ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12, Issue 6, June 2025

Drug-Target and Drug-Drug Interaction Prediction using Knowledge Graph and Graph Neural Network

[1] V. Vijaya Bhaskar Reddy, [2] T. Priyadharshini, [3] T. Amrutha Varsha, [4] Arshiya Mehnaz, [5] V. Bhavana

[1] CSE Department, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology (VNRVJIET), Hyderabad, Telangana, India

[2] [3] [4] [5] B.Tech CSE, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology (VNRVJIET), Hyderabad, Telangana, India

Corresponding Author Email: [1] vijayabhaskarareddy_v@vnrvjiet.in, [2] tanetipriyadharshini@gmail.com, [3] amruthatarigopula002@gmail.com, [4] 2004arsh@gmail.com, [5] vaddepallibhavana@gmail.com

Abstract*— Understanding drug-drug and drug-target interactions is crucial for developing safe and effective therapeutic strategies. With the rapid growth of pharmaceutical research, traditional methods for identifying these interactions, which heavily rely on experimental data, are often inefficient, costly, and difficult to scale. These limitations can hinder the discovery of new drug candidates and the prediction of harmful drug combinations. Despite advancements in computational techniques, there are still significant challenges in accurately predicting interactions due to the complexity of integrating vast biological datasets. Leveraging advanced data mining, machine learning, and bioinformatics approaches can improve the prediction of DDIs and DTIs. Ultimately, these advancements will help optimize drug development processes and ensure safer, more effective treatments. This paper represents the ways to bridge gaps in current computational techniques by providing insights into safer drug combinations and facilitating personalized medicine, ultimately optimizing the drug development process and contributing to more effective therapeutic strategies.

Index Terms*— Attention Mechanism, Machine Learning, Multi-layer Framework, Pharmaceutical research.

I. INTRODUCTION

The study of drug-drug interactions (DDIs) and drug-target interactions (DTIs) has become increasingly important in the rapidly evolving landscape of pharmaceutical research. These interactions play a pivotal role in ensuring the safety and efficacy of therapeutic applications. DDIs occur when multiple drugs interact, potentially leading to reduced efficacy or harmful side effects, while DTIs focus on how drugs interact with specific biological targets, such as proteins and receptors. Traditional methods for identifying these interactions rely heavily on experimental techniques, which, while reliable, can be inefficient, costly, and challenging to scale. This reliance can result in missed opportunities for discovering new drug candidates and predicting harmful interactions.

To overcome these limitations, there is a growing need for advanced computational approaches that leverage data mining, machine learning, and bioinformatics. These methods can efficiently analyze large and complex biological datasets, integrating diverse information such as chemical structures and pharmacokinetic properties. By focusing on various biological targets, including enzymes and GPCRs, computational models can facilitate drug repurposing, enhance patient safety, and support personalized medicine by tailoring treatments to individual profiles. Ultimately, these innovations have the potential to streamline drug development processes and improve therapeutic outcomes, making significant strides in the field of medicine.

A. Related Work

The framework used in [18,21] utilizes Graph Neural Networks, machine learning, and meta-learning techniques like MAML for zero-shot drug-target interaction predictions. Future research should investigate additional data sources and applications in other biological tasks. The research done in [20,23] combines self-supervised learning and Graph Attention Network techniques to enhance drug-target interaction predictions with limited labeled data. Recommendations include integrating more biological features and improving model adaptability. The model presented in [2,22] utilizes a weighted heterogeneous multi-layer network and enhanced random walk algorithms to predict candidate drugs for aging diseases. Future enhancements should integrate disease-specific information and non-genetic factors. The study in [1,5,14,24] employs a hybrid model combining Similarity Network Fusion and Hybrid Convolutional Neural Networks. Future research should focus on multi-drug interactions and clinical data integration. The framework in [2,3,9,26] uses Drug Association and Drug Substructure encoders, along with an attention mechanism. Future research should refine feature extraction methods for better prediction accuracy. The framework in [25] utilizes adaptive graph structures, attention mechanisms, and pseudo-label supervision to enhance drug-target interaction predictions. Future research should aim for end-to-end learning frameworks to improve computational efficiency.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12, Issue 6, June 2025

B. Organization of the Paper

The rest of the paper is organized as follows: Section 2 presents the methodology for Drug-Target and Drug-Drug Prediction. In section 3, we have discussed the research directions and challenges. Details of the DDI and DTI applications are discussed in section 4. In section 5, we have concluded our survey research and the respective references are given in section 6.

II. DRUG-TARGET AND DRUG-DRUG INTERACTIONS PREDICTION TECHNIQUES

Various drug-target and drug-drug interaction prediction techniques exist in literature are summarized as below:

A. Graph Neural Network approaches for DTI prediction

a. DTI-MvSCA Framework for DTI Prediction

In 2024, Lihong Peng. al. [20] introduced a multi-view neural network framework called DTI-MvSCA to improve drug-target interaction (DTI) prediction. This framework addresses challenges in DTI prediction, such as limited negative DTI samples and the limitations of current graph neural networks (GNNs). DTI-MvSCA leverages graph topological feature learning using a self-attention mechanism and the SHADOW graph attention network. Node features are also learned using 1D convolutional neural networks (1D-CNN), enhancing DTI prediction accuracy. The model was validated on DrugBank datasets and showed improved performance over existing methods. The paper emphasizes the importance of advanced DTI prediction techniques for drug discovery and cost reduction, proposing DTI-MvSCA as a more precise method due to its anti-over-smoothing approach and multi-view architecture.

b. Multi-layer Graph Neural Network

In 2024, Yuansheng Liu et al. propose SSR-DTA, a multi-layer graph neural network designed to enhance DTA prediction accuracy by capturing complex substructural details in drug molecules and target proteins. SSR-DTA incorporates the BiGNN module to simultaneously process sequence-based and structure-based features, effectively handling protein data. Additionally, the SSA-Fusion module is introduced to capture intricate substructure relationships, further improving prediction accuracy. SSR-DTA outperforms previous models on benchmark datasets (Davis, KIBA) with up to a 20% reduction in mean squared error, showcasing its potential as a robust tool for DTA prediction in drug discovery.

B. Deep Learning Approach for DTI Prediction

In 2024, Hao Zhang et al.[5] proposed a deep learning-based approach to predict drug-target binding affinity(DTA). Drug-target binding affinity plays a crucial role in drug repurposing, aimed at identifying potential drug-target pairs with high binding affinity. DTA prediction is pivotal in understanding drug mechanisms and facilitating drug repurposing, which helps streamline the drug repurposing process. Various DL models, including Convolutional Neural Networks (CNNs), Graph Convolutional Networks (GCNs), and Recurrent Neural Networks (RNNs), are utilized to model DTA, transitioning away from traditional machine learning models. The paper discusses the types of input representations for DL models, which include SMILES strings to represent drug compounds and amino acid sequences for protein targets. Some models also integrate the three-dimensional structure of drug molecules, using graph-based structures for detailed representation of atomic interactions and physicochemical properties.

C. Multi-Head Attention and Heterogenous Attribute Graph Learning Approach

In 2024, Lin-Xuan Hou et. al.[5] proposed the MathEagle, which is a framework designed to improve predictions of drug-drug interaction events (DDIs) by leveraging multi-head attention and heterogeneous attribute graph learning. This approach constructs a heterogeneous information graph, which captures both chemical properties and interaction histories of drugs through various node and relationship types. MathEagle's methodology combines message passing neural networks (MPNNs) and heterogeneous graph convolution, allowing drugs to share and refine information across different types of connections.

The framework's multi-head attention mechanism assigns varying importance to features and relationships, enabling MathEagle to focus on the most relevant signals for interaction prediction. By using graph embedding algorithms, the model transforms high-dimensional attributes into efficient low-dimensional representations, making it possible to predict both the likelihood and risk levels of interactions accurately. In benchmark testing, MathEagle demonstrated high accuracy and Area Under Curve (AUC) scores, outperforming comparative models. This capability makes it a promising tool for clinicians and researchers in identifying potential risks in drug combinations, supporting safer and more informed decision-making in drug therapy management.

D. Ensemble Based Approach

In 2022, Heba El-Beheryet al.[19] proposed a ensemble-based drug–target interaction (DTI) prediction approach that incorporates multiple features and addresses data imbalance. To represent drugs and proteins accurately, specific feature extraction techniques are applied. Drug characteristics are represented using Morgan fingerprints and constitutional descriptors. Morgan fingerprints provide a binary digital sequence based on each drug's molecular structure, capturing essential chemical information. In contrast, constitutional descriptors summarize molecular

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12, Issue 6, June 2025

properties without spatial or geometric data, offering a simpler, yet informative view of the drug's characteristics.

For protein feature extraction, two main descriptors are used: amino acid composition (AAC) and dipeptide composition (DC). AAC is calculated by determining the frequency of each amino acid type within a protein sequence, while DC represents the frequency of all possible two-residue combinations in the sequence. Together, these features provide a detailed molecular representation of the protein's sequence composition and local structural patterns, which are essential for understanding binding interactions. To address the challenge of data imbalance, the methodology uses a one-class Support Vector Machine (SVM) classifier for negative sample prediction. The classifier constructs a hyperplane around positive interactions, estimating distances for unknown interactions. These distances are used to identify likely negative interactions, which are included as balanced inputs to the prediction model. This step is crucial because DTI datasets often contain significantly more unknown or negative interactions than positive ones.

Finally, ensemble machine learning algorithms—such as Random Forest (RF), AdaBoost, XGBoost, and Light Boost—are employed to classify potential DTIs. The prediction algorithms are evaluated using a 10-fold cross-validation process to ensure robustness. Combining feature-rich drug and protein representations with data balancing techniques and advanced machine learning, improving accuracy in DTI prediction.

E. Knowledge Graph Based Approach

a. Knowledge Subgraph Learning

In 2024, Yaqing Wang et al.[17] proposed methodology for predicting drug-drug interactions (DDIs) combines drug-pair and biomedical knowledge graph data to generate interpretable predictions. First, the DDI and external knowledge graphs (KGs) are merged, where drugs, genes, and interactions form nodes and edges. Each node is initially represented by a generic embedding learned from this combined graph. For each target drug pair (head drug h and tail drug t), a "drug-flow subgraph" captures relevant context from the combined network by tracing relational paths up to a set length.

KnowDDI, a method within this framework, iteratively optimizes the drug-pair-specific subgraph to yield a "knowledge subgraph" by removing low-importance edges and adding new connections between similar nodes.KnowDDI, a method within this framework, iteratively optimizes the drug-pair-specific subgraph to yield a "knowledge subgraph" by removing low-importance edges and adding new connections between similar nodes. The connection strength, assesses the edge relevance within each drug-pair subgraph, retaining only paths that explain potential interactions. The method then predicts the DDI type through a classification function that leverages the refined embeddings for h and t, as well as the subgraph's overall structure. This approach maximizes interpretability by preserving only paths relevant to DDIs, facilitating identification of previously unknown interactions ning only paths that explain potential interactions. The method then predicts the DDI type through a classification function that leverages the refined embeddings for h and t, as well as the subgraph's overall structure. This approach maximizes interpretability by preserving only paths relevant to DDIs, facilitating identification of previously unknown interactions.

b. Knowledge Graph Embeddings (KGE) and Protein Sequence Pretraining (ProtBERT)

In 2023, Warith Eddine Djeddi et al.[16] introduced a methodology for predicting drug-target interactions (DTIs) combines knowledge graph embeddings (KGE) and protein sequence pretraining (ProtBERT) to integrate structural and sequence data. A heterogeneous graph is built with drugs and targets as nodes, linked by interactions, and KGE methods like DistMult encode these relationships in continuous vector space. To capture detailed biochemical context, ProtBERT generates embeddings from protein sequences, while SMILES data for drugs provides structural information as fingerprints. Protein sequences are also represented by biochemical properties (non-polar, polar, acidic, or basic). Drug-drug and target-target similarity matrices are calculated using cosine similarity or Euclidean distance. These contextual (ProtBERT) and structural (SMILES, biochemical) embeddings are combined for each drug-target pair and input into classifiers like Random Forest or Gradient Boosting to predict interactions. This approach enhances DTI prediction accuracy and interpretability, supporting drug discovery and repurposing efforts.

F. Matrix Factorization Approach

a. Graph Regularized Probabilistic Matrix Factorization Approach

In 2024, Emilie Chouzenoux et al.[13] introduced methodology involves using a Graph Regularized Probabilistic Matrix Factorization (GRPMF) approach to predict Drug-Drug Interactions (DDIs). This matrix completion framework targets completing a symmetric drug interaction matrix Y, where known interactions are initially set, and unobserved entries are masked as zeros. GRPMF operates by factorizing this interaction matrix into latent factors while incorporating graph-based regularization. Expert knowledge is encoded in a matrix, representing drug similarity scores derived from chemical structure comparisons.

The matrix factorization involves iterative optimization to minimize the error between predicted and actual interactions, regularized by constraints from the similarity graph. Hyperparameters for controlling regularization (like λR and λU) are tuned to enhance model performance. The algorithm iterates through steps for updating matrices and ensuring

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12, Issue 6, June 2025

graph-based constraints, ultimately estimating a complete matrix that predicts both observed and unobserved DDIs. This approach integrates drug-specific information to improve interpretability and accuracy in DDI prediction, setting GRPMF apart from generic matrix completion models.

b. Constrained Tensor Factorization Approach

In 2024, Guosheng Han et al.[7] proposed CTF-DDI model involves three main modules: similarity fusion, constrained tensor factorization (CTF), and deep neural networks (DNN). First, similarity matrices are generated to capture diverse relationships among drugs. These include chemical, ligand-based, side-effect, anatomical, therapeutic, and biological structure similarities from drug datasets. These matrices are then integrated using weighted averaging to create comprehensive drug similarity profiles.

In the CTF module, these similarity matrices serve as constraints in a CP tensor factorization model to reduce data sparsity and improve prediction accuracy. This factorization, enhanced by Hessian and L{2,1} regularizations, produces feature matrices for drugs and drug interactions. Finally, these matrices are fed into a DNN, which uses a fully connected network with ReLU activation to extract nonlinear features. The DNN predicts potential drug-drug interactions (DDIs) with high accuracy by learning complex patterns within the feature data, yielding a final tensor representing interaction types and probabilities.

III. RESEARCH DIRECTIONS AND CHALLENGES

The rapid growth in biomedical data, driven by advances in high-throughput screening, genomics, and electronic health records, presents unprecedented opportunities and challenges for Drug-Target Interaction (DTI) and Drug-Drug Interaction (DDI) prediction. As pharmaceutical research increasingly adopts data-driven methodologies, DTI and DDI predictions have become vital tools in drug discovery, repurposing, and personalized medicine. Some of the research challenges are given:

A. Improving Efficiency and Scalability for Large-Scale Prediction

Developing scalable and efficient models for Drug-Target Interaction (DTI) and Drug-Drug Interaction (DDI) prediction presents a significant challenge due to the computational resources required for processing large biomedical datasets. High-throughput screening, molecular simulations, and complex graph-based models often generate vast amounts of data, which can be difficult to handle with standard computing resources. Techniques like parallel processing, distributed computing, and optimized graph algorithms can help mitigate this issue by splitting workloads across multiple processors or nodes. Additionally, advanced memory management and efficient data retrieval mechanisms can reduce latency, especially for real-time applications. These optimizations are particularly important when handling large molecular databases or datasets that include detailed structural and interaction information, ensuring that predictions remain practical for large-scale or high-frequency usage.

B. Data Quality and Availability

The availability of high-quality and comprehensive datasets is crucial for reliable DTI and DDI prediction, but the biomedical field often suffers from data limitations, including gaps and inconsistencies. Many existing datasets lack verified negative samples, which are essential for effective training, and often have imbalances that skew model performance toward over-represented classes. This scarcity of well-annotated, negative interaction data complicates the creation of robust models and can introduce bias, leading to unreliable predictions. Addressing this issue requires expanding open-access databases and improving data annotation methods to capture a wider array of interactions and experimental outcomes. Collaboration among researchers to create more standardized, well-balanced datasets, along with the use of synthetic data generation techniques, can further improve prediction accuracy and generalizability.

C. Incorporating Temporal Data

Drug interactions can vary over time due to factors like drug dosage and metabolic changes. Exploring time-series or longitudinal data analysis for DDI and DTI prediction could improve predictions for dynamic interaction profiles, especially in chronic treatments.

D. Enhancing Interpretability through Domain Knowledge Integration

Enhancing the interpretability of Drug-Target Interaction (DTI) and Drug-Drug Interaction (DDI) prediction models is crucial for their application in clinical and research contexts. Incorporating pharmacological and biochemical domain knowledge can clarify predictions by relating them to established biological processes, such as drug metabolic pathways and target binding sites. This integration can improve model accuracy by focusing on biologically relevant patterns during data preprocessing and feature selection. Additionally, frameworks that facilitate expert feedback can ensure predictions align with current scientific understanding, making models more trustworthy and actionable for clinicians and researchers.

E. Overcoming Data Sparsity and Imbalance

DTI and DDI datasets are often sparse, with significantly fewer positive interaction samples compared to unknown or negative ones. Addressing this imbalance is essential, as biased training data can reduce model accuracy. Developing methods to handle sparse data, such as synthetic data generation, semi-supervised learning, or active learning, could enhance prediction models' performance.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12, Issue 6, June 2025

IV. APPLICATIONS

Drug-Drug Interaction (DDI) and Drug-Target Interaction (DTI) prediction has numerous applications across scientific research and the healthcare industry. These applications are as follows:

A. Drug Repurposing

DTI predictions can identify potential new uses for existing drugs, accelerating the repurposing process and reducing development time and costs. This approach leverages known drug-target associations to discover alternative therapeutic targets, particularly for complex or rare diseases.

B. Adverse Drug Reaction (ADR) Prevention

By predicting interactions, DDI tools assist in identifying combinations that might cause adverse reactions. This is critical in multi-drug therapies, especially for chronic conditions, where complex drug combinations are common.

C. Drug Discovery and Development

DTI prediction assists pharmaceutical research in narrowing down drug candidates, focusing on those most likely to interact with desired targets. This accelerates the initial phases of drug discovery, allowing researchers to prioritize compounds with higher efficacy potential.

D. Healthcare and Clinical Decision Support

Integration of DDI prediction models in healthcare systems aids clinicians in real-time decision-making. These models help in selecting safer drug combinations, thereby enhancing patient safety and improving overall treatment outcomes.

E. Precision in Polypharmacy Management

In treating patients with multiple medications, such as the elderly, DDI prediction models are essential for assessing polypharmacy risks. This application is vital for maintaining therapeutic efficacy while minimizing toxicological risks in complex treatment plans.

V. CONCLUSION

In this Paper, we presented a comprehensive approach toward drug target prediction and analysis of drug-drug interaction based on advanced data processing for pharmaceutical research. We studied complex interactions within drug networks using predictive models on large datasets, providing a more accurate identification of the target and prediction of the interaction. Our framework tackles major challenges of data integration handling diverse biological datasets to yield insights useful for drug discovery and repurposing. Future work may focus on improvements in accuracy and scalability, and it may explore novel approaches toward unifying different datasets in support of applications.

REFERENCES

[1] Y. Liu, X. Xia, Y. Gong, B. Song, and X. Zeng, “SSR-DTA: Substructure-aware multi-layer graph neural networks for drug-target binding affinity prediction,” Artif. Intell. Med., vol. 157, 2024.

[2] L.-X. Hou, H.-C. Yi, Z.-H. You, S.-H. Chen, and J. Zheng, “MathEagle: Accurate prediction of drug-drug interaction events via multi-head attention and heterogeneous attribute graph learning,” Comput. Biol. Med., vol. 177, 2024.

[3] M. Nejati and A. Lakizadeh, “GADNN: A graph attention-based method for drug-drug association prediction considering the contribution rate of different types of drug-related features,” Inform. Med. Unlocked, vol. 44, 2024.

[4] H. Luo et al., “Drug-drug interactions prediction based on deep learning and knowledge graph: A review,” iScience, vol. 27, p. 109148, Mar. 2024.

[5] H. Zhang, X. Liu, W. Cheng, T. Wang, and Y. Chen, “Prediction of drug-target binding affinity based on deep learning models,” Comput. Biol. Med., vol. 174, 2024.

[6] Z. Zhu et al., “Drug–target binding affinity prediction model based on multi-scale diffusion and interactive learning,” Expert Syst. Appl., vol. 255, 2024.

[7] G. Han, L. Peng, A. Ding, Y. Zhang, and X. Lin, “CTF-DDI: Constrained tensor factorization for drug-drug interactions prediction,” Future Gener. Comput. Syst., vol. 161, Dec. 2024.

[8] Z. Yang et al., “Advancing Drug-Target Interaction prediction with BERT and subsequence embedding,” Comput. Biol. Chem., vol. 110, 2024.

[9] A. Schulman, J. Rousu, T. Aittokallio, and Z. Tanoli, “Attention-based approach to predict drug-target interactions across seven target superfamilies,” Bioinformatics, 2024.

[10] W. H. Ali et al., “Drug-Drug Interaction Checker,” J. Clin. Eng., 2019.

[11] H. Yang et al., “MINDG: a drug-target interaction prediction method based on an integrated learning algorithm,” Bioinformatics, 2024.

[12] W. Song et al., “Drug-target interaction predictions with multi-view similarity network fusion strategy and deep interactive attention mechanism,” Bioinformatics, 2024.

[13] E. Chouzenoux, K. Kumar, A. Majumdar, and S. Jain, “Graph Regularized Probabilistic Matrix Factorization for Drug-Drug Interactions Prediction,” IEEE J. Biomed. Health Inform., May 2023.

[14] Y. Sun, Y. Y. Li, C. K. Leung, and P. Hu, “iNGNN-DTI: prediction of drug-target interaction with interpretable nested graph neural network and pretrained molecule models,” Bioinformatics, 2024.

[15] Q. Ye et al., “A unified drug-target interaction prediction framework based on knowledge graph and recommendation system,” Nat. Commun., 2021.

[16] W. E. Djeddi, K. Hermi, S. B. Yahia, and G. Diallo, “Advancing drug-target interaction prediction: a comprehensive graph-based approach integrating knowledge graph embedding and ProtBert pretraining,” BMC Bioinformatics, 2023.

[17] Y. Wang, Z. Yang, and Q. Yao, “Accurate and interpretable drug-drug interaction prediction enabled by knowledge subgraph learning,” Commun. Med., 2024.

[18] T. Zhao et al., “Identifying drug-target interactions based on

graph convolutional network and deep neural network," Brief. Bioinform., vol. 22, Mar. 2021.

[19] H. El-Behery, A.-F. Attia, N. El-Fishawy, and H. Torkey, "An ensemble-based drug-target interaction prediction approach using multiple feature information with data balancing," J. Biol. Eng., Aug. 2022.

[20] L. Peng et al., "DTI-MvSCA: An Anti-Over-Smoothing Multi-View Framework with Negative Sample Selection for Predicting Drug-Target Interactions," IEEE J. Biomed. Health Inform., 2024.

[21] Y. Xu et al., "A Protein-Context Enhanced Master Slave framework for zero-shot drug target interaction prediction," IEEE J., Aug. 2015.

[22] S. Muniyappan, A. X. A. Rayan, and G. T. Varrieth, "DRADTiP: Drug repurposing for aging disease through drug-target interaction prediction," Comput. Biol. Med., vol. 182, 2024.

[23] S. Mishra, A. Chinthala, and M. Bhattacharya, "Drug-target prediction through self-supervised learning with dual task ensemble approach," Comput. Biol. Chem., Dec. 2024.

[24] D. Kumari et al., "A study on improving drug-drug interactions prediction using convolutional neural networks," Appl. Soft Comput., vol. 166, 2024.

[25] Y. Xie, X. Wang, P. Wang, and X. Bi, "A pseudo-label supervised graph fusion attention network for drug-target interaction prediction," Expert Syst. Appl., vol. 259, 2024.

[26] S. Liu et al., "Enhancing Drug-Drug interaction Prediction using deep attention neural networks," IEEE/ACM Trans. Comput. Biol. Bioinform., Mar./Apr. 2023.