DDI Prediction With Heterogeneous Information Network - Meta-Path Based Approach

Farhan Tanvir, Khaled Mohammed Saifuddin, Muhammad Ifte Khairul Islam, and Esra Akbas

***Abstract*—Drug-drug interaction (DDI) indicates where a particular drug's desired course of action is modified when taken with other drug (s). DDIs may hamper, enhance, or reduce the expected effect of either drug or, in the worst possible scenario, cause an adverse side effect. While it is crucial to identify drug-drug interactions, it is quite impossible to detect all possible DDIs for a new drug during the clinical trial. Therefore, many computational methods are proposed for this task. This paper presents a novel method based on a heterogeneous information network (HIN), which consists of drugs and other biomedical entities like proteins, pathways, and side effects. Afterward, we extract the rich semantic relationships among these entities using different meta-path-based topological features and facilitate DDI prediction. In addition, we present a heterogeneous graph attention network-based end-to-end model for DDI prediction in the heterogeneous graph. Experimental results show that our proposed method accurately predicts DDIs and outperforms the baselines significantly.**

***Index Terms*—Chemical structure, drug-drug interaction, graph neural network, link prediction, representation learning.**

I. Introduction

DRUG discovery is crucial in healthcare and pharmaceutical research to develop new drugs to treat diseases and improve patient outcomes. Addressing drug-drug interactions (DDIs), drug-target interactions (DTIs) [1], microRNA-drug associations [2], and drug-pathway associations [3] are essential for patient care and drug development. Adverse drug reactions have become a significant health concern, particularly in the United States, where ADRs are the fourth leading cause of death. One-third of adverse drug reactions occur due to drug-drug interactions (DDI). Effective and early DDI prediction is required to reduce the impact of unwanted pharmaceutical side effects. However, due to the numerous possible drug combinations and comorbidities, conducting experiments on all drugs in a clinical setting is impractical. Therefore, computational models are needed to detect new DDIs for drugs.

Computational models have been proposed to detect DDIs ([4], [5], [6], [7], [8], [9]) by integrating drug-related information from public databases like DrugBank, STITCH, SIDER, PubChem, and KEGG. To leverage the increasing relational information about drugs, current methods integrate multiple data sources to extract drug features, such as side effects and target proteins [10]. Recently, network-based methods using graphs have gained attention, capturing complex interactions between drugs and other entities in various fields, including social networks [11], [12], citation networks [13], [14], and biological networks [15].

Despite the achievements of existing models in DDI prediction, several limitations still need to be addressed. First, drug-centric interaction data are varied and diverse, as different types of drugs can interact with other entities, resulting in various events. Second, heterogeneous graphs, widely used in data mining applications, offer detailed information and rich semantics. Meta-paths and meta-path-based neighbors are core features of heterogeneous graphs, where meta-paths [16] serve as composite relations connecting two objects, measuring higher-level similarity and semantics between items. However, no previous works have employed meta-paths and meta-path-level attention to denote drug similarity and perform DDI prediction. Thirdly, most prior studies have focused on a limited number of data sources for predicting DDIs. To accurately predict possible DDIs, it is essential to incorporate information from multiple data sources [5], [17]. Lastly, assessing model effectiveness for new drugs needs to be addressed. Models tested on current drugs may give different effectiveness for new drugs [17], [18].

We propose a novel Heterogeneous Graph Attention Network (`HAN-DDI`) for DDI prediction to address these limitations. Our approach integrates rich drug-centric information from multiple data sources, creating a comprehensive drug-based interaction dataset. We represent drugs and their interactions with other entities using a Heterogeneous Information Network (HIN). Lastly, two essential features of a heterogeneous information network (HIN) are the meta-path and the meta-path-based neighbors. Meta-paths are used on the HIN to measure higher-level similarity among drugs.

As a baseline solution for DDI, we use meta-paths to extract handcrafted topological features. Topological features consider the entire graph structure and node connectivity during feature extraction. However, handcrafted topological features may be enough to learn the complex relationship between drugs and biomedical entities. Also, each meta-path conveys different semantics for drug pairs and may have different impacts on the interactions of drugs. Therefore, we must determine which

Manuscript received 2 March 2023; revised 27 May 2024; accepted 11 June 2024. Date of publication 21 June 2024; date of current version 9 October 2024. This work was supported by National Science Foundation under Grant 2308206. *(Corresponding author: Farhan Tanvir.)*

The authors are with the Department of Computer Science, GA State University, Atlanta, GA 30332 USA (e-mail: ftanvir@gsu.edu; ksaifuddin1@student.gsu.edu; mislam29@student.gsu.edu; eakbas1@gsu.edu).

Digital Object Identifier 10.1109/TCBB.2024.3417715

meta-paths are more crucial and incorporate their significance into our model.

We introduce a novel graph encoder-decoder framework for predicting drug-drug interactions based on a heterogeneous graph attention network ([19]). In contrast, prior works for link prediction tasks in biology (e.g., [20], [21]) use a two-stage pipeline consisting of a graph feature extraction model and a link prediction model that are both trained independently. Our approach utilizes several meta-paths to indicate different relations between drug pairs within the HIN. Our model employs a hierarchical attention mechanism, combining node-level and meta-path-level attention, to learn node representation from diverse meta-paths efficiently. This model propagates information from local neighbors using a meta-path. For each meta-path, we integrate node-level attention to obtain representations for nodes. We also use meta-path-level attention to learn the importance of distinct meta-paths to aggregate node representations from different meta-paths efficiently. Afterward, the pair-wise representations of drugs are passed through decoder functions to predict a binary score, expressing the likelihood of interaction for each drug pair.

To demonstrate the efficacy of our model, we perform extensive experiments, considering imbalanced data distribution and evaluating performance using various metrics. The results demonstrate our model's ability to predict DDIs for new drugs without known interactions.

The following are the primary contributions of our models:

- *Combining multiple data sources on Heterogeneous Information Network:* We integrate various data sources to generate a comprehensive drug-based interaction dataset, enabling more accurate information about drug interactions. Our HIN captures the interactions of drugs with proteins, diseases, and pathways, considering the involvement of protein species and pathway subjects.
- *DDI Prediction on HIN with Meta-paths:* We create multiple meta-paths representing different relations between drug pairs within the HIN. The `HAN-DDI` model, incorporating node-level and meta-path-level attention, efficiently learns node features from diverse meta-paths, enhancing the accuracy of DDI prediction.
- *Heterogeneous Graph Encoder-Decoder:* We propose a novel graph encoder-decoder framework for predicting drug-drug interactions based on a heterogeneous graph attention network. Our hierarchical attention mechanism contributes to improved prediction performance.
- *Extensive experiment:* We conduct comprehensive experiments, considering imbalanced data distribution and employing various evaluation metrics, including F_1-score, Recall, Precision, AUROC, and AUPR. The results demonstrate that our model predicts DDIs for new drugs without known interactions.

The paper is organized as follows: Section II summarizes related works. SectionIII details data integration, heterogeneous graphs, and meta-paths. Section IV presents our HAN-DDI model. Section V outlines the experimental setup and results. Finally, we conclude in Section VI.

II. Related Work

Previous DDI prediction research can be categorized into two main approaches: Similarity-based methods and graph deep learning-based techniques.

A. Similarity-Based Methods

Traditionally, statistical learning methods based on pharmacological, topological, or semantic similarity have been employed to predict adverse drug reactions (ADRs) and drug-drug interactions (DDIs) [22], [23]. These similarity-based approaches rely on the assumption that drugs with similar characteristics are likely to interact. Various research studies [17], [24], [25], [26] have utilized diverse similarity metrics to predict DDIs successfully. Notably, some researchers [10] have addressed the challenge of imbalanced and skewed datasets by constructing a heterogeneous graph using multiple data sources. Additionally, [27], [28] utilized hypergraphs to represent the chemical structure-based similarity between drugs, connecting multiple drugs through a single hyper-edge if they share a similar chemical substructure. However, many of these methods focus on a limited number of datasets and drug-centric interactions.

B. Graph Deep Learning-Based Methods

Knowledge graphs, represented as heterogeneous networks with entities and various relations, have been widely studied using graph neural networks (GNNs) [29]. In recent years, these GNN-based approaches have been applied to various problems, including DDI prediction ([30], [31], [32]). GNN-based models construct knowledge graphs based on drug-centric interactions and extract relations among drugs using neural networks. For instance, Decagon [5] developed a graph convolutional network with encoding, decoding, and model training phases for DDI prediction, incorporating protein-protein interactions, drug-drug interactions, and drug-protein interactions. KGNN [33] utilized GNNs to learn drug embeddings from a knowledge graph and DDI information. HyGNN [28], based on hypergraphs, employed an encoder-decoder architecture with an attention-based edge encoder to obtain drug embeddings and predict interactions between drug pairs. Additionally, MUFFIN [34] combines a message-passing neural network with TransE to capture drug structure representation from the molecular map and semantic features from the knowledge graph, ensuring powerful drug representation for DDI prediction. SkipGNN [35] employs a graph neural network method that aggregates data from second-order interactions and direct interactions to predict molecular interactions. Moreover, MRCGNN [36] applies GNN on the multi-relational DDI event graph after extracting drug features from drug molecular graphs. Then, they implement a multi-relational contrastive learning to capture rare DDI events. While these techniques demonstrate excellent performance, they often treat DDIs as independent data samples, disregarding their relationships within the knowledge graph. Our approach sets itself apart by focusing on extracting drug interactions and DDIs using various meta-paths and introducing an attention

TABLE I
STATISTICS OF DATASET

Nodes/Edges	Number of nodes/edges
Drug	481
Protein	1266
Disease (Indication and Side Effect)	1602
Pathway	48,703
Pathway Subject	7

mechanism to determine the importance of drug nodes and meta-paths, effectively leveraging semantic information from nodes in heterogeneous networks.

Drug-drug interaction prediction involves identifying potential interactions between drug combination, leading to unwanted side-effects. On the other hand, drug combination prediction focuses on identifying potential synergistic combinations of drugs, where the combined effect of two or more drugs is greater than the sum of their individual effects, leading to improved therapeutic outcomes and reduced drug resistance. Both areas play a significant role in pharmacological research, and they are related since drug combinations can also lead to drug interactions, influencing their individual effects and overall therapeutic outcome. Studies like SNRMPACDC [37] and NLLSS [20] have emerged as efficient tools to predict potential synergistic drug combinations. SNRMPACDC utilizes Siamese convolutional networks and random matrix projection for predicting anticancer synergistic drug combinations. On the other hand, NLLSS integrates different types of information, including known synergistic drug combinations, drug-target interactions, and drug chemical structures, to predict antifungal synergistic drug combinations with excellent accuracy.

III. Data Processing

This section outlines the comprehensive process of preparing the data for drug-drug interaction (DDI) prediction, involving three main steps: Data Integration, Heterogeneous Graph Construction, and Meta-path Construction. The following subsections provide a more detailed explanation of each of these data processing steps.

A. Data Integration

Our dataset encompasses diverse entities and their interactions from publicly available sources, including DrugBank,[1] KEGG ,[2] and DEB2 [3] (it combines DrugBank, MedLine, MedLinePlus, Sider2, and NDRFT) and TWOSIDES. We have uploaded our pre-processed datasets to the following GitHub repository: https://github.com/farhantanvir1/HIN-DDI. Table I summarizes the fundamental statistics of the dataset used.

The dataset includes drugs, proteins, pathways, chemical substructures, ATC codes, and diseases with the following interactions:

- *Drug-Protein Interactions:* Different drugs target specific proteins in the human body, known as target proteins, to induce therapeutic changes. DrugBank provides data on drugs interacting with specific target proteins, involving 1266 target proteins, 89 species, and 481 drugs.
- *Drug-Pathway Interactions:* The pathways of drugs provide valuable information about their mechanisms of action and metabolism and also include pathway subjects, such as diseases, proteins, and physiological processes. Interaction data among drug pathways, pathway subjects, and drugs are sourced from DrugBank, yielding relational data for 481 drugs, 48703 pathways, and seven pathway subjects.
- *Drug-Indication Interactions and Side effect data:* DEB2 integrates data from various sources, associating 481 drugs with 1602 indication/side effect instances.
- *The chemical substructure of Drugs:* The simplified molecular-input line-entry system (SMILES) representation of 481 drugs from DrugBank and KEGG is converted into MACCS keys, a binary fingerprint with 167 bits, each corresponding to a chemical substructure indicating its presence or absence.
- *ATC code of drugs:* ATC codes, categorizing drugs based on their operating organs and chemical, therapeutic, and pharmacological characteristics, are obtained from DrugBank and KEGG.
- *Drug-drug interactions:* Both DrugBank and TWOSIDES contain information on drug-drug interactions among the 481 drugs, curated from adverse drug effect reports.

B. Heterogeneous Graph Construction

In our DDI prediction approach, we operate under the hypothesis that similar drugs interact with similar biomedical entities. To model these interactions effectively, we construct a Heterogeneous Information Network (HIN) consisting of diverse entity types. The definition of a heterogeneous information network is as follows:

Definition 1 (Heterogeneous information network). [38] A heterogeneous information network (HIN) is defined as a graph $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ with an entity type mapping ϕ: $\mathrm{V} \rightarrow \mathrm{A}$ and a relation type mapping $\psi : \epsilon \rightarrow \mathrm{R}$, where V denotes the entity set and E is the relation set, A denotes the entity type set and R is the relation type set and the number of entity types $|\mathrm{A}| > 1$ or the number of relation type $|\mathrm{R}| > 1$.

Our constructed HIN comprises eight node types: drugs, proteins, species, pathways, pathway subjects, chemical substructures, ATC codes, and diseases. Leveraging the datasets described in subsection III-A, we establish relations among these distinct node types, which are elaborated below:

- *I1:* T matrix represents the drug-target protein interaction where each element $t_{i,j}$ states whether drug i targets protein j.
- *I2:* R matrix denotes the relationship between species and proteins, with each element $r_{i,j}$ indicating whether protein j can be found in species i.
- *I3:* W matrix captures the relationship between drugs and pathways, with each element $w_{i,j}$ describes whether pathway describing whether drug i is associated with pathway j.

[1] https://go.drugbank.com
[2] https://www.kegg.jp
[3] https://www.vumc.org/cpm/deb2

- *I4:* The type of activities of the drug pathway may vary, such as metabolic, protein, and drug action. B matrix describes the association of pathway subjects with pathways, where each element $b_{i,j}$ shows whether pathway subject *j* is related to pathway *i*.
- *I5:* IND matrix depicts the drug-indication relation where each element $ind_{i,j}$ shows whether drug *i* is used to treat indication *j*.
- *I6:* SE matrix represents the drug-side effect relation where each element $se_{i,j}$ describes whether drug *i* causes side effect *j*.
- *I7:* H matrix outlines the drug-chemical substructure relation where each element $h_{i,j}$ refers to whether the drug *i* has chemical substructure *j*.
- *I8:* AS matrix demonstrates the drug-anatomical subgroup of ATC code relation where each element $as_{i,j}$ refers to whether drug *i* affects organ or system *j*.
- *I9:* ATS matrix shows the interaction between drugs and the anatomical and therapeutic subgroup of ATC codes. Each element $ats_{i,j}$ indicates whether drug *i* impacts a specific organ and its corresponding therapeutic subgroup *j*.
- *I10:* $ATPS$ matrix illustrates the relationship between drugs and the ATC code's anatomical, therapeutic, and pharmacological subgroup. Each element $atps_{i,j}$ indicates whether drug *i* acts on a particular organ and possesses its corresponding therapeutic and pharmacological subgroup *j*.

C. Meta-Path Construction

After constructing the heterogeneous network, the next step involves creating meta-paths to extract relationships between drugs through other entities. Meta-paths [16] are crucial in measuring relationships and similarities within a heterogeneous graph. Furthermore, meta-paths are represented by a commuting matrix. Meta-path and commuting matrix are defined below.

Definition 2 (Meta-path). A meta-path P is a path on the network schema diagram $T_G = (A, R)$, and is represented in the shape of $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} A_{L+1}$, describing a composite relationship $R = R1 \circ R2 \circ \cdots \circ R$ between entities A_1 and A_{L+1}, where $\circ$ denotes composition operator association, and length of P is L.

Definition 3 (Commuting matrix). Given a network G, a commuting matrix M_P for a meta-path $P = (A_1 A_2 \cdots A_{L+1})$ is defined as $M_P = (G_{A_1 A_2} G_{A_2 A_3} \cdots G_{A_l A_{L+1}})$, where $G_{A_i A_j}$ is the adjacency matrix between types A_i and A_j. $M_P(i, j)$ represents the number of path instances between entity $x_i \in A_1$ and entity $y_i \in A_{L+1}$ under meta-path P.

To design meta-paths for our DDI prediction model, we conduct a rigorous literature survey to identify key biomedical entities and relationships involved in DDIs. This knowledge contribute to the development of six meta-paths that encapsulate the most crucial and frequently occurring patterns of drug interactions with these entities. These meta-paths represent the collective understanding of the mechanisms involved in DDIs.

For instance, let's consider a meta-path between two drugs: *drug* $\xrightarrow{se}$ *disease* $\xleftarrow{se^T}$ *drug* indicating that two drugs cause the same side effect. We can construct the adjacency matrix between drugs and side effects as $G_{drug,side-effect}$. Then, the commuting matrix computed using the meta-path is $G_{drug,side-effect}$ and $G^T_{drug,side-effect}$, represented as $SE * SE^T$, where each element denotes the number of side effects caused by this pair of drugs. We can generate six meta-paths based on a given network schema, considering the different types of entities and their interactions described in Section III-B.

Different meta-paths assess the similarities between two drugs from various perspectives. PID-1 calculates the similarity of two drugs based on their common target proteins and species. If two drugs share the same target protein, there will be a path between them through that protein. This meta-path starts and ends with a drug node but traverses through an intermediate species node, enabling exploration of the influence of species-specific protein interactions on drug interactions. The choice of PID-1 is motivated by the fact that species-specific protein interactions can play a crucial role in drug interactions, as drugs may interact differently with proteins of different species, leading to distinct interaction patterns. Meta-path PID-2 measures how similar two drugs are based on their shared pathways and pathway subjects. In addition, PID-3 and PID-4 meta-paths measure drug pairs' similarity based on their relation with diseases. Meta-path PID-3 considers common diseases cured by drug pairs, whereas PID-4 considers common side effects caused by drug pairs. Moreover, PID-5 connects two drugs based on their shared chemical substructures, indicating their structural similarity. Chemical substructures of drugs are represented as SMILES strings and converted into MACCS keys, a binary fingerprint representing 167 keys. Finally, PID-6 determines the ATC code-based similarity of drug pairs. ATC code illustrates the affected organs of drugs and the therapeutic, chemical, and pharmacological properties of drugs. If drug pairs affect the same organ and have the same therapeutic, chemical, and pharmacological properties, then the drug pair will be connected through a meta-path. Thus, HIN can naturally provide different similarities between drugs with different meta-path-based semantics. Meta-paths used in our network are illustrated in Fig. 1.

IV. Methodology

This section introduces our methodology for DDI prediction, starting with the baseline model, `HAN-DDI`. In `HAN-DDI`, we extract meta-path topological features to predict drug interactions. Subsequently, we present the limitations of this model and propose an improved approach, `HAN-DDI`, which incorporates a multi-layer Heterogeneous Graph Attention Network consisting of an encoder for drug feature generation and a decoder for DDI prediction. Fig. 2 illustrates the system architecture of our `HAN-DDI` model.

A. Meta-Path Topological Features-Based Model

We first explain our model `HIN-DDI`, where we extract meta-path topological features to perform DDI prediction. Four topological features of meta-paths on heterogeneous networks are utilized in our model:

- *Path count:* This feature calculates the number of path instances between two entities for a given meta-path R.

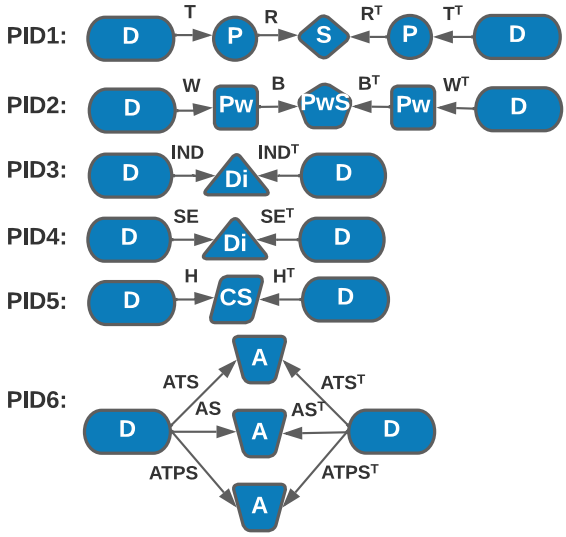

Fig. 1. Meta-paths used in HAN-DDI.

The path count is stored in the corresponding meta-path's commuting matrix.

- *Normalized path count:* The normalized path count computes the number of paths between two network entities through the entire communication network and divides it by the full connectivity of each network entity.
- *Random walk-based normalized path count:* In this feature, Random Walk (RW) is used to normalize the number of path occurrences based on the overall connectedness of the network. The Random Walk is calculated as $PC_R(a_i, a_j)/PC_R(a_i, \circ)$, where $PC_R(a_i)$ are row-wise summations.
- *Symmetric random walk-based normalized path count:* This feature involves a symmetric random walk, considering the random two-way walk between entities.

After extracting the topological features with each meta-path for drug pair relations, we obtain 24 features for each drug pair. The objective is to predict whether two drugs interact or not. For each drug pair relation, we obtain 24 features, and various ML algorithms such as SVM, Logistic Regression, Random Forest, and Neural Network are used for DDI prediction.

However, the complexity of drug relations may require additional features beyond handcrafted meta-paths. Each meta-path conveys distinct semantics and similarity values, potentially affecting drug interactions differently. Thus, it becomes essential to determine the crucial meta-paths and incorporate their relevance into our model.

B. Heterogeneous Graph Attention Network Model

In response to the limitation mentioned above, we propose the Heterogeneous Graph Attention Network (HAN-DDI) for DDI prediction, drawing inspiration from HAN [19]. HAN-DDI consists of an encoder responsible for generating drug embeddings and a decoder for predicting drug interactions.

Our heterogeneous graph contains comprehensive information and rich semantics, incorporating various biological entities and their interactions. The encoder learns drug embeddings in two steps: firstly, drug embeddings are learned for each meta-path using the Heterogeneous Graph Attention Model, resulting in diverse meta-path-based embeddings that capture different connection semantics. To obtain a high-quality embedding, we compute meta-path-level attention scores to determine the significance of each meta-path and then aggregate the embeddings and attention scores. This process generates the final output embedding. The decoder uses these drug embeddings to assign scores to drug-drug edges, indicating the likelihood of a DDI.

TABLE II
NOTATIONS AND EXPLANATIONS OF HAN-DDI

Explanation	Notation
Meta-path	ν
Initial node feature	h
Type-specific transformation matrix	M_ν
Projected node feature	$h^{'}$
Importance of meta-path based node pair (i, j)	$e^{\nu}_{i,j}$
Weight of meta-path based node pair (i, j)	$\alpha^{\nu}_{i,j}$
Meta-path based neighbors	N^{ν}
Semantic-specific node embedding	z^{ν}
Semantic-level attention vector	q
Importance of meta-path ν	w^{ν}
Weight of meta-path ν	β^{ν}
The final embedding	Z

The HAN-DDI model is designed to handle heterogeneous graph data effectively, capturing subtle differences between nodes and meta-paths. Table II presents the notations used throughout the article for ease of reference.

1) Node's Feature Extraction: In the feature extraction step, we extract node features for the drugs. Specifically, we focus on the chemical substructures of drugs, represented as SMILES strings. To create drug features, we employ the ESPF [39] algorithm. ESPF decomposes the SMILES string into frequent substructures, selecting the most significant ones based on a frequency threshold. These substructures provide informative features for the drugs, which will be utilized in the subsequent steps of the HAN-DDI model.

2) Encoder: Drug Representation Learning: The encoder layer in our HAN-DDI model aims to learn drug embeddings by utilizing weighted neighborhood aggregation. However, since different types of nodes (e.g., drugs, proteins, pathways, etc.) have diverse feature spaces, we first create type-specific transformation matrices (M_i) for each kind of node (e.g., a node of type ν_i) to project their features into the same feature space. The projection procedure is defined as follows:

$$h'_i = M_{\nu_i} \circ h_i \tag{1}$$

Here, h and h' represent the original and projected features of node i, respectively, and $\circ$ denotes the inner product between the two matrices.

Next, we introduce node-level attention to learn the importance of meta-path-based neighbors and aggregate them to obtain meta-path-specific node embeddings. Additionally, we define meta-path level attention to determine the importance of different meta-paths and their weights in combining multiple

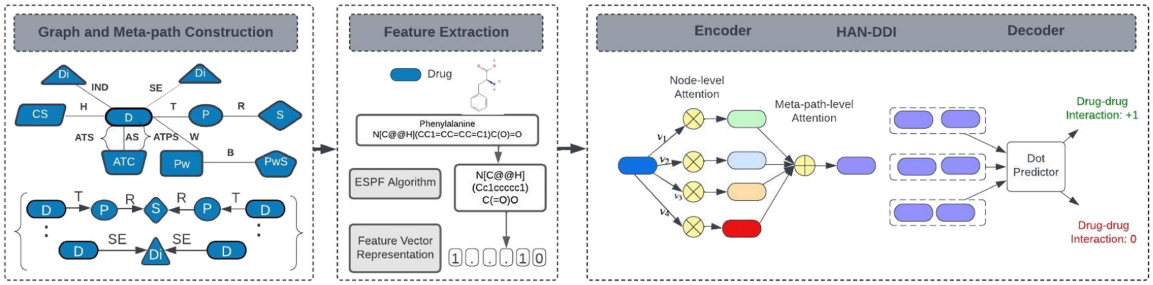

Fig. 2. System architecture of HAN-DDI.

meta-path-specific drug embeddings into one comprehensive drug embedding.

Node Level Attention In our model, meta-paths are used to propagate information from local neighbors, but each node's meta-path-based neighbors may have varying impacts on the target node. Each node's meta-path-based neighbors play different roles and affect the learning of the node's embedding differently. To address this, we introduce node-level attention for each meta-path to learn node representations.

Given a node pair (i, j) connected by a meta-path ν, the node-level attention e_{ij}^{ν} determines node j's significance for node i. The significance of the meta-path-based node pair (i, j) can be expressed as follows:

$$e_{ij}^{\nu} = att_{node}(h_i', h_j'\nu) \tag{2}$$

where Att_{node} represents the deep neural network responsible for node-level attention. For a given meta-path, Att_{node} is shared by all meta-path-based node pairs. Masked attention is employed to consider structural information, and we compute e_{ij}^{ν} only for nodes $j \in N_i^{\nu}$, where N_i^{ν} denotes the meta-path-based neighbors of node i (including itself). We then normalize the weights of all meta-path-based neighbors using a Softmax function to obtain the attention coefficient α_{ij}:

$$\alpha_{ij}^{\nu} = \text{softmax}_j(e_{ij}^{\nu}) \tag{3}$$

In the subsequent $(i+1)$-layer, the meta-path-based embedding of node i is aggregated by the embeddings of its neighbors at the i-layer, with the corresponding attention coefficients applied as follows:

$$z_i^{\nu} = \delta\left(\sum_{j \in N^{\nu}(i)} (\alpha_{ij}^{\nu} \circ h_j')\right) \tag{4}$$

where z_i^{ν} represents the meta-path ν-learned embedding of node i and δ is a non-linear activation function, such as RELU.

To make the learning process more robust, multi-head attention is utilized. Specifically, K attention mechanisms are applied individually to perform the feature transformation specified by Equation 4. The resulting modified features are concatenated (represented as $||$), resulting in the output feature representation as a vector:

$$z_i^{\nu} = ||_{k=1}^{K} \delta\left(\sum_{j \in N^{\nu}(i)} (\alpha_{ij}^{\nu} \circ h_j')\right) \tag{5}$$

This allows the model to dynamically assign greater aggregate weights to nearby nodes that are more relevant to the DDI prediction task, making the embedding of nodes more effectively aggregated based on dynamic weighting. This property makes our technique highly effective for representation learning.

Meta-path Level Attention To combine multiple meta-path-specific representations for each node, we use a meta-path-level attention approach. The objective is to learn the weight of each meta-path ν based on the following equation:

$$w^{\nu} = \sum_{i \in V} q^T tanh(W \circ z_i^{\nu} + b) \tag{6}$$

where W is the weight matrix, b denotes the bias vector, and q represents the meta-path-level attention vector. The attention scores for a meta-path ν are then normalized using the softmax function as follows:

$$\beta^{\nu} = \frac{exp(w^{\nu})}{\sum_{t=1}^{T} w^t} \tag{7}$$

where T is the number of meta-paths. The final representation for each node i is then obtained by aggregating the meta-path-specific representations as follows:

$$z_i = \sum_{t=1}^{T} \beta^t z_i^{\nu} \tag{8}$$

3) Decoder: DDI Learning: The objective of the decoder is to learn whether drug pairs interact using the drug representations obtained from the encoder. The decoder assigns a score to each drug pair (v_i, v_j), expressing the probability of interaction. The dot predictor function is used as the decoder:

$$\gamma(z_x, z_y) = z_x \cdot z_y. \tag{9}$$

After performing an element-wise dot product between the corresponding drug features, a scalar score is obtained for each edge.

Following that, we apply the decoder output to a sigmoid function: $y_{x,y} = \sigma(\gamma(z_x, z_y))$, which generates a prediction score, Y', ranging from 0 to 1. A score close to 1 indicates a high likelihood of interaction between two drugs, whereas a score close to 0 indicates a less likely interaction.

4) Model Training: The entire encoder-decoder architecture is trained as a binary classification problem by minimizing a

binary cross-entropy loss function defined as:

$$L = -\sum_{i=1}^{N} Y_i \log Y_i' + (1 - Y_i) \log(1 - Y_i') \qquad (10)$$

where N is the total number of samples, Y_i is the actual label, and Y_i' is the predicted score.

V. Experimental Results

In this section, we conduct a comprehensive evaluation of our proposed model for drug-drug interaction (DDI) prediction through extensive experiments. The primary objective is to compare the performance of our model with state-of-the-art baseline methods using several accuracy metrics. The model used to predict the DDIs of existing drugs may not be as effective as those used to predict the DDIs of new drugs. Therefore, we assess our model's performance for new and existing drugs. For DDI prediction, we explore two main graph representations: the DDI graph and the heterogeneous graph. Heterogeneous graph construction is described in Section III-B. The DDI graph is a traditional graph representation commonly used in drug-drug interaction studies. In this graph, nodes represent individual drugs, and edges between nodes indicate known drug-drug interactions.

It is important to note that the dataset is imbalanced and skewed, where the number of positive samples (DDIs) is significantly smaller than the number of positive negative samples (non-interactions). To address this issue, we employ negative sampling. For each positive example, we use one negative sample. In addition, we conduct k-fold cross-validation (k = 5). In each fold, the dataset is randomly divided into five subsets, and the model is trained and evaluated five times, each time using a different subset as the testing set and the remaining subsets as the training set. The cross-validation experiment is conducted to ensure `HAN-DDI` model's robustness and consistency across different data splits, reaffirming its generalization ability. Our experiments employ several accuracy metrics, including precision, recall, F-1 score, and area under the receiver operating characteristic curve (AUC-ROC). The use of k-fold cross-validation ensures that our evaluation is statistically sound and provides reliable performance estimates. In the subsequent sections, we provide a detailed description of the parameters used, explain our experimental setup, present the results of the subset analysis, and finally, provide a thorough analysis of our overall results.

A. Parameters Used

For model training and optimization, we employ an end-to-end optimization method for our `HAN-DDI` model. We simultaneously optimize all trainable parameters and propagate loss function gradients through the model's encoder and decoder components. We use the Adam optimizer with a learning rate of 0.005 and a dropout rate of 0.6. The number of attention heads and hidden units used are set to 8 and 16, respectively. These parameter values were chosen as they demonstrated a good balance between model performance and computational efficiency. The model does not take a long time to run and shows promising accuracy results, which indicates that the chosen hyperparameters contribute to an efficient and effective DDI prediction process. The parameters used in this model are outlined in Table III.

TABLE III
Hyper-Parameter Settings

Learning rate	0.005
Number of heads	8
Hidden units	8
Dropout	0.6
Weight decay	0.001
Number of epochs	200
Patience	100

B. Baseline Methods

We compare the performance of `HAN-DDI` against several baseline methods, covering different types of models. After generating drug node embeddings from these baseline methods, we concatenate them and use the concatenated embeddings as features for drug pairings. We then feed these concatenated embeddings to a machine learning classifier. To measure the performance of each model, we use various accuracy metrics, including precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Details of the baselines are summarized below based on their types.

- *Graph Embedding on DDI:* We use DeepWalk [40] and node2vec [41] to generate low-dimensional feature representations of drug nodes based on neighborhood information in the DDI network.
- *Graph Neural Network on DDI:* We use GNN architectures on DDI graphs to learn the representation of drugs. We select three standard GNN-based methods: GCN [42], GAT [43], and GraphSAGE [44].
- *Graph Neural Network on Heterogeneous Graph:* We apply these common GNN models to our heterogeneous graph to learn drug embeddings.
- *Graph Neural Network on Homogeneous Graphs:* We construct various homogeneous graphs consisting of drug nodes, where edges among drug nodes are constructed based on their relation to other entities if they share target proteins, cause side effects or possess similar chemical substructures. We have three different graphs as described below-
  - HG1: Node type: Drugs; Edge: Drug Nodes sharing the same target proteins
  - HG2: Node type: Drugs; Edge: Drug Nodes causing the same side effects
  - HG3: Node type: Drugs; Edge: Drug Nodes possessing the same chemical substructures

  To learn drug node embedding, we apply GCN to these homogeneous graphs.
- *HIN-DDI:* We use HIN-DDI [10], our meta-path topological feature-based model, as a baseline. HIN-DDI first constructs a DDI-related HIN, and then the instance numbers of selected meta-paths are used as features to represent the interactions between drugs; finally, a neural network is trained as the classifier for predicting DDIs.

TABLE IV
PERFORMANCE COMPARISONS OF HAN-DDI WITH BASELINE MODELS FOR EXISTING DRUGS

Model	Method	F1	RECALL	PRECISION	AUROC
GE on DDI graph	DeepWalk	70.81	69.73	71.92	69.80
	Node2Vec	69.96	68.56	71.42	67.82
	HIN-DDI	74.08	74.02	74.15	74.05
GNN on DDI graph	GAT	82.17	82.95	81.41	82.12
	GraphSAGE	81.49	83.22	79.84	81.17
	GCN	82.39	84.55	80.34	82.88
GNN on Heterogeneous graph	GAT	84.15	84.38	83.92	84.33
	GraphSAGE	85.62	86.39	84.87	83.23
	GCN	85.91	87.16	84.69	87.22
Homogeneous Graphs	HG1	85.85	88.21	83.62	86.78
	HG2	85.01	88.57	81.72	85.62
	HG3	87.17	86.72	87.63	86.26
	DDIMDL	77.73	71.82	84.71	95.12
ML Classifier on drugs' FR	Concatenated Drug Features	82.19	83.18	81.22	78 .91
	SkipGNN	87.12	87.93	86.32	88.6
	Decagon	89.48	90.88	88.12	86.52
	MUFFIN	92.67	93.6	91.75	95.26
	HAN-DDI	**96.63**	**97.92**	**95.38**	**95.44**

TABLE V
PERFORMANCE COMPARISONS OF HAN-DDI WITH BASELINE MODELS FOR NEW DRUGS

Model	Method	F1	RECALL	PRECISION	AUROC
GE on DDI graph	**Node2Vec**	50.86	51.48	50.26	50.27
	DeepWalk	47.48	44.92	50.35	45.63
	HIN-DDI	66.90	65.72	68.12	66.82
GNN on DDI graph	GAT	75.15	75.83	74.49	74.32
	GraphSAGE	75.02	75.96	74.11	74.52
	GCN	75.58	76.81	74.39	75.92
GNN on Heterogeneous graph	GAT	77.15	77.58	76.72	78.04
	GraphSAGE	76.40	76.71	76.09	75.36
	GCN	78.71	81.13	76.43	79.88
Homogeneous Graphs	HG1	75.98	75.62	76.34	75.18
	HG2	76.03	77.87	74.28	77.62
	HG3	77.39	76.38	78.42	77.44
	DDIMDL	68.41	67.68	69.16	71.48
ML Classifier on drugs' FR	Concatenated Drug Features	75.60	77.16	74.1	73.94
	SkipGNN	79.52	80.15	78.89	77.96
	MUFFIN	81.51	82.43	80.62	82.95
	HAN-DDI	**84.76**	**85.92**	**83.63**	**83.84**

- *DDIMDL:* We use DDIMDL [45], which combines various drug features, including chemical substructures, targets, enzymes, and pathways, with deep neural network (DNN) to build a model for DDI prediction.
- *ML Classifier on Drug Functional Representation (FR):* Principal Component Analysis (PCA) [46] is a dimensionality-reduction approach commonly used to reduce the dimensionality of substantial data sets. This method generates a feature vector for each drug based on the PCA representation of the drug-target protein interaction matrix, the PCA representation of the drug-chemical substructure possession matrix, and the PCA representation of the drug-side effects matrix.
- *SkipGNN:* We employ SkipGNN [35], a graph neural network method that predicts molecular interactions by aggregating data from second-order and direct interactions.
- *MUFFIN:* MUFFIN [34] combines message-passing neural networks with TransE to capture drug structure representation from molecular maps and semantic features from KG.
- *Decagon:* Decagon [5], a graph convolutional network model, is employed to predict multi-relational links in heterogeneous networks, using end-to-end learning to produce drug embeddings and predict DDIs.

C. Comparison With Baselines

In our experiments, we evaluate the performance of our HAN-DDI model and several state-of-the-art baseline models on existing and new drugs. For new drug prediction, we partition the dataset so that 20% of drugs are unseen during training and only appear in the testing set. This 20% of drugs are considered as new drugs and independent from the training set.

Experimental results for existing and new drugs are shown in Tables IV and V. For existing drugs, our model achieves the best results for F-1 score, Recall, Precision, and AUROC scores as 96.63%, 97.92%, 95.98%, and 95.44%, respectively, while the highest-scoring baseline, MUFFIN, obtain 92.67%, 93.6%, 91.75%, and 95.26% for the respective metrics. HIN-DDI yields scores of 74.08%, 74.02%, 74.15%, and 74.05% for

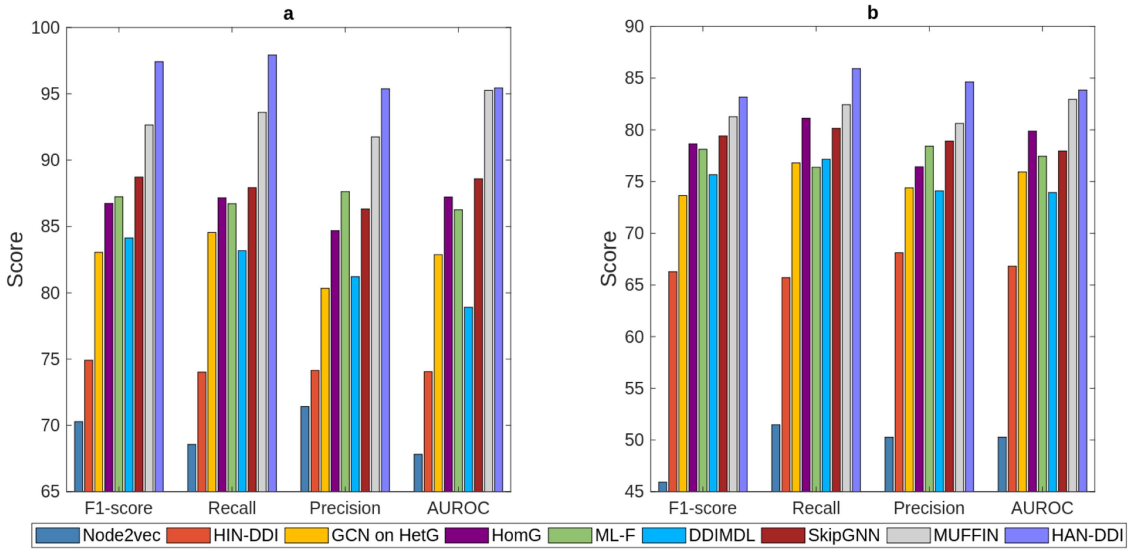

Fig. 3. Performance comparison of models for a) existing and b) new drugs.

the same metrics. Similarly, for new drugs, our model achieves scores of 84.76%, 85.92%, 84.63%, and 83.84%, while the best-performing baseline, MUFFIN, obtained 81.51%, 82.43%, 80.62%, and 82.95% for F-1 score, Recall, Precision, and AUROC, respectively. HIN-DDI generated accuracy scores of 66.9%, 65.72%, 68.12%, and 66.82% for these metrics. Our model demonstrates its generalization ability by predicting interactions for both existing and new drugs.

To statistically compare the performance of our proposed HAN-DDI model with baseline methods, we conduct paired t-tests using a significance level of $p < 0.05$. The null hypothesis is that there is no significant difference in performance between HAN-DDI and each baseline method. For existing drugs, the p-values for F-1 score, Recall, Precision, and AUROC are all less than 0.05, indicating that HAN-DDI significantly outperforms all baseline models in terms of these metrics. Similarly, for new drugs, the p-values for F-1 score, Recall, Precision, and AUROC are all less than 0.05, indicating significant improvement in the performance of HAN-DDI compared to the baselines. These statistical tests further support the superiority of our proposed model.

In further analysis of baseline models on existing drugs, Node2Vec shows the best performance on the DDI graph among graph embedding models. For GNN models, GCN performs best on both DDI and heterogeneous graphs, with the latter yielding better results due to considering more nodes and edges. Among homogeneous graphs, HG3, which forms edges for drug nodes sharing similar chemical substructures, achieves higher accuracy, indicating the significance of chemical substructures in drug node representation and DDI characterization. DDIMDL achieves comprehensive evaluation scores, while MUFFIN emerges as the top-performing baseline with an F1 score of 92.67%. However, in some cases, HAN-DDI outperforms baseline methods by a significant 19% performance gain.

Our research findings are significant for drug safety and patient care. For new drugs, Node2Vec from graph embedding models outperforms DeepWalk. GCN produces the best results for GNN models on DDI graphs, but GCN outperforms it on the more diverse heterogeneous graph. Similar to existing drugs, HG3 surpasses other homogeneous graph variants in evaluation metrics for new drugs. Furthermore, MUFFIN performs the best among the baseline models. Our in-depth analysis revealed that HAN-DDI excels in predicting DDIs for existing and new drugs, demonstrating its strong generalization ability.

Furthermore, we present our model result with the best-performing method from each baseline category. Fig. 3 compares all these models, with HAN-DDI outperforming all others. Notably, GNN-based models, including HAN-DDI, show remarkable performance because of its capacity to analyze graph structure data. The capability of GNN to represent the interactions between graph nodes is a milestone in graph analysis research. Moreover, message passing between graph nodes allows GNNs to capture graph dependence. Heterogeneous graph Attention network is a variant of GNN. Our proposed HAN-DDI model represents a significant leap in drug-drug interaction prediction, offering a new and superior approach compared to state-of-the-art baseline methods. This superiority can be attributed to the innovative multi-layer heterogeneous graph attention network architecture, combined with node-level and meta-path level attention mechanisms. The experimental results validate the effectiveness of HAN-DDI and underscore its potential as a game-changing tool for advancing drug safety research and patient care.

D. Case Study: Prediction and Validation of Novel DDI Predictions

We conduct an evaluation to determine the effectiveness of our HAN-DDI model in predicting novel drug-drug interactions (DDIs). To do this, we select 10 specific drug pairs that lack DDI information in TWOSIDES.

TABLE VI
NOVEL DDI PREDICTIONS AND THEIR VALIDATION

Drug1	Drug2	TWOSIDES Label	Predicted Score	DrugBank Label
Carbamazepine	Cimetidine	0	0.9549	1
Ampicillin	Tacrolimus	0	0.9931	1
Sildenafil	Cimetidine	0	0.9683	1
Loratadine	Isradipine	0	0.9336	1
Quinolones	Citalopram	0	0.91	1
Hydroxychloroquine	Loratadine	0	0.9082	1
Fluvastatin	Metronidazole	0	0.9812	1
Bexarotene	Maprotiline	0	9.9993e-10	0
Amoxapine	Econazole	0	6.8256e-09	0
Nabilone	Oxaprozin	0	4.1440e-08	0

TABLE VII
PROPORTIONS OF CORRECTLY PREDICTED DDIS AMONG TOP 100 HIGHLY SCORED PREDICTIONS

Method	Proportion of Correctly Predicted DDIs (%)
MUFFIN	91.00
HAN-DDI	94.00

After training our HAN-DDI model using the TWOSIDES dataset, we use it to predict interactions for these selected drug pairs. The predicted scores for these drug pairs are presented in Table VI. As we see from the table, our model predicted scores above 90% for seven pairs, indicating a high probability of interaction despite the zero TWOSIDES labels. To validate these predictions, we cross-reference them with DrugBank, which is not used for the training. We confirm that DrugBank includes interactions for all seven pairs.

In contrast, the predicted scores are minimal for the three-drug pairs in Table VI. Both TWOSIDES and DrugBank data indicate that these pairs do not interact. This consistent prediction between our model and the existing datasets highlights the reliability of our HAN-DDI model in identifying DDIs.

E. Case Study: Proportions of Correctly Predicted DDIs Among Top 100 Highly Scored Predictions

To assess our model's ability to rank and identify the most relevant drug pairs among a large pool of potential interactions, we calculate the proportions of correctly predicted drug-drug interactions (DDIs) among the top 100 highly scored predictions. Our results, presented in Table VII, show that our HAN-DDI model outperforms the best-performing baseline method MUFFIN in accurately identifying relevant DDIs within the top predictions. This analysis demonstrates the effectiveness of our model in prioritizing potential interactions, which can streamline experimental validation efforts and contribute to safer and more effective drug combination strategies.

F. Case Study: Performance Across Disease Categories

To assess our model's performance across different disease categories, we experiment on drugs associated with cardiovascular diseases, infectious diseases, and cancer. The results, presented in Table VIII, show that our HAN-DDI model achieves high accuracy in predicting DDIs for drugs used in these disease categories. These findings suggest that our model has broad applicability and can contribute significantly to drug discovery and development efforts across diverse disease types.

In order to assess the performance of our proposed HAN-DDI model across different disease categories, we experiment on drugs associated with cardiovascular diseases, infectious diseases, and cancer. For each disease, we extract the drugs used for these diseases. Then, we evaluate using k-fold cross-validation, and the model's accuracy metrics are calculated for each disease.

Table VIII presents the performance of HAN-DDI across different disease categories. The results indicate that our HAN-DDI model demonstrates consistent and promising performance. It achieves high accuracy in predicting DDIs for drugs used in cardiovascular diseases, infectious diseases, and cancer. Furthermore, it maintains robustness and effectiveness in other categories. These findings suggest that our model has a broad application scope and can be valuable in predicting DDIs across diverse disease types, facilitating drug discovery and development.

G. Ablation Study

In this section, we conduct an ablation study to assess the impact of node-level and meta-path-level attention on the overall performance of our HAN-DDI model. We develop two HAN-DDI variants for this study:

- HAN-DDI-MP: We use meta-path-level attention acquired from HAN-DDI while a random matrix was used for node-level attention.
- HAN-DDI-N: We use node-level attention generated by HAN-DDI, and for meta-path-level attention, we assigned equal weights to each meta-path.

For HAN-DDI-MP, the average F1-score, Recall, Precision, and AUROC scores are 93.65%, 94.18%, 90.72%, and 80.88%, respectively, for existing drugs. For HAN-DDI-N, our average F1-score, Recall, Precision, and AUROC scores are 92.87%, 93.19%, 89.75%, and 78.48%, respectively.

Performance analysis in Fig. 4 shows that HAN-DDI outperforms both HAN-DDI-MP and HAN-DDI-N. This indicates that both node-level and meta-path-level attention play important roles in effectively capturing different meta-path information of nodes. However, meta-path-level attention has a more significant impact on performance than node-level attention, as seen by the lower performance of HAN-DDI-N compared to HAN-DDI-MP. This finding demonstrates the significance of meta-path-level attention in our model.

TABLE VIII
PERFORMANCE OF HAN-DDI ACROSS DISEASE CATEGORIES

Disease Category	F1	RECALL	PRECISION	AUROC
Cardiovascular	93.65	94.11	92.58	92.03
Infectious	95.08	94.72	93.39	93.90
Cancer	94.29	93.18	92.86	92.47

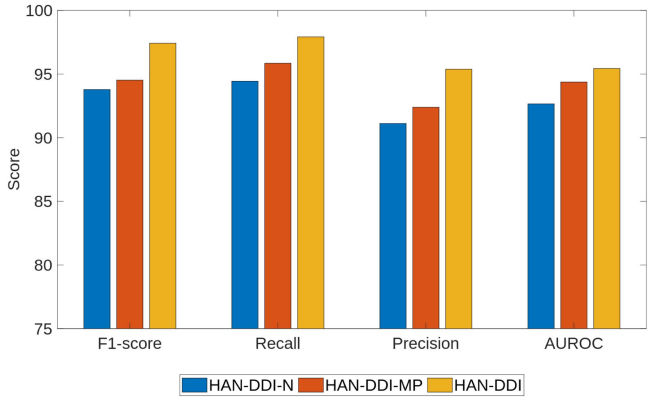

Fig. 4. Performance comparison of HAN-DDI with its variants.

VI. CONCLUSION

Our study introduces HAN-DDI, a novel Heterogeneous Graph Attention Network for predicting drug-drug interactions (DDIs). By leveraging heterogeneous graphs and meta-paths, HAN-DDI performs better in predicting existing and new drug interactions. The model's ability to process complex heterogeneous graphs enables comprehensive and precise DDI predictions. We anticipate that HAN-DDI will pave the way for more effective drug safety research and personalized medicine approaches.

This approach opens up a promising avenue for DDI predictions, especially as we progress toward personalized medicine. We anticipate a growing integration of graph-based models in DDI prediction due to their effective handling of multifaceted drug interactions. In addition to its potential for improving drug safety, HAN-DDI could also be used to address other challenges in drug discovery and development, such as identifying new drug combinations, polypharmacy side effect prediction, and developing personalized treatment plans.

REFERENCES

[1] X. Chen et al., "Drug-target interaction prediction: Databases, web servers and computational models," *Brief. Bioinf.*, vol. 17, no. 4, pp. 696–712, 2016.
[2] X. Chen, N.-N. Guan, Y.-Z. Sun, J. Li, and J. Qu, "MicroRNA-small molecule association identification: From experimental results to computational models," *Brief. Bioinf.*, vol. 21, no. 1, pp. 47–61, 2020.
[3] C.-C. Wang, Y. Zhao, and X. Chen, "Drug-pathway association prediction: From experimental results to computational models," *Brief. Bioinf.*, vol. 22, no. 3, 2020, Art. no. bbaa061.
[4] H.-C. Yi, Z.-H. You, D.-S. Huang, and C. K. Kwoh, "Graph representation learning in bioinformatics: Trends, methods and applications," *Brief. Bioinf.*, vol. 23, no. 1, 2022, Art. no. bbab340.
[5] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, 2018.
[6] Y. Qiu, Y. Zhang, Y. Deng, S. Liu, and W. Zhang, "A comprehensive review of computational methods for drug-drug interaction detection," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, pp. 1968–1985, 2021.
[7] Y. Deng et al., "META-DDIE: Predicting drug-drug interaction events with few-shot learning," *Brief. Bioinf.*, vol. 23, 2021, Art. no. bbab514.
[8] C. Han, C.-C. Wang, L. Huang, and X. Chen, "MCFF-MTDDI: Multi-channel feature fusion for multi-typed drug-drug interaction prediction," *Brief. Bioinf.*, vol. 24, 2023, Art. no. bbad215.
[9] W. Zhang et al., "SFLLN: A sparse feature learning ensemble method with linear neighborhood regularization for predicting drug-drug interactions," *Inf. Sci.*, vol. 497, pp. 189–201, 2019.
[10] F. Tanvir, M. I. K. Islam, and E. Akbas, "Predicting drug-drug interactions using meta-path based similarities," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol.*, 2021, pp. 1–8.
[11] E. Akbas and P. Zhao, "Truss-based community search: A truss-equivalence based indexing approach," in *Proc. VLDB Endowment*, vol. 10, no. 11, pp. 1298–1309, 2017.
[12] M. I. Islam, F. Tanvir, G. Johnson, E. Akbas, and M. E. Aktas, "Proximity-based compression for network embedding," *Front. Big Data*, vol. 3, 2020, Art. no. 54.
[13] W. Tanner, E. Akbas, and M. Hasan, "Paper recommendation based on citation relation," in *Proc. IEEE Int. Conf. Big Data*, 2019, pp. 3053–3059.
[14] G. R. Lopes, M. M. Moro, L. K. Wives, and J. P. M. De Oliveira, "Collaboration recommendation on academic social networks," in *Proc. Int. Conf. Conceptual Model.*, 2010, pp. 190–199.
[15] R. Sharan, I. Ulitsky, and R. Shamir, "Network-based prediction of protein function," *Mol. Syst. Biol.*, vol. 3, no. 1, 2007, Art. no. 88.
[16] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "Pathsim: Meta path-based top-k similarity search in heterogeneous information networks," in *Proc. VLDB Endow.*, vol. 4, pp. 992–1003, 2011.
[17] I. Abdelaziz, A. Fokoue, O. Hassanzadeh, P. Zhang, and M. Sadoghi, "Large-scale structural and textual similarity-based mining of knowledge graph to predict drug-drug interactions," *J. Web Semantics*, vol. 44, pp. 104–117, 2017.
[18] Y. Zheng, H. Peng, X. Zhang, Z. Zhao, X. Gao, and J. Li, "DDI-PULearn: A positive-unlabeled learning method for large-scale prediction of drug-drug interactions," *BMC Bioinf.*, vol. 20, 2019, Art. no. 661.
[19] X. Wang et al., "Heterogeneous graph attention network," in *Proc. World Wide Web Conf.*, 2019, pp. 2022–2032.
[20] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, "NLLSS: Predicting synergistic drug combinations based on semi-supervised learning," *PLoS Comput. Biol.*, vol. 12, 2016, Art. no. e1004975.
[21] N. Zong, H. E. Kim, V. Ngo, and O. Harismendy, "Deep mining heterogeneous networks of biomedical linked data to predict novel drug–target associations," *Bioinf.*, vol. 33, pp. 2337–2344, 2017.
[22] R. Çelebi, V. Mostafapour, E. Yasar, Ö. Gümüs, and O. Dikenelli, "Prediction of drug-drug interactions using pharmacological similarities of drugs," in *proc. 26th Int. Workshop Database Expert Syst. Appl.*, 2015, pp. 14–17.
[23] A. Kastrin, P. Ferk, and B. Leskoek, "Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning," *PLoS ONE*, vol. 13, 2018, Art. no. e0196865.
[24] A. Gottlieb, G. Y. Stein, Y. Oron, E. Ruppin, and R. Sharan, "INDI: A computational framework for inferring drug interactions and their associated recommendations," *Mol. Syst. Biol.*, vol. 8, pp. 592–592, 2012.
[25] S. Vilar et al., "Similarity-based modeling in large-scale prediction of drug-drug interactions," *Nat. Protoc.*, vol. 9, pp. 2147–2163, 2014.
[26] W. Zhang, Y. Chen, F. Liu, F. Luo, G. Tian, and X. Li, "Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data," *BMC Bioinf.*, vol. 18, 2017, Art. no. 18.
[27] B. Bumgardner, F. Tanvir, K. M. Saifuddin, and E. Akbas, "Drug-drug interaction prediction: A purely smiles based approach," in *Proc. IEEE Int. Conf. Big Data*, 2021, pp. 5571–5579.
[28] K. M. Saifuddin, B. Bumgardnerr, F. Tanvir, and E. Akbas, "HyGNN: Drug-drug interaction prediction via hypergraph neural network," 2022, *arXiv:2206.12747*.

[29] K. M. Saifuddin, M. I. K. Islam, and E. Akbas, "Drug abuse detection in twitter-sphere: Graph-based approach," in *Proc. IEEE Int. Conf. Big Data*, 2021, pp. 4136–4145.
[30] R. Çelebi, H. Uyar, E. Yasar, Ö. Gümüs, O. Dikenelli, and M. Dumontier, "Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings," *BMC Bioinf.*, vol. 20, 2019, Art. no. 726.
[31] F. Tanvir, K. M. Saifuddin, M. I. K. Islam, and E. Akbas, "Predicting drug-drug interactions using heterogeneous graph attention networks," in *Proc. 14th ACM Int. Conf. Bioinf. Comput. Biol. Health Inform.*, 2023, Art. no. 8.
[32] X. Lin et al., "Comprehensive evaluation of deep and graph learning on drug-drug interactions prediction," *Brief. Bioinf.*, vol. 24, 2023, Art. no. bbad235.
[33] X. Lin, Z. Quan, Z. Wang, T. Ma, and X. Zeng, "KGNN: Knowledge graph neural network for drug-drug interaction prediction," in *Proc. Int. Joint Conf. Artif. Intell.*, 2020, pp. 2739–2745.
[34] Y. Chen, T. Ma, X. Yang, J. Wang, B. Song, and X. Zeng, "MUFFIN: Multi-scale feature fusion for drug-drug interaction prediction," *Bioinf.*, vol. 37, pp. 2651–2658, 2021.
[35] K. Huang, C. Xiao, L. Glass, M. Zitnik, and J. Sun, "SkipGNN: Predicting molecular interactions with skip-graph networks," *Sci. Rep.*, vol. 10, 2020, Art. no. 21092.
[36] Z. Xiong et al., "Multi-relational contrastive learning graph neural network for drug-drug interaction event prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 5339–5347.
[37] T. Li, C.-C. Wang, L. Zhang, and X. Chen, "SNRMPACDC: Computational model focused on Siamese network and random matrix projection for anticancer synergistic drug combination prediction," *Brief. Bioinf.*, vol. 24, 2022, Art. no. bbac503.
[38] C. Shi, Y. Li, J. Zhang, Y. Sun, and P. S. Yu, "A survey of heterogeneous information network analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 29. no. 1, pp. 17–37, Jan. 2017.
[39] L. Glass, J. Sun, K. Huang, and C. Xiao, "Explainable substructure partition fingerprint for protein, drug, and more," in *Proc. NeurIPS Learn. Meaningful Representation Life Workshop*, 2019.
[40] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2014, pp. 701–710.
[41] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 855–864.
[42] T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017, *arXiv:1609.02907*.
[43] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lioì, and Y. Bengio, "Graph attention networks," 2018, *arXiv: 1710.10903*.
[44] W. L. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017.
[45] Y. Deng, X. Xu, Y. Qiu, J. Xia, W. Zhang, and S. Liu, "A multimodal deep learning framework for predicting drug-drug interaction events," *Bioinf.*, vol. 36, pp. 4316–4322, 2020.
[46] H. T. Shen, "Principal component analysis," in *Encyclopedia of Database Systems*, Berlin, Germany: Springer, 2009.

Farhan Tanvir received the PhD degree in computer science from Oklahoma State University. He is a lecturer with the Department of Computer Science, Georgia State University. His broad research interests include machine learning and data mining, specifically bioinformatics/computational biomedicine. He is focusing on knowledge discovery from large network data of interaction among biomedical entities.

Khaled Mohammed Saifuddin received the Bachelor of Science degree in electronics and communication engineering from the Khulna University of Engineering and Technology, Bangladesh. He is currently working toward the Ph.D. degree in computer science with Georgia State University. His research interests include the intersection of graph mining and graph machine learning, where he uncovers intricate patterns and develops intelligent algorithms to reveal new insights. He develops and applies cutting-edge techniques in the fields of social network analysis, bioinformatics, and natural language processing (NLP).

Muhammad Ifte Khairul Islam is working toward the PhD degree with the Department of Computer Science, Georgia State University. His broad research area is data mining and graph mining. Currently, he is working on graph neural network, graph pooling and graph compression. He is building models to retrieve information from social and biological networks.

Esra Akbas received the PhD degree in computer science from Florida State University. She is an assistant professor with the Department of Computer Science, Georgia State University. Her broad research interest is the development of algorithms for mining and analyzing large-scale data with particular emphasis on text and graph-structured data. She mainly focuses on social networks and health and biomedical as the application area.