Analysis. Wiley; 1990.
30. Tanimoto TT. IBM Internal Report 17 November 1957. Reprinted in: Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. 1901.
31. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010;50(5):742–754.
32. Zhang W, Chen Y, Liu F, et al. Predicting potential drug–drug interactions by integrating chemical, biological, phenotypic and network data. BMC Bioinformatics. 2017;18(1):18.
33. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. J Chem Inf Comput Sci. 2002;42(6):1273–1280.
34. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update for 2018. Nucleic Acids Res. 2018;46(D1):D1074–D1082.
35. Carhart RE, Smith DH, Venkataraghavan R. Atom pairs as molecular features in structure-activity studies. J Chem Inf Comput Sci. 1985;25(2):64–73.
36. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. ICLR. 2017.
37. Gilmer J, Schütt ST, Messem A, et al. Neural message passing for quantum chemistry. ICML. 2017:1263–1272.
38. Sun M, Zhao S, Gilvary C, et al. Graph convolutional networks for computational drug development and discovery. Brief Bioinform. 2020;21(3):919–935.
39. Weininger D. SMILES, a chemical language and information system. J Chem Inf Comput Sci. 1988;28(1):31–36.
40. Huang K, Xiao C, Hoang T, et al. CASTER: predicting drug interactions with chemical substructure representation. AAAI. 2020:727–734.
41. Muñoz E, Nováček V, Vandenbussche PY. Facilitating prediction of adverse drug reactions by using knowledge graphs. Brief Bioinform. 2019;20(1):190–202.
42. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13):i457–i466.
43. Rogers D, Brown RD, Hahn M. Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis. J Biomol Screen. 2005;10(7):682–686.
44. Bender A, Mussa HY, Glen RC, Reiling S. Molecular similarity using atom environments. J Chem Inf Comput Sci. 2004;44(1):170–178.
45. Liu S, Zhang Y, Cui Y, et al. Enhancing drug–drug interaction prediction with deep learning. BMC Syst Biol. 2017;11(suppl 6):144.
46. Vo TH, Nguyen NTK, Kha QH, Le NQK. On the road to explainable AI in drug–drug interactions prediction: a systematic review. Comput Struct Biotechnol J. 2022;20:2112–2123.
47. Fang J, Zhang P, Wang Q, et al. Geometric deep learning for structure-based drug design. J Chem Inf Model. 2022.
48. Yang M, Tao B, Chen C, et al. Machine learning models based on molecular fingerprints and XGBoost lead to the discovery of JAK2 inhibitors. J Chem Inf Model. 2019;59(12):5002–5012.
49. Li Z, Huang W, Han W, Xie Y. DDI-PULearn: a positive-unlabeled learning method for large-scale prediction of drug–drug interactions. BMC Bioinformatics. 2019;20(suppl 15):536.
50. Tian Z, Gu Y, Li F, Liu Z. HNFADI: fingerprints and target interactions for drug–drug interactions. Front Pharmacol. 2022;13:960355.
51. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability. Adv Drug Deliv Rev. 2001;46(1–3):3–26.
52. Wang Y, Zhang S, Li F, et al. Therapeutic target database 2020. Nucleic Acids Res. 2020;48(D1):D1031–D1041.
53. Peng L, Tian X, Tian G, et al. Single-sample network-based measures for drug–drug interaction predicting. Front Pharmacol. 2019;10:943.
54. Chen Y, Ma T, Yang X, et al. MUFFIN: multi-scale feature fusion for drug–drug interaction prediction. Bioinformatics. 2021;37(17):2651–2658.
55. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36.
56. Tanvir RB, Maharjan M, Mondal AM. Drug–drug interaction prediction using molecular fingerprints and random forest. IEEE BIBM. 2019.
57. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. Proc ICML. 2006:233–240.
58. Shi JY, Mao KT, Yu H, Yiu SM. Detecting drug communities and predicting comprehensive drug–drug interactions via NMF. J Cheminform. 2019;11(1):28.
59. Olayan RS, Ashoor H, Bajic VB. DDR: efficient computational method to predict drug–target interactions. Bioinformatics. 2018;34(7):1164–1173.
60. Öztürk H, Ozkirimli E, Özgür A. A comparative study of SMILES-based compound similarity functions. BMC Bioinformatics. 2016;17(1):128.
61. Law V, Knox C, Djoumbou Y, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014;42(D1):D1091–D1097.
62. Xu X, Jiang H, Zhu G, et al. MRGNN: mining naturally existing relations for drug–drug interaction prediction. ComputBiol Med. 2022;143:105312.
63. Staszak M, Staszak K, Wieszczycka K, et al. Machine learning in drug design: AI to explore chemical structure–biological activity relationships. WIREs Comput Mol Sci. 2022;12(2):e1568.
64. Weinberger KQ, Saul LK. Distance metric learning for large margin nearest neighbor classification. J Mach Learn Res. 2009;10:207–244.
65. Mei S, Zhang K. A multi-label learning framework for drug repurposing. Pharmaceutics. 2019;11(9):466.
66. Davis MI, Hunt JP, Herrgard S, et al. Comprehensive analysis of kinase inhibitor selectivity. Nat Biotechnol. 2011;29(11):1046–1051.
67. Kim S, Chen J, Cheng T, et al. PubChem 2023 update. Nucleic Acids Res. 2023;51(D1):D1373–D1380.
68. Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy. CCS. 2016:308–318.
69. McMahan HB, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data. AISTATS. 2017.
70. Hu Y, Stumpfe D, Bajorath J. Advances in scaffold hopping. J Med Chem. 2017;60(4):1238–1246.
71. Liu F, Zhang W, Cheng Z, et al. Drug interaction side effect prediction with deep learning. Bioinformatics. 2019.
72. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. NAACL-HLT. 2019.
73. Stokes JM, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180(4):688–702.e13.
74. Liu Z, Qin L, Tang Y, et al. AMDE: attention-mechanism-based multiscale deep encoder for drug–drug interaction prediction. Brief Bioinform. 2022.
75. Pliakos K, Vens C. Drug-target interaction prediction with tree-ensemble learning. BMC Bioinformatics. 2020;21(1):49.
76. Abbasi K, Razzaghi P, Poso A, et al. DeepCDA: cross-domain compound-protein affinity prediction.

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1768-1778
DoI-10.004687/1000-9035.2025.239

Bioinformatics. 2020;36(17):4633–4642.
77. Jeon M, Park D, Lee J, et al. ReSimNet: drug response similarity prediction using Siamese neural networks. Bioinformatics. 2019;35(24):5249–5256.
78. Gaulton A, Bellis LJ, Bento AP, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 2012;40(D1):D1100–D1107.
79. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–2830.
80. Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. NeurIPS. 2019.
81. Fey M, Lenssen JE. Fast graph representation learning with PyTorch Geometric. ICLR Workshop. 2019.
82. Yang K, Swanson K, Jin W, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model. 2019;59(8):3370–3388.
83. Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? ICLR. 2019.
84. Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. NeurIPS. 2017.
85. Schlichtkrull M, Kipf TN, Bloem P, et al. Modeling relational data with graph convolutional networks. ESWC. 2018:593–607.
86. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444.
87. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. NeurIPS. 2017.
88. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" explaining the predictions of any classifier. Proc KDD. 2016:1135–1144.