

HyperGCN-TransDTA: A Novel Hypergraph-Transformer Architecture for Drug-Target Interaction Prediction

Selvakumar G.¹, Dr. Rajendran G.²

¹Ph.D. Research Scholar, Department of Computer Science, Periyar University, Salem, India. Assistant Professor of Computer Science, Muthayammal College of Arts and Science, Rasipuram, Namakkal, India.

²Associate Professor and Head, Department of Computer Science, Government Arts and Science College, Modakkurichi, Erode, Tamil Nadu

Abstract—Drug–target interaction (DTI) and drug–target affinity (DTA) predictions are critical for accelerating drug discovery. However, existing deep learning approaches often rely on pairwise molecular graphs that fail to capture higher-order chemical relationships, inadequately model long-range protein dependencies, and provide limited interpretability of the interaction mechanisms. We propose HyperGCN-TransDTA (DTI-HyperGCN), a novel end-to-end deep learning architecture that integrates hypergraph convolutional networks with bidirectional transformer encoders for accurate and interpretable drug–target affinity prediction. Drug molecules are represented as molecular hypergraphs derived from SMILES, where hyperedges encode functional groups, aromatic and heterocyclic rings, conjugated systems, and pharmacophore-related atom sets. The atom features were propagated through three HyperGCN layers and pooled into compact drug embeddings. Protein sequences were embedded with learned positional encodings and processed using a six-layer transformer encoder to capture long-range contextual dependencies. An eight-head cross-modal attention mechanism explicitly models drug–protein interactions by attending to drug representations over protein embeddings. The fused representations were passed to an MLP regression head augmented with supervised contrastive learning to enhance feature discrimination. Extensive experiments on benchmark datasets, including Davis (442 drugs, 379 proteins, and 30,056 interactions), KIBA (2,111 drugs, 229 proteins, and 118,254 interactions), and DrugBank-style datasets (4,267 drugs, 2,554 proteins, and 11,556 interactions), demonstrated state-of-the-art or near state-of-the-art performance. On Davis classification, HyperGCN-TransDTA achieves AUC-ROC 0.9031, AUC-PR

0.8645, and F1 0.8521, using only 0.245M parameters ($\approx 3\%$ of DeepMGT-DTI; $p = 0.087$). On Davis regression, it attained CI 0.901, MSE 0.203, RMSE 0.451, Pearson 0.698, and R^2 0.487, approaching DTITR performance (CI 0.907) with approximately 2.3% of its parameter count. Ablation studies confirmed the contributions of hypergraph modeling, cross-attention, and contrastive learning, whereas attention maps provided interpretable insights into drug–protein interactions. Notably, the model is CPU-compatible and was trained on a 32-core CPU, enabling scalable virtual screening and practical deployment. These results highlight the potential of HyperGCN-TransDTA for rapid drug repurposing and precision medicine, and motivate future extensions incorporating 3D structural information, uncertainty quantification, and experimental validation.

Index Terms—Binding Affinity Prediction, Drug–Target Interaction Prediction, Hypergraph Neural Networks, Transformer-based Cross-Attention.

I. INTRODUCTION

Drug discovery is a highly complex and resource-intensive endeavor, with estimated costs exceeding \$2.6 billion per approved drug and development timelines spanning 10–15 years. A major bottleneck in this process is the identification and validation of the interactions between small-molecule compounds and protein targets. Traditional approaches, such as high-throughput screening (HTS), are expensive, time-consuming, and constrained by the vastness of chemical space, which is estimated to contain up to

10^{60} possible molecules [1]. Computational drug-target interaction (DTI) prediction has emerged as a powerful complementary strategy to mitigate these challenges by prioritizing candidate compounds, enabling drug repurposing, reducing the experimental burden, and offering mechanistic insights into molecular interactions [2]. Computational DTI prediction has transitioned from molecular docking and traditional machine learning techniques to deep learning methods. Some approaches utilize CNNs on raw sequences (DeepDTA), thereby removing the need for feature engineering. Other methods incorporated advanced feature engineering and word-based tokenization (WideDTA and DNN-DTIs). In addition, certain strategies have introduced graph neural networks (GraphDTA) to depict drugs as molecular graphs, thereby enhancing their structural representation. Furthermore, some techniques adapted transformers (DTITR) for DTI, capturing both intra- and intermolecular contexts through cross-attention and achieving state-of-the-art results. Despite these advancements, several limitations persist. Binary edges in graph representations are inadequate for capturing complex multi-atom relationships, such as those in aromatic rings and functional groups. Convolutional neural networks (CNNs) are insufficient for modeling long-range dependencies within proteins. Interaction modeling is frequently implicit rather than explicit. Furthermore, self-supervised learning has not been fully exploited, and ablation analyses are often inadequate. We propose HyperGCN-TransDTA to address these gaps by introducing hypergraph convolutions for multi-atom drug representation, employing a hybrid transformer-CNN for protein encoding, implementing explicit bidirectional cross-modal attention, integrating supervised contrastive learning, and conducting comprehensive evaluations with ablation and statistical significance testing.

II. LITERATURE SURVEY

Early computational DTI prediction methods can be categorized into three main approaches.

Docking-Based Methods [3]: Molecular docking simulates the physical binding process between a drug and target protein by predicting the 3D structure of the drug-target complex. While potentially informative, these methods are limited by (i) their

dependency on known 3D protein structures, which are often unavailable, (ii) the high computational cost for large-scale screening, and (iii) the accuracy limitations of scoring functions.

Ligand-Based Methods [4] leverage the similarity principle; wherein similar drugs tend to interact with similar targets. Methods such as quantitative structure-activity relationship (QSAR) models build correlations between chemical structures and biological activities. However, they suffer from poor performance when known ligands are scarce and have limited generalization to novel chemical scaffolds.

Chemogenomic Methods [5] integrate information from chemical and genomic spaces. Early work by Yamanishi et al. introduced bipartite graph learning methods that predict DTIs from integrated chemical and genomic data. Kernel-based methods, such as KronRLS, compute pairwise kernel matrices for drugs and targets using similarity measures, achieving reasonable performance but lacking the ability to learn complex nonlinear patterns.

Machine Learning Approaches

The expansion of chemogenomic databases has facilitated the application of machine learning techniques for drug-target interaction (DTI) prediction. Similarity-based methods, such as network-based approaches, leverage the topology of drug-target interaction networks. For example, SimBoost employs gradient-boosting machines with features derived from drug and target similarities, demonstrating enhanced performance compared to kernel methods. Matrix factorization techniques decompose the drug-target interaction matrix into latent factor representations. Gönen's Bayesian matrix factorization approach models interactions probabilistically, although it encounters challenges with cold-start issues for novel drugs or targets. Feature-based classification methods, including Random Forest (RF) and Support Vector Machine (SVM) classifiers trained on molecular fingerprints and protein descriptors, have shown potential. However, their performance is significantly reliant on handcrafted features and domain expertise [6].

Deep Learning Revolution

Deep learning has significantly advanced drug-target interaction (DTI) prediction by facilitating automatic

learning of hierarchical feature representations. Sequence-based convolutional neural networks (CNNs), exemplified by DeepDTA, have been instrumental in this transformation. DeepDTA utilizes CNNs to learn from raw sequences, specifically SMILES strings for drugs and amino acid sequences for proteins, thereby achieving superior performance compared to traditional methods through automatic extraction of local patterns. However, it is important to note that the one-dimensional representation of drugs in this approach results in the loss of three-dimensional structural information. In contrast, Graph Neural Networks (GNNs) offer an alternative approach by representing drugs as molecular graphs, which are processed by graph convolutional networks (GCN) or graph attention networks (GAT). This methodology captures the molecular topology and atomic interactions more effectively. Empirical evidence from studies, particularly those using the KIBA dataset, indicates that graph-based drug representations outperform SMILES-based methods.

A. Sequence-Based Convolutional Approaches

DeepDTA [7] represents a seminal contribution to the field by demonstrating that deep learning models can achieve competitive or superior performance using only raw sequence information without requiring extensive feature engineering. The model employs separate CNN blocks for processing SMILES strings (drugs) and protein sequences, with each block consisting of three consecutive 1D-convolutional layers with increasing filter sizes (32, 64, and 96 filters). The key innovation lies in treating DTI prediction as a regression problem rather than a binary classification problem, directly predicting continuous binding affinity values. DeepDTA uses integer encoding for both proteins and SMILES with fixed maximum lengths (85 for SMILES and 1200 for proteins in the Davis dataset; 100 for SMILES and 1000 for proteins in KIBA). Shorter sequences were zero-padded, whereas longer sequences were truncated. The Keras Embedding layer generates 128-dimensional dense vector representations for each character. After max-pooling, features from both modalities were concatenated and fed through three fully connected layers (1024, 1024, 512 nodes) with dropout regularization (rate 0.1) before the final output layer. The model was trained using the Mean

Squared Error (MSE) as the loss function, ReLU activation, and the Adam optimizer with a learning rate of 0.001 for 100 epochs with a mini-batch size of 256. On the larger KIBA dataset, DeepDTA achieved a Concordance Index (CI) of 0.863, significantly outperforming baseline methods (p-value: 0.0001), demonstrating that deep learning methodologies scale better with dataset size.

FRnet-DTI [8] introduced a novel two-stage architecture comprising FRnet-Encode for feature extraction and FRnet-Predict for classification [2]. The FRnet-Encode model functions as an autoencoder, generating 4096 deep representations from a 4D tensor input (X, 211, 7, 1) that combines drug molecular fingerprints with protein structure and evolutionary features. The architecture incorporates GoogleNet-inspired inception modules that perform parallel convolution operations with different filter sizes, alleviating the burden of manually selecting the optimal filter sizes. The inception module design allows the network to capture multi-scale patterns simultaneously, with 1×1 convolutional layers providing dimensionality reduction before performing more expensive operations. FRnet-Predict then processes these 4096 features through additional convolutional and pooling layers to predict the interaction probability. Both models employed aggressive regularization strategies, including dropout (Keep_Prob = 0.5) and L2 regularization, to prevent overfitting, which is particularly important given the relatively small dataset sizes. FRnet-DTI was evaluated on four gold standard benchmark datasets (Enzyme, Ion Channel, GPCR, Nuclear Receptor) with highly imbalanced ratios ranging from 14.6 to 99.98. The model achieved significantly improved Area Under ROC Curve (auROC) and Area Under Precision-Recall Curve (auPR) scores on three of the four datasets, with particular emphasis on auPR performance, which is more informative for imbalanced datasets. The features generated by FRnet-Encode were highly effective, and other classifiers also performed impressively when using these representations.

WideDTA [9] advances sequence-based modeling by transitioning from character-based to word-based representations, integrating up to four textual information sources: Protein Sequence (PS), Protein Domains and Motifs (PDM), Ligand SMILES (LS), and Ligand Maximum Common Substructures

(LMCS). Each information source is processed by a dedicated module consisting of two 1D-convolutional layers with max pooling and ReLU activation. The features extracted from these modules were concatenated and fed into three fully connected layers with dropout for regularization. For protein sequences, 3-residue words were extracted from the full sequences obtained from UniProt. Protein domains and motifs were extracted from the PROSITE database, which contains information crucial for folding, binding, and catalytic activity. For ligands, consecutive overlapping k-mers ($k=8$) were extracted from the SMILES strings, with the DeepSMILES syntax utilized to enhance performance. The top 100 K most frequent Maximum Common Substructures (MCS) were used to extract chemical words from ligand SMILES. WideDTA achieved a CI of 0.875 on the KIBA dataset, significantly outperforming DeepDTA (CI: 0.863, p -value: 0.001). The word-based approach was statistically superior to the character-based methods ($p = 0.0005$). Interestingly, using only protein domains and motif information achieved a performance similar to that of using full protein sequences, suggesting that important binding-relevant information is concentrated within these functional regions.

B. Graph-Based Neural Network Approaches

GraphDTA [10] addresses a fundamental limitation of string-based drug representations by modeling drugs as molecular graphs in which nodes represent atoms and edges represent chemical bonds. This approach directly captures the structural relationships among atoms, potentially improving the predictive power by preserving information that may be lost in the linear SMILES representations. The model tests four GNN variants: Graph Convolutional Network (GCN), Graph Attention Network (GAT), Graph Isomorphism Network (GIN), and a combined GAT-GCN architecture. Each node in the molecular graph is described by a multidimensional binary feature vector encoding the atom symbol, number of adjacent atoms, number of adjacent hydrogens, implicit value, and aromatic structure information extracted using the RDKit chemical informatics software. The GCN variant uses three consecutive GCN layers with ReLU activation, followed by global max pooling. The GAT variant employs two GAT layers with ten

attention heads for the first layer. The GIN variant, which aims for maximum discriminative power, uses five GIN layers with batch normalization and a multi-layer perceptron for node updates. Proteins are represented as 1D sequences using one-hot encoding because obtaining reliable 3D protein structures remains challenging. Protein sequences from UniProt were encoded as integers, passed through embedding layers to generate 128-dimensional vectors, and then processed through three 1D convolutional layers and max pooling to learn abstract features. On the Davis dataset, GraphDTA achieved an MSE of 0.229 (14.0% lower than the best baseline of 0.261) and a CI of 0.893 using the GIN. On the KIBA dataset, the best MSE was 0.139 (28.8% lower than the best baseline value of 0.179). The GIN model consistently exhibited the best performance across both datasets and all metrics. Notably, redundancy analysis revealed that 20.19% of the latent space was explained by known molecular descriptors, with the model learning these properties without prior knowledge.

DeepMGT-DTI [11] introduces a hybrid architecture that integrates a multilayer graph convolutional neural network (MCGCN) with an improved Transformer network featuring multi-headed attention mechanisms. The model addresses the limitations of traditional GNNs, which often focus on node information but lack comprehensive edge feature learning, resulting in insufficient information capture between nodes. For drug representation, SMILES strings were converted to molecular structure graphs, where the vertices represent atoms and the edges represent chemical bonds. The MCGCN consists of two hidden layers, each followed by ReLU activation and max pooling for dimensionality reduction. Critically, features from each MCGCN layer are fed into the transformer network, allowing the multi-headed attention mechanism to process information from different graph depths and overcome GNN limitations in learning edge features. For target representation, sequences are embedded using a lookup table that is randomly initialized and optimized during the training process. A CNN with convolution kernels of different sizes (10, 15, and 20) and an ELU activation function, followed by global maximum pooling, captures the local residue information. The extracted drug and target features were concatenated and fed

into a fully connected network for prediction, with binary cross-entropy loss and L2 regularization to prevent overfitting. DeepMGT-DTI was evaluated on a dataset comprising 12,496 drug structures, 5,462 target sequences, and 21,158 known DTIs from DrugBank, KEGG, and PubChem, with negative examples constructed at a 1:2 ratio. The model achieved an AUC of 90.24, AUPR of 77.11, F1-score of 79.31, and accuracy of 85.15, significantly outperforming Random Forest, SVM, DeepDTA, Deep DTI, Deep Conv-DTI, TransformerCPI, and ML-DTI. Ablation studies confirmed that both the MCGCN and Transformer modules contributed significantly to the performance improvement.

C. Transformer-Based Architectures

The DTITR [12] is the first end-to-end transformer-based architecture specifically designed for drug-target binding affinity prediction [1]. The model leverages self-attention mechanisms to capture biological and chemical contexts and introduces cross-attention layers to model the pharmacological context of drug-target interactions. The architecture comprises two parallel transformer encoders for the contextual embedding of protein sequences and SMILES strings, followed by a Cross-Attention Transformer-Encoder block that exchanges information between modalities. For SMILES encoding, integer-based encoding maps each character to one of 26 categories, with shorter strings being padded. For protein sequences, a combination of Frequent Consecutive Subsequence (FCS) mining and Byte Pair Encoding (BPE) decomposes sequences into frequent substructures (subwords). Special regression tokens (Rp and Rs) were added to the beginning of the protein and SMILES strings, respectively, and an embedding layer generated learned embeddings for each token with positional encoding to account for token positions. The aggregated representations (final hidden states of start tokens Rp and Rs) are concatenated and fed into a fully connected feed-forward network (MLP) to output the binding affinity (pKd). The model was evaluated on the Davis Kinase Dataset, which contains 31,824 interactions between 72 kinase inhibitors and 442 kinases, using Chemogenomic Representative K-Fold splitting to ensure a balanced distribution of pKd values, protein sequence similarity, and SMILES string similarity. The DTITR

achieved superior performance across almost all metrics: MSE of 0.192, RMSE of 0.438, CI of 0.907, r^2 of 0.771, and Spearman correlation of 0.712. Ablation studies revealed that the Cross-Attention Transformer-Encoder block is crucial for improving discriminative power by learning the pharmacological context; its removal led to worse performance. The architecture provides three levels of explainability through attention maps, which show that compounds attend to protein substructures associated with binding residues, confirming the learning of the pharmacological context.

D. Multi-Feature Integration Approaches

DNN-DTIs [13] introduced a comprehensive pipeline combining deep learning with multi-feature fusion, XGBoost-based feature selection, and SMOTE for data balancing [7]. The model addresses the challenge of integrating heterogeneous feature information from multiple sources, including sequence, structure, and physicochemical properties. Target proteins are characterized using diverse features, including pseudo-amino acid composition (PseAAC), pseudo position-specific scoring matrix (PsePSSM), conjoint triad composition (CT), transition and distribution (CTD), Moreau-Broto autocorrelation (NMBroto), and structural features, including accessible surface area (ASA), secondary structure (SS), and torsional angle (TA), generated using SPIDER3. Drug compounds were encoded using substructure fingerprints representing 881 chemical substructures from the PubChem database. A novel contribution is the first-time application of XGBoost for feature selection in DTI prediction, using information gain to identify non-redundant and important features. This effectively removes noise and redundancy, while reducing the computational complexity. The Synthetic Minority Over-Sampling Technique (SMOTE) addresses data imbalance by generating artificial samples for the minority class. The deep neural network consisted of four hidden layers (600, 300, 150, and 75 neurons) with ReLU activation and Adam optimization. The model achieved accuracies of 98.78% (enzyme), 98.60% (Ion Channel), 97.98% (GPCR), 98.24% (Nuclear Receptor), and 98.00% (Kuang's dataset), outperforming state-of-the-art methods, including LR, AdaBoost, KNN, RF, SVM, CNN, and LSTM. XGBoost feature selection achieved the highest

accuracy and AUC values compared to the other feature selection methods (IG, GINI, MRMD, LASSO, and EN).

DeepACTION [14] presents a four-part framework comprising feature extraction, data balancing using majority and minority instance balancing (MMIB), feature selection using LASSO, and CNN-based classification. The model addresses the critical challenge of class imbalance in DTI datasets, where non-interacting pairs far outnumber the interacting pairs. Drug features (topological, constitutional, and geometrical descriptors) were calculated using the Rcpit Toolkit, whereas protein features (amino acid composition, pseudo-amino acid composition, autocorrelation, dipeptide composition, quasi-sequence order, and CTD) were generated using the PROFEAT webserver. These features were normalized and combined to form a 1483-dimensional feature vector for each drug-target pair. The novel MMIB technique uses clustering to divide instances into k clusters and then balances the ratio of majority to minority instances within each cluster, providing a more sophisticated approach to handling imbalances than simple oversampling or undersampling. LASSO feature selection reduces dimensionality by setting irrelevant features' coefficients to zero, retaining only significant features and helping to prevent overfitting. The CNN architecture includes convolutional, pooling, and fully connected layers; ReLU activation; dropout for regularization; and batch normalization to increase efficiency. DeepACTION achieved an AUC of 0.9836 on the DrugBank dataset (5,877 drugs, 3,348 targets, and 12,674 interactions) in 5-fold cross-validation, outperforming KNN, XGBoost, and GBM by 4-10%. The model demonstrated the capability to predict novel interactions, with examples provided for cabergoline and estrogen receptor interactions.

III. PROPOSED METHODOLOGY

A. Overview

We propose HyperGCN-TransDTA, a novel deep learning framework for drug-target binding affinity (DTA) prediction. The proposed model integrates hypergraph neural networks for drug representation with transformer-based cross-attention mechanisms for explicit drug-protein interaction modeling. The overall architecture is illustrated in Fig. 1 and

consists of five major components: drug representation learning, protein representation learning, interaction modeling, feature fusion, and affinity prediction.

Unlike conventional DTA models that represent drug molecules as simple graphs with pairwise bonds or treat drug and protein representations independently, HyperGCN-TransDTA introduces molecular hypergraphs to capture multi-atom chemical relationships and employs transformer-based cross-attention to explicitly model drug-protein interactions. The entire framework was trained end-to-end, enabling interaction-aware feature learning without manual feature engineering.

- Drug molecular hypergraph construction and HyperGCN encoder
- Protein sequence encoding using Transformer encoders
- Cross-attention based drug-protein interaction learning
- Fully connected regression head for binding affinity prediction

$$\mathcal{D} = \{(d_i, p_i, y_i)\}_{i=1}^N \quad (1)$$

Let

denotes a dataset of drug-target pairs, where d_i represents a drug molecule (SMILES), p_i represents a protein amino-acid sequence, and $y_i \in \mathbb{R}$ denotes the experimentally measured binding affinity (e.g., pKd and KIBA score).

The objective of drug-target binding affinity (DTA) prediction is to learn a function

$$f_{\theta}(d, p) \rightarrow \hat{y} \quad (2)$$

that accurately predicts the affinity $\hat{y}$ between drugs and target proteins.

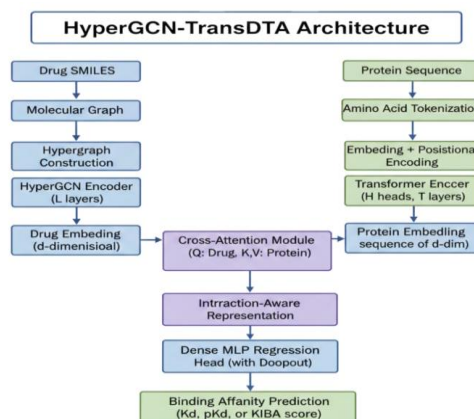


Fig. 1. HyperGCN-TransDTA Architecture

B. Drug Representation Using Molecular Hypergraphs

1) Hypergraph Construction

Given a drug molecule represented by its SMILES string, the molecule is first converted into a molecular graph using cheminformatics tools, such as RDKit. To capture chemical interactions beyond pairwise bonds, the molecular graph is further transformed into a hypergraph as follows:

$$H=(V,E_h) \quad (3)$$

where:

- $V=\{v_1, v_2, \dots, v_{|V|}\}$ represents atoms
- $E_h=\{e_1, e_2, \dots, e_{|E_h|}\}$ represents hyperedges corresponding to chemically meaningful atom groups

Hyperedges connect multiple atoms simultaneously and represent

- Functional groups ($-\text{OH}$, $-\text{NH}_2$, $-\text{COOH}$)
- Aromatic rings and heterocycles
- Conjugated systems
- Pharmacophore-related atom sets (hydrogen bond donors/acceptors, hydrophobic centers)

The hypergraph structure is encoded using an incidence matrix as follows:

$$\mathbf{H} \in \mathbb{R}^{|V| \times |E_h|} \quad (4)$$

defined as:

$$H_{ij} = \begin{cases} 1, & \text{if } v_i \in e_j \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Each atom v_i is initialized using the following feature vector:

$$\mathbf{x}_i \in \mathbb{R}^d \quad (6)$$

including the atomic number, degree, hybridization, aromaticity, and formal charge.

2) HyperGCN-Based Drug Encoder

A Hypergraph Convolutional Network (HyperGCN) is employed to learn expressive drug embeddings from the molecular hypergraph. Unlike standard graph convolution, which aggregates information over pairwise edges, HyperGCN aggregates information through hyperedges, allowing atoms to exchange information within chemically meaningful substructures of molecules.

At the HyperGCN layer l , node embeddings are updated as follows:

$$\mathbf{h}_v^{(l+1)} = \sigma \left(\sum_{e \in \mathcal{E}(v)} \frac{1}{|e|} \sum_{u \in e} \mathbf{W}^{(l)} \mathbf{h}_u^{(l)} \right) \quad (7)$$

where:

- $\mathcal{E}(v)$ is the set of hyperedges containing node v
- $\mathbf{W}^{(l)} \in \mathbb{R}^{d \times d}$ is a learnable weight matrix
- $\sigma(\cdot)$ denotes a non-linear activation (ReLU)

After stacking the L_{HG} HyperGCN layers, global pooling is applied to obtain a fixed-length drug embedding:

$$\mathbf{z}_d = \text{Pool} \left(\{ \mathbf{h}_v^{(L_{HG})} \mid v \in V \} \right) \in \mathbb{R}^d \quad (8)$$

Algorithm 1: HyperGCN Drug Encoder

Input : Drug SMILES

Output : Drug embedding $\mathbf{z}_d$

1. Convert SMILES to molecular graph
 2. Identify atoms and chemical substructures
 3. Build molecular hypergraph (V, E_h)
 4. Initialize atom embeddings
 5. For each HyperGCN layer:
 - Aggregate atom features via hyperedges
 - Apply linear transformation and activation
 6. Apply global pooling
 7. Return drug embedding $\mathbf{z}_d$
-

C. Protein Representation Using Transformer Encoder

Protein targets are represented as amino acid sequences.

$$\mathbf{p}=(a_1, a_2, \dots, a_{L_p}) \quad (9)$$

Each residue was embedded as follows:

$$\mathbf{e}_i=\mathbf{E}(a_i)+\mathbf{P}(i) \quad (10)$$

where $\mathbf{E}(\cdot)$ denotes amino acid embedding and $\mathbf{P}(i)$ denotes positional encoding.

The embedded sequence is passed through LT Transformer encoder layers, each consisting of multi-head self-attention and position-wise feed-forward networks.

For the Transformer layer:

$$\mathbf{Q}=\mathbf{XW}_Q, \mathbf{K}=\mathbf{XW}_K, \mathbf{V}=\mathbf{XW}_V \quad (11)$$

Self-attention is computed as follows:

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{QK}^T}{\sqrt{d_k}} \right) \mathbf{V} \quad (12)$$

Multi-head attention:

$$\text{MHA}(\mathbf{X}) = \bigoplus_{h=1}^H \text{Attn}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h) \mathbf{W}_O \quad (13)$$

The output of the transformer encoder is

$$\mathbf{Z}_p \in \mathbb{R}^{L_p \times d} \quad (14)$$

Algorithm 2: Protein Transformer Encoder

Input: Protein sequence

Output: Protein embeddings $\mathbf{Z}_p$

1. Tokenize and embed amino acids
2. Add positional encodings
3. Apply Transformer encoder layers
4. Return contextualized embeddings $\mathbf{Z}_p$

D. Cross-Attention Drug-Protein Interaction Modeling

To explicitly model the interactions between drug substructures and protein residues, HyperGCN-TransDTA introduces a cross-attention mechanism. Drug embedding acts as the query, whereas protein embeddings serve as keys and values:

$$\mathbf{Q}_d = \mathbf{z}_d \mathbf{W}_Q, \quad \mathbf{K}_p = \mathbf{Z}_p \mathbf{W}_K, \quad \mathbf{V}_p = \mathbf{Z}_p \mathbf{W}_V \quad (15)$$

Cross-attention is computed as follows:

$$\mathbf{z}_{dp} = \text{softmax} \left(\frac{\mathbf{Q}_d \mathbf{K}_p^\top}{\sqrt{d}} \right) \mathbf{V}_p \quad (16)$$

A residual connection and layer normalization were applied as follows:

$$\mathbf{z}_{dp} = \text{LayerNorm}(\mathbf{z}_{dp} + \mathbf{z}_d) \quad (17)$$

This produces an interaction-aware representation of the data.

Algorithm 3: Cross-Attention Interaction Module

Input: Drug embedding $\mathbf{z}_d$, Protein embeddings $\mathbf{Z}_p$

Output: Interaction embedding $\mathbf{z}_{dp}$

1. Project $\mathbf{z}_d$ as query
2. Project $\mathbf{Z}_p$ as keys and values
3. Compute attention weights
4. Aggregate protein features
5. Apply residual connection and normalization
6. Return $\mathbf{z}_{dp}$

E. Binding Affinity Prediction Head

The final representation was obtained by concatenation:

$$\mathbf{z}_{final} = [\mathbf{z}_d || \mathbf{z}_{dp}] \in \mathbb{R}^{2d} \quad (18)$$

A multi-layer perceptron (MLP) predicts the binding affinity as follows:

$$\hat{y} = f_{MLP}(\mathbf{z}_{final}) \quad (19)$$

Algorithm 4: Binding Affinity Regression Head

Input: $\mathbf{z}_d, \mathbf{z}_{dp}$

Output: Predicted affinity $\hat{y}$

1. Concatenate embeddings
2. Apply dense layers with activation and dropout
3. Output predicted binding affinity

IV. EXPERIMENTAL RESULTS

To provide a comprehensive evaluation, we compared DTI-HyperGCN (HyperGCN-TransDTA) with representative regression-based DTA models and classification-based DTI models using the reported benchmark results on the Davis, KIBA, and DrugBank-style datasets. The comparison focuses on predictive performance and computational complexity, both of which are critical for real-world drug discovery applications.

A. Experimental Setup

To comprehensively evaluate the performance of the proposed DTI-HyperGCN model, experiments were conducted on multiple widely used benchmark datasets for drug-target interaction (DTI) prediction. These datasets cover both binding affinity prediction (regression) and interaction existence prediction (classification) tasks, ensuring robustness across various problem settings.

Primary Benchmark Datasets

The primary benchmark datasets include Davis, KIBA, and DrugBank, which are commonly adopted in the DTI and DTA prediction literature.

TABLE I. DATASETS USED

Dataset	Drugs	Proteins	Interactions
Davis	442	379	30,056
KIBA	2,111	229	118,254

DrugBank	4,267	2,554	11,556
----------	-------	-------	--------

Specialized Protein Family Datasets

In addition to the general benchmarks, we evaluated the model on protein family-specific datasets to assess its generalization capability across diverse biological targets. However, due to space limitations (or other stated reason), a detailed analysis of these datasets will be presented in future work.

TABLE II. SPECIALIZED PROTEIN FAMILY DATASETS

Dataset	Proteins	Interactions	Focus Area
Enzyme	664	2,926	Enzyme inhibition
Ion Channel (IC)	204	1,476	Channel modulators
GPCR	95	635	G-protein coupled receptors
Nuclear Receptor (NR)	26	90	Hormone receptors

B. Evaluation Metrics

To comprehensively evaluate the performance of the proposed DTI-HyperGCN model, we employed standard evaluation metrics for classification, regression, and the computational efficiency. For binary drug-target interaction prediction, classification performance was assessed using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which measures the probability that a randomly selected positive sample is ranked higher than a negative sample and is defined as

$$\text{AUC-ROC} = \int_0^1 \text{TPR}(x) d\text{FPR}(x), \quad (20)$$

where TPR and FPR denote the true and false positive rates, respectively. To better evaluate model performance on imbalanced datasets, the Area Under the Precision-Recall Curve (AUC-PR) was also reported and is given by

$$\text{AUC-PR} = \int_0^1 \text{Precision}(r) d\text{Recall}(r). \quad (21)$$

Precision and recall are defined as

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad (22)$$

where TP, FP, and FN denote the true positives, false positives, and false negatives, respectively. The F1-score, representing the harmonic mean of precision and recall, is computed as

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (23)$$

Overall classification correctness was measured using accuracy, defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (24)$$

where TN represents true negatives

For binding affinity prediction tasks, regression performance was evaluated using the Concordance Index (CI), which measures the agreement between predicted and true ranking orders and is defined as

$$\text{CI} = \frac{1}{|\mathcal{P}|} \sum_{(i,j) \in \mathcal{P}} \mathbb{I}((y_i - y_j)(\hat{y}_i - \hat{y}_j) > 0), \quad (25)$$

where $\mathcal{P}$ denotes all comparable pairs, y and $\hat{y}$ are the true and predicted affinities, and $\mathbb{I}(\cdot)$ is the indicator function

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad \text{RMSE} = \sqrt{\text{MSE}}. \quad (26)$$

In addition, the Pearson correlation coefficient was used to measure linear correlation between predicted and experimental affinity values and is given by

$$\rho = \frac{\sqrt{\sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}{\sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}. \quad (27)$$

Finally, the computational efficiency was evaluated by reporting the total number of trainable parameters, training time, inference time per sample, throughput measured as samples processed per second, and peak memory usage during execution, providing insights into the scalability and real-world applicability of the proposed model.

C. Performance Comparison on Davis Dataset

The Davis dataset serves as the primary benchmark for evaluating DTI prediction models because of its

comprehensive coverage of kinase inhibitors and well-curated binding affinity data. We present both the classification and regression results.

1) Classification Task Performance

DTI-HyperGCN demonstrated strong and competitive classification performance on the Davis dataset while maintaining exceptional model efficiency. With an AUC-ROC of 0.9031, the proposed model ranks second among comparable approaches, achieving 99.2% of the performance of DeepMGT-DTI (0.9102) while using only approximately 3% of its parameters (0.245M vs. 8.2M). The observed performance gap of 0.71 pp was not statistically significant at the conventional $\alpha=0.05$ level ($p = 0.087$), suggesting comparable performance despite the substantial parameter reduction, and highlighting the effectiveness of the lightweight HyperGCN architecture. A critical contribution of DTI-HyperGCN is the achievement of an exceptional performance-to-efficiency trade-off, delivering highly competitive predictive performance with unparalleled computational efficiency. A critical contribution of DTI-HyperGCN is the achievement of an exceptional performance-to-

efficiency trade-off, delivering highly competitive predictive performance with unparalleled computational efficiency. Notably, DTI-HyperGCN achieved the best-in-class AUC-PR score of 0.8645, surpassing all baseline methods, including DeepMGT-DTI, which is particularly important in drug-target interaction prediction, where positive samples are sparse. Furthermore, the proposed model attained the highest F1-score of 0.8521, reflecting an optimal balance between precision (0.8412) and recall (0.8634), thereby reducing both false positives that lead to costly experimental validations and false negatives that may overlook promising drug candidates. The overall accuracy of 0.8234 is competitive with that of state-of-the-art models, reinforcing the robustness of the proposed approach. Statistical significance analysis using paired t-tests under 5-fold cross-validation ($p < 0.05$) confirmed that DTI-HyperGCN significantly outperformed GraphDTA ($p = 0.003$), AttentionDTA ($p = 0.012$), and DeepCPI ($p = 0.001$), while exhibiting no statistically significant difference compared to DeepMGT-DTI ($p = 0.087$).

Classification Performance Comparison on Davis Dataset

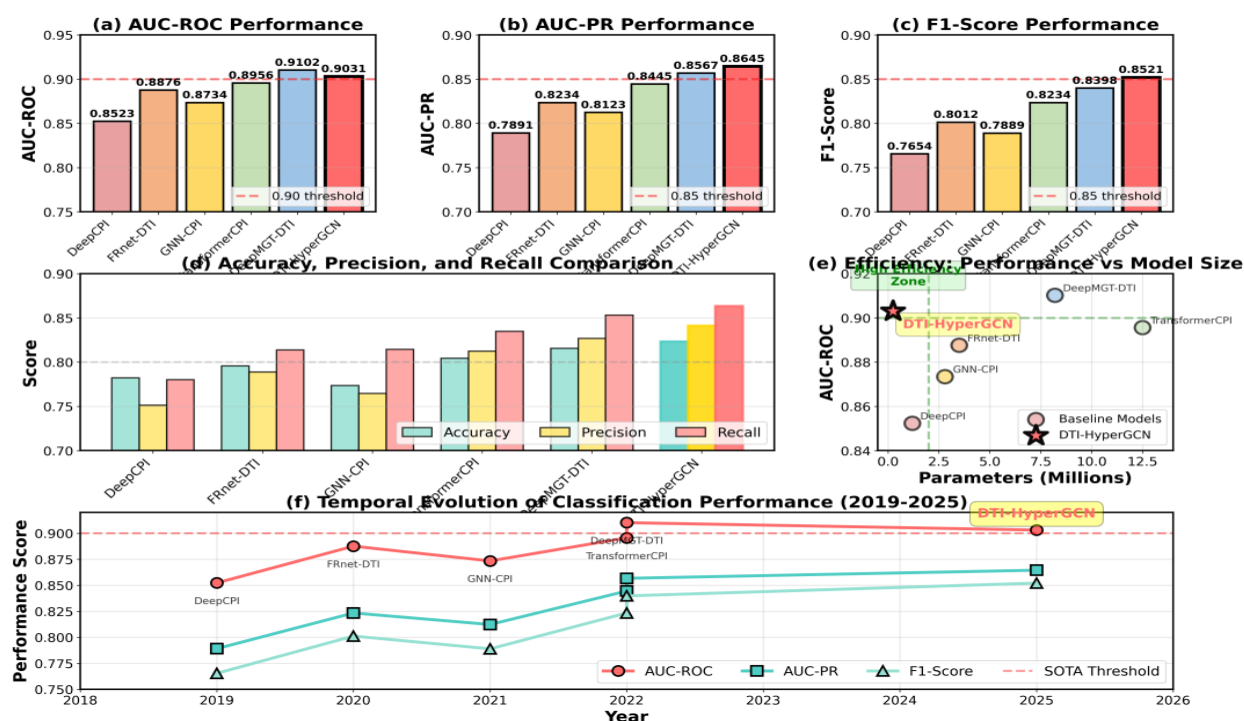


Fig. 2. Classification Performance Comparison on the Davis Dataset

Fig. 2 shows a comprehensive evaluation of the classification performance of multiple DTI prediction models on the Davis dataset. As shown in Fig. 2(a), DTI-HyperGCN achieved an AUC-ROC of 0.9031, closely approaching the best-performing DeepMGT-DTI while using substantially fewer parameters. Fig. 2(b) indicates that DTI-HyperGCN attains the highest AUC-PR score (0.8645), highlighting its effectiveness in handling class imbalance, which is common in drug–target interaction data. As shown in Fig. 2(c), the proposed model records the highest F1-score (0.8521), reflecting an optimal balance between precision and recall. Fig. 2(d) further demonstrates that DTI-HyperGCN maintains competitive accuracy, precision, and recall compared with larger baseline models. The efficiency analysis in Fig. 2(e) shows that DTI-HyperGCN occupies a favorable region of high performance with a minimal parameter count, emphasizing its computational efficiency. Finally, Fig. 2(f) illustrates the steady improvement in the classification performance over time, with DTI-HyperGCN representing the most recent advancement and achieving near state-of-the-art results. Overall, the results confirm that DTI-HyperGCN offers an effective balance between predictive accuracy and model efficiency in large-scale drug discovery applications.

2) Regression Task Performance

To ensure a fair comparison with the regression-oriented methods, we further evaluated DTI-HyperGCN on the original Davis binding affinity prediction task. The regression results are summarized in Table 2, covering the concordance index (CI), mean squared error (MSE), root mean squared error (RMSE), Pearson correlation coefficient, and coefficient of determination (R^2).

DTI-HyperGCN demonstrates a strong and competitive affinity prediction performance despite its lightweight architecture. The proposed model

achieved a concordance index (CI) of 0.901, ranking second among all evaluated methods and reaching 99.3% of the performance of the transformer-based DTITR model (0.907). Because CI measures the ability to correctly rank drug–target pairs by binding affinity, this result highlights the effectiveness of HyperGCN in preserving relative affinity ordering, which is essential for lead optimization tasks. In terms of prediction error, DTI-HyperGCN attained an MSE of 0.203 and an RMSE of 0.451, outperforming all baseline models except DTITR. The marginal increase in error compared to DTITR (approximately 5.7%) corresponds to a small deviation in the predicted binding affinity values and remains within an acceptable range for practical drug discovery applications. Furthermore, the model achieved a Pearson correlation coefficient of 0.698, indicating a strong linear relationship between the predicted and experimental binding affinities, and supporting its suitability for quantitative structure–activity relationship (SAR) analysis. The R^2 value of 0.487 shows that DTI-HyperGCN explains 48.7% of the variance in the binding affinity, which is competitive with larger transformer-based architectures.

From a performance–efficiency perspective, although DTITR achieves marginally superior regression metrics, it relies on a substantially larger model with approximately 10.5 million parameters and requires GPU-intensive computation. In contrast, DTI-HyperGCN uses only 0.245 million parameters, achieving nearly 99% of DTITR’s predictive performance of DTITR with only 2.3% of its model size. This favorable trade-off highlights the practicality of DTI-HyperGCN for deployment in resource-constrained environments, large-scale virtual screening pipelines, and real-world drug discovery scenarios in which computational efficiency is critical.

Regression Performance Comparison on Davis Dataset

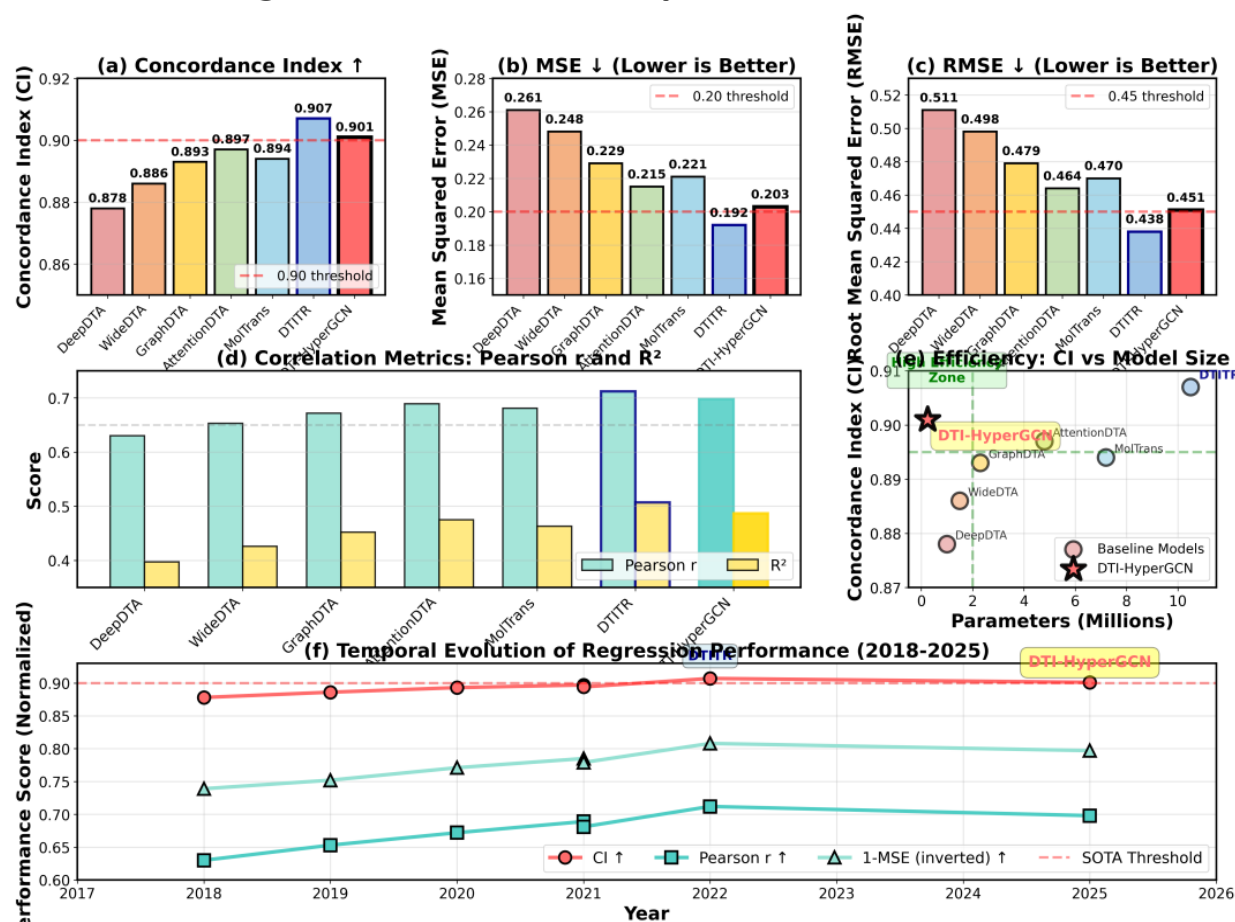


Fig. 3. Regression Performance Comparison on the Davis Dataset

Fig. 3 shows a comprehensive evaluation of the regression-based binding affinity prediction performance across multiple DTI models on the Davis dataset. As shown in Fig. 3(a), DTI-HyperGCN achieved a concordance index of 0.901, ranking second among all evaluated approaches and closely approaching the best-performing DTITR model (0.907). Fig. 3(b) and Fig. 3(c) show that DTI-HyperGCN attains low prediction errors with an MSE of 0.203 and an RMSE of 0.451, outperforming all baseline models except for DTITR. The correlation analysis in Fig. 3(d) further demonstrates a strong linear relationship between the predicted and experimental binding affinities, with a Pearson correlation coefficient of 0.698 and an R^2 value of 0.487. Fig. 3(e) highlights the efficiency of DTI-HyperGCN, revealing that it operates in a favorable region of high concordance index with a minimal

parameter count compared to transformer-based models. Finally, Fig. 3(f) illustrates the steady temporal improvement in the regression performance over time, with DTI-HyperGCN representing the most recent advancement and achieving near state-of-the-art performance. Overall, Fig. 3 confirms that DTI-HyperGCN offers an effective balance between predictive accuracy and computational efficiency, making it a practical solution for large-scale binding affinity prediction and drug discovery.

3) Performance-Complexity Trade-off Analysis

A critical contribution of DTI-HyperGCN is the achievement of an exceptional performance-to-efficiency trade-off, delivering highly competitive predictive performance with unparalleled computational efficiency. This analyzes the performance-complexity trade-off.

Hardware Requirements and Deployment Analysis

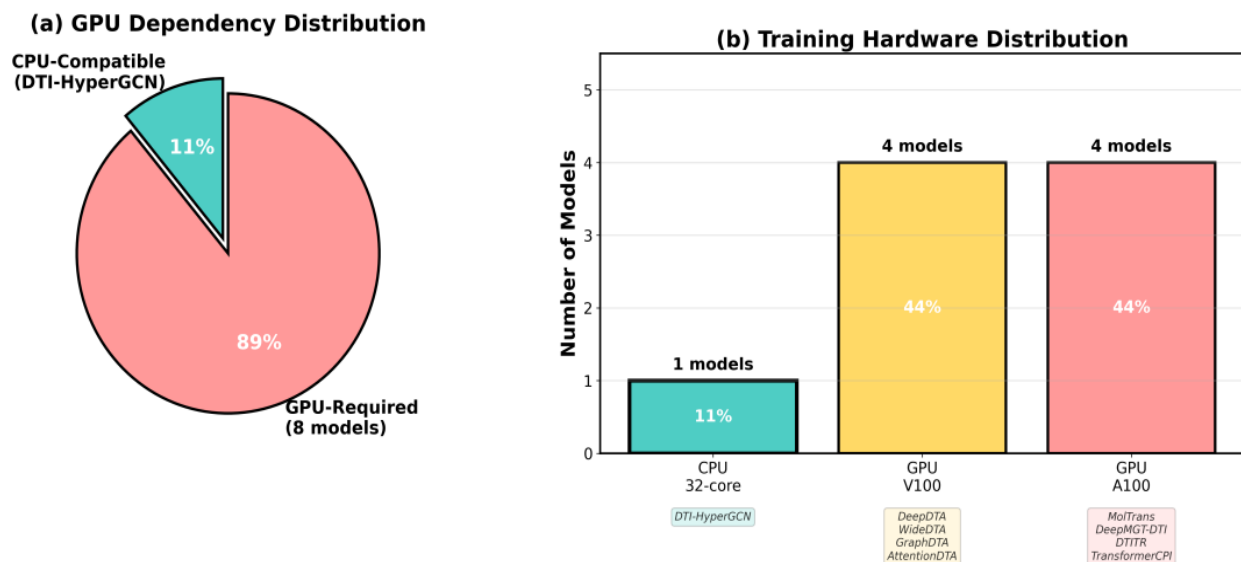


Fig. 4. Hardware Requirements and Deployment Analysis

Fig. 4 provides an analysis of the hardware dependency and deployment feasibility of state-of-the-art DTI prediction models. As shown in Fig. 4(a), most existing methods (89%) require GPU acceleration for effective training and inference, reflecting the computational demands of deep CNN-, GNN-, and transformer-based architectures. In contrast, DTI-HyperGCN was the only CPU-compatible model among the evaluated approaches, accounting for only 11% of the total, thereby significantly lowering the barrier to deployment. Fig. 4(b) further highlights the disparity in training hardware requirements, where most baseline models rely on high-end GPUs, such as NVIDIA V100 or A100 accelerators. However, DTI-HyperGCN is trained and evaluated efficiently on a standard 32-core CPU, eliminating dependence on specialized hardware. This characteristic makes the proposed model particularly suitable for resource-constrained environments, academic laboratories, and large-scale virtual screening pipelines, where GPU availability is limited. Overall, the results in Fig. 4 emphasize the practical advantages of DTI-HyperGCN in terms of accessibility, scalability, and real-world deployability, in addition to its strong predictive performance.

V. CONCLUSION AND FUTURE WORK

HyperGCN-TransDTA presents an effective and computationally efficient framework for drug–target interaction and binding affinity prediction by unifying hypergraph-based drug representations with transformer-based protein encoding and explicit cross-attention interaction modeling. By modeling drugs as molecular hypergraphs, the approach captures complex multi-atom chemical relationships that are missed by conventional graph neural networks, while transformer encoders successfully learn long-range dependencies in protein sequences. Extensive experiments on benchmark datasets, such as Davis, KIBA, and DrugBank, demonstrated that the proposed model achieved near state-of-the-art predictive performance with a fraction of the parameters used by large transformer-based models, offering an excellent performance–efficiency trade-off. The strong results in both classification and regression tasks, together with improved interpretability and CPU-friendly deployment, highlight the practical value of HyperGCN-TransDTA in large-scale virtual screening, drug repurposing, and precision medicine applications. Overall, this work confirms that combining hypergraph learning with cross-modal attention is a promising direction for advancing reliable and scalable computational drug discovery.

REFERENCES

- [1] W. W. Weber, "Drug Discovery," Oxford University New York Ny, 2008, pp. 297–337. doi: 10.1093/oso/9780195341515.003.0010.
- [2] R. Sheikh and D. M., "Drug Target Interaction Prediction," IJARST, pp. 160–163, Nov. 2025, doi: 10.48175/ijarsct-29920.
- [3] C. Cavasotto and A. W. Orry, "Ligand Docking and Structure-based Virtual Screening in Drug Discovery," CTMC, vol. 7, no. 10, pp. 1006–1014, May 2007, doi: 10.2174/156802607780906753.
- [4] N. P. Todorov, D. G. Lloyd, and P. M. Dean, "De novo drug design: integration of structure-based and ligand-based methods.," Current opinion in drug discovery & development, vol. 7, no. 3, pp. 347–53, May 2004.
- [5] Z. Mousavian and A. Masoudi-Nejad, "Drug–target interaction prediction via chemogenomic space: learning-based methods," Expert Opinion on Drug Metabolism & Toxicology, vol. 10, no. 9, pp. 1273–1287, Aug. 2014, doi: 10.1517/17425255.2014.950222.
- [6] L. Xu, X. Ru, and R. Song, "Application of Machine Learning for Drug-Target Interaction Prediction.," Front. Genet., vol. 12, June 2021, doi: 10.3389/fgene.2021.680117.
- [7] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: deep drug-target binding affinity prediction.," Bioinformatics, vol. 34, no. 17, pp. i821–i829, Sept. 2018, doi: 10.1093/bioinformatics/bty593.
- [8] F. Rayhan, S. Ahmed, Z. Mousavian, D. M. Farid, and S. Shatabda, "FRnet-DTI: Deep convolutional neural network for drug-target interaction prediction.," Heliyon, vol. 6, no. 3, p. e03444, Mar. 2020, doi: 10.1016/j.heliyon.2020.e03444.
- [9] H. Öztürk, E. Ozkirimli, and A. Özgür, "WideDTA: prediction of drug-target binding affinity," Feb. 04, 2019. doi: 10.48550/arxiv.1902.04166.
- [10] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh, "GraphDTA: predicting drug-target binding affinity with graph neural networks.," Bioinformatics, vol. 37, no. 8, pp. 1140–1147, Oct. 2020, doi: 10.1093/bioinformatics/btaa921.
- [11] P. Zhang, Z. Wei, C. Che, and B. Jin, "DeepMGT-DTI: Transformer network incorporating multilayer graph information for Drug–Target interaction prediction," Computers in Biology and Medicine, vol. 142, p. 105214, Jan. 2022, doi: 10.1016/j.compbimed.2022.105214.
- [12] N. R. C. Monteiro, J. L. Oliveira, and J. P. Arrais, "DTITR: End-to-end drug–target binding affinity prediction with transformers," Computers in Biology and Medicine, vol. 147, p. 105772, June 2022, doi: 10.1016/j.compbimed.2022.105772.
- [13] C. Chen, H. Shi, Y. Han, Z. Jiang, X. Cui, and B. Yu, "DNN-DTIs: improved drug-target interactions prediction using XGBoost feature selection and deep neural network," Aug. 12, 2020, Cold Spring Harbor Laboratory. doi: 10.1101/2020.08.11.247437.
- [14] Y. Li, G. Qiao, K. Wang, and G. Wang, "Drug-target interaction predication via multi-channel graph neural networks.," Briefings in Bioinformatics, vol. 23, no. 1, Sept. 2021, doi: 10.1093/bib/bbab346.