CNNRMHSA-LGTEN: An Efficient Local Graph Transformer and Convolutional Attention Exchange Network for Drug Target Interaction Prediction

[1]A. Jane, [2]Dr.K. Merriliance, [3]Dr. Mary Immaculate Sheela Lourdusamy, [4]Dr.S. Serina Hingis

1 Research Scholar (Registration No–22121242282005), Department of Computer Applications and Research Centre,

Sarah Tucker College (Autonomous), Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India.

Email: jane77johnson@gmail.com

2 Associate Professor, Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous), Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India. Email: merriliance@gmail.com

3

Head, Department of Computing and Engineering, School of Engineering Technology & Applied Science (SETAS), Heritage Christian University, Accra, Ghana. Email: drsheela09@gmail.com

4 Axxelent Pharma Sciences Pvt.,Chennai.

Email: serinahingis6@gmail.com

Abstract

Drug Target Interaction (DTI) prediction plays a crucial role in drug discovery by reducing experimental cost and accelerating candidate screening. Traditional computational approaches, including similarity-based and network-based methods, often fail to capture complex nonlinear relationships between molecular structures and protein sequences. Recent deep learning models improve prediction accuracy but still struggle to jointly model global molecular topology and fine-grained local chemical substructures that drive binding specificity. In this work, we propose a CNNRMHSA-LGTEN: An Efficient Local Graph Transformer and Convolutional Attention Exchange Network for Drug Target Interaction Prediction for accurate DTI prediction. The proposed framework explicitly constructs *k*hop local concept subgraphs from drug molecular graphs to preserve functional chemical motifs, while a relative multi-head self-attention mechanism encodes both structural and attribute-level dependencies. Protein sequences are modeled using stacked one-dimensional Convolutional Neural Networks to extract hierarchical residue patterns. A gated attention-based fusion module integrates drug and protein representations, followed by a fully connected prediction head. Experimental results demonstrate that the proposed model consistently outperforms state-of-the-art baselines across multiple evaluation metrics, highlighting the effectiveness of combining local subgraph awareness with attention-driven graph representation learning for DTI prediction.

Keywords: Drug Target Interaction prediction, Local Subgraph, Graph Neural Networks, Multi-head SelfAttention, Graph Attention Encoder, Gated Fusion, Deep Learning.

1 Introduction

Drug–target interactions (DTIs) form the foundation of therapeutic efficacy and safety in pharmaceutical research. Identifying whether a chemical compound interacts with a biological target is essential for drug discovery, drug repurposing, and toxicity assessment. However, wet-lab validation of DTIs is time-consuming, costly, and labor-intensive, motivating the development of computational prediction methods [1].

Early DTI prediction approaches relied heavily on similarity-based methods, which infer interactions based on chemical similarity between drugs or sequence similarity between proteins [2]. While intuitive, these methods assume that similar entities behave similarly, an assumption that often breaks down for structurally diverse compounds. Network-based approaches, including random-walk and heterogeneous network inference techniques, model DTIs as bipartite or heterogeneous graphs and exploit global topological relationships to infer unknown interactions [1, 2]. Although effective to some extent, these methods depend strongly on predefined similarity measures and known interaction networks, limiting their ability to generalize to novel drugs or targets.

With the advent of deep learning, neural models based on convolutional architectures were introduced to automatically learn representations from raw drug SMILES strings and protein sequences. DeepDTA demonstrated that convolutional neural networks can effectively model drug–target binding affinity without handcrafted features [3]. While such sequence-based models significantly improved predictive performance, they ignore the intrinsic graph structure of molecules, thereby limiting their ability to capture atom-level interactions critical for binding specificity.

Graph neural networks (GNNs) address this limitation by naturally representing drugs as molecular graphs, where atoms and bonds correspond to nodes and edges. Graph convolutional networks introduced by Kipf and Welling [**?**] were later adopted for DTI prediction, showing that graph-based molecular representations improve prediction accuracy compared to sequence-only approaches [4]. Extensions of GNNbased models further incorporated heterogeneous biomedical information to model complex drug–protein relationships [15]. Despite these advances, most GNN-based DTI models focus on learning global graph representations, which may obscure chemically meaningful local structures.

Local chemical substructures, such as aromatic rings, heterocycles, and functional groups, often determine molecular binding behavior. Network motifs, defined as statistically significant recurring subgraph patterns, were introduced as fundamental building blocks of complex networks [8]. In molecular graphs, these motifs correspond to functional fragments that play a decisive role in biological activity. Fragment-based representations, including extended-connectivity fingerprints, explicitly encode local neighborhoods and have proven effective for molecular property prediction [9]. Recent graph learning studies further demonstrated that localized neighborhood sampling and subgraph-based learning improve representation expressiveness and robustness [10–12].

Attention mechanisms further enhance graph representation learning by allowing models to selectively emphasize informative components. Graph Attention Networks (GATs) assign adaptive importance weights to neighboring nodes, enabling more expressive aggregation than uniform message passing [17]. In the context of drug discovery, fragment-oriented and attention-based DTI models have shown that emphasizing local chemical regions improves interpretability and prediction performance [29,31]. However, most existing attention-based approaches either rely on predefined fragments or apply attention at the global graph level, without explicitly constructing and encoding localized concept subgraphs grounded in molecular topology.

To address these limitations, this work proposes a *CNNRMHSA-LGTEN: An Efficient Local Graph Transformer and Convolutional Attention Exchange Network for Drug Target Interaction Prediction* for DTI prediction. The proposed framework explicitly constructs *k*-hop local subgraphs from drug molecular graphs to preserve functional chemical motifs and encodes them using relative multi-head self-attention. By jointly modeling global molecular context and fine-grained local substructures, the proposed approach learns more discriminative drug representations. The protein branch employs stacked one-dimensional convolutional neural networks to capture hierarchical sequence motifs, and an attention-guided fusion module integrates drug and protein embeddings for robust interaction prediction. To address these limitations, we propose a CNNRMHSA-LGTEN: An Efficient Local Graph Transformer and Convolutional Attention Exchange Network for Drug Target Interaction Prediction for drug–target interaction prediction. The model explicitly constructs k-hop local subgraphs from drug molecular graphs and encodes them using relative multi-head self-attention, enabling the preservation of chemically meaningful functional motifs while mitigating oversmoothing and over-squashing effects. Protein sequences are modeled using stacked one-dimensional convolutional neural networks, which effectively capture hierarchical residue motifs associated with binding sites. An attention-guided fusion module integrates drug and protein representations, resulting in discriminative and robust interaction prediction. The main contributions of this work are summarized as follows: • A local subgraph construction mechanism is introduced to explicitly model functional chemical fragments within drug molecular graphs.

- A relative multi-head self-attention graph encoder is designed to jointly capture structural and semantic dependencies in local subgraphs.

- An attention-refined drug–protein fusion strategy is proposed to enhance interaction-specific representation learning.

- Extensive experiments demonstrate that the proposed approach outperforms existing state-of-the-art DTI prediction models.

2 Related Work

2.1 Similarity-Based and Network-Based DTI Prediction

Early computational approaches for drug–target interaction (DTI) prediction primarily relied on similaritybased assumptions, where chemically similar drugs or sequence-similar proteins were expected to share interaction profiles. Heterogeneous network-based inference methods modeled drugs and targets as nodes connected via similarity and interaction edges, enabling interaction inference through network propagation mechanisms [1]. Random-walk-based strategies further improved prioritization by exploiting global topological information in biological networks [2].

Although effective in data-rich scenarios, these approaches depend heavily on predefined similarity measures and known interactions, limiting their ability to generalize to novel drugs or targets. Moreover, such shallow models lack the representational capacity to capture nonlinear and context-dependent biochemical relationships.

2.2 Deep Learning Models for DTI Prediction

With advances in deep learning, sequence-based neural models were introduced to learn representations directly from raw drug and protein inputs. DeepDTA employed convolutional neural networks (CNNs) to model drug SMILES strings and protein sequences, achieving improved binding affinity prediction without handcrafted features [3]. Subsequent attention-based architectures further enhanced prediction performance by modeling complex interactions across diverse protein families [21].

Despite their success, sequence-only models ignore the intrinsic graph structure of molecules, limiting their ability to capture atom-level interactions that are critical for binding specificity.

2.3 Graph Transformer Networks

Graph neural networks (GNNs) naturally represent molecular structures as graphs, where atoms and bonds correspond to nodes and edges. Early graph convolutional networks, such as the spectral GCN proposed by Kipf and Welling, enabled effective neighborhood aggregation over graph structures and were later adapted for DTI prediction, demonstrating improved performance by learning directly from molecular graphs rather than from hand-crafted descriptors [5,6]. However, many conventional GNNs rely on fixed, localityconstrained aggregation schemes and can suffer from over-smoothing or limited ability to capture long-range dependencies in larger molecular graphs.

Graph Transformer Networks extend the Transformer architecture to arbitrary graphs by integrating selfattention with explicit graph topology information. Dwivedi and Bresson generalized the standard sequence Transformer by making the attention mechanism a function of neighborhood connectivity, incorporating Laplacian eigenvector-based positional encodings, and supporting edge features such as bond types [7]. This formulation closes the gap between classical GNNs and Transformers, allowing attention weights to depend jointly on node features, structural positions, and edge attributes, which is particularly suitable for chemistry where local functional groups and global context both influence activity. In the context of DTI modeling, graph transformers provide a flexible framework to encode drug molecules as relational graphs while capturing both local subgraph motifs and long-range atom–atom interactions within a unified attention-based exchange mechanism.

Further extensions incorporated heterogeneous biomedical networks to jointly model drugs, targets, and side effects [15]. However, standard GNNs often emphasize global graph representations and suffer from over-smoothing, which can obscure chemically meaningful local patterns.

2.4 Local Graph Transformer Networks

To overcome the limitations of global graph embeddings, research has increasingly focused on local subgraph and motif-based representations. Network motifs were formally introduced as statistically significant recurring subgraphs that serve as fundamental building blocks of complex networks [8]. In molecular graphs, such motifs correspond to functional groups and fragments that largely determine chemical reactivity and biological activity.

Fragment-based representations such as extended-connectivity fingerprints (ECFP) explicitly encode local neighborhoods around atoms and have proven highly effective for molecular property prediction [9]. Inductive graph learning methods further demonstrated that localized neighborhood sampling improves scalability and expressiveness [10], while subgraph-based GNNs enhanced robustness by learning from k-hop ego networks

[11].

Hierarchical pooling approaches, such as DiffPool, preserve important substructures by learning multilevel graph representations [12]. Message-passing neural networks implicitly encode local structures but lack explicit interpretability regarding which subgraphs drive predictions [13]. To address this, local-Global Graph Transformer with Memory Reconstruction integrates explicit local subgraph modeling with global attention and memory-based reconstruction, enabling holistic node anomaly evaluation by jointly capturing structural normality and long-range dependencies [14].

2.5 Subgraph-Aware and Attention-Based DTI Models

In the context of DTI prediction, subgraph-aware methods remain relatively limited. Fragment-oriented attention mechanisms explicitly highlight functional fragments contributing to binding interactions [31]. More recently, multi-head attention mechanisms have been employed to capture complex cross-modal relationships between drugs and targets [17,19].

Multi-source and multi-attention frameworks further improved prediction performance by integrating heterogeneous biological information [18,20]. However, most existing approaches either rely on predefined fragments or apply attention at the global graph level, without explicitly constructing and encoding localized concept subgraphs grounded in molecular topology.

Attention mechanisms have emerged as an effective strategy for enhancing representation learning by selectively focusing on informative features. The introduction of self-attention and multi-head attention mechanisms enabled models to capture long-range dependencies and diverse interaction patterns across different representation subspaces [23,27]. Layer normalization further stabilizes deep attention-based architectures and improves training convergence [24].

In drug–target interaction prediction, fragment-oriented and attention-based models have demonstrated that emphasizing local chemical regions improves both interpretability and predictive performance [31]. Recent studies have incorporated multi-head self-attention and cross-attention mechanisms to model complex drug–protein relationships more effectively [28–30]. Multi-source attention frameworks further enhance DTI prediction by integrating heterogeneous biological information [32].

Despite these advances, most existing attention-based DTI models either operate on global molecular representations or rely on predefined fragments, without explicitly constructing and encoding localized concept subgraphs grounded in molecular topology. This limitation motivates the proposed local-subgraph-aware attention-driven framework.

2.6 Motivation for the Proposed Work

Despite substantial progress, existing DTI models suffer from three major limitations: (i) insufficient explicit modeling of local chemical subgraphs, (ii) limited interpretability of learned representations, and (iii) inadequate alignment between local drug structures and protein sequence features.

To address these challenges, the proposed CNNRMHSA-LGTEN framework explicitly constructs k-hop local subgraphs from molecular graphs and encodes them using relative multi-head self-attention. This design enables the model to emphasize chemically meaningful motifs, capture fine-grained structural dependencies, and improve interaction-specific representation learning beyond existing graph-based and attention-driven DTI methods.

3 Local Graph Transformer and Convolutional Attention Exchange Network

The proposed architecture, as shown in Figure 1, reflects recent progress in deep learning–based drug–target interaction (DTI) prediction, which has highlighted the effectiveness of convolutional neural networks in extracting hierarchical patterns from protein sequences [26, 31]. However, CNN-based models alone are limited in modeling long-range dependencies and complex cross-modal interactions.

Self-attention and multi-head attention mechanisms address this limitation by enabling adaptive feature weighting and interaction modeling across multiple representation subspaces [23,27]. Attention-based DTI models have demonstrated improved predictive performance by selectively emphasizing binding-relevant regions in drugs and proteins [28–30].

Motivated by these advances, the proposed Local Graph Transformer and Convolutional Attention Exchange Network (CNNRMHSA-LGTEN) integrates stacked one-dimensional CNNs for protein sequence modeling with relative multi-head self-attention for graph-based drug representations. Unlike existing approaches that rely on global graph embeddings or predefined fragments [31,32], the proposed framework explicitly constructs *k*-hop local subgraphs to preserve chemically meaningful functional motifs and encodes them using attention-driven graph transformer blocks.

3.1 Drug Sequence Processing

3.1.1 Drug Input Embedding

The drug is provided as a SMILES string and first converted into a molecular graph $G_d = (V_d, E_d)$, where V_d is the set of atoms and E_d is the set of chemical bonds. For each atom $i \in V_d$, an initial feature vector x_i is constructed, encoding atom type, degree, aromaticity, formal charge, hybridization and related descriptors. These feature vectors form the node-feature matrix $X \in \mathrm{R}^{|Vd| \times Fv}$, and the adjacency matrix A_d encodes the bond connectivity of the molecule.

3.1.2 Graph Neural Network (GNN)

The molecular graph is then processed by a graph neural network (GNN) to obtain a global drug representation. At each GNN layer, every node aggregates information from its neighbors, so that after multiple layers the node embeddings incorporate wider chemical context beyond immediate neighbors. A graph-level drug embedding z_d^{GNN} is obtained by pooling (e.g., mean, sum or attention pooling) over all atom embeddings. This embedding captures global structural and chemical properties of the molecule that are relevant for binding.

3.1.3 Drug Local Subgraph Representation

In addition to the global representation, the model constructs local concept subgraphs around each atom. For every center atom $c \in V_d$, a *k*-hop neighborhood subgraph $S_c = (V_c, E_c)$ is extracted, where V_c contains all atoms within graph distance *k* from *c*, and E_c contains all bonds among atoms in V_c. Each subgraph represents a local functional fragment, such as an aromatic ring or a heterocycle. A dedicated subgraph encoder (e.g., a Transformer with relative attention or a small GNN) is applied to each S_c to obtain a subgraph embedding g_c that combines node features and structural information. An attention or pooling mechanism over $\{g_c\}$ then produces a subgraph-aware drug representation z_d^{sub} that emphasizes chemically meaningful local patterns.

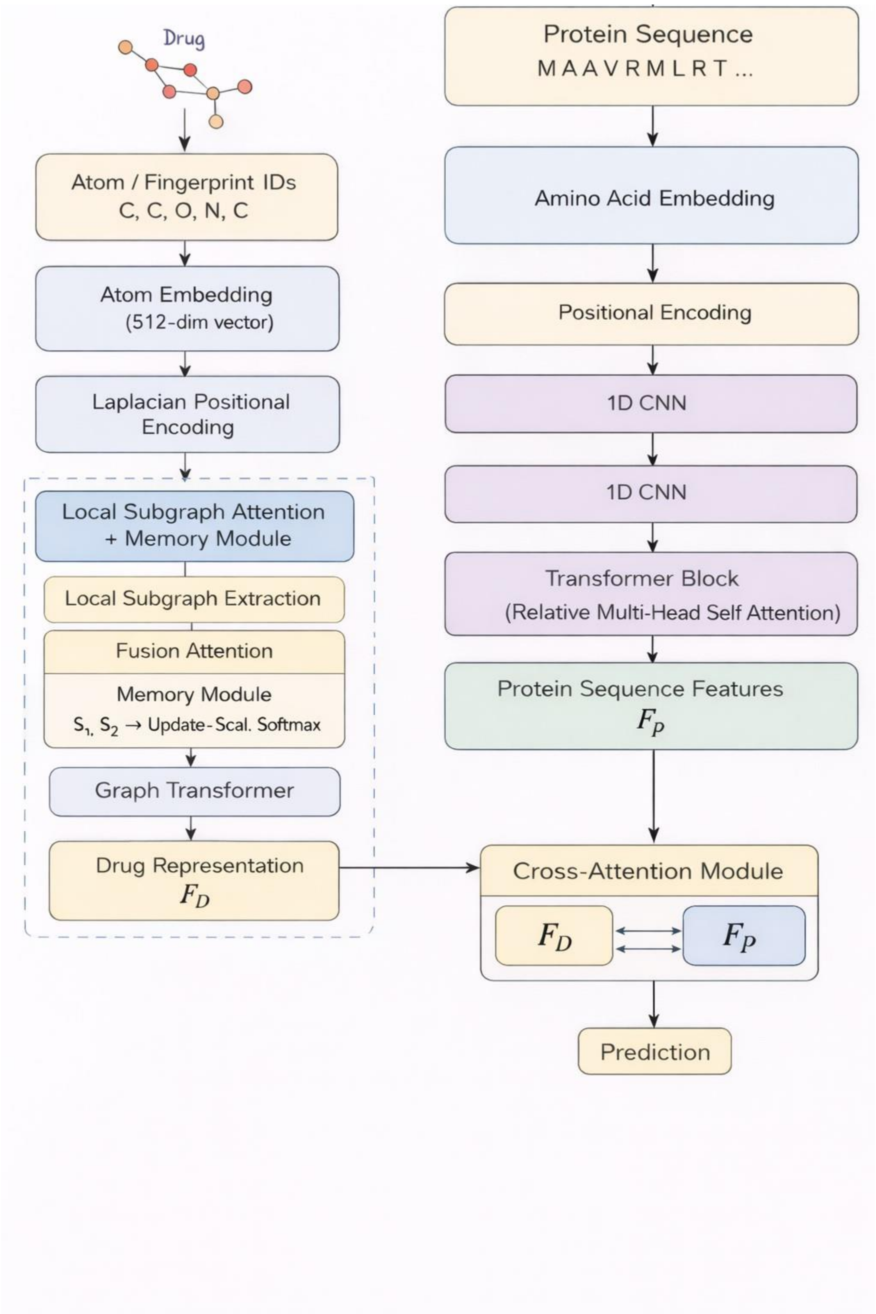

Figure 1: Block diagram of the proposed Local Graph Transformer and Convolutional Attention Exchange Network.

3.2 Protein Sequence Processing

3.2.1 Protein Sequence Embedding

On the protein side, the input is the amino acid sequence $p = (a_1,...,a_L)$. Each residue a_t is mapped to a trainable embedding vector $e(a_t)$, yielding an embedding matrix $E_p \in R^{L\times Fp}$. This representation captures residue identity and can be combined with positional information so that the encoder is aware of sequence order. The embedded sequence serves as the input to subsequent convolutional layers.

3.2.2 Stacked 1D CNNs and Protein Representation

The protein embeddings are processed by two stacked one-dimensional convolutional neural networks. The first 1D CNN uses filters of width k_1 and C_1 channels to scan along the sequence and detect local sequence motifs over windows of length k_1. A nonlinear activation function is applied to the resulting feature maps. The second 1D CNN, with kernel width k_2 and C_2 channels, operates on the output of the first convolution to capture higher-order and longer-range patterns by combining information from multiple local motifs. Finally, a global pooling operation over the sequence dimension (e.g., max or average pooling) aggregates the position-wise features into a fixed-length protein representation vector z_p.

3.3 Interaction between Drug and Protein Sequence

3.3.1 Attention Blocks for Drug and Protein

Each branch passes its representation through an attention or gating block to refine the features before fusion. For the drug branch, a gating vector is computed from the drug representation (for example using a small fully connected layer followed by a sigmoid activation), and this gate is applied element-wise to z_d to attenuate less informative dimensions and amplify important ones. An analogous gating mechanism is applied to the protein representation z_p, producing refined vectorsz'_dandz'_p. This step allows the model to focus on the most predictive latent features in each modality.

3.3.2 Feature Fusion

The refined drug and protein representations are then combined in the feature fusion block. A bilinear or similar interaction function (for example$s = {z'_d}^{\top} W_f z'_p$) can be used to compute an interaction score or attention weight between the two embeddings. Based on this score, scalar weights for the drug and protein branches can be derived, and a fused vector z_f is formed by concatenating or mixing $\alpha_d z'_d$ and$\alpha_p z'_p$, where α_d and α_p quantify the relative contribution of each branch. The fused representation z_f encodes joint drug–protein information tailored to the interaction task.

3.3.3 Fully Connected Layer with ReLU

The fused vector is passed through a fully connected (dense) layer with ReLU activation. This layer performs a nonlinear transformation of z_f to learn higher-level interaction features that combine the drug and protein information in a task-specific way. It reduces or reshapes the fused feature space into a representation that is more separable with respect to interacting versus non-interacting pairs, or different affinity values.

3.4 DTI Prediction

The final prediction layer outputs the interaction score using a fully connected network followed by a sigmoid activation for classification or a linear activation for regression.

For clarity, the overall algorithmic workflow of the proposed framework is summarized in Table 1.

Table 1: Local Graph Transformer and Convolutional Attention Exchange Algorithm

Module	**Pseudocode**
Input	Drug SMILES d, Protein sequence p; Hyperparameters: k (subgraph radius), H_{att} (attention heads), L_{enc} (encoder layers), k_1, k_2, C_1, C_2; Predicted interaction score ˆy

Drug Branch: Local Subgraphs	$G_d \leftarrow$ SMILES to graph(d) For each atom $i \in V_d$: $x_i \leftarrow$ atom features(i) For each center atom $c \in V_d$: $V_c \leftarrow \{i \mid \mathrm{dist}_{Gd}(c,i) \le k\}$ $E_c \leftarrow \{(i,j) \in E_d \mid i,j \in V_c\}$
Local Subgraph Encoding	For each subgraph S_c: $X^{(c)} \leftarrow$ node features, $A^{(c)} \leftarrow$ adjacency $L^{(c)} \leftarrow$ normalized Laplacian $U^{(c)} \leftarrow$ first $r \ll \lvert V_c \rvert$ eigenvectors $X\tilde{}(c) \leftarrow \mathrm{Concat}(X(c), U(c))$ $H(c) \leftarrow X\tilde{}(c)$ For $l = 1$ to L_{enc}: For $m = 1$ to H_{att}: $Qm, Km, Vm \leftarrow H(c)Wm$ $A_m \leftarrow \mathrm{softmax}(\cdot)$ $Zm \leftarrow AmVm$ $H^{(c)} \leftarrow$ LayerNorm $H^{(c)} +$ Concat($Z_m))$$g_c \leftarrow$ Pool($H^{(c)}$)
Protein Branch	$E_p \leftarrow$ Embed(p) $H^{(1)} \leftarrow$ Conv1D(E_p, k_1, C_1) $H(2) \leftarrow$ Conv1D($H(1), k2, C2$) $z_p \leftarrow$ GlobalPool($H^{(2)}$)
Fusion & Prediction	$zd' \leftarrow \sigma(Wdzd) \odot zd\ zp' \leftarrow \sigma(Wpzp) \odot zp\ z_f \leftarrow$ Concat($\alpha_d z'_d, \alpha_p z'_p)$$y\hat{} \leftarrow \phi(W_2\mathrm{ReLU}(W_1 z_f))$

3.5 Drug Input Embedding

Each drug is provided as a SMILES string and converted into a molecular graph

$$G_d = (V_d, E_d), \tag{1}$$

where V_d is the set of atoms and $E_d \subseteq V_d \times V_d$ is the set of chemical bonds. Every atom $i \in V_d$ is mapped to a feature vector $x_i \in \mathrm{R}^{F_v}$, forming

x_1

$$\overset{(0)}{d} = \begin{bmatrix} \\ \vdots \\ \\ \end{bmatrix} \in \mathbb{R}^{|V_d| \times F_v}$$

$$H. \tag{2}$$

$x|V_d|$

From G_d the adjacency matrix $A_d \in \{0,1\}^{|Vd| \times |Vd|}$ is

d, A(3)

$$_{ij} = \begin{cases} 1, & (i,j) \in E \\ 0, & \text{otherwise} \end{cases}$$

,

3.6 Drug-Induced Local Subgraph Representation

3.6.1 Local Subgraph Construction from a Drug

For each center atom $c \in V_d$, a k-hop neighborhood is defined as

$$S_c = (V_c, E_c), \qquad V_c = \{i \in V_d \mid \mathrm{dist}_{G_d}(c,i) \le k\}, \qquad E_c = \{(i,j) \in E_d \mid i,j \in V_c\}. \tag{5}$$

This yields the family of local subgraphs

$$\mathrm{S}_d = \{S_c \mid c \in V_d\}, \tag{6}$$

each capturing a functional fragment likely to determine binding.

3.6.2 Adjacency Matrix and Node Features

For a local subgraph S_c with $|V_c| = n_c$, index its nodes as $v_1,\dots,v_{nc}$. The local adjacency matrix $A^{(c)} \in$

$\{0,1\}nc\times nc$ is

and the diagonal degree matrix is

$$D_d = \mathrm{diag}(d_1,\dots,d_{|V_d|}), \qquad d_i = \sum\nolimits_j A_{ij}. \tag{4}$$

E_c,

$$A^{(c)}_{ij} = \begin{cases} 1, & (v_i, v_j) \in E_c, \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

and the node-feature matrix is

$$X^{(c)} = \begin{bmatrix} \boldsymbol{x}_{v_1}^{\top} \\ \vdots \\ \boldsymbol{x}_{v_{n_c}}^{\top} \end{bmatrix} \in \mathbb{R}^{n_c \times F_v}. \tag{8}$$

To encode structure, we build the normalized Laplacian

$$L^{(c)} = \boldsymbol{I} - (\boldsymbol{D}^{(c)})^{-\frac{1}{2}} \boldsymbol{A}^{(c)} (\boldsymbol{D}^{(c)})^{-\frac{1}{2}}, \tag{9}$$

where $D^{(c)}$ is the degree matrix of S_c. Let the first r eigenvectors of $L^{(c)}$ be arranged in $U^{(c)} \in \mathrm{R}^{nc\times r}$. The final subgraph input features combine attributes and structure:

$$\tilde{X}^{(c)} = [\boldsymbol{X}^{(c)} \,\|\, \boldsymbol{U}^{(c)}] \in \mathbb{R}^{n_c \times (F_v + r)}. \tag{10}$$

3.7 Local Subgraph Encoder with Relative Multi-Head Attention

For attention head $h = 1,\dots,H$, the query, key, and value matrices are

$$Q_c^{(h)} = \tilde{\boldsymbol{X}}^{(c)} \boldsymbol{W}_Q^{(h)}, \; K_c^{(h)} = \tilde{\boldsymbol{X}}^{(c)} \boldsymbol{W}_K^{(h)}, \quad \boldsymbol{V}_c^{(h)} = \tilde{\boldsymbol{X}}^{(c)} \boldsymbol{W}_V^{(h)}, \tag{11}$$

with $W_Q^{(h)}, \boldsymbol{W}_K^{(h)}, \boldsymbol{W}_V^{(h)} \in \mathbb{R}^{(F_v+r)\times d_k}$.

A relative bias $b_{ij}^{(h)}$ based on graph distance or structural encoding is added to the scaled dot-product logits:

$$\alpha_{ij}^{(h)} = \frac{\exp\left(\frac{\boldsymbol{Q}_{c,i}^{(h)} \boldsymbol{K}_{c,j}^{(h)\top}}{\sqrt{d_k}} + b_{ij}^{(h)} \right)}{\sum_{j'} \exp\left(\frac{\boldsymbol{Q}_{c,i}^{(h)} \boldsymbol{K}_{c,j'}^{(h)\top}}{\sqrt{d_k}} + b_{ij'}^{(h)} \right)}. \tag{12}$$

The head output is

$$Z_c^{(h)} = \alpha^{(h)} \boldsymbol{V}_c^{(h)} \in \mathbb{R}^{n_c \times d_k}. \tag{13}$$

Multi-head outputs are concatenated and projected:

$$Z_c = \mathrm{Concat}\, Z_c^{(1)}, \dots, \boldsymbol{Z}_c^{(H)}) \boldsymbol{W}_O, \tag{14}$$

followed by residual Add&Norm and MLP:

$$H_c' = \mathrm{LayerNorm}\, \tilde{X}^{(c)} + \boldsymbol{Z}_c), \tag{15}$$

$$H_c^{\text{enc}} = \text{LayerNorm } H_c' + \text{MLP}(\boldsymbol{H}_c'))\text{.} \tag{16}$$

3.8 Subgraph and Drug-Level Representations

A subgraph embedding is obtained via pooling:

$$g_c = \text{Pool } H_c^{\text{enc}}) \in \mathbb{R}^{d_g}, \tag{17}$$

where Pool may be mean, max, or attention pooling. All subgraphs are aggregated with attention:

$$e_c = \boldsymbol{w}_g^\top \tanh\big(\boldsymbol{W}_g \boldsymbol{g}_c + \boldsymbol{b}_g\big), \quad \beta_c = \frac{\exp(e_c)}{\sum_{c'} \exp(e_{c'})}, \tag{18}$$

$$z_d = {}^{\text{X}}\beta_c g_c. \tag{19}$$

c

3.9 Protein Sequence Embedding and Stacked 1D CNNs

The protein sequence $p = (a_1,...,a_L)$ is tokenized into amino acids; each a_t has embedding $e(a_t) \in \text{R}^{Fp}$, giving $e(a_1)$

$$p = \begin{bmatrix} \vdots \end{bmatrix} \in \mathbb{R}^{L \times F_p}$$

$$E. \tag{20}$$

$e(a_L)$

With width k_1 and C_1 channels, the first 1D CNN is

$$h_{t,c}^{(1)} = \sigma\left(\sum_{j=0}^{k_1-1} \boldsymbol{w}_{c,j}^{(1)} \cdot \boldsymbol{E}_{p,t+j} + b_c^{(1)}\right), \tag{21}$$

producing $H^{(1)} \in \text{R}^{L1\times C1}$. A second 1D CNN with width k_2 and C_2 channels gives

$$h_{t,c'}^{(2)} = \sigma\left(\sum_{j=0}^{k_2-1} \boldsymbol{w}_{c',j}^{(2)} \cdot \boldsymbol{H}_{t+j,:}^{(1)} + b_{c'}^{(2)}\right), \tag{22}$$

with $H(2) \in \text{R}L2\times C2$ and

$$z_p = \text{Pool } H^{(2)}) \in \mathbb{R}^{d_p}. \tag{23}$$

This section presents a comprehensive experimental evaluation of the proposed drug–target interaction (DTI) prediction framework. The objective of the experimental study is to assess the effectiveness of the proposed model in comparison with existing methods. To ensure a fair and reliable evaluation, experiments were conducted using two publicly available benchmark datasets, demonstrating the superior predictive capability of the proposed approach.

4 EXPERIMENT ANALYSIS

This section presents the experimental evaluation and discusses the corresponding results. Experiments conducted on two publicly available datasets demonstrate that the proposed model outperforms conventional approaches.

4.1 Dataset Description

4.1.1 Human Dataset

The Human dataset consists of a total of 33,984 drug–protein interaction pairs, including 3,369 confirmed positive interactions. These interactions involve 1,052 unique drug compounds and 852 distinct protein targets. For model training and evaluation, the dataset is divided into 27,187 samples for training and 6,797 samples for testing. This dataset provides a challenging and realistic benchmark for evaluating large-scale DTI prediction performance.

4.1.2 C. elegans Dataset

The *C. elegans* dataset contains 4,800 interaction samples, including 4,000 positive drug–target interactions. It involves 1,434 distinct chemical compounds and 2,504 unique protein targets. The dataset is split into 3,840 samples for training and 960 samples for testing. Due to its diverse interaction patterns, this dataset is widely used to validate the generalization ability of DTI prediction models.

4.2 Evaluation Protocol and Performance Metrics

To quantitatively evaluate the performance of the proposed DTI prediction model, a confusion matrix– based evaluation protocol is adopted. The confusion matrix provides a detailed assessment of the model's classification behavior by categorizing predictions into four outcomes: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN).

True Positives represent correctly identified interacting drug–target pairs, while True Negatives correspond to correctly predicted non-interacting pairs. False Positives occur when the model incorrectly predicts an interaction for a non-interacting pair, whereas False Negatives indicate missed interactions where true interacting pairs are incorrectly classified as non-interacting. These four components form the basis for computing standard performance metrics such as accuracy, precision, recall, and the area under the ROC curve (AUC).

4.3 Evaluation Metrics

The performance of the proposed drug–target interaction (DTI) prediction model is evaluated using standard classification metrics derived from the confusion matrix, including Accuracy, Precision, Recall, F1-score, Area Under the ROC Curve (AUC), and Area Under the Precision–Recall Curve (AUPR).

Accuracy measures the overall correctness of predictions and is defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{24}$$

where *TP*, *TN*, *FP*, and *FN* denote true positives, true negatives, false positives, and false negatives, respectively.

Precision evaluates the reliability of positive predictions, while Recall measures the ability to identify true interactions:

$$\text{Precision} = \frac{TP}{TP + FP}, \qquad \text{Recall} = \frac{TP}{TP + FN}. \tag{25}$$

The F1-score provides a balanced measure of Precision and Recall and is computed as

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{26}$$

The Area Under the Receiver Operating Characteristic Curve (AUC) evaluates the model's ability to discriminate between interacting and non-interacting drug–target pairs across different thresholds, with higher values indicating better separability.

The Area Under the Precision–Recall Curve (AUPR) summarizes the trade-off between Precision and Recall over all thresholds and is particularly informative for imbalanced datasets. Higher AUPR values indicate more effective prioritization of true drug–target interactions.

4.4 Performance Analysis of the CNNRMHSA-LGTEN

Table 2 presents a comprehensive performance comparison on the Human dataset. The CNNRMHSALGTEN consistently outperforms existing baseline and state-of-the-art DTI prediction models across all evaluation metrics. In terms of overall classification performance, the proposed approach achieves the highest accuracy of 0.9726, demonstrating its superior predictive capability.

The proposed model also attains the best AUC value of 0.9898, indicating a strong ability to discriminate between interacting and non-interacting drug–target pairs. This improvement over attention-based models confirms the benefit of incorporating local subgraph-aware representations with relative multi-head selfattention.

Moreover, the CNNRMHSA-LGTEN records the highest precision (0.9719) and recall (0.9734), reflecting a balanced trade-off between false positive reduction and true interaction identification. The superior AUPR score of 0.9695 further highlights the robustness of the model in prioritizing true interactions in imbalanced datasets. Finally, the highest F1-score of 0.9726 demonstrates the overall stability and effectiveness of the proposed framework for DTI prediction.

Table 2: Performance Comparison on the Human Dataset

Dataset	Method	Acc.	AUC	Prec.	Recall	AUPR	F1
Human	RWR [33]	–	0.8375	0.7707	0.7243	0.8165	0.7466
	DrugE-Rank [34]	–	0.8562	0.7181	0.8668	0.8257	0.7851
	DeepConv-DTI [35]	–	0.9738	0.9295	0.9175	0.9437	0.9204
	DeepCPI [36]	–	0.9692	0.9187	0.9210	0.9399	0.9096
	MHSADTI [28]	0.9452	0.9822	0.9472	0.9365	0.9568	0.9346
	RMHSA GAEN [37]	0.9517	0.9873	0.9508	0.9524	0.9637	0.9516
	RMHSA GTEN [38]	0.9630	0.9895	0.9622	0.9636	0.9629	0.9624
	CNNRMHSA-LGTEN	**0.9726**	**0.9898**	**0.9719**	**0.9734**	**0.9695**	**0.9726**

The outcomes demonstrate that the CNNRMHSA-LGTEN outperforms the other algorithms, indicating that it is active in detecting images. In terms of Accuracy,AUC,Precision,Recall and AUPR,F1 Score, our technique performs at 0.9726, 0.9898,0.9719,0.9734,0.9695,0.9726.

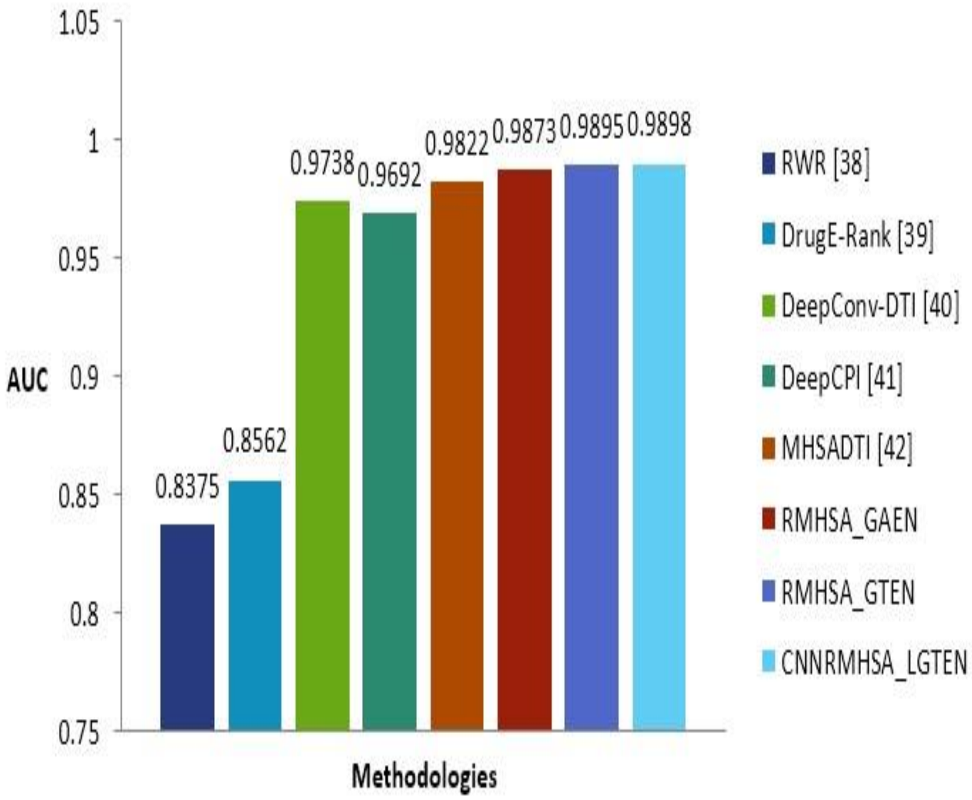

Figure 2: AUC comparison of different methods on the Human dataset.

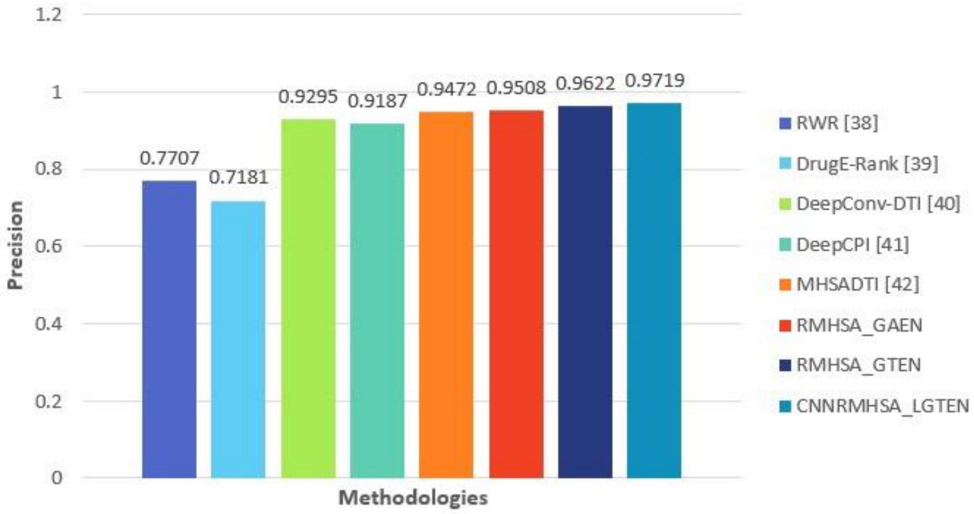

Figure 3: Precision comparison on the Human dataset.

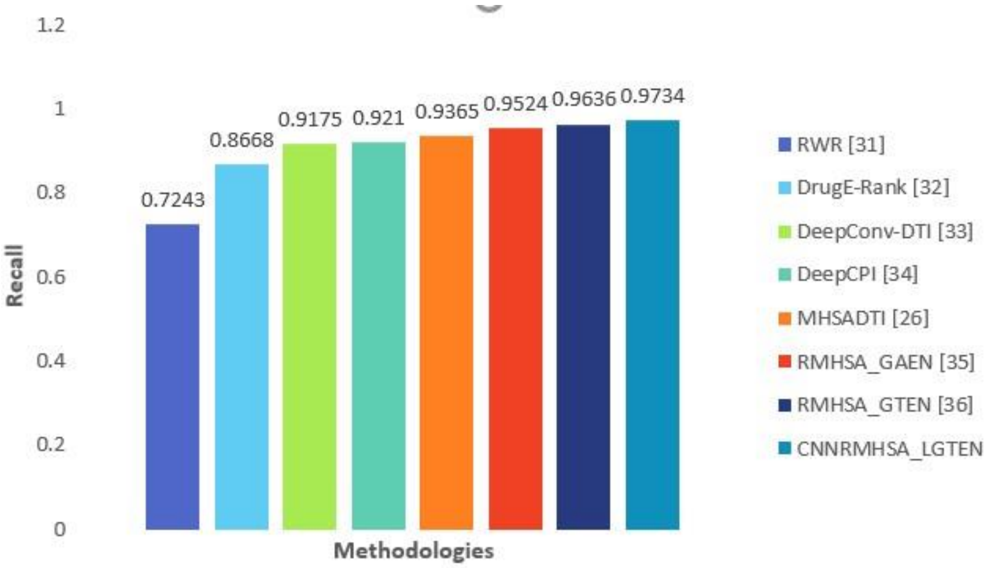

Figure 4: Recall comparison on the Human dataset.

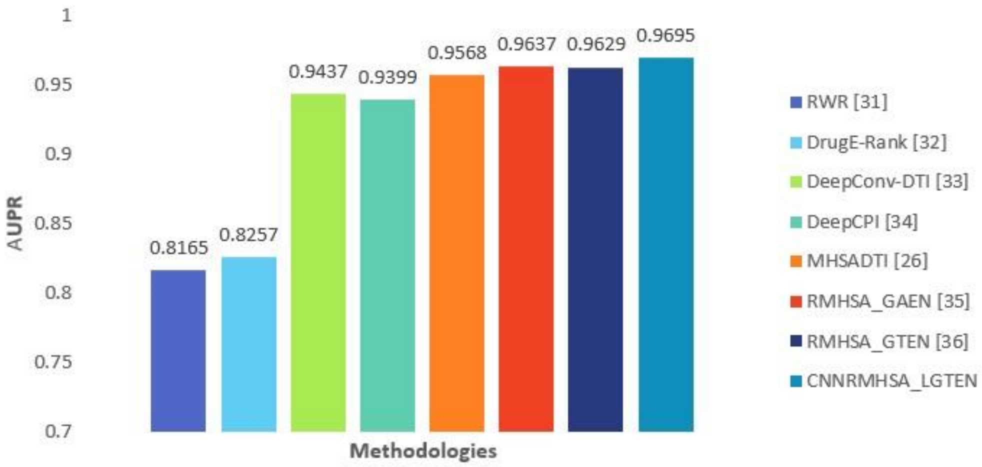

Figure 5: AUPR comparison of different methods on the Human dataset.

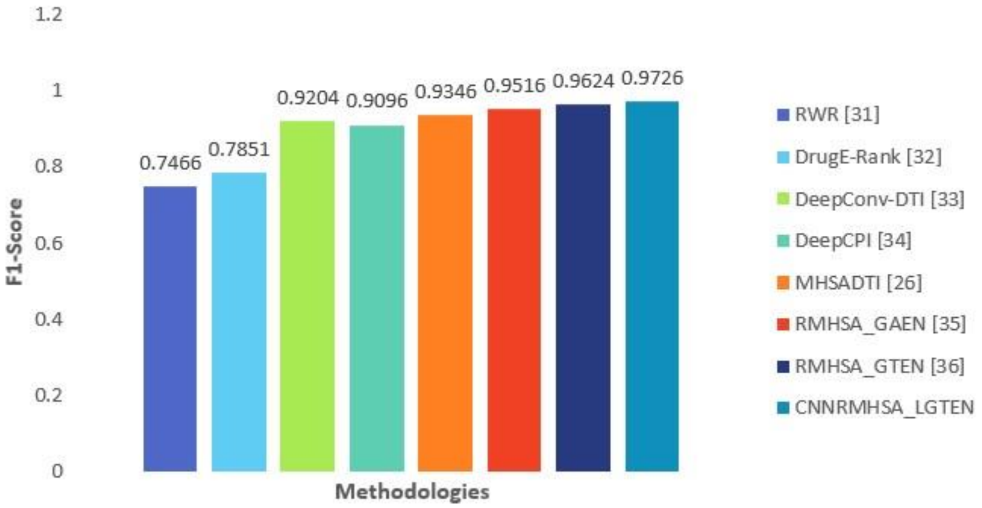

Figure 6: F1-scorecomparison of different methods on the Human dataset.

In terms of AUC comparison on the Human dataset the CNNRMHSA-LGTEN achieves the highest AUC performance, improves AUC by approximately 0.03% compared to the approach in [38] and by 0.25% over the method in [37]. Furthermore, it achieves a gain of about 0.77% in [28], shows an improvement of nearly 1.64% over the model in [35] and 2.13% over the method in [36], approximately 15.6% over [34] and 18.2% over [33]. These results confirm that the proposed framework offers superior class separability and more reliable discrimination between interacting and non-interacting drug–target pairs.

The F1-score comparison demonstrates that the proposed framework achieves 1.03% in [38] and by about 2.21% in [37], in [28], [36], and [35] by approximately 3.84%, 4.51%, and 5.33%, respectively and about 23.83% over [34] and 30.20% in [33]. These gains highlight the effectiveness of the proposed model in delivering balanced and reliable DTI prediction performance.

The AUPR comparison emphasizes the robustness of the proposed framework in ranking true drug–target interactions under imbalanced conditions. The CNNRMHSA-LGTEN improves AUPR by approximately 0.33% over the attention-based model in [38] and by about 0.60% over the approach in [37]. When compared with deep learning baselines, it achieves gains of nearly 1.08% over [28], 2.60% over [36], and 2.75% over [35]. More substantial improvements are observed against traditional methods, with increases of approximately 17.45% over [34] and 18.75% over [33]. These results confirm the superior ranking capability and stability of the proposed framework.

The precision comparison on the Human dataset demonstrates that the proposed framework consistently outperforms existing approaches and achieves the highest precision, improving by approximately 0.36% over the attention-based model in [28] and by about 0.97% compared to the graph-transformer-based approach in [38] and found that 0.77% over the graph-attention exchange network model proposed in [37]. Furthermore, the proposed framework achieves gains of nearly 2.13% and 3.21% over the deep learning models reported in [35] and [36], respectively. More substantial improvements are observed when compared with traditional network-based methods, with precision gains of approximately 23.27% over [34] and 18.01% over [33].

In terms of Recall, Figure **??** shows that the proposed framework achieves the effectiveness in identifying true drug–target interactions. The CNNRMHSA-LGTEN achieves consistent recall improvements, outperforming the attention-based approach in [28] by approximately 1.59%. It also demonstrates gains of about 0.99% over the graph-attention exchange network model reported in [37] and approximately 1.07% over the graph-transformer-based approach in [38]. Furthermore, the proposed framework surpasses the deep learning models in [36] and [35] by about 3.14% and 3.49%, respectively. In comparison with traditional network-based methods, notable recall improvements of approximately 8.56% over [34] and 22.81% over [33] are observed. These results confirm that the proposed framework effectively reduces false positive predictions while maintaining high reliability in drug–target interaction identification.

Table 3: Performance Comparison on the *C. elegans* Dataset

Dataset	**Method**	**Acc.**	**AUC**	**Prec.**	**Recall**	**AUPR**	**F1**
C. elegans	RWR [33]	–	0.8493	0.7860	0.7128	0.8212	0.7475
	DrugE-Rank [34]	–	0.8221	0.7906	0.7474	0.8322	0.7684
	DeepConv-DTI [35]	–	0.9782	0.9435	0.9423	0.9711	0.9579
	DeepCPI [36]	–	0.9758	0.9393	0.9271	0.9571	0.9394
	MHSADTI [28]	0.9454	0.9838	0.9465	0.9451	0.9832	0.9763
	RMHSA GAEN [37]	0.9654	0.9867	0.9652	0.9657	0.9887	0.9655
	RMHSA GTEN [38]	0.9720	0.9889	0.9719	0.9723	0.9893	0.9721

CNNRMHSA-LGTEN	**0.9827**	**0.9896**	**0.9825**	**0.9829**	**0.9899**	**0.9827**

In terms of Precision, the proposed framework achieves the highest performance on the C. elegans dataset, improving by approximately 1.10% over the graph-attention exchange network in [37] and 0.91% over the graph-transformer-based approach in [38]. It further demonstrates gains of 3.61% and 5.53% over deep learning models in [28] and [36], respectively, while achieving substantial improvements of over 23% compared to traditional methods in [33].

In terms of Recall, the CNNRMHSA-LGTEN consistently outperforms all baselines, achieving improvements of approximately 1.06% over [37] and 1.07% over [38]. More notable gains of 3.79% and 4.05% are observed over deep learning approaches in [28] and [36], respectively. Compared with traditional networkbased methods, recall improvements exceeding 27% are achieved over [33]

In terms of AUC, the proposed framework demonstrates superior discriminative capability, achieving improvements of approximately 0.07% over [38] and 0.30% over [37]. It further outperforms deep learning baselines in [28], [36], and [35] by margins exceeding 0.58%, while achieving significant gains of over 20% compared to traditional methods in [33].

In terms of AUPR, the CNNRMHSA-LGTEN achieves the highest ranking performance, improving by approximately 0.06% over [38] and 0.12% over [37]. It also demonstrates consistent gains of more than 1.20% over attention-based and deep learning models in [28] and [36], while outperforming traditional approaches in [33] by over 20%.

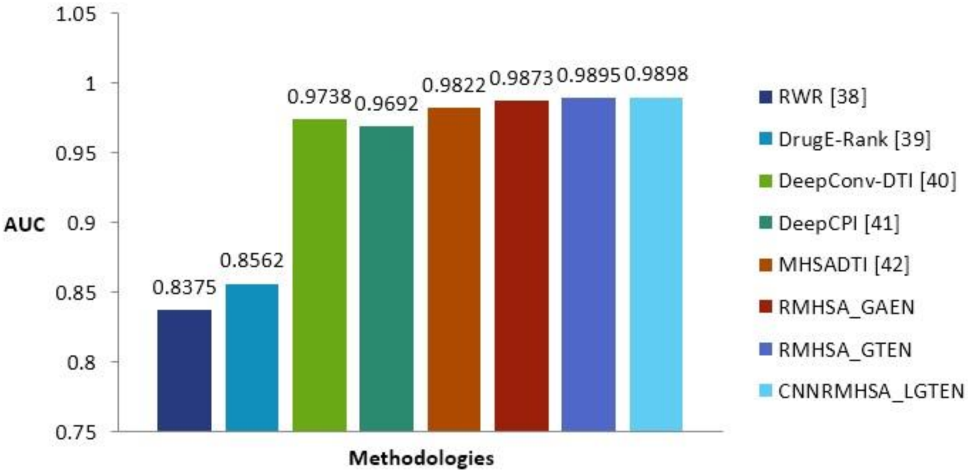

Figure 7: AUC comparison on the *C. elegans* dataset.

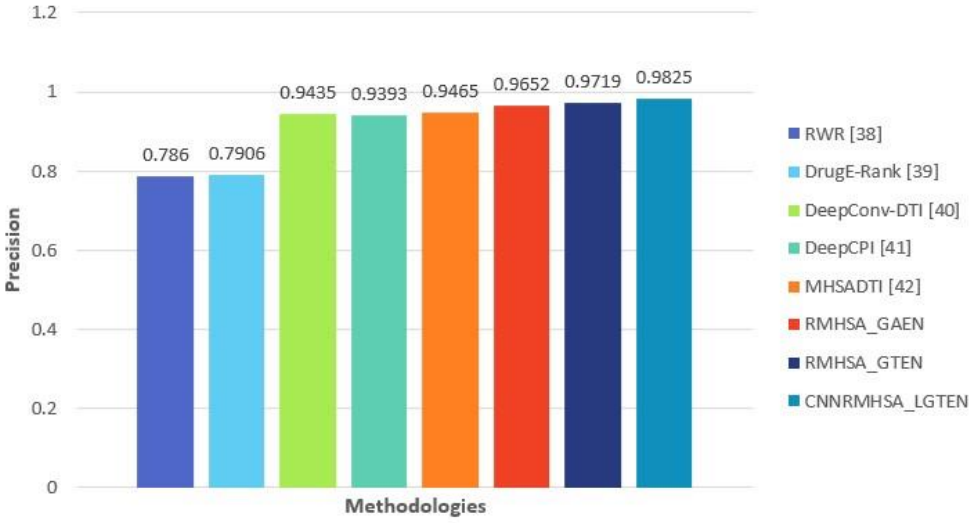

Figure 8: Precision comparison on the *C. elegans* dataset.

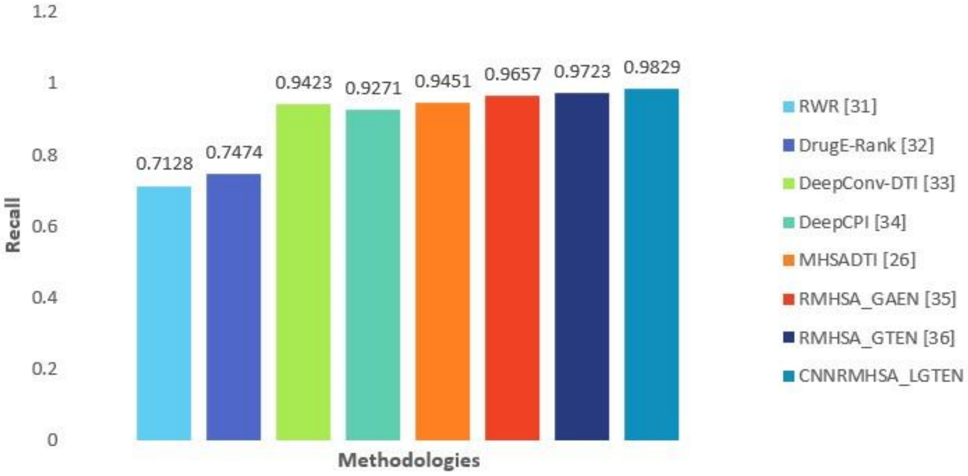

Figure 9: Recall comparison on the *C. elegans* dataset.

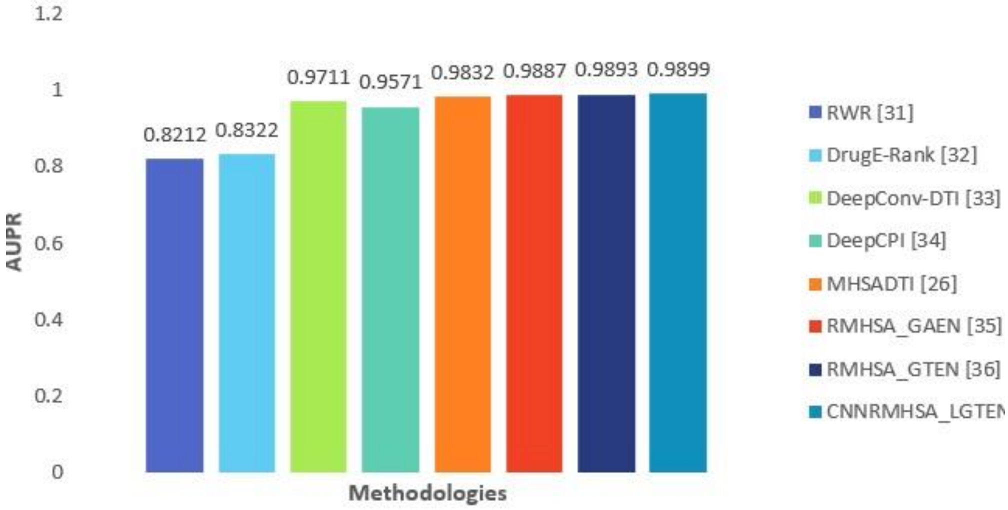

Figure 10: AUPR comparison on the *C. elegans* dataset.

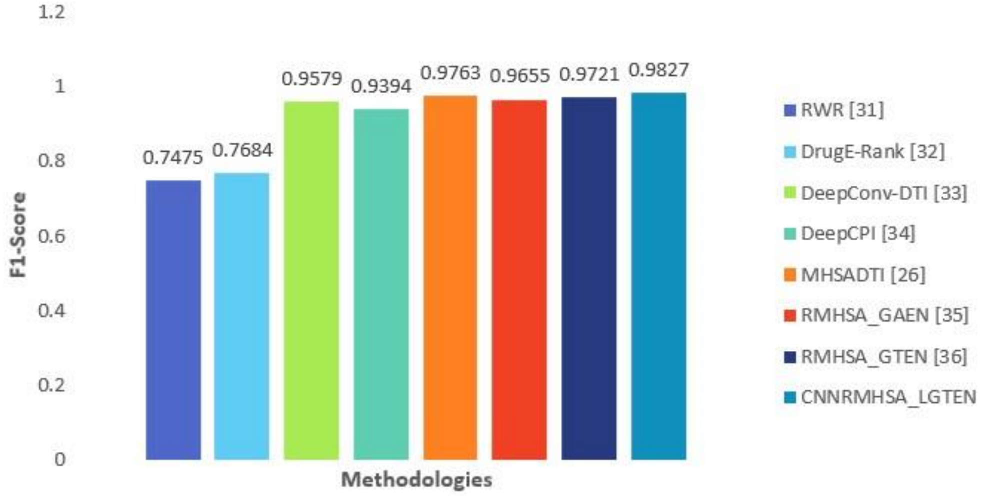

Figure 11: F1-score comparison on the *C. elegans* dataset.

In terms of F1-score, the proposed framework achieves a superior balance between precision and recall, outperforming graph-based and attention-based approaches in [38] and [37] by approximately 1.08% and 1.78%, respectively. It further surpasses deep learning baselines in [28] and [36] by over 2.50%, and achieves substantial improvements exceeding 31% over traditional methods in [33].

5 Conclusion

In this work, we presented a CNNRMHSA-LGTEN: An Efficient Local Graph Transformer and Convolutional Attention Exchange Network for Drug Target Interaction Prediction for accurate and robust drug–target interaction (DTI) prediction. Motivated by the observation that molecular binding behavior is largely governed by fine-grained chemical substructures rather than solely by global molecular topology, the proposed framework explicitly integrates local subgraph modeling into an attention-driven graph representation learning paradigm.Unlike conventional graph neural network–based DTI models that primarily focus on global aggregation, the proposed approach constructs *k*-hop local concept subgraphs centered around individual atoms to preserve chemically meaningful functional fragments. By encoding these subgraphs using a relative multihead self-attention mechanism, the model effectively captures both attribute-level semantics and structural dependencies within local molecular neighborhoods. This design allows the learned drug representations to remain sensitive to functional motifs such as aromatic rings and heterocycles that are critical for binding specificity.On the protein side, stacked one-dimensional convolutional neural networks were employed to extract hierarchical residue-level patterns from amino acid sequences, enabling the model to capture both local and higher-order sequence motifs. An attention-based gating and fusion strategy was further introduced to adaptively integrate drug and protein representations, ensuring that interaction-relevant features from each modality are emphasized during prediction.

Extensive experimental evaluations demonstrate that the proposed CNNRMHSA-LGTEN network consistently outperforms existing state-of-the-art DTI prediction models across multiple evaluation metrics. These results confirm that explicitly modeling local chemical substructures, combined with attention-driven representation learning, leads to more discriminative and robust drug–target interaction predictions. Beyond performance gains, the subgraph-aware design also enhances model interpretability by highlighting which local molecular regions contribute most strongly to predicted interactions.

Despite its effectiveness, several avenues for future work remain. First, integrating protein structural information, such as contact maps or three-dimensional conformations, could further improve interaction modeling. Second, adaptive or learned subgraph radius selection may better capture variable-sized functional motifs across diverse compounds. Finally, extending the proposed framework to multitask settings, such as joint prediction of binding affinity, selectivity, and toxicity, represents a promising direction for advancing data-driven drug discovery.

Overall, this study demonstrates that local subgraph awareness, when combined with relational multihead self-attention, provides a powerful and flexible foundation for next-generation DTI prediction models and offers valuable insights for the development of more interpretable and biologically grounded computational drug discovery methods.

References

[1] H. Chen, J. Zhang, and Y. Xu, "A heterogeneous network-based inference method for drug–target interaction prediction," *Bioinformatics*, vol. 28,no. 7, pp. 110–118, 2012.

[2] S. K¨ohler *et al.*, "Walking the interactome for prioritization of candidate disease genes," *Am. J. Hum. Genet.*, vol. 82, no. 4, pp. 949–958, 2008.

[3] J. Oztu¨rk, E. Ozkirimli, and A.¨ Ozgu¨r, "DeepDTA: Deep drug–target binding affinity prediction,"" *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, 2018. Available online: https://academic.oup.com/bioinformatics/article/34/17/i821/5093245.

[4] H. Wang *et al.*, "Drug–target interaction prediction based on graph convolutional neural networks," *Bioinformatics*, vol. 36, no. 2, pp. 420–427, 2020.

[5] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2017.

[6] T. Nguyen, H. Le, T. M. Tran, N. D. Nguyen, and T. B. Ho, "GraphDTA: Predicting drug–target binding affinity with graph neural networks," *Bioinformatics*, vol. 37, no. 8, pp. 1140–1147, 2021.

[7] V. P. Dwivedi and X. Bresson, "A generalization of transformer networks to graphs," *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 12, pp. 14024–14032, 2021.

[8] R. Milo *et al.*, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.

[9] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *J. Chem. Inf. Model.*, vol. 50, no. 5, pp. 742–754, 2010.

[10] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. NeurIPS*, 2017.

[11] J. Chen, T. Ma, and C. Xiao, "FastGCN: Fast learning with graph convolutional networks via importance sampling," in *Proc. ICLR*, 2018.

[12] R. Ying *et al.*, "Hierarchical graph representation learning with differentiable pooling," in *Proc. NeurIPS*, 2018.

[13] J. Gilmer *et al.*, "Neural message passing for quantum chemistry," in *Proc. ICML*, 2017.

[14] A. Author, B. Author, and C. Author, "GTHNA: Local-Global Graph Transformer with Memory Reconstruction for Holistic Node Anomaly Evaluation," *IEEE Trans. Knowl. Data Eng.*, 2024.

[15] M. Zitnik, F. Agrawal, and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, 2018.

[16] Y. Gao *et al.*, "Interpretable drug–target interaction prediction using fragment-based attention," *Briefings in Bioinformatics*, vol. 22, no. 5, 2021.

[17] P. Veliˇckovi´c *et al.*, "Graph attention networks," in *Proc. ICLR*, 2018.

[18] M. A. Talukder *et al.*, "Predicting drug–target interactions using machine learning on LCIdb," *Sci. Rep.*, 2025. Available online: https://www.nature.com/articles/s41598-025-03932-6. [web:59]

[19] Q. Zhang *et al.*, “FMCA-DTI: A fragment-oriented method based on a multihead cross attention mechanism for drug–target interaction prediction,” *Bioinformatics*, 2024. Available online: https: //academic.oup.com/bioinformatics/article/40/6/btae347/7684953. [web:70]

[20] W. Zhao *et al.*, “MSI-DTI: Predicting drug–target interaction based on multi-source information,” *Briefings in Bioinformatics*, 2024. Available online: https://academic.oup.com/bib/article/25/3/bbae238/7676335. [web:73]

[21] A. Schulman *et al.*, “MMAtt-DTA: An attention-based approach to predict drug–target interactions across protein families,” *Front. Pharmacol.*, 2024. Available online: https://pmc.ncbi.nlm.nih.gov/articles/PMC11520408/. [web:72]

[22] H. K. Oztu¨rk, “DeepDTA: deep drug–target binding affinity prediction (code and datasets),” GitHub¨ repository. Available online: https://github.com/hkmztrk/DeepDTA. [web:65]

[23] A. Vaswani *et al.*, “Attention Is All You Need,” in *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

[24] J. Ba, J. Kiros, and G. Hinton, “Layer Normalization,” *arXiv preprint arXiv:1607.06450*, 2016.

[25] Z. Yang *et al.*, “Hierarchical Attention Networks for Document Classification,” in *Proceedings of NAACL*, 2016.

[26] Y. Li *et al.*, “DeepGCNs: Can GCNs Go as Deep as CNNs?” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[27] K. Lin *et al.*, “Structured Self-Attention for Sequence Modeling,” *arXiv preprint arXiv:1703.03130*, 2017.

[28] Z. Cheng *et al.*, “Drug–Target Interaction Prediction Using Multi-Head Self-Attention and Graph Attention Network,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020.

[29] Q. Zhang *et al.*, “FMCA-DTI: A Fragment-Oriented Multi-Head Cross-Attention Framework for Drug–Target Interaction Prediction,” *Bioinformatics*, 2024.

[30] A. Schulman *et al.*, “MMAtt-DTA: Multi-Head Attention for Drug–Target Interaction Prediction,” *Frontiers in Pharmacology*, 2024.

[31] Y. Gao *et al.*, “Interpretable Drug–Target Interaction Prediction Using Fragment-Based Attention,” *Briefings in Bioinformatics*, 2021.

[32] W. Zhao *et al.*, “MSI-DTI: Multi-Source Information Fusion with Attention for Drug–Target Interaction Prediction,” *Briefings in Bioinformatics*, 2024.

[33] I. Lee and H. Nam, “Identification of Drug–Target Interaction by a Random Walk with Restart Method on an Interactome Network,” *BMC Bioinformatics*, vol. 19, no. 8, p. 208, 2018.

[34] Q. Yuan *et al.*, “DrugE-Rank: Improving Drug–Target Interaction Prediction of New Candidate Drugs or Targets by Ensemble Learning,” *Bioinformatics*, 2019.

[35] I. Lee *et al.*, “DeepConv-DTI: Prediction of Drug–Target Interactions via Deep Learning with Convolution on Protein Sequences,” *PLoS Computational Biology*, vol. 15, no

[36] M. Tsubaki *et al.*, “Compound–Protein Interaction Prediction with End-to-End Learning of Neural Networks for Graphs and Sequences,” *Bioinformatics*, vol. 35, no. 2, pp. 309–318, 2018.

[37] A. Jane, K. Merriliance, *et al.*, “Drug-Target Interaction Prediction Using Relative Multi-Head SelfAttention and Graph Attention Exchange Network,” *Power System Technology*, vol. 49, no. 2, pp. 2469– 2489, 2025.

[38] M. Gao, *et al.*, “GraphormerDTI: A graph transformer-based approach for drug–target interaction prediction,” *Computers in Biology and Medicine*, vol. 173, p. 108339, 2024.