IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510154

REVIEW PAPER ON APLICATIONS OF GRAPH NEURAL NETWORKS AND BIOMEDICAL LINK PREDICTION

Atharva Prafulla Pagey[1], Diya Baldota[2], Rajlaxmi Take[3]

Student, Computer Department, PICT, Pune, India[1]

Student, Computer Department, PICT, Pune, India[2]

Student, Computer Department, PICT, Pune, India[3]

Abstract: Graph neural networks (GNN) provide a natural framework for learning from relational biomedical data, where drugs, proteins, diseases and adverse events are connected through heterogeneous interactions. This review examines twelve papers spanning foundational Graph Neural Network (GNN) architectures, drug-drug interaction (DDI) prediction models, biomedical knowledge graphs and recent projects in attention, multimodal learning, contrastive learning and cold-start prediction. Architectures like Graph Convoluted Networks (GCN) [1] establish an efficient neighborhood propagation method; Graph Attention Network (GAT) [2] introduces learned neighbor attention; architectures like Relational Graph Convoluted Network (R-GCN) [3] and Relational Graph Attention Networks (r-GAT) [12] incorporate relation-specific transformations. Decagon [4] extends relational graph learning to polypharmacy side effects; and Multimodal Deep Neural Network (MDNN) [5] combines graph and heterogeneous drug features. More recent approaches further incorporate multiple features to current approaches like the Deep Attention Neural Network for Drug-Drug Interations (DANN-DDI) [6], GNN which uses a novel approach of Simplified Molecular Input Line Entry System (SMILES) strings (SmileGNN) [7], or a cross-attention GNN [8], a flow based GNN for emerging drug interations (EmerGNN) [9], a Knowledge Graph based DDI model with contrastive learning (KG-CLDDI) [10], and models which use large-scale biomedical knowledge graphs such as PrimeKG [11]. Rather than comparing numerical scores directly across incompatible datasets, the review focuses on how each architectural choice addresses a specific limitation of earlier approaches. The comparison highlights four recurring requirements for adverse drug event prediction: preserving relation semantics, integrating complementary biological views, handling sparse or unseen entities, and maintaining interpretability and scalability. These observations motivate a three-branch GAT architecture for the proposed ADE project.

Keywords: Drug-Drug Interaction Prediction, Graph Neural Networks, Knowledge Graphs, Link Prediction, Multimodal Learning.

I. INTRODUCTION

Biomedical data is inherently relational: proteins interact with each other, drugs bind to targets, and diseases associated with genes, can produce adverse effects. Representing these relationships as graphs preserves information that is difficult to capture with conventional tabular models. Graph Neural Networks (GNNs) exploit this structure by iteratively aggregating information from neighboring nodes, making them suitable for node classification, link prediction and event prediction.

The literature has progressed from relatively homogeneous graphs to heterogeneous and multi-relational biomedical networks. Graph Convoluted Network (GCN) [1] introduced a simple normalized propagation rule; Graph Attention Network (GAT) [2] made neighbor importance learnable through attention; and Relational-Graph Convoluted Network (R-GCN) [3] extended message passing to multi-labeled relations. These foundations enabled biomedical systems such as Decagon [4] and Multimodal Deep Neural Network [5], which moved from generic graph learning toward specific Drug Drug Interations (DDI) and polypharmacy predictions. Subsequent work has focused on combining different graph topologies with molecular structures and multiple biological feature networks, improving prediction for rare or emerging drugs, and learning more robust representations through contrastive objectives.

This review therefore considers twelve papers as a connected progression rather than twelve isolated methods. The next section compares their approaches, datasets used, problem formulations, graph inputs, advantages and limitations. The Third Section defines the performance parameters, summarizes reported results. The review will end in proposal of new architecture which could an improvement on the current structures.

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

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510154

II. RELATED WORK

The twelve papers are organized into 5 stages. GCN [1], GAT [2] provide explaination of core mechanisims and R-GCN [3] and r-GAT [12] establish the core mechanisms for propagation, attention and relation-specific message passing. Decagon [4], SmileGNN [7], CrossAttention GNN [8]. MDNN [5], Deep Attetion Neural Network for DDI [6], and EmerGNN [9] adapt these ideas to increasingly specific DDI settings. Knowledge Graph-Contrastive Learning DDI model [10], and PrimeKG [11] represent the broader move toward richer knowledge graphs, contrastive representations and relation-aware attention. The important comparison is therefore not simply which model reports the largest metric, but what additional information or inductive bias each model introduces.

A. Foundations of Graph Neural Networks

Kipf and Welling[1] proposed a scalable approach to semi-supervised node classification on graph-structured data, such as documents in a citation network where only a small subset of nodes carries labels. Rather than smoothing label information over the graph through a Laplacian regularization term in the loss, they encoded the graph structure directly in a neural network model f (X, A) and trained it on a supervised loss system over the labeled nodes only. Their GCN [1] stacks layers that follow the propagation rule:

$$H^{(l+1)} = \sigma\left(\tilde{\mathrm{D}}^{-\frac{1}{2}}\tilde{\mathrm{A}}\tilde{\mathrm{D}}^{-\frac{1}{2}}H^{l}W^{l}\right) \qquad (1)$$

where $\tilde{\mathrm{A}} = \mathrm{A} + I_N$ is the adjacency matrix with added self-connections, $\tilde{\mathrm{D}}$ is its degree matrix, W^l is a trainable layer-specific weight matrix, and σ (·) is an activation function such as ReLU. Stacking several such layers yields a rich class of convolutional filters without the explicit parameterization, and the model reduced overfitting on local neighborhoods in graphs with wide node degree distributions. However, the framework has several limitations: it has high memory requirements, it does not support edge features, and is restricted to undirected graphs. It also implicitly assumes locality, meaning each node depends only on its K^{th}-order neighborhood in a K-layer GCN, and treats self-connections and edges to neighboring nodes as equally important.

Building on these limitations, Veličković et al. [2] introduced Graph Attention Networks (GATs), which replaced the fixed neighborhood normalization of GCNs with masked self-attentional layers. In each layer, a shared linear transformation W is applied to every node, and a shared attention mechanism computes a coefficient $e_{ij} = a\left(Wh_i, Wh_j\right)$ that expresses how important node j's features are to node i. These coefficients are normalized over each node's neighborhood with a softmax function,
$\alpha_ij = \exp\,(e_ij\,)/\,\Sigma_{\{k\in N_i\}}\exp(e_{ik})$ and each node's new representation is the attention-weighted combination of its neighbor's transformed features. The model further has a multi-head attention to stabilize learning. As a result, nodes within the same neighborhood receive different implicit weights, which increases model capacity. The computation is parallelizable across edges and needs no eigen decompositions or other costly matrix operations. Because attention is applied in a shared manner to every edge, the model does not depend on upfront access to the global graph structure, so it handles nodes of varying degree and applies directly to inductive problems, including generalization to completely unseen graphs. The paper show results which match or exceed the state of the art in 2018 on the Cora, Citeseer, and Pubmed citation benchmarks and on an inductive protein-interaction dataset.

B. Relational Graph Neural Networks

Schlichtkrull et al. [3] introduce the relational graph convolutional network (R-GCN) for highly multi-relational data in incomplete knowledge bases, which store facts as directed (subject, predicate, object) triples. The model infers missing information from a node's local neighborhood and is applied to entity classification and link prediction. Each layer updates a node as

$$h_i^{l+1} = \sigma\left(\Sigma_r\Sigma_{\{j\in N_i^r\}}\left(\frac{1}{c_{\{i,r\}}}\right)W_r^l h_j^l + W_0^l h_i^l\right) \qquad (2)$$

aggregating messages from neighbors under each relation r through relation-specific weight matrices W_r, normalizing them by a constant $c_{\{i,r\}}$, and adding a self-loop transformation W_0. To keep the parameter count manageable with many relation types, the model regularizes the weights through basis decomposition or block diagonal decomposition. For link prediction, they pair an R-GCN encoder with a DistMult tensor factorization decoder, which yields a 29.8% relative improvement in mean reciprocal rank (MRR) on FB15k-237. However, thousands of relations can still cause rapid parameter growth and overfitting if the regularization is poorly tuned, and fixed normalization constants are problematic for high-degree hub nodes, which suggests a need for data-dependent attention.

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

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510154

Chen et al. [12] propose the relational graph attention network (r-GAT) for multi-relational graphs such as knowledge graphs, where directed edges carry different relation types. They argue that standard graph attention networks ignore relation features and learn only generic, static representations. To address this, r-GAT projects entity and relation features into K independent semantic channels, which separates different aspects of an entity, such as location and career. Each layer computes

$$e_v^{\{k(l)\}} = \sigma\left(\Sigma_{\{u\in N_v\}}\Sigma_{\{i\in R_{vu}\}}\alpha^k_{\{vi\,u\}}\left[e^k_u * r^k_i\right]\right) \qquad (3)$$

where α is a normalized attention score for neighbor u connected through relation i, and e_u^k and r_i^k are the entity and relation features in the k-th subspace. A query-aware attention mechanism then lets the model focus on the channels most relevant to the query relation, correcting the "query-ignorant" behavior of conventional GNNs while avoiding the over-parameterization of other multi-relational models. The authors report state-of-the-art results on link prediction (e.g., FB15k-237) and entity classification, with interpretability through attention analysis. The number of channels must be tuned per dataset by grid search, since adding channels beyond the true number of semantic aspects brings no gain, and extremely high-degree hub entities remain challenging.

C. Graph Neural Networks for Drug-Drug Interation Prediction

Zitnik et al. [4] present Decagon, an approach for modeling polypharmacy side effects, which are difficult to identify manually because they are rare, because testing every drug pair is practically impossible, and because small clinical trials usually do not observe them. The authors build a large multimodal graph of protein-protein interactions, drug-protein target interactions, and drug-drug interactions, where each side effect is a separate edge type, and cast the task as multirelational link prediction. They note that earlier approaches typically use a two-stage pipeline in which a feature extraction model and a link prediction model are trained separately, and that polypharmacy side effects do not occur independently of one another in co-prescribed drug pairs. Decagon is therefore designed to reuse what it learns about the molecular basis of one side effect to understand another. It is a graph auto-encoder, which makes it end-to-end trainable, and has two components. The encoder is a graph convolutional network that gives each node (drug or protein) a d-dimensional embedding. Each layer propagates latent features across edges with a separate processing channel for each relation type, following the relational GCN of Schlichtkrull et al. [3]:

$$h_i^{k+1} = \varphi\left(\Sigma_r\Sigma_{\{j\in N_i^r\}}c_{ij}^r W_r^k h_j^k + c_i^r h_i^k\right) \qquad (4)$$

The decoder is a tensor factorization model that scores a triplet (v_i, r, v_j) to estimate how likely drugs v_i and v_j are to interact through side effect r. For two drugs the score is $z_i^T D_r R\, D_r z_j$, and for protein-protein or protein-drug pairs it is $z_i^T M_r z_j$, followed by a sigmoid to give a probability. In this way, Decagon predicts side effects that arise from a drug combination and cannot be attributed to either drug alone [4].

Han et al. [7] present SmileGNN, a model that predicts drug-drug interactions (DDIs) by combining the structural features of drugs with topological features drawn from biomedical knowledge graphs. The work is motivated by the fact that combination therapy, though essential for complex diseases, often causes unexpected adverse reactions that are impractical to screen clinically at scale. The structural pipeline converts SMILES strings into a 251-element word bag, applies one-hot encoding, and uses principal component analysis (PCA) to compress the representation to 64 dimensions. The topological pipeline builds knowledge graphs from the KEGG and PDD databases and applies a graph neural network (KGNN) with a two-hop receptive field and concatenation aggregation to capture multi-hop relationships. The two feature sets are fused, and a sigmoid function produces an interaction score, with scores above 0.5 classed as predicted DDIs. As validation, 5 of the top 10 novel predictions were later verified and added to DrugBank within 2.5 years. The approach has two weaknesses: negative samples are unreliable, because drug pairs recorded as non-interacting may simply lack clinical verification, and SMILES data are missing for many drug pairs (notably in KEGG), while the structural feature pipeline remains relatively rough and independent of the graph component.

Dietrich [8] investigates whether atom-level comparison of molecular structures can improve DDI prediction and help distinguish interaction types. The approach converts drug Simplified Molecular Input Line Entry System (SMILES) strings into molecular graphs and encodes each drug with a shared-weight Siamese message-passing neural network. Three designs are compared: simple concatenation, bidirectional multi-head atom-level cross-attention, which lets each drug's representation focus on relevant atoms in the other drug, and a ternary model that adds an interaction graph built from selected similar atom pairs. Cross-attention reaches a binary AUC of 0.899 against 0.888 for concatenation and a macro-F1 of 0.596 against 0.410, while the ternary model performs poorly (AUC 0.761, macro-F1 0.072). The paper also reports correct DDI-type predictions for 10 of 10 held-out acetylsalicylic-acid pairs. Nevertheless, the results are specific to the datasets and configurations used, this small validation subset should not be read as evidence of general clinical reliability, and the work is an arXiv preprint.

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



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510154

D. Multimodal and Attention Based Deep Learning for DDI Prediction

Lyu et al. [5] propose MDNN, a multimodal deep neural network for predicting drug-drug interaction (DDI) events. They observe that most existing methods pay little attention to the correlations between DDI events and other multimodal drug data, such as targets and enzymes. MDNN uses a two-pathway framework. The first pathway is based on a drug knowledge graph (DKG), where a graph neural network layer learns drug representations by extracting both topological structure and semantic relations. The second pathway is based on heterogeneous features (HF) and supplies the multimodal drug information. Because the two pathways provide complementary information, a multimodal fusion neural layer considers their coherence and complementarity together: the embeddings from both pathways are concatenated into a final multimodal embedding for each drug, which is used to predict DDI events. This design lets MDNN mine similarities across modalities from multiple sources while also exploiting the DKG's topological and semantic information. On a real-world dataset, MDNN outperforms the DDIMDL baseline by 3.23% in accuracy, 4.6% in AUPR, 0.08% in AUC, 7.16% in F1, 1.51% in precision, and 10.2% in recall. The authors attribute these gains to using both the topological embeddings from the knowledge graph and the cross-modality embeddings from the multimodal data. As with other knowledge-graph-based approaches, its performance depends on the availability and quality of these data sources [5].

The authors propose DANN-DDI [6], a deep attention neural network that predicts drug-drug interactions (DDIs) by combining several types of drug information and learning which features matter for each drug pair. The model builds five feature networks based on chemical substructures, drug targets, enzymes, biological pathways, and known DDIs, and it uses Structural Deep Network Embedding (SDNE) to learn numerical representations from each network. These representations are combined, and a pair-specific attention mechanism weights the feature information for each candidate pair before a deep neural network predicts whether an interaction exists or which DDI-associated event may occur. For event prediction, the study reports an accuracy of 0.8874, a micro-AUPR of 0.9088, a micro-AUC of 0.9943, and a macro-F1 of 0.7781. However, performance depends on the coverage and quality of the input networks and labels, and results on the reported datasets do not guarantee equally strong performance on new or clinically untested drug pairs.

E. Knowledge Graph Based Approaches

The EmerGNN work [9] targets the prediction of interactions for emerging drugs, for which almost no historical interaction data exist. The model instead draws on a large biomedical network linking drugs, genes, and diseases, and its flow-based GNN traces step-by-step biological paths between two drugs to infer how they interact. The method has four steps. First, it merges the drug interaction network with the biomedical network and adds reverse edges so that information can flow in both directions. Second, for a given drug pair, it extracts all connecting paths up to a fixed length to form a small local subgraph. Third, it propagates information from the first drug to the second along these paths, using attention weights to determine which connections matter most. Fourth, it combines the information flowing in both directions and passes it to a classifier that outputs the interaction type. EmerGNN outperforms baseline models on new drugs and makes its predictions interpretable by showing which paths drove them. Its limitations are that an emerging drug must connect to entities in the biomedical network, that building a local subgraph for every drug pair raises training and computing cost, and that overly long paths introduce irrelevant noise and reduce accuracy.

Zhong et al. [10] present KG-CLDDI, a method that improves DDI prediction by combining biomedical knowledge-graph information, DDI graph structure, and contrastive learning. A biomedical knowledge graph supplies contextual information about drugs, and a DDI graph supplies known interactions. Attention-based knowledge aggregation learns from related entities and relations, while LightGCN-style topological aggregation captures patterns in the DDI graph. The two representations are fused and trained with contrastive-learning strategies, including graph-view changes such as dropout and horizontal reversal, before a classifier predicts the interaction type. The method is evaluated on multi-class prediction with 1,569 drugs, 172,426 DDIs, and 81 types, and on multi-label prediction on TWOSIDES with 1,276 mapped drugs, 98,327 DDIs, and 200 types. In inductive settings, it improves AUC by 17.49% and AUPR by 24.97% over the previous best comparison cited in the study. The model depends on the coverage and correctness of the knowledge graph and known DDI records, and the gains depend on the dataset and evaluation setting, so they do not establish clinical safety or causality

Chandak et al. [11] build PrimeKG, a knowledge graph that combines biomedical information from multiple sources to support precision-medicine research. It integrates 20 biomedical resources and links drugs, diseases, proteins, pathways, phenotypes, and biological processes, giving 129,375 nodes and 4,050,249 relationships, including drug-disease indications, contraindications, and off-label-use relationships. The graph includes 17,080 disease nodes and covers ten major biological scales, so researchers can trace links among entities to study disease mechanisms, drug effects, and possible adverse events. For DDI research, it serves as a large, connected source of contextual information for downstream prediction systems. It is a knowledge resource rather than a standalone DDI model, however, so a separate

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510154

model is still needed to predict or classify drug interactions, and its coverage and accuracy depend on its source databases and their update dates.

III. COMPARATIVE ANALYSIS

The comparison shows a shift from learning on a single graph toward combining several complementary representations. GCN and GAT primarily improve how local neighborhoods are aggregated; R-GCN improves how relation semantics are preserved. Decagon and MDNN demonstrate that these mechanisms become more useful when the graph represents clinically meaningful interactions. DANN-DDI and SmileGNN add independent biological or chemical evidence, while CrossAtt and EmerGNN target mechanism-level and cold-start limitations. KG-CLDDI adds representation-level robustness through contrastive learning, whereas PrimeKG and r-GAT address the scale and heterogeneity of the underlying knowledge graph.

Biomedical approaches such as Decagon and MDNN extend these foundations to drug interaction and adverse-event prediction by incorporating relational and heterogeneous drug information.

More recent approaches address limitations related to multimodal representation, molecular structure, interaction mechanisms and unseen drugs. DANN-DDI integrates multiple heterogeneous feature networks using attention, while SmileGNN combines molecular SMILES information with graph topology. CrossAtt introduces atom-level cross-attention for mechanism-level prediction, and EmerGNN specifically addresses cold-start prediction for emerging drugs. KG-CLDDI improves representation robustness through contrastive learning, whereas PrimeKG provides a large heterogeneous biomedical knowledge graph for integrating information from multiple biomedical resources. r-GAT combines relation-aware message passing with multi-channel attention.

Overall, existing approaches demonstrate that no single representation is sufficient for comprehensive adverse drug-event prediction. Graph topology, relation semantics, molecular structure and heterogeneous biological information provide complementary evidence. However, challenges remain in handling unseen drugs, noisy relationships, sparse interactions, relation diversity and model interpretability. These limitations motivate the proposed multi-branch GAT architecture for the ADE prediction task.

Paper	Dataset / Graph Source	Input Structure	Approach Used	Performance Parameter	Advantages	Limitations
GCN [1]	Cora	Homogeneous graph	Graph convolution with normalized neighborhood aggregation	Accuracy: 81.5%	Simple and computationally efficient graph propagation	Mainly designed for homogeneous and transductive settings
GAT [2]	Cora, Citeseer, PPI	Homogeneous graph	Multi-head graph attention	Cora accuracy: 83.0%; PPI F1 improvement: 20.5%	Learns importance of individual neighbors; supports inductive learning	Higher computational cost due to attention mechanism
R-GCN [3]	FB15k-237	Multi-relational knowledge graph	Relation-specific graph convolution with basis/block decomposition	MRR: 0.248 vs. 0.191 for DistMult	Preserves relation semantics and handles multiple relation types	Large relation sets can increase model complexity
Decagon [4]	Drug-side-effect / drug-protein interaction graph	Drug-protein multimodal graph	Relational GCN encoder with tensor-factorization decoder	AUROC: 0.872; AUPRC: 0.832	Models multiple polypharmacy side effects simultaneously	Relies primarily on observed graph structure; cold-start performance is challenging
MDNN [5]	65-type DDI event data	Knowledge graph + drug features	Dual pathway combining graph and heterogeneous drug features	Accuracy: 0.9175; F1: 0.8301	Combines graph information with target, substructure and enzyme features	Multiple feature sources increase complexity
DANN-DDI [6]	DDI feature networks	Five heterogeneous feature networks	Structural embedding with dimension-wise attention	AUROC: 0.9763; AUPR: 0.9709; Accuracy: 0.9962	Integrates multiple biological views and identifies important feature networks	Depends on multiple heterogeneous feature networks
SmileGNN [7]	KEGG	SMILES + biomedical knowledge graph	Molecular structure and graph-topology fusion	Accuracy: 0.8936; AUC: 0.9521	Combines chemical structure with relational information	Requires molecular SMILES and graph information
CrossAtt [8]	Molecular graph / aspirin validation	Molecular graphs	Atom-level cross-attention	F1-macro: 0.596; +45% vs. Concat	Captures fine-grained atom-level interaction mechanisms	More complex than simple pair-level models
EmerGNN [9]	DrugBank	Biomedical path subgraphs	Flow/path-based graph representation	F1: 62.0% single cold-start; 31.9% dual cold-start	Addresses emerging-drug cold-start prediction; provides paths for explanation	Cold-start performance remains lower than in conventional settings
KG-CLDDI [10]	DDI knowledge graph	Knowledge graph + DDI graph	Knowledge-graph aggregation with contrastive learning	+17.49% AUC; +24.97% AUPR (inductive setting)	Improves robustness under incomplete evidence and inductive settings	Contrastive training adds complexity
PrimeKG [11]	20 biomedical resources	Large heterogeneous biomedical KG	Multi-source biomedical knowledge-graph integration	129,375 nodes; 4,050,249 relationships	Rich connections among drugs, diseases, proteins, pathways and phenotypes	Resource itself is not a DDI prediction model
r-GAT [12]	Multi-relational KG	Multi-relational knowledge graph	Multi-channel relational attention	Link prediction / classification performance	Combines relation-aware message passing with adaptive attention	Multiple semantic channels can increase computational complexity

Fig. 1 Comparative analysis table of selected research papers.

IV. RESULT ANALYSIS

A. Performance Parameters

The performance of the reviewed models is evaluated using different metrics depending on the prediction task. Accuracy measures the proportion of correctly classified instances, while precision and recall provide information about the correctness and coverage of positive predictions. F1-score provides a balance between precision and recall and is particularly useful for imbalanced classification tasks. AUROC measures the ability of a model to distinguish between

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



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510154

positive and negative samples across different classification thresholds, whereas AUPR is particularly informative when positive interactions are relatively sparse. For knowledge-graph link prediction, Mean Reciprocal Rank (MRR) evaluates the ranking quality of predicted entities or relations.

Since the reviewed papers use different datasets, tasks and evaluation protocols, their reported values should not be interpreted as a single performance ranking. Instead, the metrics should be used to understand the effectiveness of each architectural contribution within its respective experimental setting

B. Result Comparision

The reported metrics should not be interpreted as a single leaderboard because the papers use different datasets, tasks, class counts and evaluation protocols. A 0.95 AUC for binary DDI detection, for example, measures a different problem from an F1 score for 86-way mechanism classification. The useful comparison is therefore the improvement associated with each architectural contribution and the setting in which that contribution was evaluated.

Several consistent trends emerge. First, adding graph structure to feature-only or decoder-only models generally provides a measurable gain, as illustrated by R-GCN and Decagon. Second, attention becomes useful when neighborhood quality is uneven: GAT learns which neighbors matter, while DANN-DDI and r-GAT extend this principle to multiple semantic or feature channels. Third, multimodal fusion repeatedly improves performance because chemical structure, biological targets, pathways and observed interactions contain complementary information. MDNN and SmileGNN provide direct examples, while PrimeKG offers a broader substrate for such fusion.

Fourth, generalization beyond observed drug pairs is increasingly important. Decagon is largely transductive, whereas EmerGNN and KG-CLDDI explicitly investigate

cold-start or inductive conditions. This distinction is particularly important for an ADE system intended to reason about newly introduced drugs. Finally, interpretability is moving from post-hoc inspection toward structurally meaningful evidence: EmerGNN exposes predictive paths, attention-based systems expose influential neighbors or feature networks, and CrossAtt provides atom-level attention over chemically relevant regions.

The progression across the twelve papers can also be viewed according to the type of missing information each method attempts to recover. GCN primarily recovers local structural information that conventional feature models ignore. GAT addresses the assumption that all neighbors are equally useful by learning adaptive weights. R-GCN then addresses a different problem: in a knowledge graph, two edges connecting the same entities can have completely different meanings, so the message-passing operation must preserve relation identity.

The biomedical models add increasingly specialized sources of evidence. Decagon represents side effects directly as typed relations and therefore moves beyond the binary question of whether two drugs interact. MDNN and DANN-DDI broaden the feature space by adding target, enzyme, pathway and other heterogeneous information. SmileGNN adds a chemical view through SMILES, demonstrating that graph topology alone does not fully describe a drug. These models suggest that ADE prediction should not depend on a single representation of a drug; molecular structure, biological function and observed interaction history describe different aspects of the same entity.

The newer papers focus on generalization and fine-grained reasoning. CrossAtt moves from pair-level representations toward atom-level communication when the objective is to identify a particular mechanism. EmerGNN addresses a practical deployment problem in which an emerging drug may have little historical interaction data. KG-CLDDI attempts to make representations more robust to incomplete evidence by contrasting related and unrelated drug-pair representations. PrimeKG supplies a much broader heterogeneous context, while r-GAT combines relation-aware propagation with attention over multiple semantic channels. These directions are complementary rather than mutually exclusive.

C. Knowledge Graph Based Approaches

The reviewed approaches provide several important advantages, including effective graph-based representation learning, relation-aware message passing, adaptive attention, multimodal feature integration and improved handling of heterogeneous biomedical information. However, these approaches also have limitations. Several methods depend strongly on observed interactions and therefore may struggle with newly introduced drugs and rare interactions. Other methods require multiple heterogeneous feature networks or molecular information, increasing data and computational requirements. In addition, differences in datasets, prediction tasks and evaluation protocols make direct comparison of reported performance difficult.

Despite substantial progress, several limitations remain. First, many DDI datasets are dominated by common drugs and known interactions, making performance on rare or emerging drugs difficult to estimate from random train-test splits. EmerGNN and the inductive experiments in KG-CLDDI directly address this issue, but cold-start performance remains substantially harder than conventional transductive prediction. Second, heterogeneous biomedical graphs combine information with different levels of reliability, granularity and provenance. A dense graph can improve connectivity while also introducing noisy paths or spurious correlations.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510154

Third, binary interaction prediction and multi-class event prediction should be treated as different objectives. A model that accurately detects that two drugs interact does not necessarily identify the clinical mechanism or adverse event associated with that interaction. CrossAtt, Decagon and MDNN illustrate different responses to this problem, ranging from typed graph edges to event classification and atom-level attention. Finally, large relation vocabularies create a trade-off between expressiveness and parameter efficiency. R-GCN's parameter sharing, Decagon's shared interaction factors and r-GAT's multi-channel design all attempt to manage this trade-off, but further work is needed to combine relation specificity with scalable training.

V. CONCLUSION

The twelve reviewed papers show a clear evolution of GNN-based biomedical prediction. GCN provides efficient propagation, GAT makes neighbor importance adaptive, and R-GCN preserves relation semantics in knowledge graphs. Biomedical systems then extend these foundations through polypharmacy side-effect modeling, multimodal feature fusion, molecular structure integration, atom-level interaction modeling, cold-start path reasoning and contrastive representation learning. PrimeKG and r-GAT further demonstrate that future ADE systems must operate on large heterogeneous graphs while retaining relation-aware attention. These observations motivate the proposed three-branch GAT architecture, which combines DPI-PPI, DDI and unified graph inputs through attention-based branches and a fusion layer for adverse drug-event prediction.

REFERENCES

[1] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2017.
[2] P. Veličković *et al.*, "Graph Attention Networks," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2018.
[3] M. Schlichtkrull *et al.*, "Modeling Relational Data with Graph Convolutional Networks," *arXiv preprint arXiv:1703.06103*, 2018.
[4] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling Polypharmacy Side Effects with Graph ConvolutionalNetworks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, 2018.
[5] T. Lyu *et al.*, "MDNN: A Multimodal Deep Neural Network for Predicting Drug-Drug Interaction Events," in *Proc. IJCAI*, 2021, pp. 3536–3542.
[6] "Enhancing Drug-Drug Interaction Prediction Using Deep Attention Neural Networks," 2022.
[7] X. Han, R. Xie, X. Li, and J. Li, "SmileGNN: Drug–Drug Interaction Prediction Based on the SMILES and GraphNeural Network," *Life*, vol. 12, no. 2, p. 319, 2022.
[8] J. Dietrich, "From Detection to Mechanism: Cross-Attention Graph Neural Networks Enable Drug-Drug InteractionType Prediction: An Ablation Study with Acetylsalicylic Acid Validation," *arXiv preprint arXiv:2605.27861*, 2026.
[9] "Emerging drug interaction prediction enabled by a flow-based graph neural network with biomedical network,"*Nature Computational Science*, 2024.
[10] J. Zhong, H. Zhao, Q. Zhao, and J. Wang, "A Knowledge Graph-Based Method for Drug-Drug Interaction PredictionWith Contrastive Learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2024.
[11] P. Chandak, K. Huang, and M. Zitnik, "Building a knowledge graph to enable precision medicine," *Scientific Data*, 2023.
[12] M. Chen, Y. Zhang, X. Kou, Y. Li, and Y. Zhang, "r-GAT: Relational Graph Attention Network for Multi-RelationalGraphs," *arXiv preprint arXiv:2109.05922*, 2021.

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