

HBDTA: Hierarchical Bi-LSTM Networks for Drug-target Binding Affinity Prediction

Yongqing Wu *, Yao Jin, Peng Sun, Zhichen Ding

Abstract—Accurately predicting drug-target binding affinity (DTA) is crucial for advancements in drug repositioning. In this paper, we present HBDTA, an innovative predictive methodology that harnesses the capabilities of graph deep learning and multi-layer networks for DTA prediction. The HBDTA approach encompasses a comprehensive framework with three distinct graph neural network (GNN) algorithms: multi-head graph attention networks (Multi-head GAT), generalized aggregation networks (GENConv), and graph convolutional networks (GCNConv) designed to extract drug features. Additionally, we employ a multi-layer bi-directional long short-term memory (MBLSTM) with residual blocks to extract protein features. After deriving feature vectors for drugs and proteins, they independently pass through fully connected layers before integration into a self-attention layer. Subsequently, the resulting feature vectors are concatenated and passed through four layers of fully secured networks to facilitate prediction. Finally, we assess our model's performance on the Davis, KIBA, Metz, and DTC datasets. Comparative analysis against state-of-the-art methodologies, such as DeepGLSTM, DeepNC, and GraphDTA, among others, underscores the effectiveness of HBDTA. The results suggest that HBDTA holds significant potential for practical drug discovery and personalized medicine applications.

Index Terms—Drug-target binding affinity, Deep learning, Proteins, Residual blocks, Multi-layer networks

I. INTRODUCTION

THE accurate prediction of drug-target binding affinity (DTA) is a critical technology in both drug development and the domain of drug repositioning [1]. The pursuit of discovering novel drugs is accompanied by substantial financial investments, with annual costs estimated at approximately \$2.6 billion, as reported in the existing literature [2]. Moreover, it is well-known that the regulatory approval process overseen by the Food and Drug Administration (FDA) is renowned for its lengthy duration [3]. Despite these challenges, the continuous advancement of knowledge at the intersection of computer science and biology has opened up new avenues for drug development. Notably, the convergence of these fields has led to the emergence of graph deep learning, a powerful aspect of deep learning that has garnered significant attention in the pharmaceutical research domain. The availability of extensive datasets linking drugs and targets has sparked significant interest in harnessing

machine learning and deep learning techniques for drug development.

Machine learning techniques entail extracting insights from extensive datasets comprising known drug-target pairs, the extraction of relevant features, and the formulation of prediction models. As an illustrative example, SimBoost [4] employs a gradient-boosting methodology utilizing features derived from drugs, targets, and drug-target pairs to predict binding affinity. In contrast, KronRLS [5] uses a kernel-based approach to determine similarity between drugs or proteins. Deep learning methodologies, on the other hand, leverage multi-layer neural networks to capture high-order features of drugs and targets, thereby enhancing prediction accuracy. For instance, DeepDTA [6] focuses on one-dimensional representations of targets, utilizing a three-layer convolutional neural network (CNN) [7] for processing drugs and extracting representations. Its evolution, WideDTA [8], refines and extends this approach by incorporating four text-based information sources. WideDTA leverages four text-based information sources, including the protein sequence, ligand Simplified Molecular Input Line Entry System (SMILES) [9], protein domains and motifs, and maximum common substructure words to predict binding affinity. Taking a novel approach, DeepCPI [10] harnesses the power of attention mechanisms within neural networks for improved performance through an end-to-end methodology. Attention mechanisms enable the identification of crucial proteins in drug-target interactions. Attention mechanisms enable the identification of crucial proteins in drug-target interactions (DTI). MT-DTI [11] introduces an original molecular representation based on a self-attention tool, accompanied by a unique drug-target interaction model. GANsDTA [12] advances a semi-supervised generative adversarial network-driven approach, divided into feature extraction and prediction regression networks. AttentionDTA [13] culminates in a comprehensive deep learning-driven framework, combining an attention mechanism with the prediction of DTI binding affinity.

However, the scope of deep learning methodologies is limited when it comes to handling non-Euclidean data. In the field of drug discovery, where molecules are represented as graphs, graph neural networks (GNNs) emerge as a promising solution for tackling the challenge of modeling non-Euclidean data. GraphDTA [14], for instance, configures drugs as graph structures and employs GNNs [15] to predict drug-target binding affinity. This method utilizes four variants of GNNs to predict drugs while processing protein sequences using one-dimensional convolution. DeepGS [16] harnesses the power of deep neural networks to extract localized chemical context from amino acid and SMILES sequences, combining them with graph neural networks to capture the intricate molecular structures within drugs. DeepGLSTM [17], on the other hand, introduces a graph

Manuscript received May 24, 2023; revised December 7, 2023.

This work was supported by the National Natural Science Foundation of China (Grant No. 52174184).

Y. Q. Wu is an Associate Professor of School of Software, Liaoning Technical University, Huludao 125105, China (e-mail: yqwuyywu@163.com).

Y. Jin is a postgraduate student of School of Software, Liaoning Technical University, Huludao 125105, China (e-mail: jy937005748@163.com).

P. Sun is a postgraduate student of School of Software, Liaoning Technical University, Huludao 125105, China (e-mail: sp_9885@126.com).

Z. C. Ding is a postgraduate student of School of Software, Liaoning Technical University, Huludao 125105, China (e-mail: dingzhichen1997@outlook.com).

convolution networks (GCN) block to process drug compounds using power graph representation and a Bi-directional Long Short-Term Memory Network (Bi-LSTM) [18] layer to process protein sequences. DeepNC [19] uses generalized aggregation networks (GENConv) [20], graph convolutional networks, and hypergraph convolution-hypergraph attention (HypergraphConv) [21] to extract drug features, along with a one-dimensional CNNs (1D-CNNs) architecture to extract protein attributes. Above prediction methods are summarized in Table I.

However, regardless of whether these models are rooted in graph deep learning or traditional machine learning paradigms, they tend to rely on shallow networks [22] when analyzing features embedded in protein sequences. This often involves the use of either 1D-CNNs or simple LSTM [23] structures, resulting in suboptimal extraction of sequence features and consequently impacting the model's accuracy. The use of 1D-CNNs for protein feature extraction overlooks the intricate spatial intricacies inherent in proteins, which can lead to less precise predictions. Furthermore, since excessive layer stacking can lead to gradient vanishing, while insufficient layering may result in inadequate information extraction, the incorporation of a self-attention mechanism [24] within the model presents a viable solution to address this issue.

In this study, we introduce a novel graph deep learning model named HBDTA for predicting the binding affinity between drugs and targets. The HBDTA model integrates a variety of advanced techniques, creating a unique combination of innovations. Within this framework, we employ three distinct variants of graph neural network models to extract features from drugs, demonstrating their suitability for subsequent experimental endeavors. Furthermore, we propose the MBLSTM deep network to capture features from protein sequences. The MBLSTM architecture is characterized by the inclusion of three layers of network connections, each interleaved with a residual block. This strategic addition addresses the common challenge of gradient vanishing often encountered in deep networks. Through MBLSTM, the intricate dependency relationships inherent in protein sequences are effectively captured, encompassing the inherent characteristics of the original signals across various temporal scales, while efficiently managing memory requirements. The orchestrated fusion of drug and protein features is further enriched by our customized self-attention layer. To assess our model's performance, we compare it with other representative methods using four established datasets: Davis [25], KIBA [26], Metz [27], and DTC [28]. Consistently, our HBDTA model outperforms alternative approaches, demonstrating its effectiveness across both datasets.

II. CORRELATION ALGORITHM

A. Multi-head Graph Attention Network

To augment the generalization efficacy of the attention mechanism, we opted for the integration of a multi-head attention [29]. The specific formulation is delineated as follows:

The graph attention layer takes as input a set of node feature vectors, $\mathbf{h} = (\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N), \vec{h}_i \in R^F$, where N is the number of nodes, and F represents the number

of features per node. The matrix \mathbf{h} , spanning dimensions $N \times F$, encapsulates the features of all nodes. Conversely, R signifies the feature of an individual node, thereby assuming a size of $F \times 1$. After a graph attention layer, a fresh feature vector is engendered, its dimension denoted as F' (which might not necessarily align with F). The novel set of node feature vectors assumes the configuration $\mathbf{h}' = (\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N), \vec{h}_i \in R^{F'}$.

To effectuate the requisite input-output transformation, a minimum of one linear conversion predicated on input features is imperative. Hence, a weight matrix $W \in R^{F' \times F}$ is trained for all nodes. This matrix elucidates the interplay between the input's F features and the output's F' features. We then perform self-attention on the nodes- a shared attentional mechanism $a: R^{F'} \times R^{F'} \rightarrow R$ computes attention coefficients. The self-attention mechanism is executed for each node, with the corresponding attention coefficients as follows:

$$e_{ij} = a(W\vec{h}_i, W\vec{h}_j) \quad (1)$$

The significance of node j 's features concerning node i is represented by e_{ij} . In this context, i and j as subscripts refer to the i -th and j -th nodes, respectively. Computation of e_{ij} is limited to nodes j belonging to N_i , which defines the neighborhood of node i within the graph. To facilitate cross-node comparability of coefficients, we employ the *softmax* function:

$$\alpha_{ij} = softmax_j(e_{ij}) = \frac{exp(e_{ij})}{\sum_{k \in N_i} exp(e_{ik})} \quad (2)$$

In our experiments, we employ the attention mechanism a as a single-layer feedforward neural network, which is characterized by a weight vector $\vec{a} \in R^{2F'}$ and utilizes the *LeakyReLU* nonlinearity. The coefficients computed by the attention mechanism may then be expressed as:

$$\alpha_{ij} = \frac{exp(LeakyReLU(\vec{a}^T [W\vec{h}_i || W\vec{h}_j]))}{\sum_{k \in N_i} exp(LeakyReLU(\vec{a}^T [W\vec{h}_i || W\vec{h}_k]))} \quad (3)$$

where $.^T$ represents transposition and $||$ is the concatenation operation. The regularization of attention coefficients across distinct nodes is achieved through the above operations. These coefficients, once obtained, serve as the basis for predicting the output features of each node:

$$\vec{h}_i = \sigma\left(\sum_{j \in N_i} \alpha_{ij} W\vec{h}_j\right) \quad (4)$$

where $\sigma(\cdot)$ represents a non-linear activation function.

In our model, a multi-head graph attention mechanism was implemented, with the number of heads set at 10.

$$\vec{h}_i = \parallel_{k=1}^K \sigma\left(\sum_{j \in N_i} \alpha_{ij}^k W^k \vec{h}_j\right) \quad (5)$$

where K is equivalent to the count of attention mechanisms under consideration, with k signifying the k -th instance within this set. The term α_{ij}^k denotes the weight coefficient computed by the attention mechanism of the k -th group, while W^k embodies the weight coefficient associated with the k -th module.

TABLE I
COMPARISON MODEL

Method	Published time	Model	Summary
SimBoost	2016	Gradient boosting regression trees	Predicting continuous values of binding affinities for compounds and proteins
KronRLS	2018	Multiple kernel learning	Utilizing multiple pairwise kernels for time and memory-efficient learning in the first approach
DeepDTA	2018	CNN	Using CNNs to process protein sequences and 1D compound structures
WideDTA	2019	CNN	Combining four different textual pieces of information related to proteins and ligands
DeepCPI	2019	GNN+CNN	Mining valuable compound and protein features from vast unlabeled corpora using NLP techniques
GANsDTA	2019	GAN+CNN	Effectively learning valuable features from labeled and unlabeled data
DeepGS	2020	GAT+Bi-GRU	Extracting the topological information of the molecular map and the local chemical context of the drug
GraphDTA	2021	GCN/GAT+GCN/GIN/GAT	Introducing multiple models of graph neural networks
Attention-DTA	2022	CNN	A multi-perspective molecular feature attention fusion based deep learning method for predicting DTA
DeepNC	2022	GENconv/HypergraphConv	Two algorithms for extracting drug characteristics
DeepGLSTM	2022	GCN+LSTM	A method based on graph convolutional network and LSTM

Finally, we utilize averaging and postpone the application of the ultimate nonlinearity until that point:

$$\vec{h}_i = \sigma\left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i} \alpha_{ij}^k W^k \vec{h}_j\right) \quad (6)$$

B. Graph Convolutional Network

Graph convolutional neural networks [7] constitute a potent tool for addressing the challenge of data alignment by leveraging vertices and edges to construct topological graphs that encapsulate the associated relationships. Their widespread implementation spans domains such as social networks, information networks, and various other spheres. The fundamental objective of GCNs resides in the acquisition of node-level representations derived from a given input graph $G = (V, E)$. Within this context, V denotes a node feature matrix encompassing attributes for N nodes, while E signifies the assortment of edges connecting nodes. Additionally, $A \in R^{N \times N}$ represents the adjacency matrix that encapsulates the structural composition of the graph. To effectuate this, GCNs deploy weight matrices W^l across each layer l , whereby the graph convolution operation is delineated as follows:

$$H^1 = X \quad (7)$$

$$H^l = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{l-1} W^{l-1}) \quad (8)$$

where $\tilde{A} = A + I_N$ symbolizes the adjacency matrix enriched with self-loops for each node, and the identity matrix is denoted as I_N . The notation W^{l-1} designates the weight matrix of the GCN at the $(l-1)$ -th layer. $\tilde{D} = \sum_j \tilde{A}_{ij}$, and \tilde{D} represents the degree matrix. $\sigma(\cdot)$ is a non-linear function which is a *ReLU* in our later experiments.

C. Residual Blocks

Residual blocks [30], also referred to as skip connections, assume a pivotal role within our model's architecture. Mathematically, the operation of a residual block can be represented as follows:

$$y = F(x) + x \quad (9)$$

where x signifies the input introduced to the block, $F(x)$ signifies the alterations executed by the block, and y designates the outcome emerging from the block. The operator $+$ signifies element-wise addition. To ensure the validity of the element-wise addition, x and $F(x)$ must share identical dimensions. Hence, in scenarios where the dimensions of x and $F(x)$ diverge, a linear projection can be employed on x to harmonize its dimensions with those of $F(x)$, preparatory to the addition operation.

D. Generalized Aggregation Graph Networks

A Graph Convolutional Network (GCN) predicated on the principles of message-passing that adheres to message-passing criteria has been developed [20]. Within the context of the l -th layer, we consider $m_v^{(l)} \in R^D$ as the aggregated message for node v , while $m_{vu}^{(l)} \in R^D$ is designated as an individual message corresponding to each neighbor $u \in N_v$ of node v . The messages from neighbors, alongside the aggregated message for node v , and its associated features, undergo updates as per the subsequent equations:

$$\begin{aligned} m_{vu}^{(l)} &= p^{(l)}(h_v^{(l)}, h_u^{(l)}, h_{e_{vu}}^{(l)}) \\ m_u^{(l)} &= \zeta^{(l)}(m_{vu}^{(l)}) \\ h_v^{(l+1)} &= \phi^{(l)}(h_v^{(l)}, m_v^{(l)}) \end{aligned} \quad (10)$$

where $p^{(l)}$, $\zeta^{(l)}$, and $\phi^{(l)}$ are functions that can be learned or differentiated, intended for the tasks of constructing messages, aggregating messages, and updating nodes at the l -th layer.

The function $\zeta^{(l)}$ responsible for message aggregation is presented with two alternative forms: *SoftMax* aggregation and *PowerMean* aggregation. These generalized functions are defined as follows:

$$\begin{aligned} \text{SoftMax_Agg}_\beta(\cdot) &= \sum_{u \in N(v)} \frac{\exp(\beta m_{vu})}{\sum_{i \in N(v)} \exp(\beta m_{vi})} \\ \text{PowerMean_Agg}_p(\cdot) &= \left(\frac{1}{|N(v)|} \sum_{u \in N(v)} m_{vu}^p\right)^{\frac{1}{p}} \end{aligned} \quad (11)$$

where β represents a continuous variable referred to as the inverse temperature, while p is a non-zero continuous variable indicating the exponent of p .

To facilitate the utilization of $SoftMax_Agg_\beta$ and $PowerMean_Agg_p$, a prerequisite is to ensure that the node message value m_{vu} remains non-negative. This requirement mandates that the message construction function $p^{(l)}$ take the form:

$$m_{vu}^{(l)} = p^{(l)}(h_v^{(l)}, h_u^{(l)}, h_{e_{vu}}^{(l)}) = ReLU(h_u^{(l)} + \varphi(h_{e_{vu}}^{(l)}) \cdot h_{e_{vu}}^{(l)}) + \epsilon \quad (12)$$

where $ReLU$ denotes a rectified linear unit that yields values greater than or equal to 0. The function $\varphi(\cdot)$ is an indicator function that returns 1 when edge features are present and 0 otherwise. The parameter ϵ signifies a minute positive constant selected as 10^{-7} .

During the node update stage, an additional layer for message normalization was incorporated into the node update function:

$$h_v^{(l+1)} = \phi^{(l)}(h_v^{(l)}, m_v^{(l)}) = MLP[h_v^{(l)} + s \cdot \|h_v^{(l)}\|_2 \cdot \left(\frac{m_v^{(l)}}{\|m_v^{(l)}\|_2} \right)] \quad (13)$$

where $MLP[\cdot]$ denotes a multi-layer perceptron, while s signifies a scaling factor that is amenable to learning. In practical implementations, s is assigned the role of a learnable scalar and is initialized with a value of 1.

E. Multi-layer Bi-Directional Long Short-Term Memory Network

The multi-layer bi-directional long short-term memory (MBLSTM) represents a profound deep learning architecture that combines the strengths of Bi-LSTMs and deep neural networks, offering several advantages over traditional LSTM models.

First and foremost, MBLSTM harnesses the potential of multiple layers of LSTM units. Each layer progressively learns more complex representations of the input sequence. The hierarchical structure of these multiple layers empowers the model to capture intricate dependencies and patterns within the data, thereby enhancing its performance in tasks requiring the understanding of long-term dependencies.

Furthermore, MBLSTM benefits from the presence of residual connections that link the layers. These residual blocks facilitate the seamless flow of information from one layer to the next, mitigating the challenge of vanishing gradients and strengthening the model's learning capabilities. These connections also enable the model to retain important information from previous layers, contributing to improved information propagation and gradient flow.

Additionally, MBLSTM exhibits the ability to handle input sequences of variable lengths, making it suitable for tasks involving natural language processing and speech recognition, where input lengths can vary significantly. The bidirectional nature inherent in MBLSTM allows it to generate predictions at each time step, providing essential contextual insights for tasks requiring sequential predictions.

Consider an input sequence denoted as x , where the hidden state for each layer of the MBLSTM network is represented by h , and the cell state is indicated by c . The forward

propagation process for the initial layer of the MBLSTM network can be formally articulated as follows:

$$\begin{aligned} \vec{h}_t^1, \vec{c}_t^1 &= \overrightarrow{LSTM}(\vec{x}_t, \vec{h}_{t-1}^1, \vec{c}_{t-1}^1) \\ \overleftarrow{h}_t^1, \overleftarrow{c}_t^1 &= \overleftarrow{LSTM}(\overleftarrow{x}_t, \overleftarrow{h}_{t+1}^1, \overleftarrow{c}_{t+1}^1) \\ h_t^1 &= \vec{h}_t^1 + \overleftarrow{h}_t^1 \\ c_t^1 &= \vec{c}_t^1 + \overleftarrow{c}_t^1 \\ x_t &= h_t^1 \end{aligned} \quad (14)$$

The forward propagation process for the second layer of the MBLSTM network can be articulated as follows:

$$\begin{aligned} \vec{h}_t^2, \vec{c}_t^2 &= \overrightarrow{LSTM}(\vec{x}_t, \vec{h}_{t-1}^2, \vec{c}_{t-1}^2) \\ \overleftarrow{h}_t^2, \overleftarrow{c}_t^2 &= \overleftarrow{LSTM}(\overleftarrow{x}_t, \overleftarrow{h}_{t+1}^2, \overleftarrow{c}_{t+1}^2) \\ h_t^2 &= \vec{h}_t^2 + \overleftarrow{h}_t^2 + h_t^1 \\ c_t^2 &= \vec{c}_t^2 + \overleftarrow{c}_t^2 + c_t^1 \\ x_t &= h_t^2 \end{aligned} \quad (15)$$

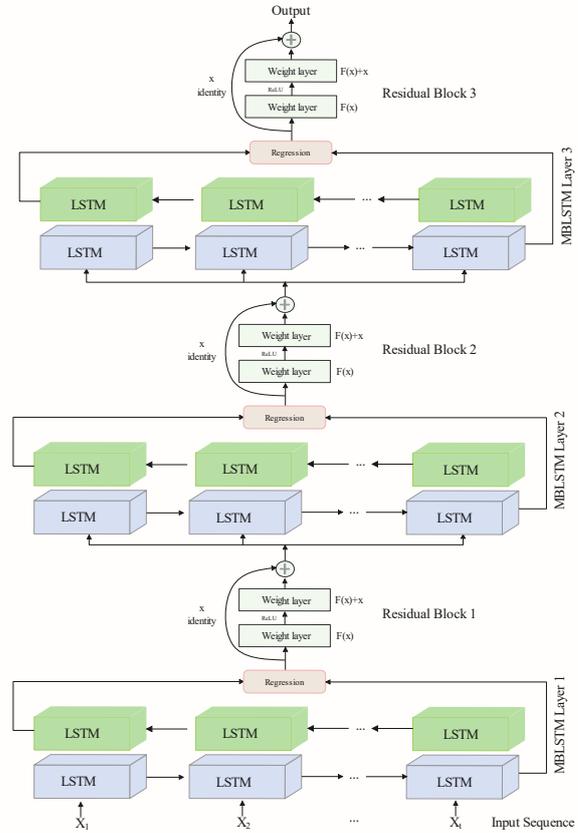


Fig. 1. Schematic diagram illustrating the architecture of the MBLSTM network

The forward propagation process for the third layer of the MBLSTM network can be articulated as follows:

$$\begin{aligned} \vec{h}_t^3, \vec{c}_t^3 &= \overrightarrow{LSTM}(\vec{x}_t, \vec{h}_{t-1}^3, \vec{c}_{t-1}^3) \\ \overleftarrow{h}_t^3, \overleftarrow{c}_t^3 &= \overleftarrow{LSTM}(\overleftarrow{x}_t, \overleftarrow{h}_{t+1}^3, \overleftarrow{c}_{t+1}^3) \\ h_t^3 &= \vec{h}_t^3 + \overleftarrow{h}_t^3 + h_t^2 \\ c_t^3 &= \vec{c}_t^3 + \overleftarrow{c}_t^3 + c_t^2 \\ x_t &= h_t^3 \end{aligned} \quad (16)$$

Among these elements, \overrightarrow{LSTM} and \overleftarrow{LSTM} correspond to the forward and reverse LSTM units, respectively, with

TABLE II
EIGHT ATOMIC CHARACTERIZATIONS

Feature	Description	Size
Atom Type	C, N, O, S, F, Si, P, Cl, Br, Mg, Na, Ca, Fe, As, Al, I, B, V, K, Tl, Yb, Sb, Sn, Ag, Pd, Co, Se, Ti, Zn, H, Li, Ge, Cu, Au, Ni, Cd, In, Mn, Zr, Cr, Pt, Hg, Pb, or "Unknown" (one-hot)	44
Degree	Number of directly bonded neighbors	11
Num of H	Number of H bound to the atom	11
Implicit Valence	Number of implicit H bound to the atom	11
Total Valence	Number of total H bound to the atom	11
Charge	Number of charge bound to the atom	11
Aromaticity	Whether the atom is aromatic	1
InRing	Whether the atom is in a ring	1
Total		101

t representing the time step. It is crucial to highlight that each layer in the MBLSTM network incorporates LSTM units in both forward and reverse directions. The mathematical expressions provided elucidate the forward propagation process within an MBLSTM network consisting of three layers. Each layer includes bi-directional LSTM units and seamlessly integrates the hidden state as well as the cell state from the previous layer into the current layer through residual connections. The comprehensive model architecture is visually depicted in Fig. 1.

III. MATERIALS AND METHODS

A. Methodology

HBDTA serves as a predictive framework designed for the task of drug-target binding affinity prediction through regression. The initial phase involves feature extraction for drug molecules, where the first step is the conversion of Simplified Molecular Input Line Entry System (SMILES) notations of drug molecules into molecular graphs. Subsequently, we utilize the RDKit [31] software to extract atomic attributes and establish the molecular graph structure. Feature representations are then obtained by subjecting the molecular graphs to graph neural networks, which facilitate the extraction of underlying molecular features. The resulting feature vector passes through two fully connected layers, culminating in the definitive feature representation for the drug. For proteins, the feature extraction process involves processing the amino acid sequences through MBLSTM to discern characteristic attributes. These learned feature vectors undergo further processing through two fully connected layers to arrive at the ultimate feature representation for the protein. After passing through the self-attention layer, the resulting embedding vectors for the drug and protein are concatenated and subsequently pass through a sequence of four fully connected layers to generate binding affinity predictions. An illustration of the comprehensive model architecture is presented in Fig. 2.

B. Input Features of Drugs

To enable the computational analysis of drug compounds, the initial step involves transforming them into the SMILES format. This standardized format is then compatible with molecular editing tools, allowing for the generation of 2D or 3D molecular models. To represent pharmacological compounds as graphs that capture atom interactions, we employ the RDKit program to extract atomic-level details. The

diagonal elements of the adjacency matrix are set to 1 to account for self-connections within the graph. Node features are conveyed as 101-dimensional feature vectors, following the approach used in Embed-DTI [32]. Each node's feature vector comprises eight distinct facets of atom information, including the atomic symbol, degree, count of explicit and implicit hydrogen atoms linked to the atom, total valence, charge, aromaticity indicator, and whether the atom forms part of a ring structure. A concise summary of these atomic feature representations applied within our model is provided in Table II.

C. Input Features of Proteins

Proteins are represented using the one-hot encoding technique applied to the target amino acid sequence. The protein sequence, obtained from the UniProt database [33], is provided as a string of ASCII letters corresponding to amino acids. To ensure optimal training conditions, we follow the practice of standardizing the sequence length to 1000 residues through either padding or truncation. In cases where the sequences are shorter than the designated length, we apply padding by introducing zero values. Subsequently, these resulting integer sequences are passed through embedding layers, resulting in the acquisition of a 128-dimensional vector representation. After vectorization, the 128-dimensional vectors undergo further processing through the MBLSTM architecture, ultimately yielding abstract features designed to capture the inherent patterns embedded within the protein sequence data.

D. Datasets

The evaluation of our proposed HBDTA model was conducted using four well-established benchmark datasets: Davis, KIBA, Metz, and DTC datasets. Within this evaluation framework, 80% of the data was allocated for training purposes, while the remaining 20% were set aside for validation. Furthermore, to further assess the model's effectiveness, we conducted "cold experiments" on the Davis and KIBA datasets. In addition, we selected three proteins—namely, Nsp14, ACE2, and Spike—from the DrugBank database [34] to evaluate the potential efficacy of drugs in COVID-19 treatment strategies.

These datasets cover a diverse range of drug-target pairs, each characterized by varying degrees of binding affinity. The evaluation of binding affinity between drug and target pairs relies on different metrics, including the dissociation

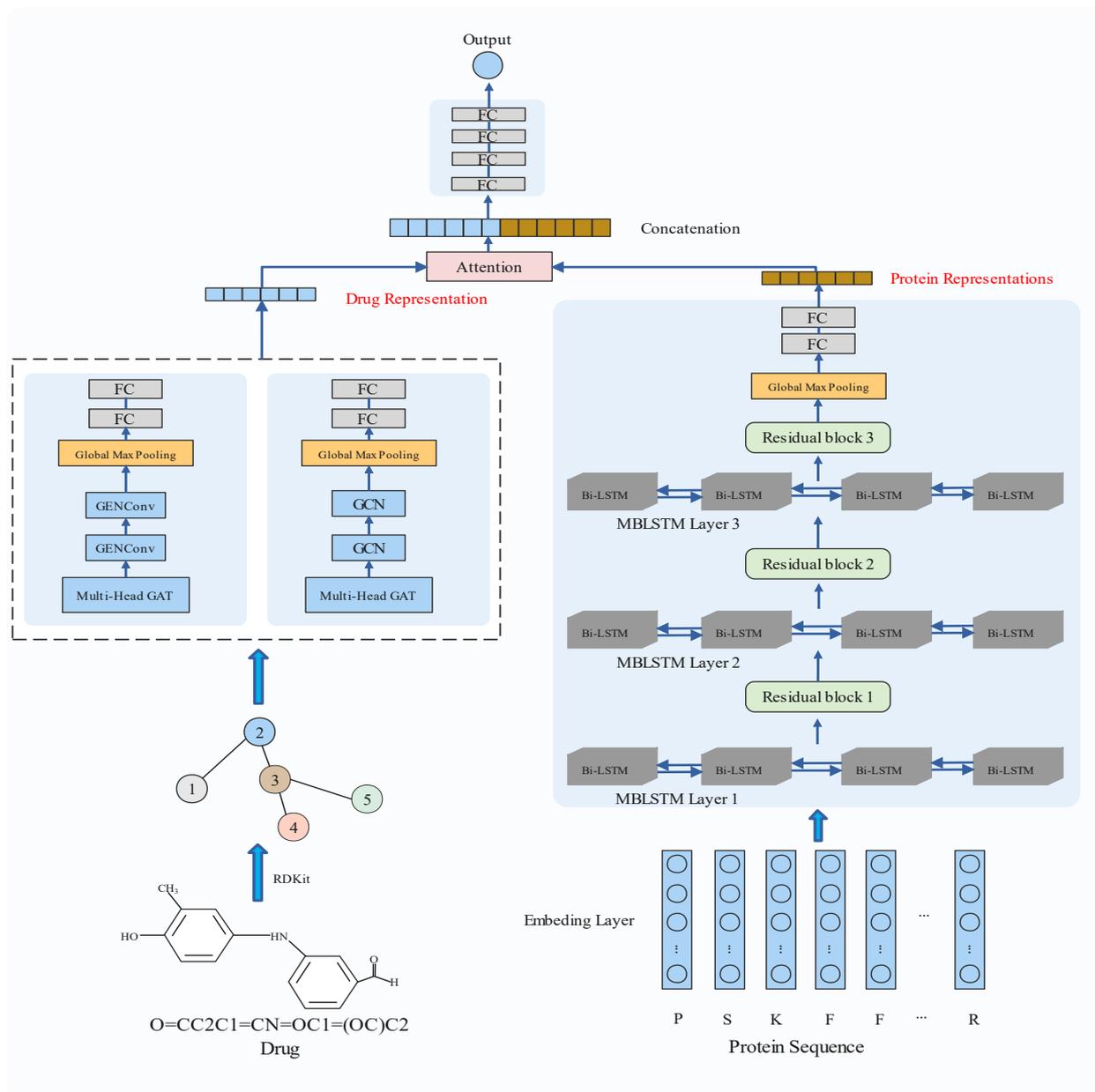


Fig. 2. Comprehensive illustration of the overall architecture of the drug-target binding affinity prediction model (HBDTA)

constant (K_d) [35], half-maximal inhibitory concentration (IC_{50}) [35], and inhibition constant (K_i) [36]. These metrics serve as critical indicators of the strength of interactions between drugs and targets. Specifically, the K_d value quantifies the interaction potency, with higher values indicating lower affinity. Similarly, the K_i value represents the inhibitor’s strength in inhibiting the target, while the IC_{50} value signifies the concentration required to inhibit half of the specified biological process. In our model, we use the log-transformed K_d as the model’s output, in line with established practices in the field of drug-target affinity prediction. This approach is widely accepted and enhances modeling effectiveness. For a summary of the key details and characteristics of the two datasets, please refer to Table III.

$$pK_d = -\log_{10}\left(\frac{K_d}{1e9}\right) \quad (17)$$

TABLE III
SUMMARY OF THE DATASETS

Dataset	Drugs	Targets	Binding Entities
Davis(pK_d)	68	442	30,056
KIBA	2,111	229	118,254
Metz	1,471	170	35,307
DTC	5,983	118	67,894

E. Experimental Settings

We standardized the protein sequence length to 1000 residues and employed MBLSTM with a hidden vector dimension of 128. To mitigate potential overfitting, we introduced dropout regularization with a probability rate (p) of 0.2. In terms of optimization, we utilized the ADAM optimizer [37] with a learning rate of 0.0005 for the Davis and Metz datasets and 0.001 for the KIBA and DTC datasets. Throughout the training process, all models underwent 1000 epochs of iteration. The batch size was set to 128 for the

Davis and Metz datasets and 256 for the KIBA and DTC datasets. The devices that are used for the experiments are an Intel(R) Xeon(R) Platinum 8260 CPU @ 2.30 GHz and an NVIDIA GeForce RTX 3090 GPU. We also followed the experimental methodology of NGGNDTA [38]. For the “cold drug” setting, the Davis dataset was partitioned into 54, 7, and 7 non-overlapping drugs for the training, validation, and test sets, respectively. Similarly, the KIBA dataset was divided into 1654, 207, and 207 non-overlapping drugs. Under the “cold target” setting, the Davis dataset was split into 354, 44, and 44 non-overlapping proteins for the training, validation, and test sets, while the KIBA dataset was separated into 182, 23, and 23 non-overlapping proteins. The model’s performance evaluation included several key metrics, such as Mean Square Error (MSE) [39], r_m^2 Index [6], and CI [6]. These metrics provide comprehensive insights into the model’s predictive capabilities. For specific details on the hyperparameter settings used, please refer to Table IV.

TABLE IV
HYPERPARAMETER SETTINGS

Hyperparameter	Settings
Epoch	1000
Dropout	0.2
Optimizer	Adam
Residual blocks in model	3
Fully connected layers after GNN	2
Hidden units in final Fully connected layers	1024,512
Bi-LSTM layers in model	3
Max length of protein sequences	1000

F. Evaluation Metrics

The consideration of drug-target binding affinity prediction as a regression problem is a fundamental aspect of our proposed approach. In this context, we employ a set of three evaluation criteria to assess the performance of our model. Among these, Mean Square Error (MSE) stands out as a widely accepted and commonly used evaluation metric for regression models. Importantly, MSE also serves as the loss function within our methodology. The formulation of MSE is articulated as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2 \quad (18)$$

where N is the number of samples, y_i is the actual target binding affinity value for the i -th sample, and p_i is the predicted binding affinity value for the i -th sample. The MSE computation serves as a means to quantify the average of the squared discrepancies between the actual and predicted binding affinity values. It is important to underscore that a lower MSE value is indicative of superior model performance, signifying a more accurate prediction alignment with the ground truth values.

The second evaluation approach we employ is the Concordance Index (CI). This metric gauges the probability that, for any pair of randomly chosen samples, the sample boasting the higher predicted binding affinity score simultaneously holds the higher actual binding affinity score. A CI value of 1 indicates perfect agreement between the ranking of predicted and actual binding affinity scores. The CI can be calculated utilizing the following equation:

$$CI = \frac{1}{Z} \sum_{\delta_x > \delta_y} h(b_x - b_y) \quad (19)$$

where b_x represents the predicted binding affinity concerning the actual higher binding affinity δ_x , while b_y pertains to the predicted binding affinity about the actual lower binding affinity δ_y . To normalize the value to the interval $[0, 1]$, a normalizing constant labeled as Z is employed. The equation is characterized by a step function termed $h(x)$.

$$h(x) = \begin{cases} 0, & \text{if } x < 0 \\ 0.5, & \text{if } x = 0 \\ 1, & \text{if } x > 0 \end{cases} \quad (20)$$

The r_m^2 index, a metric utilized in DeepDTA and also adopted in our proposed approach for assessing model performance, gauges the correlation existing between the anticipated binding affinity values and the factual binding affinity values. The calculation of r_m^2 is articulated in the ensuing equation:

$$r_m^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (21)$$

where n signifies the count of samples, y_i corresponds to the true binding affinity value for the i -th sample, \hat{y}_i represents the predicted binding affinity value for the i -th sample, and \bar{y} denotes the mean of the actual binding affinity values. A r_m^2 value of 1 serves as an indicator of perfect alignment between predicted and actual values. Conversely, a r_m^2 value of 0 implies that model predictions do not surpass the efficacy of using the mean of the actual values as a prediction. A negative r_m^2 value implies the model’s inferiority compared to employing the mean value as a prediction.

IV. ANALYSIS OF EXPERMENTS AND RESULTS

A. Comparison Experiment

We evaluated two distinct algorithms, denoted as HBDTA(GAT_GCN) and HBDTA(GAT_GENConv), using three different evaluation metrics: MSE, CI, and r_m^2 . In the Davis dataset, our HBDTA(GAT_GCN) outperformed all the models we compared it to, achieving the best overall performance and superior results in the testing phase. When compared to the top-performing model, DeepGLSTM, HBDTA(GAT_GCN) exhibited a notable reduction of 0.004 in MSE, an improvement of 0.004 in CI, and a substantial enhancement of 0.031 in r_m^2 . In the KIBA dataset, the HBDTA(GAT_GCN) model showed a reduction of 0.005 in MSE, an advancement of 0.003 in CI, and a notable increase of 0.019 in r_m^2 . These empirical outcomes highlight the effectiveness of combining GAT and GCN layers within our model.

In both the Metz and DTC datasets, our model consistently yielded superior results across all performance metrics, providing strong validation for the efficacy of combining deep neural networks and graph neural networks in HBDTA for improving DTA prediction accuracy. While HBDTA(GAT_GENConv) did not surpass the performance achieved by HBDTA(GAT_GCN), it still exhibited superior performance compared to a significant portion of the models we compared it to. This consistent performance underscores

TABLE V
COMPARISON OF MSE, CI AND r_m^2 SCORES ON DATASETS

Method	Davis			KIBA			Metz			DTC		
	MSE ↓	CI ↑	r_m^2 ↑	MSE ↓	CI ↑	r_m^2 ↑	MSE ↓	CI ↑	r_m^2 ↑	MSE ↓	CI ↑	r_m^2 ↑
SimBoost	0.282	0.873	0.644	0.222	0.836	0.629	0.376	0.683	0.557	0.304	0.771	0.727
KronRLS	0.379	0.869	0.407	0.411	0.782	0.342	0.459	0.675	0.411	0.482	0.694	0.412
DeepDTA	0.261	0.878	0.630	0.194	0.863	0.673	0.353	0.703	0.537	0.274	0.791	0.698
WideDTA	0.262	0.886	0.633	0.179	0.875	0.675	0.359	0.721	0.545	0.270	0.807	0.702
DeepCPI	0.293	0.867	0.607	0.211	0.852	0.657	0.391	0.696	0.514	0.307	0.774	0.673
Attention-DTA	0.245	0.887	0.657	0.162	0.882	0.735	0.326	0.724	0.573	0.262	0.825	0.728
DeepGS	0.252	0.880	0.686	0.193	0.860	0.684	0.337	0.714	0.548	0.285	0.792	0.717
GANsDTA	0.276	0.881	0.653	0.224	0.866	0.775	0.347	0.718	0.658	0.283	0.771	0.715
GraphDTA(GCN)	0.254	0.880	0.663	0.139	0.889	0.691	0.317	0.801	0.620	0.317	0.878	0.812
GraphDTA(GAT_GCN)	0.245	0.881	0.677	0.139	0.891	0.701	0.333	0.795	0.602	0.200	0.857	0.790
GraphDTA(GIN)	0.229	0.893	0.649	0.147	0.882	0.684	0.317	0.800	0.645	0.176	0.876	0.798
GraphDTA(GAT)	0.232	0.892	0.662	0.179	0.866	0.671	0.393	0.775	0.549	0.195	0.859	0.788
DeepNC(GEN)	0.233	0.887	0.653	0.133	0.897	0.695	0.385	0.770	0.538	0.187	0.886	0.813
DeepNC(HGC_GCN)	0.243	0.881	0.686	0.172	0.872	0.624	0.407	0.759	0.561	0.223	0.872	0.805
DeepGLSTM	0.232	0.895	0.680	0.133	0.897	0.792	0.294	0.810	0.640	0.149	0.895	0.841
HBDTA(GAT_GCN)	0.226	0.899	0.711	0.128	0.900	0.811	0.283	0.872	0.653	0.142	0.899	0.891
HBDTA(GAT_GENConv)	0.234	0.895	0.690	0.134	0.897	0.763	0.291	0.831	0.632	0.149	0.894	0.887

the versatility and adaptability inherent in our model, even when operating with a diverse combination of Graph Neural Network (GNN) algorithms. A comprehensive summary of these experimental findings is concisely provided in Table V, offering a succinct overview of the comparative performance of the models under various configurations.

Furthermore, to investigate the impact of the stacked deep network architecture on the quality of experimental results, we conducted tests on the number of Bi-LSTM layers employed within the HBDTA model, specifically within the Davis and KIBA datasets. The range of configurations included a single-layer Bi-LSTM, double-layer Bi-LSTM, three-layer Bi-LSTM (as proposed in our study), and four-layer Bi-LSTM. The outcomes of these tests are depicted in Fig. 3 and Fig. 4 for reference and further insights.

From Fig. 3 and Fig. 4, we observe that when MBLSTM is configured with three layers, the experiments yield the best results. However, when MBLSTM is extended to four layers, there is a decline in experimental outcomes. We attribute this phenomenon to the increase in network depth, which leads to diminishing or exploding gradients, rendering the weights unstable and making the network challenging to train, consequently impacting the results. Additionally, the propagation of gradients across multiple layers may result in information loss and confusion, hindering the ability of subsequent layers to learn effectively.

B. Deeper Evaluation

Randomly dividing the dataset into training, validation, and test sets carries the risk of inadvertently leaking information about drugs and proteins into the test set, potentially compromising the integrity of the evaluation process. To address this concern, we further employ three distinct partitioning methods to assess the performance of the DTA model, aiming to demonstrate the model's effectiveness: drug cold-start partitioning, protein cold-start partitioning, and drug-protein cold-start partitioning. Drug cold-start partitioning involves separating individual drugs into training, validation, and test sets. This partitioning ensures that drugs appearing in the training and validation sets do not overlap with those in

the test set (and vice versa), thereby preserving the integrity of the test set. A similar approach is applied to protein cold-start partitioning, which is based on different proteins.

TABLE VI
THE PERFORMANCE COMPARISON BETWEEN HBDTA AND OTHER MODELS ON THE DAVIS DATASET

Scenario	Method	MSE ↓	CI ↑	r_m^2 ↑
Cold drug	GraphDTA	0.920(0.030)	0.678(0.032)	0.160(0.019)
	DeepGLSTM	0.861(0.019)	0.708(0.026)	0.173(0.015)
	HBDTA	0.853(0.047)	0.724(0.014)	0.179(0.016)
Cold target	GraphDTA	0.510(0.075)	0.729(0.032)	0.154(0.019)
	DeepGLSTM	0.461(0.020)	0.810(0.023)	0.198(0.013)
	HBDTA	0.458(0.027)	0.829(0.019)	0.224(0.017)
All cold	GraphDTA	0.968(0.085)	0.579(0.017)	0.026(0.016)
	DeepGLSTM	0.915(0.090)	0.610(0.033)	0.057(0.041)
	HBDTA	0.907(0.096)	0.626(0.039)	0.063(0.048)

TABLE VII
THE PERFORMANCE COMPARISON BETWEEN HBDTA AND OTHER MODELS ON THE KIBA DATASET

Scenario	Method	MSE ↓	CI ↑	r_m^2 ↑
Cold drug	GraphDTA	0.471(0.045)	0.713(0.002)	0.342(0.009)
	DeepGLSTM	0.447(0.032)	0.724(0.006)	0.349(0.015)
	HBDTA	0.431(0.029)	0.736(0.004)	0.358(0.011)
Cold target	GraphDTA	0.469(0.089)	0.610(0.034)	0.368(0.057)
	DeepGLSTM	0.457(0.035)	0.616(0.024)	0.381(0.015)
	HBDTA	0.446(0.022)	0.633(0.031)	0.382(0.031)
All cold	GraphDTA	0.676(0.103)	0.601(0.030)	0.149(0.067)
	DeepGLSTM	0.634(0.081)	0.621(0.023)	0.164(0.039)
	HBDTA	0.623(0.076)	0.625(0.019)	0.173(0.021)

To facilitate a fair and consistent comparison, our dataset-splitting methodology remains uniform throughout the evaluation process. Consequently, we evaluate HBDTA (GAT_GCN) against GraphDTA and DeepGLSTM under the cold-start scenarios for drugs and proteins using the Davis and KIBA datasets as benchmarks. Our model consistently demonstrates superior performance across all three metrics, as evidenced by the results presented in Tables VI and VII. This superiority can be attributed to the inherent advantages of leveraging deep neural networks, which excel at effectively extracting valuable information from complex

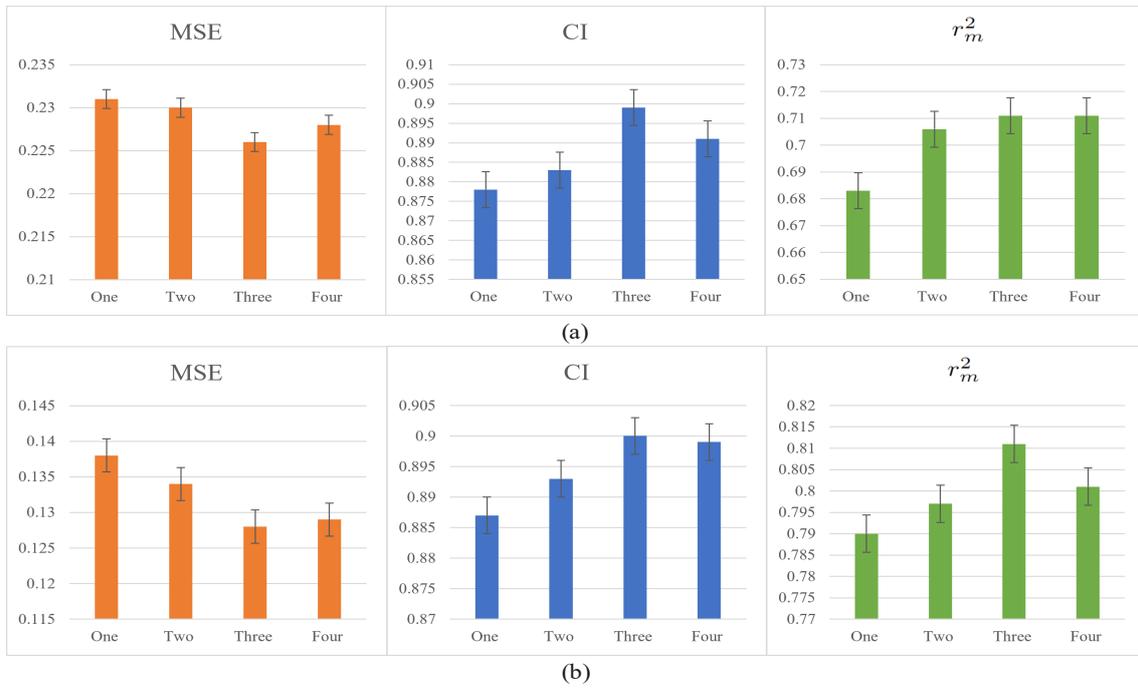


Fig. 3. The influence of MBLSTM layers in the HBDTA(GAT_GCN) model on the results from the Davis(a) and KIBA(b) datasets

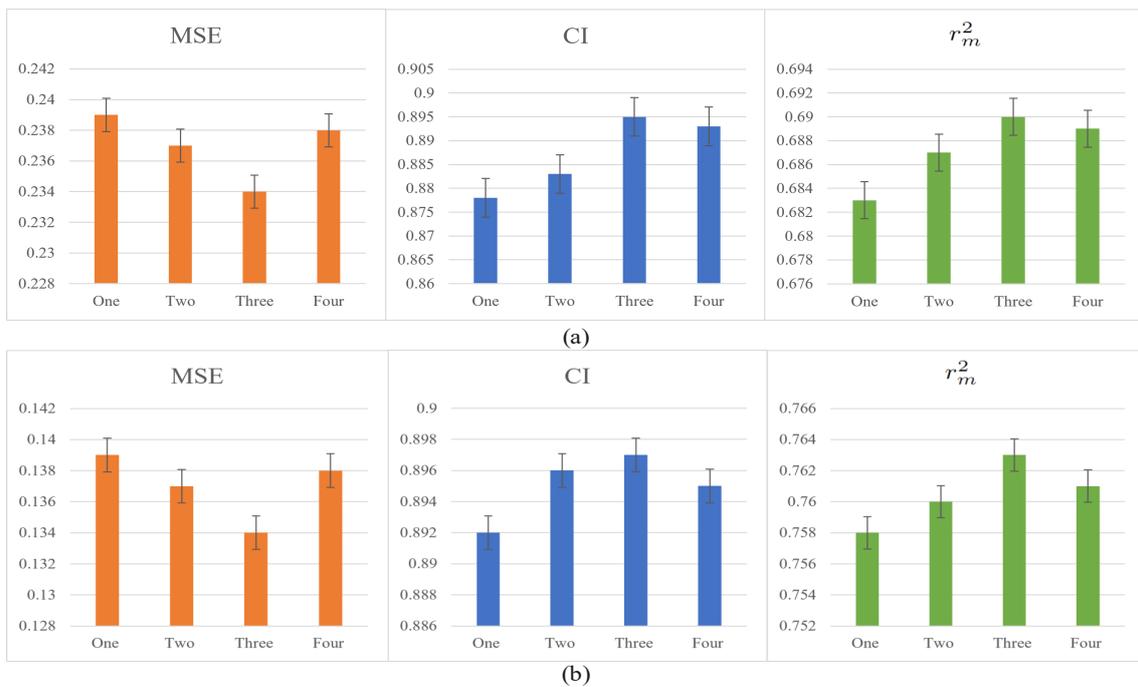


Fig. 4. The influence of MBLSTM layers in the HBDTA(GAT_GENConv) model on the results from the Davis(a) and KIBA(b) datasets

samples. Furthermore, Fig. 5 shows the MSE changes exhibited by the GraphDTA(GAT), GraphDTA(GAT_GCN), HBDTA(GAT_GCN), and HBDTA(GAT_GENConv) models on Davis, KIBA, Metz, and DTC datasets from 0 to 1000 epochs. Gray and yellow represent our proposed models with faster-decreasing trends.

C. Case Study

To visually evaluate the efficacy of our model, HBDTA was utilized to assess the potential effectiveness of various drugs against COVID-19. We identified five specific proteins (NSP14, ACE2, Spike, 3CLPro, Envelope) from DrugBank

and predicted five drugs with the highest binding affinities. A significant discovery emerged in this study: Oxytetracycline, a potent protein synthesis inhibitor, stood out as a prominent candidate, demonstrating the highest affinity for key proteins Nsp14, ACE2, Spike, 3CLPro, and Envelope. Oxytetracycline achieves this by disrupting the binding of aminoacyl-tRNA to the complex ribosomal RNA, thereby interfering with the protein synthesis process. In addition to Oxytetracycline, four other selected drug compounds exhibit significant antimicrobial and anti-inflammatory properties. However, it is crucial to emphasize that while these experimental findings may not directly translate into practical

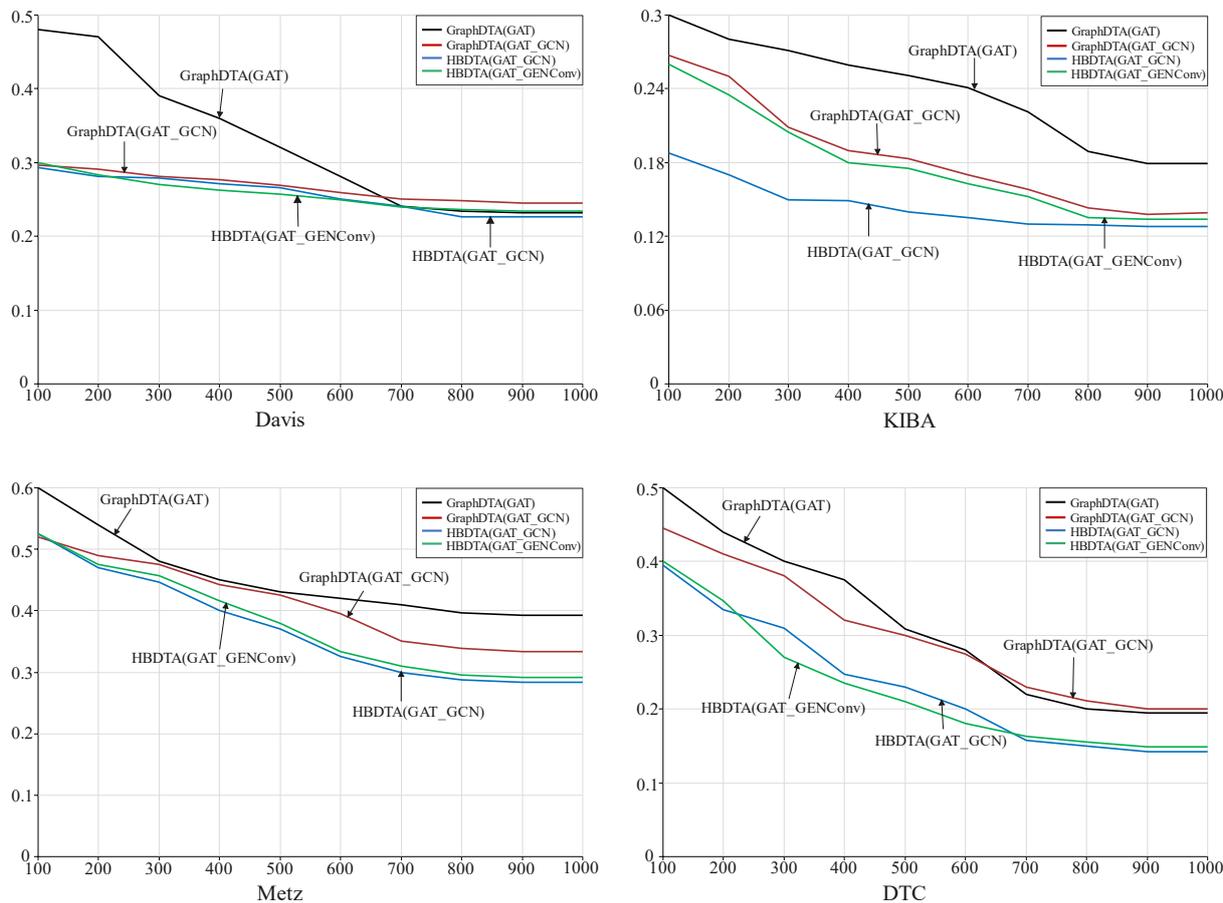


Fig. 5. Comparing MSE trends between HBDTA and GraphDTA on Davis, KIBA, Metz, and DTC datasets

TABLE VIII
PREDICTION OF COVID-19 TREATMENT DRUGS

	Nsp14	ACE2	Spike	3CLPro	Envelope
1	Oxytetracycline(13.93)	Oxytetracycline(13.81)	Oxytetracycline(13.70)	Oxytetracycline (13.77)	Oxytetracycline (13.84)
2	Doxycycline(13.68)	Clomocycline(13.48)	Rifabutin (13.51)	Doxycycline (13.43)	Doxycycline (13.67)
3	Clomocycline(13.21)	Rifabutin (12.94)	Clomocycline(13.34)	Clomocycline (13.41)	Clomocycline (13.55)
4	Demecloicycline(12.89)	Tetracycline (12.65)	Doxycycline(13.12)	Rifabutin (13.03)	Rifabutin (13.21)
5	Tetracycline(12.71)	Doxycycline(12.64)	Demecloicycline(12.86)	Tetracycline (12.51)	Demecloicycline (13.05)

efficacy against COVID-19, they provide valuable directional insights that warrant thorough consideration and further in-depth investigation. For detailed information regarding these selected compounds, please refer to Table VIII, which offers a comprehensive analysis arranged in descending order of binding affinity values.

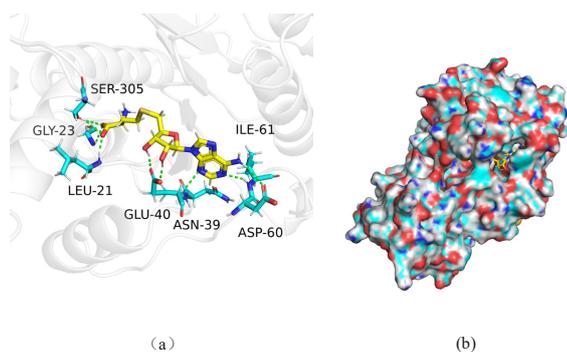


Fig. 6. Residue-ligand interaction between P05102 and Bosutinib predicted by blind docking

Furthermore, we employed a visualization approach to elucidate potential binding sites of drug-target interactions, aiming to enhance the interpretability of our model. In Fig. 6 (a), we utilized HBDTA to randomly select the P01502 protease from the UniProt database and predict potential interaction sites with Bosutinib. The results indicate that, according to the model predictions, proteins SER-305, Gly-23, Leu-21, GLU-40, ASN-39, ASP-60, ILE-61 within P01502 have potential binding sites with Bosutinib. This discovery provides robust support for our understanding of the interaction between the drug and its target. Given that small molecules bind within hydrophobic pockets, in Fig. 6 (b), we presented the protein structure in a shell format to demonstrate that the small molecule is indeed located within the pocket and spatially rationalizes its presence. Such visualization methods contribute to an intuitive display of the drug's positioning within the protein structure, providing visual cues for further exploration of the binding mechanism between the drug and the target. In conclusion, while the model predictions of these potential sites serve as a starting point for our research, it is crucial to carefully consider

the limitations and uncertainties of the model. Subsequent research should focus on validating the accuracy and biological significance of these potential sites through biophysical experiments and bioactivity assays.

Lastly, in Fig. 7, graphical representations are provided to illustrate the relationship between predicted values (p) and measured values (m) across both the Davis and KIBA datasets. This comparison serves as a crucial metric for evaluating the model's performance. A close alignment with the black line ($p = m$) indicates a robust model. Therefore, based on the distribution of predicted values and observed values in the dataset, we conclude that HBDTA demonstrates excellent predictive capabilities.

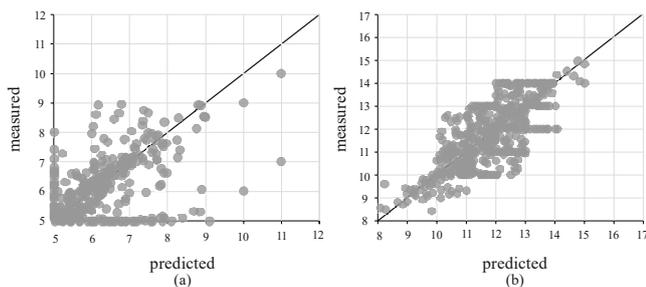


Fig. 7. Plots comparing HBDTA-predicted values with measured binding affinity values for the Davis dataset (a) and KIBA dataset (b)

V. CONCLUSIONS

The field of graph neural network models has gained traction due to their improved interpretability in graph processing. Our study presents a new GNN-based approach for predicting drug-target binding affinity, utilizing three unique GNN algorithms: multi-head graph attention network, graph convolutional network, and generalized aggregation graph network to construct drug molecular representations. Furthermore, we propose a deep architecture, MBLSTM, to extract features from protein sequences. The MBLSTM structure enhances processing and more efficiently allocates spatial parameters to effectively address the insufficiency of most existing shallow models for data mining. This highlights the efficacy of deeper network architectures in specific contexts. We validated four datasets to predict drugs with the greatest potential to treat COVID-19 and visually interpreted drug-target interactions. The experimental findings demonstrate the superior performance of HBDTA compared to alternative models in terms of both speed and the quality of generated predicted structures. Future research should explore the potential of HBDTA in diverse drug-related tasks such as drug repurposing or toxicity prediction.

REFERENCES

- Xing Chen, Chenggang Clarence Yan, Xiaotian Zhang, et al, "Drug-target interaction prediction: databases, web servers and computational models," *Briefings in Bioinformatics*, vol. 17, no. 4, pp696-712, 2016
- Asher Mullard, "New drugs cost US \$2.6 billion to develop," *Nature Reviews Drug Discovery*, vol. 13, no. 2014, pp877-877, 2014
- Ted T. Ashburn and Karl B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nature Reviews Drug discovery*, vol. 3, no. 8, pp673-683, 2004
- Tong He, Marten Heidemeyer, Fuqiang Ban, Artem Cherkasov, and Martin Ester, "SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines," *Journal of Cheminformatics*, vol. 9, no. 1, pp1-14, 2017
- Anna Cichonska, Balaguru Ravikumar, Elina Parri, Sanna Timonen, et al, "Computational-experimental approach to drug-target interaction mapping: a case study on kinase inhibitors," *PLoS Computational Biology*, vol. 13, no. 8, ppe1005678, 2017
- Hakime Öztürk, Arzucan Özgür, and Elif Ozkirimli, "DeepDTA: deep drug-target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, ppi821-i829, 2018
- Xiaoyan Xu, "Research on a Small Target Object Detection Algorithm for Electric Transmission Lines Based on Convolutional Neural Network," *IAENG International Journal of Computer Science*, vol. 50, no. 2, pp375-380, 2023
- Hakime Öztürk, Elif Ozkirimli, and Arzucan Özgür, "WideDTA: prediction of drug-target binding affinity," *arXiv preprint: 1902.04166*, 2019
- David Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology, and encoding rules," *Journal of Chemical Information and Computer Sciences*, vol. 28, no. 1, pp31-36, 1988
- Masashi Tsubaki, Kentaro Tomii, and Jun Sese, "Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences," *Bioinformatics*, vol. 35, no. 2, pp309-318, 2019
- Bonggun Shin, Sungsoo Park, Keunsoo Kang, and Joyce C. Ho, "Self-attention based molecule representation for predicting drug-target interaction," *Machine Learning for Healthcare Conference*, PMLR, pp230-248, 2019
- Lingling Zhao, Junjie Wang, Long Pang, Yang Liu, and Jun Zhang, "GANsDTA: Predicting drug-target binding affinity using GANs," *Frontiers in Genetics*, vol. 10, pp1243-1243, 2020
- Qichang Zhao, Guihua Duan, Mengyun Yang, Zhongjian Cheng, Yao-hang Li, and Jianxin Wang, "AttentionDTA: prediction of drug-target binding affinity using attention model," *In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, pp64-69, 2019
- Thin Nguyen, Hang Le, Thomas P. Quinn, Tri Nguyen, Thuc Duy Le, and Svetha Venkatesh, "GraphDTA: predicting drug-target binding affinity with graph neural networks," *Bioinformatics*, vol. 37, no. 8, pp1140-1147, 2021
- Ruoxi Sun, Hanjun Dai, and Adams Wei Yu, "Does gnn pretraining help molecular representation?" *Advances in Neural Information Processing Systems*, vol. 35, pp12096-12109, 2022
- Wenlong Ma, Zhixu Qiu, Jie Song, Qian Cheng, and Chuang Ma, "DeepGS: Predicting phenotypes from genotypes using Deep Learning," *arXiv preprint: pp241414*, 2017
- Shrimon Mukherjee, Madhusudan Ghosh, and Partha Basuchowdhuri, "DeepGLSTM: Deep Graph Convolutional Network and LSTM based approach for predicting drug-target binding affinity," *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, Society for Industrial and Applied Mathematics, pp729-737, 2022
- Farah Shahid, Aneela Zameer, and Muhammad Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons & Fractals*, vol. 140, pp110212, 2020
- Huu Ngoc Tran Tran, J. Joshua Thomas, and Nurul Hashimah Ahamed Hassain Malim, "DeepNC: a framework for drug-target interaction prediction with graph neural networks," *PeerJ*, vol. 10, ppe13163, 2022
- Guohao Li, Chenxin Xiong, Ali Thabet, and Bernard Ghanem, "Deepergcn: All you need to train deeper gcns," *arXiv preprint: 2006.07739*, 2020
- Song Bai, Feihu Zhang, and Philip H.S. Torr, "Hypergraph convolution and hypergraph attention," *Pattern Recognition*, vol. 110, pp107637, 2021
- Monica Bianchini and Franco Scarselli, "On the complexity of neural network classifiers: A comparison between shallow and deep architectures," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 8, pp1553-1565, 2014
- J. Adisa, Pius Samuel Ojo, Adewale Owolawi, Agnieta Pretorius, et al, "Application of an Improved Optimization Using Learning Strategies and Long Short Term-Memory for Bankruptcy Prediction," *IAENG International Journal of Computer Science*, vol. 50, no. 2, pp512-524, 2023
- Zhaoyang Niu, Guoqiang Zhong, and Hui Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp48-62, 2021
- Mindy I. Davis, Jeremy P. Hunt, Sanna Herrgard, et al, "Comprehensive analysis of kinase inhibitor selectivity," *Nature Biotechnology*, vol. 29, no. 11, pp1046-1051, 2011
- Jing Tang, Agnieszka Szwarzda, Sushil Shakyawar, Tao Xu, et al, "Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis," *Journal of Chemical Information and Modeling*, vol. 54, no. 3, pp735-743, 2014
- James T. Metz, Eric F. Johnson, Niru B. Soni, et al, "Navigating the kinome," *Nature Chemical Biology*, vol. 7, no. 4, pp200-202, 2011

- [28] Jing Tang, Zia-ur-Rehman Tanoli, Balaguru Ravikumar, et al. "Drug target commons: a community effort to build a consensus knowledge base for drug-target interactions," *Cell Chemical Biology*, vol. 25, no. 2, pp224-229, 2018
- [29] Zhaorui Meng, Yanqi Xie, and Jinhua Sun, "Detecting Credit Card Fraud by Generative Adversarial Networks and Multi-head Attention Neural Networks," *IAENG International Journal of Computer Science*, vol. 50, no. 2, pp381-387, 2023
- [30] Dongxian Wu, Yisen Wang, Shu-Tao Xia, James Bailey, and Xingjun Ma, "Skip connections matter: On the transferability of adversarial examples generated with resnets," *arXiv preprint: 2002.05990*, 2020
- [31] A. Patrícia Bento, Anne Hersey, Eloy Félix, et al. "An open source chemical structure curation pipeline using RDKit," *Journal of Cheminformatics*, vol. 12, pp1-16, 2020
- [32] Yuan Jin, Jiarui Lu, Runhan Shi, and Yang Yang, "EmbedDTI: Enhancing the Molecular Representations via Sequence Embedding and Graph Convolutional Network for the Prediction of Drug-Target Interaction," *Biomolecules*, vol. 11, no. 12, pp1783, 2021
- [33] UniProt Consortium, "UniProt: the universal protein knowledgebase in 2021," *Nucleic Acids Research*, vol. 49, no. D1, ppD480-D489, 2021
- [34] David S. Wishart, Yannick D. Feunang, An C. Guo, et al. "DrugBank 5.0: a major update to the DrugBank database for 2018," *Nucleic Acids Research*, vol. 46, no. D1, ppD1074-D1082, 2018
- [35] Jacques Barbet and Sandrine Huclier-Markai, "Equilibrium, affinity, dissociation constants, IC50: Facts and fantasies," *Pharmaceutical Statistics*, vol. 18, no. 5, pp513-525, 2019
- [36] Yung-Chi Cheng and William H. Prusoff, "Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 percent inhibition (I50) of an enzymatic reaction," *Biochemical Pharmacology*, vol. 22, no. 23, pp3099-3108, 1973
- [37] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint: 1412.6980*, 2014
- [38] Haohuai He, Guanxing Chen, and Calvin Yu-Chian Chen, "NHGNN-DTA: a node-adaptive hybrid graph neural network for interpretable drug-target binding affinity prediction," *Bioinformatics*, vol. 39, no. 6, ppbtad355, 2023
- [39] David M. Allen, "Mean square error of prediction as a criterion for selecting variables," *Technometrics*, vol. 13, no. 3, pp469-475, 1971



Yongqing Wu received the PhD degree in computational mathematics from Lanzhou University, China, in 2011. He worked as a visiting scholar with the University of Texas at Arlington, Texas, USA, from 2019 to 2020. Currently, he is an Associate Professor in the the School of Software, Liaoning Technical University, Huludao, China. His current research interests include recommendation systems, machine learning algorithms, complex systems and complex networks.



Yao Jin M.S. candidate. His research interests include graph neural network, drug-target binding affinity prediction.



Peng Sun M.S. candidate. His research interests include graph neural network, recommendation systems.



Zhichen Ding M.S. candidate. His research interests include graph neural network, recommendation systems.