Received: January 17, 2025. Revised: February 12, 2025. 528

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Improving Drug-Target Affinity Prediction Using Dynamic Graph Attention Network with Multi-scale Features and Attention Mechanism

Muhammad Rizky Yusfian Yusuf[1]* **Isman Kurniawan[1]**

[1]*School of Computing, Telkom University, Bandung, Indonesia*
* Corresponding author's Email: rzkysfian@student.telkomuniversity.ac.id

Abstract: Drug-target affinity (DTA) prediction is essential in drug discovery because traditional methods are time-consuming and expensive. Yet, recent computational approaches often struggle with limitations in representing the structural and sequential complexities of drugs and proteins, resulting in inferior prediction performance. Therefore, this study proposes a method to enhance DTA prediction accuracy using Dynamic Graph Attention Networks (GATv2) and Bidirectional Long Short-Term Memory (BiLSTM). The model incorporates multi-scale features, which include drug motif graphs, and a three-way multi-head attention mechanism to capture complex interactions between drug and protein representations. Evaluated on Davis and KIBA datasets, the proposed model outperformed baseline models (e.g., GCN, GAT, 1DCNN, LSTM) and benchmark methods (e.g., GraphDTA, MSGNN-DTA, and DGDTA) across three evaluation metrics, achieving MSE of 0.3209 and 0.1864, CI of 0.8646 and 0.8616, and r2m of 0.5046 and 0.6672, respectively. This approach addresses limitations in static attention mechanisms, lack of multi-scale representation, and simplified interaction modeling in existing methods, offering a more robust method for DTA prediction.

Keywords: Drug target affinity, Drug graph, Protein sequences, Regression, Dynamic graph attention network, Attention mechanism.

1. Introduction

The process of binding between a drug and a target that modifies the target's function or behavior is known as drug-target interaction (DTI) [1]. A drug is any chemical substance that, when ingested, alters the organism's chemical composition. The term "target" refers to any biological component of the organism that interacts with the drug, resulting in alterations in a chemical environment. The common target itself is a protein or nucleic acid, including enzymes, G-protein coupled receptors, nuclear receptors, and ion channels [1]. DTI is an important aspect of the drug development process, which may take up to 2.6 billion US dollars and at least 17 years to complete the process of one drug from the original hypothesis to official marketing [2]. The process is long in time, complex, costly, and has a low chance of success. In addition, because of their unknown interactions, the majority of known chemical substances have yet to be utilized as drugs [1]. Therefore, in recent years, there has been a great deal of interest in the process to understand how drugs interact with their targets and how to predict drug-target interactions [2].

In general, direct laboratory experiments using techniques like high throughput screening (HTS) can be used for the DTI prediction process [2]. However, the laboratory process in the lab takes a lot of time and costs a lot of money. Therefore, a new in silico approach is needed to address these problems [1]. One approach that can be used as an alternative to predict the DTI is the computational method. Also, the availability of large volumes of data on drug compounds with hundreds of potential targets makes computational methods available in DTI prediction [2]. Computational methods for drug-target interaction (DTI) are divided into three approaches, namely ligand-based, docking simulation-based, and chemogenomic-based [3]. Ligand-based approaches and docking simulations are conventional approaches. Ligand-based approaches are developed with the

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

assumption that similar molecules share similar properties, so drug molecules should be able to bind to proteins that have similar molecules. However, this approach has drawbacks because the interaction predictions made are only limited to known drug molecules and proteins [4]. The second approach is based on docking simulation by utilizing the 3D structure of the protein. However, this approach has disadvantages when the 3D structure of the protein is unknown, and also the simulation is quite complex and requires large computational power [4]. The chemogenomic approach offers better solutions to answer the shortcomings of the previous two approaches. This approach uses drug chemical space information and protein genomic space and unifies them in the same subspace to infer possible interactions [1]. Several methods are often used in chemogenomic approaches, such as statistical methods [5, 6], machine learning methods [7, 8], and deep learning methods [9, 10].

In recent years, the use of deep learning has been frequently used in DTI prediction. This is related to the deep learning architecture that can identify hidden or complex patterns or data representations. As such, creating effective deep-learning models is essential for discovering hit compounds, which are identified as potential drugs for therapeutic use [11]. Generally, DTI prediction is commonly classified as a classification task [11-13]. This means the task only predicts whether a drug interacts with a specific target or not. However, one important piece of information is missing from the prediction results, which is a binding affinity value that represents the strength of a drug's interaction with a target pair that is measured by a continuous number [14]. Thus, the prediction of drug-target affinity (DTA) offers the advantage of predicting the strength value of the interaction between a drug and its target. Hence, this approach reduces the extensive search space for compounds in drug discovery research [14].

Several previous studies have performed DTA prediction by implementing deep-learning models with various representations of drugs and targets. Some studies represent drugs and targets as string sequences using the Simplified Molecular Input Line Entry System (SMILES) for drug compounds and protein amino acid sequences for targets, employing models such as 1DCNN and RNN-based models to do the feature learning [15, 16].

Despite that, the current state of representing drugs as sequence strings is unsuitable. This is due to the possibility of losing structural details when employing string representation, which may have an impact on the binding affinity prediction [17]. Therefore, other studies used graphs as a drug representation by converting SMILES strings into graph-based representation using Graph Neural Network (GNN) models to do the representation learning [17, 18]. Although several studies have performed DTA prediction with drug representation as a graph, there are still some issues that need to be addressed. For example, many studies have used Graph Attention Networks (GATs) for drug representation learning because of their attention mechanism. However, GAT models have a static attention mechanism that applies the same attention weights regardless of the query node's context. This limitation results in the model treating all node relationships equally, without adapting to the specific context of each query node. To overcome this limitation, GATv2 was introduced as a dynamic attention variant, providing more expressive representations by adapting attention weights to the query node's context [19].

Also, even though current graph-based methods have already implemented GATv2 models for DTA prediction, these studies commonly represent drugs as a single graph structure. They neglect multi-scale structural information within drugs, such as structural features of individual amino acids, motifs, and various scales of structural features, including atoms and molecular fragments. These interactions and correlations between different structural levels play a pivotal role in drug–target protein interactions [20]. Furthermore, substructures such as motifs carry special meanings in drug molecules, such as NO2 and carbon ring groups that are prone to mutagenesis [21]. Thus, motifs deserve more attention as additional drug representations. By integrating drug multi-scale features using both the overall molecular graph and specific motifs, DTA models can achieve a more comprehensive representation of the drug, leading to improved prediction accuracy.

Besides, protein sequence representations are long, with each character describing an amino acid. Conventional models are unable to process contextual relationships within sequences, missing critical relationships between preceding and following amino acids. Bi-directional LSTM (BiLSTM) networks provide an alternative approach by considering both past and future contexts, enabling a more comprehensive understanding of the protein's structural and functional properties [22]. Moreover, recent interaction modeling often relies on straightforward concatenation of drug and protein representations. This simplified approach overlooks the complex relationships between drug graphs and protein sequences, potentially ignoring the essential interaction information that might influence binding affinity. Moreover, many current methods only focus

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

on capturing interactions between two representations, drug graphs, and protein sequences without accounting for other potential representations (e.g., drug motifs). This limitation hinders the model's ability to fully capture the complex interactions between various drug and target features, possibly affecting binding affinity value. To address these issues, alternative approaches can leverage advanced attention mechanisms to incorporate additional representations and employ interaction models capable of processing multiple inputs in a more context-aware manner.

To address the limitations mentioned above, this study proposes an enhanced GATv2 model to obtain more informative features from the drug graph node and combine it with drug multi-scale features to dynamically adapt and selectively fuse features from different drug feature representation scales. For protein sequences, BiLSTM is utilized to capture long-term dependencies and contextual associations, leveraging its ability for sequence data. We also incorporate a new attention mechanism inspired by the AttentionDTA study [23] called a three-way multi-head attention mechanism, which enables each representation to focus on critical regions of the other, effectively highlighting important cross-interactions. This comprehensive framework aims to overcome existing limitations and improve the accuracy and interpretability of DTA predictions.

The main contributions of this study are:

- Employed GATv2 as a dynamic attention mechanism to overcome the limitations of static attention in traditional GATs, enabling more context-aware graph representations.
- Introduced a multi-scale representation for drug compounds, incorporating both molecular graph and motifs graph structures to enhance the granularity and context of drug representations.
- Implemented a three-way multi-head attention mechanism to capture complex interactions between drug and target features, enabling the model to consider multiple aspects of inputs and improve prediction accuracy.
- Utilized Bi-directional LSTM (BiLSTM) to effectively capture long-term dependencies and contextual relationships within protein amino acid sequences, providing a more robust understanding of protein properties.

The rest of this research paper is organized as follows. Section 2 discussed several literature studies related to DTA prediction. Section 3 describes the research methodology, implementation, and experiment design. Section 4 presents the implementation results and discussion. Finally, the conclusion and future works are presented in section 5.

2. Related works

Several previous studies have performed Drug-target affinity (DTA) prediction using many different approaches. Many of these studies represent drug compounds and target proteins as sequences. In 2018, Öztürk et al. [14] proposed DeepDTA, a model that utilizes a deep learning architecture, specifically CNN, to predict binding affinity. This model learns a 1D representation based on SMILES strings for drugs and protein sequences for targets, which are then combined and forwarded to the fully connected (FC) layer for predictions. DeepDTA bypassed the need for engineered features or structural data by directly leveraging raw sequence information and effectively outperformed traditional methods, establishing its utility for binding affinity prediction. However, CNN models struggle with capturing the ordered relationships within protein sequences. In the following year in 2019, Öztürk et al. [24] further explored DTA prediction by introducing additional representations, such as protein domain and motif (PDM) sequences and ligands' maximum common substructure (LMCS). The WideDTA model retained the same CNN architecture for data representation learning and the FC layer for predictions but differed from DeepDTA in its use of letter-based rather than character-based models. Despite its improved architecture, WideDTA did not significantly outperform prior models, and in some cases, the addition of domain, motif, and LMCS features even decreased performance.

In 2022, Ghimire et al. introduced the CSatDTA [20], which is CNN with a self-attention mechanism to address the limitations of sequence-based DTA models. Unlike traditional CNNs, which struggle with long-range dependencies and distant atomic interactions, CSatDTA uses self-attention to capture these complex relationships effectively. CSatDTA demonstrates its robustness in DTA prediction. However, it still relies on sequence-based representations, which may miss important structural details of drug and protein molecules. In 2023, D'Souza et al. proposed the DeepPS model [26], which focuses on motif-rich protein subsequences and uses SMILES as input for drug compounds. By leveraging binding site residues instead of full-length sequences, DeepPS reduces computational complexity and improves interpretability. However, this selective approach risks missing important

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

interactions from residues outside the binding pocket, potentially affecting binding affinity predictions. The model also showed slightly lower performance on smaller datasets like Davis, possibly due to the limited diversity of binding site residues in the training data.

In recent years, several models have incorporated attention mechanisms to improve drug-target affinity (DTA) prediction by capturing interaction modeling between drug and protein features. Zhao et. Al. [27] introduced an attention-based architecture that uses 1D convolutional neural networks (1D-CNNs) to extract sequence-level features from drug SMILES and protein amino acid sequences. Its novelty lies in its ability to calculate attention scores between subsequences of drugs and proteins, focusing on their critical interactions for predicting binding affinity.

However, the attention mechanism relies on simplistic calculations, such as dot products or weighted sums, which may not fully capture the complex, non-linear relationships between drug and protein features, potentially limiting its ability to model intricate binding interactions.

Abbasi et. al. [28] introduced a model called DeepCDA, which incorporated a two-sided attention mechanism and domain adaptation, simultaneously analyzing the interactions between drug and protein substructures. By computing attention coefficients between drug and protein fragments, DeepCDA captured the binding strength of each fragment pair. This approach offered better insight into mutual interactions than models that only utilized single-sided attention. However, the model faces limitations in cases of negative transfer, where domain adaptation may degrade performance when transferring knowledge between significantly different domains. In 2023, Zhao et al. [23] improved their previous paper by introducing a two-side multi-head attention mechanism to enhance the capture of non-covalent interactions between drug atoms and protein amino acids. The two-side multi-head attention allowed the model to evaluate relationships from multiple perspectives, increasing its expressiveness and predictive accuracy. The model achieved improved performance. However, the model's reliance on sequence-based inputs and its lack of structural information may limit its ability to fully capture complex spatial interactions, which are often critical for accurate binding affinity prediction.

Many early studies represented drug compounds and target proteins as sequences for DTA prediction. However, this approach often overlooks crucial spatial and topological information about drugs. To address these limitations, several studies have represented drugs as graphs. One notable study is GraphDTA by Nguyen et al. [17], which pioneered the use of graph neural networks (GNNs) for DTA prediction. GraphDTA represents drugs as undirected molecular graphs and proteins as sequence embeddings. Experimental results showed that graph-based drug representations significantly improved predictive performance compared to sequence-based models. However, the model's inability to fully leverage multi-scale structural information and limited protein representations left room for improvement.

Liang et al. introduced GLSTM-DTA [29], which combined GNNs for drug compound features and LSTMs for protein sequences. This hybrid approach successfully captured long-term dependencies in protein sequences and improved DTA prediction accuracy. Despite its success, the study highlighted the need for better feature fusion mechanisms between drugs and proteins. Despite its success, the study highlighted the need for better feature fusion mechanisms between drugs and proteins. Another study was MSGNN-DTA proposed by Wang et al. [21]. This model integrated multi-scale topological feature fusion with gated skip connections in GNNs to enhance drug-target representation by including drug molecules, motifs, and weighted protein structures. While MSGNN-DTA achieved state-of-the-art results on benchmark datasets, it still struggled with high computational complexity because of the weighted protein graph.

Xia et al. [30] proposed a novel integration of GCN and Word2vec embeddings for DTA prediction. The model represented drugs as molecular graphs and proteins as word embeddings derived from their amino acid sequences. Although this approach improved local feature extraction, the model faced challenges in accurately modeling non-local interactions between molecular substructures. Chen et al. [31] introduced SGNet, which fused drug graph representations with conjoint triad protein encodings. This method enhanced feature extraction from both graphs and sequences, achieving competitive results on various datasets. However, SGNet struggled with integrating high-resolution structural data due to the unavailability of 3D structures for many proteins.

The DGDTA model by Zhai et al. [32] implemented a dynamic graph attention mechanism (GATv2), allowing the model to assign context-aware weights to molecular substructures. This innovation significantly improved interpretability and accuracy in DTA prediction. However, the model's reliance on a single drug graph structure limited its application on drug feature complexity. Banerjee et al. [33] introduced DeepGLSTM, which utilized a GCN model to extract molecular features

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

from drugs and a bidirectional LSTM (Bi-LSTM) to process protein sequences. This architecture demonstrated superior performance by leveraging power graph representations to capture topological information. Despite its effectiveness, the GCN model lacked expressiveness when capturing information from graph nodes and was limited in its ability to handle complex, multi-scale molecular interactions.

Based on our literature studies, significant progress has been made in DTA prediction using various computational approaches. However, the state-of-the-art model lacked expressiveness when handling drug structural information, protein contextual information, integrating drug multi-scale structural features, and faced challenges in effectively fusing drug and protein representations. Therefore, this research aims to address these limitations by proposing a comprehensive framework by leveraging GATv2, BilSTM, drug multi-scale feature fusion, and a three-way multi-head attention mechanism, our approach seeks to improve the accuracy and interpretability of DTA predictions.

3. Methodology

This research aims to develop a robust deep-learning model for Drug-Target Affinity (DTA) prediction by utilizing advanced computational methods. The study follows a structured workflow, as illustrated in Fig. 1, including dataset selection, data preprocessing to generate appropriate representations for drugs and proteins, model development, and model performance evaluation using established metrics. The Davis and KIBA datasets were used for model training and validation. Preprocessing techniques were implemented to extract relevant features from drugs and targets, such as molecular graphs, motif graphs, and protein sequences. The proposed model combines sequence-based and graph-based methods with an attention mechanism to improve predictive accuracy. Finally, several evaluation metrics were used to assess the model's ability to predict binding affinities, ensuring its reliability and robustness. In addition, the notations used in this study are as follows:

K_d	Dissociation constant that describes the binding affinity value
pK_d	Logarithmic scale of K_d
e_{ij}	Attention coefficient
a^T	A learnable vector used for scoring attention
W	A shared learnable weight matrix used to transform node feature in new feature space
h_i and h_j	Feature vectors for node i and j, respectively
h_i'	Updated feature representation for node i
σ	Non-linear activation function (ReLU)
a_k	The attention scores for each head k
W_k^x	Trainable weight matrices for molecule/motif/protein specific to each attention head, respectively
E_x	Embeddings of molecule/motif/protein
H_k	Weighted embedding for each attention k
H	Final combined embedding computed by concatenating the output of all attention heads
y	Affinity value final prediction
MSE	Mean squared error
CI	Concordance index
r_m^2	Regression toward the mean

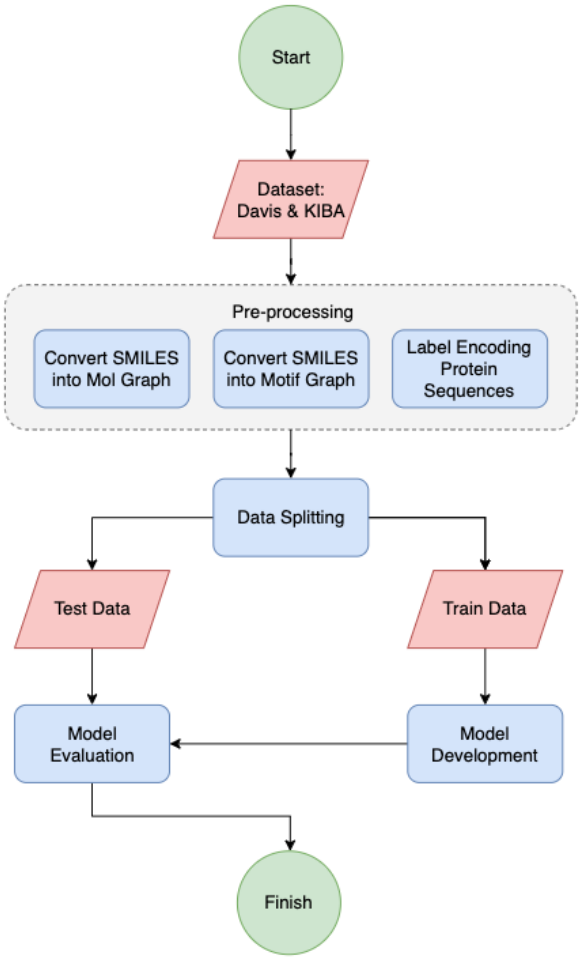


Figure. 1 Design Process Workflow

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

Table 1. Overview of Davis & KIBA Datasets

Description	Davis	KIBA
Drug	68	2,068
Protein	379	229
Total interaction	30,056	118,254
Train set (80%)	25,046	94,603
Test set (20%)	5,010	23,651

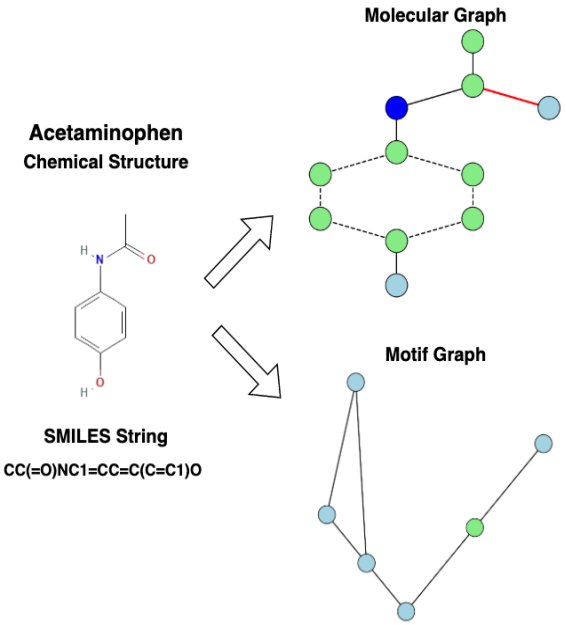


Figure. 2 Drug graphs construction illustration

3.1 Dataset

This study utilized the Davis and KIBA datasets to predict DTA as a regression task [34, 35]. The Davis and KIBA datasets are publicly available through previous study repositories or the Therapeutics Data Commons (TDC) website. These datasets are the most widely used in the DTA prediction field due to their comprehensive and diverse drug-protein interaction data [36]. The Davis dataset's affinity value measurement is represented in dissociation constant (K_d) values. To better illustrate the relationship between K_d and binding affinity values, the K_d are converted to logarithmic form as pK_d, as formulated in Eq. (1). Higher pK_d values indicate stronger binding affinity, with values ranging from 5.0 to 10.8.

$$pK_d = -log_{10}\left(\frac{K_d}{10^9}\right) \quad (1)$$

For the KIBA dataset, the affinity value measurement derived from an approach called KIBA, where inhibitor kinase bioactivity from various sources such as K_i, K_d, and IC_{50} is combined to obtain a value called the KIBA score. KIBA score values range from 0.0 to 17.2, with higher scores indicating high binding affinity values. A summary of the drug, target, and number of interactions is presented in Table 1.

3.2 Pre-processing

Before we developed the model, we performed several preprocessing steps. First, we checked for missing values and duplicate data and found there were none. Second, for the Davis dataset, we transformed the affinity value (K_d) into logarithmic form (pK_d) as mentioned earlier using Eq. (1). Next, we constructed the drug graphs for both datasets. The Simplified Molecular Input Line Entry System (SMILES) strings are commonly used to represent the three-dimensional structure of drug molecules as one-dimensional sequences. The SMILES string provides information about drug characteristics, such as the number of atomic weights or valence electrons [17].

Following this, each drug SMILES was transformed into a molecule where the nodes denoted the set of atoms in the drug and the edges represented the chemical bonds connecting the atoms [2]. To better represent the node feature in graphs, we accommodate a set of atomic features from DeepChem [37]. Here, each node in the molecular graph represents the chemical properties of its corresponding atom using a 78-dimensional feature vector, with each dimension corresponding to a particular chemical attribute. A detailed description of thesTo improve the representation of drug structure information, we incorporate a drug multi-scale feature representation by creating a motif-level graph alongside the molecular graph. Motifs in drugs, like the benzene ring, are closely linked to molecular properties. For example, the benzene ring maintains its significance as a whole structure but loses its meaning when its bonds are viewed in isolation.

Table 2. Molecular graph atom features

Feature	Dimension
Atomic features	44
Degree of atom	11
Total number of connected hydrogens atoms (implicit and explicit)	11
Implicit valence of atoms	11
Whether the atom is aromatic or not	1

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

Table 3. Motif graph atom features

Feature	Dimension
Atomic symbols contained motif	44
Number of atoms in the motif	11
Number of edges connecting motif	11
Total number of hydrogens atoms connected by motif (implicit & explicit)	12
Implicit valence of motif	12
Whether the motif is simple a simple ring	1
Whether the motif is chemically bonded or not	1

Multiple layers of graph neural networks (GNNs) often fail to capture the full information within these cyclic structures, resulting in an incomplete feature extraction [21]. Fig. 2 illustrates this transformation from SMILES to a graph. The motif-level graph was constructed by considering cyclic structures, individual chemical bonds that are not part of any cyclic structures, and their corresponding atom pairs as fundamental building blocks [21]. These elements are represented as nodes in the motif graph. Nodes corresponding to cyclic structures represent groups of atoms and bonds connected in a cycle, while other nodes represent individual chemical bonds and their associated atom pairs. The edges in the motif graph denote the chemical bonds linking these nodes. This approach captures more comprehensive structural information, aiding in motif graph generation and enhancing model training. Like the molecular graph, the features of the motif graph nodes were encoded into a 92-dimensional vector based on the DeepChem [37], which can be seen in Table 3.

For target pre-processing, the protein's primary structure was represented by an amino acid sequence. Protein sequences are composed of 25 different amino acids, each represented by a specific ASCII character that encodes its properties. Adopting the method from the GraphDTA study, these sequences were encoded using label encoding, with a maximum length set to 1,000 characters. Sequences shorter than this limit were padded with zeros, while those exceeding 1,000 characters were truncated [17]. This approach ensures a consistent input dimension size for training convenience. Moreover, this length is suitable for most proteins, as their lengths range from 200 to 2000, with a median of 700 characters.

3.3 Model development

In this study, we developed a prediction model of drug target affinity by combining two models, which are Dynamic Graph Attention Networks (GATv2) [19] and Bidirectional Long Short Term-Memory (Bi-LSTM) [38]. To enhance the performance of these combined models, multi-scale features and a three-way multi-head attention mechanism are implemented [21, 23]. The overview of the proposed model architecture can be seen in Fig. 3. For drug feature learning, drug graphs are represented as $G = (V, E)$, where V or nodes corresponding to atoms encoded as 78-dimensional feature vectors for the molecular graph and 92-dimensional feature vectors for the motif graph, and V or edges represent the chemical bonds between the atoms. Both graphs are processed through three GATv2 layers to learn node embeddings, respectively. Each GATv2 layer updates the representation of node i as described in Eq. (3).

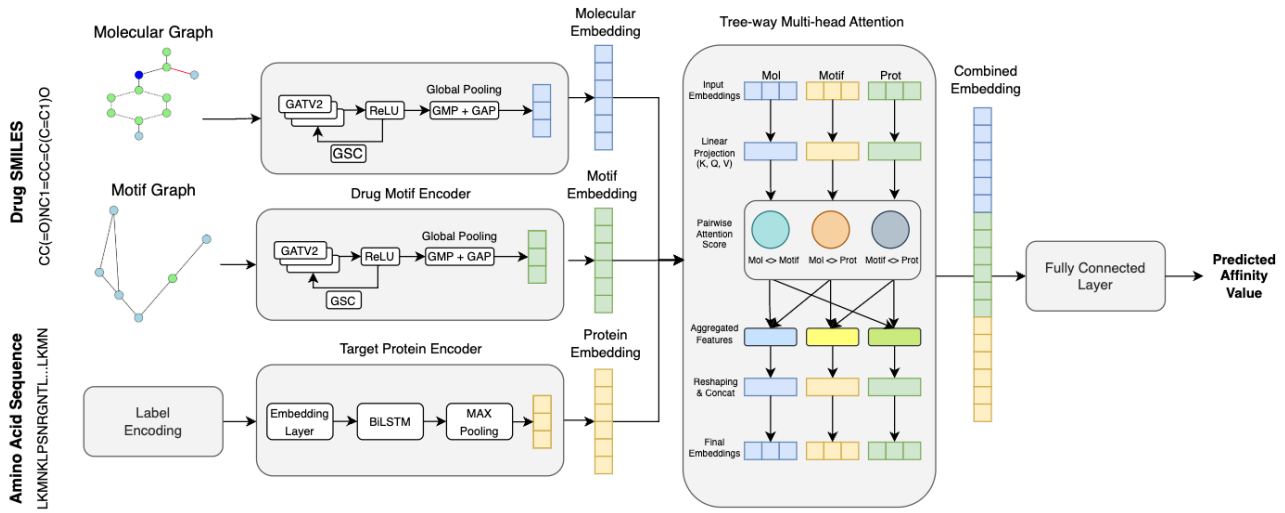


Figure. 3 Proposed model architecture overviews

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

$$e_{ij} = Softmax\left(a^T LeakyReLU\left(W\left[h_i \parallel h_j\right]\right)\right) \quad (2)$$

$$h_i' = \sigma\left(\sum_{j\in N(i)} e_{ij}.Wh_j\right) \quad (3)$$

here, h_j is the feature vector of the neighboring node j, W is a learnable weight matrix, e_{ij} is the attention coefficient, and σ is a non-linear activation (ReLU in this implementation). The attention coefficient is calculated as in Eq. (2) where a^T which is a learnable weight vector and ‖ denotes concatenation [19]. After each GATv2 layer, a gated skip connection is applied to combine features from- adjacent layers while mitigating gradient vanishing and feature degradation [21]. This process is repeated across three GAT layers, after which the node features are aggregated using mean global pooling and max global pooling to consolidate information across all nodes. The resulting pooled features are then passed through two fully connected layers to generate the final drug embeddings for the molecular graphs (E_{mol}) and motif graphs (E_{motif}).

For target feature learning, it is represented as amino acid sequences, where each amino acid is encoded into a 128-dimensional vector using an embedding layer. These sequences are processed by a Bidirectional Long Short-Term Memory (BiLSTM) network, which captures contextual relationships in both forward and backward directions. The final output from the BiLSTM is aggregated using a max pooling operation and then passed through two fully connected layers to produce the protein embedding (E_{prot}).

The embeddings E_{mol}, E_{motif}, and E_{prot} are combined using a three-way multi-head attention mechanism to capture interactions between drugs and proteins across multiple modalities [31]. For each attention head k, the attention scores are calculated as described in Eq. (4), where W_k^{mol}, W_k^{motif}, W_k^{prot} are trainable weight matrices. The weighted embeddings are then computed as shown in Eq. (5). The final combined embedding H is obtained by concatenating the outputs from all attention heads, as defined in Eq. (6). This combined embedding is subsequently passed through two fully connected layers to predict the DTA value.

$$a_k = softmax\begin{pmatrix} W_k^{mol}E_{mol} + \\ W_k^{motif}E_{motif} + \\ W_k^{prot}E_{prot} \end{pmatrix} \quad (4)$$

$$H_k = a_k \odot \left[E_{mol} \parallel E_{motif} \parallel E_{prot}\right] \quad (5)$$

$$H = Concat(H_1, H_2, \ldots, H_k) \quad (6)$$

Finally, this concatenated embedding H is passed through two fully connected layers to predict the affinity value. The affinity value final prediction is computed in Eq. (7).

$$y = \sigma\,(WH + b) \quad (7)$$

Where W and b are the output layer weights and bias, and σ is a non-linear activation function (ReLU in this implementation).

3.4 Training & hyperparameters configuration

In this study, the development of the proposed drug-target affinity (DTA) model was conducted using several libraries. For drug preprocessing, the RDKit library was utilized for converting drug SMILES strings into graph representations. For target preprocessing, we used NumPy and Scikit-learn for label encoding. For the model development, PyTorch and PyTorch Geometric were used. The experiment was conducted on a computer equipped with an Intel(R) Core (TM) i7-7850H CPU @ 2.20GHz and NVIDIA GeForce GTX 1050TI GPU. It was also conducted on Google Colaboratory Pro, which is equipped with a Tesla T4 GPU with 16GB memory capacity. The hyperparameter settings explored during the experimental design are detailed in Table 4. To identify the optimal settings, multiple configurations of learning rate and batch size were evaluated. These learning rate range values were chosen to explore both large step sizes for faster convergence and small step sizes for stability. Meanwhile, the batch size range values were varied, with smaller sizes providing precise updates at the cost of higher computational overhead and larger sizes offering stability and efficiency.

Due to limited computational resources, the hyperparameter optimization was conducted sequentially.

Table 4. Hyperparameter setting

Hyperparameters	Value
Epoch	300
Learning rate	0.001, 0,005, 0.0001, **0.0005**
Batch size	32, 64, 128, 256, **512**
Dropout rate	0.2
Loss function	MSE
Optimizer	Adam

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

The experiments started by identifying the optimal learning rate, with a base batch size of 32. After determining the best learning rate configuration (0.0005), the next step focused on finding the optimal batch size while keeping the learning rate fixed at 512. This systematic approach allowed for efficient exploration of the hyperparameter space while balancing the constraints of computational efficiency. Among the configurations tested, the values highlighted in bold in Table 4 were selected as the final hyperparameters based on prior research and experimental results. The model was trained for a maximum of 300 epochs, with 20% of the training data allocated for validation. The mean squared error (MSE) was employed as the loss function, and the Adam optimizer was used for training. These hyperparameters were chosen to balance computational efficiency and model accuracy.

3.5 Evaluation metrics

The proposed model's performance was assessed by benchmarking it against several state-of-the-art models using standard evaluation metrics for DTA predictions. Given that DTA prediction is a regression problem, three primary metrics frequently employed in DTA research are Mean Squared Error (MSE), Concordance Index (CI), and regression toward the mean (r_m^2) [39, 40]. MSE evaluates the average squared deviation between the predicted and actual values, where a lower MSE value signifies improved prediction accuracy by reducing error. The formula for MSE is provided in Eq. (8).

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(y_i^{`} - y_i)^2 \tag{8}$$

The Concordance Index (CI) is a metric used to assess the ranking accuracy of predictions by evaluating whether the predicted values for two randomly chosen drug-target pairs align with their true binding affinities. Specifically, if $y_i - y_j$, the predicted binding affinity b_i should also be greater than b_j. A higher CI value reflects better predictive performance. The CI is computed using Eq. (9).

$$CI = \frac{1}{z}\sum_{y_i - y_j} h(b_i - b_j) \tag{9}$$

Here, z is a normalization constant, and $h(u)$ represents the step function. CI measures the consistency of prediction rankings relative to the actual dataset [7]. The r_m^2 metrics measure the external predictive performance of the regression model. It evaluates how well the predicted values of a variable approach the mean in subsequent measurements, even when there are extreme values. The formula is defined in Eq. (10).

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \tag{10}$$

3.6 Experimental design

There are three main experimental scenarios conducted to evaluate the model's performance. The first experiment involved a parameter search to identify the optimal configuration for achieving the best results. Three critical parameters were explored, i.e., the hidden size of the BiLSTM, the number of heads in the GATv2 layers, and the number of heads in the three-way multi-head attention mechanism. The parameter exploration was performed sequentially. At first, the optimal hidden size for the BiLSTM was determined, while the number of heads in both the GAT layers and the attention mechanism was fixed at two. Once the best-hidden size was identified, the optimal number of heads for the GAT layers was explored. Finally, the best number of heads for the attention mechanism was determined using the previously established hidden size and GAT head values. The details of the parameter settings are provided in Table 5.

In the second experiment, after determining the optimal key parameters, the proposed model was compared against benchmark models and baseline models. For the benchmark models, the models are compared against several state-of-the-art methods, such as GraphDTA [17], MSGNN-DTA [21], and DGDTA [32]. GraphDTA was chosen because it was the first study to model DTA prediction using graph-based representations, laying the foundation for graph neural network applications in this area. MSGNN-DTA was included for its use of multi-scale features, which enable the model to capture structural details at various levels of the molecular graph.

Table 5. Parameter exploration details

Parameter	Value	Description
BiLSTM hidden size	16, 32, 64	Number of units in the hidden layer for the model to learn complex sequential dependencies
GATv2 Head	2, 4, 8	Number of attention heads. Higher values allow the model to capture diverse relationships among nodes
Attention Mechanism head	2, 4, 8	Number of concatenated attention heads. Controls the richness of modality interactions

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

Finally, DGDTA was selected because it utilizes the Graph Attention Network v2 (GATv2), which offers dynamic attention capabilities, enhancing the representation of graph structures. The model we used is GIN [41] for GraphDTA, GAT for MSGNNDTA, and GATv2 for the DGDTA as these models are the best performance models stated in each respective paper. For fair comparison, we re-executed the publicly available models with their original configuration described in their respective papers but limited the epoch size to be the same as our setting.

For the baseline comparison, the graph convolutional network (GCN) [42] and graph attention network (GAT) [43] were selected as baselines for the drug encoder. GCNs aggregate information from neighboring nodes to capture the structural properties of molecules. Meanwhile, GAT introduces an attention mechanism that assigns varying weights to neighboring nodes, allowing the model to focus on the most relevant node features during aggregation. By comparing the proposed GATv2 with GCN and GAT, this evaluation aims to determine whether GATv2's dynamic attention mechanism provides a meaningful improvement in modeling complex molecular structures. For protein embedding, 1D convolution neural network (1DCNN) and long short-term memory (LSTM) were chosen as baselines. 1DCNN captures local patterns within protein sequences by applying convolutional filters, making it effective for extracting sequential features. LSTM is capable of modeling long-term dependencies and relationships within sequences, offering a solid baseline for comparison with the proposed BiLSTM, which further improves performance by incorporating bidirectional processing to capture contextual relationships within protein sequences more comprehensively.

In the third experiment, we evaluated the performance impact of incorporating multi-scale features and the three-way multi-head attention mechanism by developing several model variants. The first variant, without multi-scale features, excludes multi-scale information (motif drugs) to assess its specific contribution to the model's performance. The second variant, without the three-way multi-head attention mechanism, excludes the attention mechanism to analyze their impact on interaction modeling between drug and target that contribute to the model prediction accuracy. Lastly, the third variant, without both multi-scale features and the three-way multi-head attention mechanism, removes both components to provide a comprehensive evaluation of their combined effect on the model's performance.

Table 6. Parameters exploration results (Davis dataset)

Parameter	Value	MSE ↓	CI ↑	r_m^2 ↑
BiLSTM hidden size	16	0.342	0.855	0.418
	32	0.325	0.868	0.493
	64	**0.317**	**0.870**	**0.539**
GATv2 head	4	0.336	0.865	0.493
	8	**0.311**	**0.871**	**0.554**
	10	0.325	0.869	0.463
Attention mechanism head	**4**	**0.254**	**0.888**	**0.604**
	8	0.326	0.847	0.445
	10	0.331	0.857	0.453

Table 7. Parameters exploration results (KIBA dataset)

Parameter	Value	MSE ↓	CI ↑	r_m^2 ↑
BiLSTM hidden size	16	0.192	0.855	0.603
	32	0.190	0.860	0.601
	64	**0.185**	**0.861**	**0.612**
GATv2 head	4	0.189	0.859	0.689
	8	**0.184**	**0.862**	**0.715**
	10	0.191	0.859	0.672
Attention mechanism head	**4**	**0.166**	**0.877**	**0.729**
	8	0.190	0.859	0.707
	10	0.188	0.859	0.694

4. Experiment result

4.1 Parameter exploration results

In this experiment, we investigated the contribution of three key parameters, i.e., the hidden size of the BiLSTM, the number of heads in the GATv2 layers, and the number of heads in the three-way multi-head attention mechanism. The Concordance Index (CI) was selected as the primary metric to identify the best-performing parameter configurations. The results for the Davis dataset are summarized in Table 6. For the BiLSTM hidden size, a hidden size of 64 achieved the best performance, with a CI of 0.870, the lowest MSE of 0.317, and the highest r_m^2 of 0.539. For the GATv2 number of heads, the optimal value was found to be 8 heads, achieving a CI value of 0.871, an MSE of 0.311, and a r_m^2 of 0.554. For the attention mechanism number of heads, the best performance was achieved with 4 heads, resulting in a CI value of 0.888, and an MSE of 0.254, and a r_m^2 of 0.604. As for the KIBA dataset, as shown in Table 7, the BiLSTM hidden size of 64 achieved the lowest MSE of 0.185, the highest CI of 0.861, and the highest r_m^2 of 0.612. for the GATv2 head, the best value obtained is 8 head with MSE of 0.184, CI of 0.862, and r_m^2 of 0.715. Finally, for the attention

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

mechanism head, the best results were obtained with 8 heads, giving an MSE of 0.166, a CI of 0.877, and the highest r_m^2 of 0.729.

The parameter exploration across both the Davis and KIBA datasets revealed several consistent trends. Increasing the BiLSTM hidden size from 16 to 64 consistently improved the model performance. These results indicated that a larger hidden size allowed the model to effectively capture complex sequential dependencies within the protein sequences. For the GATv2 heads, a smaller value of 4 heads resulted in slightly worse performance as it did not capture sufficient diversity in node relationships. On the other hand, a higher value of 10 heads showed a drop in performance likely due to overfitting or increased computational overhead, which added unnecessary complexity without meaningful improvements. Finally, for the attention mechanism heads, a higher number of heads decreased the model's performance. This was likely because increasing the number of attention heads diluted the attention weights, reducing the focus on key modality interactions and indicating diminishing returns and added computational complexity. Overall, this consistency across both datasets suggests that the chosen parameter configuration is robust and effective even for larger and more diverse datasets like KIBA. The best parameters configuration obtained for both datasets are BiLSTM hidden size of 64, Gatv2 head of 8, and attention mechanism head of 4. These configurations balance model complexity and performance for DTA prediction.

4.2 Models comparison results

4.2.1. Benchmark comparison

We compared our proposed models with state-of-the-art benchmark models. The results for both datasets are shown in Table 8. On the Davis dataset, the proposed model achieved the lowest MSE of 0.254 the highest CI of 0.888, and r_m^2 0.604 compared to benchmark models. The proposed model achieves consistent gains across all three metrics, indicating the model's accurate overall predictions for binding affinities. On the KIBA dataset, our proposed model achieved an MSE of 0.166, the highest CI of 0.877, and r_m^2 of 0.729, surpassing GraphDTA [17] and DGDTA [32]. Although MSGNNDTA [21] obtained a marginally lower MSE of 0.156, the performance gap is relatively small. It may be attributed to MSGNNDTA's weighted protein graph representation, which can capture more protein features more effectively but also increase computational complexity moderately. Nevertheless, the proposed model's stronger CI and r_m^2 values highlight its superior overall ranking ability and its predictive power for binding affinities within a large dataset. This suggests that the multi-scale drug representation using GATv2 and BiLSTM-based protein encoder with a three-way multi-head attention mechanism provides more comprehensive and robust modeling of drug-target interactions, ultimately yielding more reliable DTA predictions.

4.2.2. Baseline comparison

Table 9 presents the comparison results of the proposed model with various baselines, where different graph-based encoders (GCN, GAT, and GATv2) are paired with different sequence-based encoders (1DCNN, LSTM, and BiLSTM). On the Davis dataset, our proposed model (GATv2-BiLSTM) achieved the best results for MSE (0.254), CI (0.888), and r_m^2 (0.604), outperforming all other baselines. A similar trend is observed on a larger dataset, which is the KIBA dataset. GATv2BILSTM achieved the lowest MSE (0.166) and achieved the highest CI (0.877) and r_m^2 (0.729).

These results highlight two key factors that contribute to the superiority of the proposed model. GATv2's dynamic attention assigns context-specific weights to neighboring nodes, which more effectively captures the subtleties of multi-scale molecular structures than the uniform aggregation of GCN or the static attention of GAT. Secondly, BiLSTM's bidirectional sequence encoding more thoroughly extracts contextual dependencies across long protein sequences compared to 1DCNN or unidirectional LSTM, thereby producing richer representations of the target proteins. Consequently, the combination of GATv2 (drug) and BiLSTM (protein) enables more precise DTA predictions, as reflected by consistently lower MSE and higher CI and r_m^2 values on both datasets.

Table 8. State-of-the-art models comparison

Dataset	Model	MSE ↓	CI ↑	r_m^2 ↑
Davis	GraphDTA	0.256	0.884	0.531
	MSGNNDTA	0.330	0.863	0.444
	DGDTA	0.283	0.869	0.505
	Proposed Model	**0.254**	**0.888**	**0.604**
KIBA	GraphDTA	0.192	0.850	0.619
	MSGNNDTA	**0.156**	0.875	0.660
	DGDTA	0.257	0.824	0.493
	Proposed Model	0.166	**0.877**	**0.729**

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

Table 9. Baseline models comparison

Dataset	Model		MSE ↓	CI ↑	r_m^2 ↑
	Drug Encoder (Mol & Motif)	Target Encoder (Protein)			
Davis	GCN	1DCNN	0.302	0.868	0.514
	GCN	LSTM	0.330	0.865	0.446
	GAT	1DCNN	0.304	0.869	0.496
	GCN	LSTM	0.330	0.862	0.507
	GATV2	1DCNN	0.298	0.859	0.460
	GATV2	LSTM	0.396	0.861	0.456
	GATV2	**BiLSTM**	**0.254**	**0.888**	**0.604**
KIBA	GCN	1DCNN	0.185	0.864	0.557
	GCN	LSTM	0.184	0.861	0.588
	GAT	1DCNN	0.187	0.859	0.648
	GCN	LSTM	0.188	0.860	0.598
	GATV2	1DCNN	0.193	0.856	0.545
	GATV2	LSTM	0.191	0.862	0.467
	GATV2	**BiLSTM**	**0.166**	**0.877**	**0.729**

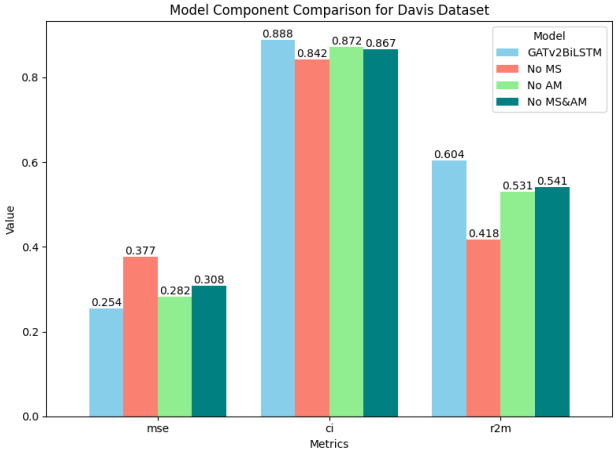
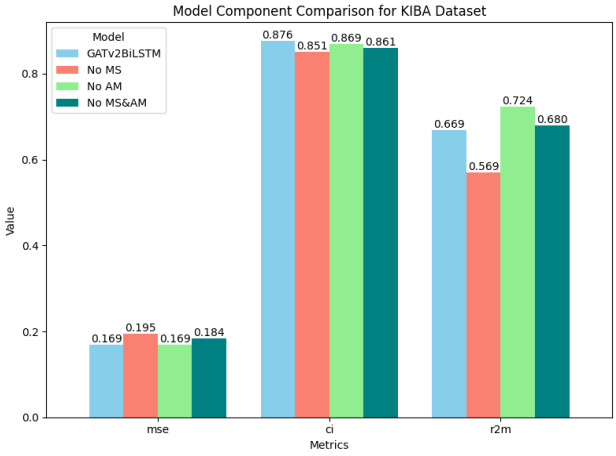


Figure. 4 Model component comparison for the Davis dataset



(b)

Figure. 5 Model component comparison for the d KIBA dataset

4.3 Ablation study results

In the third, experiment, we evaluated the impact of multi-scale features and a three-way multi-head attention mechanism on our proposed models to enhance the DTA prediction. Fig. 4 and Fig. 5 illustrate the comparative performance of four model variants on the Davis and KIBA datasets, respectively: (1) the full GATv2-BiLSTM model, (2) a variant without multi-scale features (No MS), (3) a variant without the three-way multi-head attention mechanism (No AM), and (4) a variant without both (No MS&AM). Across both datasets, the complete GATv2-BiLSTM model outperforms all ablated variants in terms of MSE, CI, and r_m^2, underline the importance of integrating both multi-scale features and the three-way multi-head attention mechanism.

On the Davis dataset, the No MS variant raises the MSE from 0.254 to 0.377, while reducing the CI from 0.888 to 0.842 and r_m^2 from 0.604 to 0.418. This indicates that the multi-scale (molecular + motif) representations significantly improve the model's ability to capture complex chemical-structural patterns relevant to binding affinity. Similarly, the No AM variant reduces the model's MSE by 0.282 and lowers both CI (0.872) and r_m^2 (0.531) compared to the full model. When both multi-scale features and attention are removed (No MS&AM variant), performance further declines (MSE = 0.308, CI = 0.867, r_m^2 = 0.541), reinforcing that these two components complement each other in boosting DTA prediction. On the KIBA dataset, the No MS variant

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

showed an MSE of 0.195, a CI of 0.851, and an r_m^2 of 0.569. Meanwhile, the No AM variant showed an MSE of 0.169, a CI of 0.869, and r_m^2 of 0.724. this result is slightly better than the NO MS variant, underlining the importance of multi-scale features for drug representation. for the No MS&AM variant, the model showed an MSE of 0.184, a CI of 0.861, and an r_m^2 of 0.680. This result further demonstrates the importance of multi-scale features and attention mechanisms in enhancing our proposed model's overall prediction accuracy.

Overall, the multi-scale drug representation effectively captures both the drug's global and local structural properties, while the three-way multi-head attention mechanism enhances interaction modeling between drug molecule, drug motif, and protein sequence representations. By jointly incorporating these components, the proposed GATv2–BiLSTM model achieves stronger predictive capabilities and more robust generalization on both the Davis and KIBA datasets.

5. Conclusion

This study introduces an enhanced method for DTA prediction by integrating GATv2 for dynamic graph representation and BiLSTM for protein sequence encoding. The incorporation of multi-scale features which include a drug motif graph, and a three-way multi-head attention mechanism significantly improves the model's accuracy and interpretability, as evaluated on the Davis and KIBA datasets. The proposed approach addresses key limitations in previous studies, such as the static nature of traditional attention mechanisms, insufficient structural drug representation, and simple interaction modeling. The proposed model achieves significant performance improvements over baselines and state-of-the-art methods. On the Davis dataset, it achieved an MSE of 0.254, a CI of 0.888, and an of 0.604, while on the KIBA dataset, it attained an MSE of 0.166, a CI of 0.877, and an r_m^2 of 0.729. These metrics demonstrate the model's ability to provide accurate and robust DTA predictions, outperforming baseline models and benchmarks such as GraphDTA, MSGNN-DTA, and DGDTA.

Although these experimental scenarios showed promising results, this study acknowledges certain limitations. The data used was limited to the Davis and KIBA datasets, leaving questions about the model's ability to generalize on broader datasets. Furthermore, the computational complexity of the GATv2 layers and three-way multi-head attention mechanisms, while contributing to high performance, presents scalability challenges for large-scale drug discovery efforts due to high computational complexity usage. Future research should evaluate the model on the more diverse dataset, with additional representations of drugs and targets, including molecular fingerprints, protein subsequences, and 3D protein structures. Moreover, several approaches such as the graph sampling method, sparse attention mechanisms, and hardware optimization should be considered to improve computational efficiency. Exploring transformer-based architectures, such as ESM or ProtBERT, may enhance protein sequence representation and reduce training time. Extending the framework with additional features, optimizing efficiency, and testing in low-resource environments could boost its practicality for real-world drug discovery, enabling faster identification of therapeutic candidates.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

"Conceptualization, Muhammad Rizky Yusfian Yusuf and Isman Kurniawan; methodology, Muhammad Rizky Yusfian Yusuf; software, Muhammad Rizky Yusfian Yusuf; validation, Isman Kurniawan and Muhammad Rizky Yusfian Yusuf; formal analysis, Muhammad Rizky Yusfian Yusuf; resources, Muhammad Rizky Yusfian Yusuf; data curation, Muhammad Rizky Yusfian Yusuf; writing—original draft preparation, Muhammad Rizky Yusfian Yusuf; writing—review and editing, Isman Kurniawan; visualization, Muhammad Rizky Yusfian Yusuf; supervision, Isman Kurniawan.

References

[1] K. Sachdev and M. K. Gupta, "A comprehensive review of feature based methods for drug target interaction prediction", *Journal of Biomedical Informatics*, Vol. 93, p. 103159, 2019, doi: 10.1016/j.jbi.2019.103159.

[2] Y. Zhang, Y. Hu, N. Han, A. Yang, X. Liu, and H. Cai, "A survey of drug-target interaction and affinity prediction methods via graph neural networks", *Computers in Biology and Medicine*, Vol. 163, p. 107136, 2023, doi: 10.1016/j.compbiomed.2023.107136.

[3] A. Suruliandi, T. Idhaya, and S. P. Raja, "Drug Target Interaction Prediction Using Machine Learning Techniques - A Review", *IJIMAI*, Vol. 8, No. 6, p. 86, 2024, doi: 10.9781/ijimai.2022.11.002.

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

[4] H. Bhargava, A. Sharma, and P. Suravajhala, "Chemogenomic Approaches for Revealing Drug Target Interactions inDrug Discovery", *CG*, Vol. 22, No. 5, pp. 328-338, 2021, doi: 10.2174/1389202922666210920125800.

[5] Z. Li *et al.*, "In silico prediction of drug-target interaction networks based on drug chemical structure and protein sequences", *Sci Rep*, Vol. 7, No. 1, p. 11174, 2017, doi: 10.1038/s41598-017-10724-0.

[6] Z. Shi and J. Li, "Drug-Target Interaction Prediction with Weighted Bayesian Ranking", In: *Proc. of the 2nd International Conference on Biomedical Engineering and Bioinformatics*, Tianjin China, pp. 19-24, 2018, doi: 10.1145/3278198.3278210.

[7] T. Pahikkala *et al.*, "Toward more realistic drug-target interaction predictions", *Briefings in Bioinformatics*, Vol. 16, No. 2, pp. 325-337, 2015, doi: 10.1093/bib/bbu010.

[8] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, and M. Ester, "SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines", *J Cheminform*, Vol. 9, No. 1, p. 24, 2017, doi: 10.1186/s13321-017-0209-z.

[9] I. Lee, J. Keum, and H. Nam, "DeepConv-DTI: Prediction of drug-target interactions via deep learning with convolution on protein sequences", *PLoS Comput Biol*, Vol. 15, No. 6, p. e1007129, Jun. 2019, doi: 10.1371/journal.pcbi.1007129.

[10] 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", *Bioinformatics*, preprint, 2020, doi: 10.1101/2020.08.11.247437.

[11] N. R. C. Monteiro, B. Ribeiro, and J. P. Arrais, "Drug-Target Interaction Prediction: End-to-End Deep Learning Approach", *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, Vol. 18, No. 6, pp. 2364-2374, 2021, doi: 10.1109/TCBB.2020.2977335.

[12] S. M. Hasan Mahmud, W. Chen, H. Jahan, B. Dai, S. U. Din, and A. M. Dzisoo, "DeepACTION: A deep learning-based method for predicting novel drug-target interactions", *Analytical Biochemistry*, Vol. 610, p. 113978, 2020, doi: 10.1016/j.ab.2020.113978.

[13] Y.-B. Wang, Z.-H. You, S. Yang, H.-C. Yi, Z.-H. Chen, and K. Zheng, "A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network", *BMC Med Inform Decis Mak*, Vol. 20, No. S2, p. 49, 2020, doi: 10.1186/s12911-020-1052-0.

[14] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: deep drug-target binding affinity prediction", *Bioinformatics*, Vol. 34, No. 17, pp. i821-i829, 2018, doi: 10.1093/bioinformatics/bty593.

[15] Y. Pu, J. Li, J. Tang, and F. Guo, "DeepFusionDTA: Drug-Target Binding Affinity Prediction With Information Fusion and Hybrid Deep-Learning Ensemble Model", *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, Vol. 19, No. 5, pp. 2760-2769, 2022, doi: 10.1109/TCBB.2021.3103966.

[16] W. Yuan, G. Chen, and C. Y.-C. Chen, "FusionDTA: attention-based feature polymerizer and knowledge distillation for drug-target binding affinity prediction", *Briefings in Bioinformatics*, Vol. 23, No. 1, p. bbab506, 2022, doi: 10.1093/bib/bbab506.

[17] 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, 2021, doi: 10.1093/bioinformatics/btaa921.

[18] G. Liu *et al.*, "GraphDTI: A robust deep learning predictor of drug-target interactions from multiple heterogeneous data", *J Cheminform*, Vol. 13, No. 1, p. 58, 2021, doi: 10.1186/s13321-021-00540-0.

[19] S. Brody, U. Alon, and E. Yahav, "How Attentive are Graph Attention Networks?", Jan. 31, 2022, *arXiv*: arXiv:2105.14491. Accessed: 11, 2023.

[20] H. Yu, W.-X. Xu, T. Tan, Z. Liu, and J.-Y. Shi, "Prediction of drug-target binding affinity based on multi-scale feature fusion", *Computers in Biology and Medicine*, Vol. 178, p. 108699, 2024, doi: 10.1016/j.compbiomed.2024.108699.

[21] S. Wang *et al.*, "MSGNN-DTA: Multi-Scale Topological Feature Fusion Based on Graph Neural Networks for Drug-Target Binding Affinity Prediction", *IJMS*, Vol. 24, No. 9, p. 8326, 2023, doi: 10.3390/ijms24098326.

[22] A. Mahdaddi, S. Meshoul, and M. Belguidoum, "EA-based hyperparameter optimization of hybrid deep learning models for effective drug-target interactions prediction", *Expert Systems with Applications*, Vol. 185, p. 115525, 2021, doi: 10.1016/j.eswa.2021.115525.

[23] Q. Zhao, G. Duan, M. Yang, Z. Cheng, Y. Li, and J. Wang, "AttentionDTA: Drug-Target Binding Affinity Prediction by Sequence-Based Deep Learning With Attention Mechanism",

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

IEEE/ACM Trans. Comput. Biol. and Bioinf., Vol. 20, No. 2, pp. 852-863, 2023, doi: 10.1109/TCBB.2022.3170365.

[24] H. Öztürk, E. Ozkirimli, and A. Özgür, "WideDTA: prediction of drug-target binding affinity", *arXiv: arXiv:1902.04166*, 2019. doi: 10.48550/arXiv.1902.04166.

[25] A. Ghimire, H. Tayara, Z. Xuan, and K. T. Chong, "CSatDTA: Prediction of Drug-Target Binding Affinity Using Convolution Model with Self-Attention", *IJMS*, Vol. 23, No. 15, p. 8453, 2022, doi: 10.3390/ijms23158453.

[26] S. D'Souza, K. V. Prema, S. Balaji, and R. Shah, "Deep Learning-Based Modeling of Drug-Target Interaction Prediction Incorporating Binding Site Information of Proteins", *Interdiscip Sci Comput Life Sci*, Vol. 15, No. 2, pp. 306-315, 2023, doi: 10.1007/s12539-023-00557-z.

[27] Q. Zhao, F. Xiao, M. Yang, Y. Li, and J. Wang, "AttentionDTA: prediction of drug-target binding affinity using attention model", In: *Proc. of 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, CA, USA, pp. 64-69, 2019, doi: 10.1109/BIBM47256.2019.8983125.

[28] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J. B. Ghasemi, and A. Masoudi-Nejad, "DeepCDA: deep cross-domain compound-protein affinity prediction through LSTM and convolutional neural networks", *Bioinformatics*, Vol. 36, No. 17, pp. 4633-4642, 2020, doi: 10.1093/bioinformatics/btaa544.

[29] Y. Liang, S. Jiang, M. Gao, F. Jia, Z. Wu, and Z. Lyu, "GLSTM-DTA: Application of Prediction Improvement Model Based on GNN and LSTM", *J. Phys.: Conf. Ser.*, Vol. 2219, No. 1, p. 012008, 2022, doi: 10.1088/1742-6596/2219/1/012008.

[30] M. Xia, J. Hu, X. Zhang, and X. Lin, "Drug-Target Binding Affinity Prediction Based on Graph Neural Networks and Word2vec", *Intelligent Computing Theories and Application*, Vol. 13394, pp. 496-506, 2022, doi: 10.1007/978-3-031-13829-4_43.

[31] P. Chen, H. Shen, Y. Zhang, B. Wang, and P. Gu, "SGNet: Sequence-Based Convolution and Ligand Graph Network for Protein Binding Affinity Prediction", *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, Vol. 20, No. 5, pp. 3257-3266, 2023, doi: 10.1109/TCBB.2023.3262821.

[32] H. Zhai, H. Hou, J. Luo, X. Liu, Z. Wu, and J. Wang, "DGDTA: dynamic graph attention network for predicting drug-target binding affinity", *BMC Bioinformatics*, Vol. 24, No. 1, p. 367, 2023, doi: 10.1186/s12859-023-05497-5.

[33] A. Banerjee, Z.-H. Zhou, E. E. Papalexakis, and M. Riondato, Eds., In: *Proc. of the 2022 SIAM International Conference on Data Mining (SDM)*, 2022, doi: 10.1137/1.9781611977172.

[34] M. I. Davis *et al.*, "Comprehensive analysis of kinase inhibitor selectivity", *Nat Biotechnol*, Vol. 29, No. 11, pp. 1046-1051, 2011, doi: 10.1038/nbt.1990.

[35] J. Tang *et al.*, "Making Sense of Large-Scale Kinase Inhibitor Bioactivity Data Sets: A Comparative and Integrative Analysis", *J. Chem. Inf. Model.*, Vol. 54, No. 3, pp. 735-743, 2014, doi: 10.1021/ci400709d.

[36] X. Zeng, S.-J. Li, S.-Q. Lv, M.-L. Wen, and Y. Li, "A comprehensive review of the recent advances on predicting drug-target affinity based on deep learning", *Front. Pharmacol.*, Vol. 15, p. 1375522, 2024, doi: 10.3389/fphar.2024.1375522.

[37] B. Ramsundar, P. Eastman, P. Walters, and V. Pande, *Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery and more*, First edition. Sebastopol, CA: O'Reilly Media, 2019.

[38] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[39] M. Gönen and G. Heller, "Concordance probability and discriminatory power in proportional hazards regression", *Biometrika*, Vol. 92, No. 4, pp. 965-970, 2005, doi: 10.1093/biomet/92.4.965.

[40] K. Roy, P. Chakraborty, I. Mitra, P. K. Ojha, S. Kar, and R. N. Das, "Some case studies on application of 'r_m^2' metrics for judging quality of quantitative structure-activity relationship predictions: Emphasis on scaling of response data", *J Comput Chem*, Vol. 34, No. 12, pp. 1071-1082, 2013, doi: 10.1002/jcc.23231.

[41] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How Powerful are Graph Neural Networks?", *arXiv: arXiv:1810.00826*, 2019.

[42] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks", *arXiv: arXiv:1609.02907*, 2017.

[43] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks", *arXiv: arXiv:1710.10903*, 2018.

This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
License details: https://creativecommons.org/licenses/by-sa/4.0/