Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 10, No. 4, December 2024, pp. 843-857
ISSN: 2338-3070, DOI: 10.26555/jiteki.v10i4.30425

Enhancing Drug-Target Affinity Prediction with Multi-scale Graph Attention Network and Attention Mechanism

Muhammad Rizky Yusfian Yusuf, Isman Kurniawan
School of Computing, Telkom University, Indonesia

ARTICLE INFO

Article history:

Received December 11, 2024
Revised December 28, 2024
Published January 02, 2025

Keywords:

Drug-target affinity;
Drug graph;
Protein sequences;
Graph attention network;
Attention mechanism

ABSTRACT

Drug-target affinity (DTA) prediction is critical to drug discovery, yet traditional experimental methods are expensive and time-consuming. Existing computational approaches often struggle with limitations in representing the structural and sequential complexities of drugs and proteins, resulting in suboptimal prediction accuracy. This study proposes a novel framework integrating Graph Attention Networks (GAT) for drug molecular and motif graphs and Bidirectional Long Short-Term Memory (BiLSTM) for protein sequences. A two-sided multi-head attention mechanism is utilized to dynamically model drug-protein interactions, enhancing robustness and accuracy. This research contribution is the development of a robust computational model that improves the accuracy of DTA predictions, reducing dependency on traditional laboratory methods. The integration of structural and sequential features provides a more comprehensive representation of drug-protein interactions. The study utilizes the Davis and KIBA, a binding affinity datasets that is widely used. the proposed model achieving the lowest Mean Squared Error (MSE) of 0.3209 and 0.1864, the highest Concordance Index (CI) of 0.8646 and 0.8616, and the highest r_m^2 of 0.5046 and 0.6672, respectively, outperforming baseline models. In conclusion, this study showed the proposed approach as a reliable method for DTA prediction, offering a faster and more accurate alternative in the drug discovery research field. However, there are still limitations, such as high computational complexity and the GAT model still uses static attention. Future work will focus on addressing this issue, testing the model across broader datasets, and implementing additional drug and target representation for richer feature extraction.

Corresponding Author:

Muhammad Rizky Yusfian Yusuf, School of Computing, Telkom University, Indonesia
Email: rzkysfian@student.telkomuniversity.ac.id

1. INTRODUCTION

Drug-target interaction (DTI) refers to the binding process between a drug and its specific target, leading to changes in the target's function or behavior [1]. A drug is defined as any chemical compound that, upon consumption, changes the organism's chemical composition. The "target" denotes any biological element within the organism that the drug interacts with, causing changes in chemical states. DTI plays a critical role in the drug development process, which can require investments of up to 2 billion US Dollars and span more than 15 years from initial concept to public distribution. Due to the unpredictable nature of their interactions, most known chemical compounds have not yet been explored for therapeutic use. As a result, there has been a growing interest in explaining the mechanisms of drug-target interactions and developing methods to predict these interactions effectively [2]. Typically, the DTI predictions can be carried out through direct laboratory experiments [2]. However, these laboratory experiments are time-consuming and expensive. To overcome these challenges, new in silico methods are needed [1]. One of the solutions is using computational approaches, that, for the case of DTI, can be categorized into three main types, i.e., ligand-based, docking simulation-based,

and chemogenomic-based approaches [3]. Ligand-based methods work by assuming that similar molecules share similar properties, allowing drug molecules to bind to proteins with analogous structures [4], [5]. However, this method has several drawbacks, such as the prediction that can be conducted only if both of the structure is known [6]. Meanwhile, docking simulation methods rely on the three-dimensional structures of proteins. However, this method also face several challenges such as the unavailability of three-dimensional structures of proteins compared to an actual number of target proteins in the human body [6], [7], [8], [9]. The chemogenomic approach integrates chemical information of drugs and genomic information of proteins within the same subspace to predict potential interactions using machine learning methods [1]. The chemogenomic approach has an advantage related to the procedure of this method that is not necessary to have the 3D structure of protein and drug and also has a large biological data that publicly available [1].

A chemogenomic approach using the machine learning method can be further classified into several methods, such as similarity-based methods [10], [11], [12], feature-based methods [13], [14], [15], and deep learning methods [16], [17]. Recently, deep learning has gained prominence in DTI prediction due to its ability to uncover hidden or complex patterns in data. Generally, DTI predictions using deep learning methods are defined as a classification task [18], [19], [20], [21] that determines the existence of interaction between drug and target. However, this approach neglects the precision value of binding affinity, which quantifies the strength of the interaction between a drug and its target as a continuous numerical value [22]. Therefore, predicting drug-target affinity (DTA) provides the additional benefit of estimating the interaction strength between a drug and its target, compared to DTI prediction. This capability helps narrow down the vast pool of potential compounds in drug discovery research, making the search process more efficient [22].

Several previous studies have been performed related to DTA prediction using deep learning models. In 2018, Öztürk *et al.* introduced DeepDTA, which uses 1D Convolutional Neural Networks (1DCNN) for feature learning on both drugs and targets sequences [22]. This approach was further refined in WideDTA by incorporating additional drug and target representations [23]. Pu *et al.* later proposed DeepFusionDTA, which combines sequence and structural information using Dilated-CNN and BiLSTM blocks to create fusion feature maps, followed by LightGBM model for prediction [24]. Another studies, such as FusionDTA, CSatDTA, and DeepMHDTA combines an attention mechanism in the feature extraction phase to better capture information from drugs and protein sequences [25], [26], [27].

Although deep learning models have demonstrated promising results in predicting drug-target affinity (DTA), many studies often only use simple concatenation between drug and protein in the prediction phase. This can lead to some lost interaction information between drugs and proteins. To address this issue, various studies have integrated attention mechanisms for better representation and interaction modeling between drugs and proteins. For instance, Zhao *et al.* introduced AttentionDTA, utilizing an attention mechanism to dynamically capture critical subsequences of drugs and proteins [28]. Abbasi *et al.* proposed DeepCDA, which introduced attention mechanism to highlight the mutual interactions between compound substructures and protein subsequences before concatenating their representations [29]. Mahdaddi *et al.* contributed to CNN-AbiLSTM, combining CNN and attention-based BiLSTM to prioritize important regions within drug-protein pairs during concatenation [30]. Zhao *et al.* enhanced their model in 2023 with a two-side multi-head attention mechanism, which calculates attention scores between subsequences of drugs and proteins [31]. These enhancements have improved the performance of DTA prediction models by addressing the limitations of simple embedding concatenation.

Even though attention mechanisms for model interaction between drugs and targets has significantly enhanced the performance of DTA prediction models by addressing the limitations of simple embedding concatenation. However, there is still a limitation when representing drugs as string sequences. This challenge arises because string representations can lose essential structural details of the drugs, affecting the accuracy of binding affinity predictions [32]. To overcome this issue, several studies have employed graphs as a drug representation. Nguyen *et al.* introduced GraphDTA, leveraging Graph Neural Networks (GNNs) for drug molecular graph representation and CNNs for protein sequence learning, enabling riche feature extraction for drugs [32]. Zhang *et al.* extended this concept by developing SAG-DTA, which utilizes a self-attention mechanism within the GNN to emphasize important atomic features in drug graphs [33]. Several models use GNN for drug graphs and also incorporate LSTM for protein sequences, addressing long-term dependencies in sequential data [34], [35]. Chen *et al.* presented SGNet, integrating drug graphs with protein sequences encoded via Conjoint Triad and convolutional layers to improve the fusion of structural and sequential features [36]. Qiu *et al.* developed LSTM-SAGDTA, employing self-attentive graph pooling and LSTM for DTA predictions [37]. Tran *et al.* introduced DeepNC, a framework using GNNs and hypergraph attention mechanisms for drug-target binding affinity prediction, focusing on comprehensive graph-based representations [38]. Wang *et al.* further advanced this field with MSGNN-DTA, integrating multi-scale topological features from motif graphs

and weighted protein graphs while employing a gated skip-connection mechanism for robust feature fusion [39]. Overall, representing the drug as a graph improved the DTA prediction performance by focusing on more structural information.

Despite these advancements, recent graph-based models for DTA prediction face several limitations. Firstly, while graph-based approaches excel at representing drugs, capturing the full complexity of a drug is challenging with a single graph structure. Standard molecular graphs may neglect important substructures, such as drug motifs within the molecule that significantly influence the drug's properties and interactions. These drug motifs, play crucial roles in biological activity and binding affinity. Alternative approaches involve incorporating motif graphs or higher-order structural representations, which can capture these meaningful biological substructures within the molecular graph. By integrating the overall molecular graph and specific motifs, the model can achieve a more comprehensive representation of the drug, leading to improved prediction accuracy. Secondly, GNN models like Graph Convolutional Networks (GCNs) aggregated information from neighboring nodes equally, which may not effectively capture the varying importance of different nodes and edges in the graph. Graph Attention Networks (GATs) address this limitation by incorporating an attention mechanism that allows the model to learn and assign different importance weights to neighboring nodes. By focusing on the most relevant parts of the graph, GATs can capture more nuanced and expressive representations, leading to improved performance in DTA prediction.

Thirdly, protein embeddings predominantly rely on sequence information but often fail to capture the contextual relationships between preceding and succeeding amino acids within the sequences. Traditional sequence models may not effectively model long-range dependencies or the full context of the sequence, which can be crucial for understanding protein function and interactions. This limitation affects the model's ability to accurately predict binding affinities, as important contextual information is missed. Alternative approaches involve using models that better capture sequence context, such as bi-directional LSTM (BiLSTM) networks. These models consider both past and future context in the sequence, enabling a more comprehensive understanding of the protein's structural and functional properties. Lastly, current interaction modeling techniques can only capture interactions between two representations—drug graphs and protein sequences—without considering additional representations. This limitation restricts the model's ability to fully capture the complex interplay between different features of drugs and targets, potentially overlooking critical interactions that could influence binding affinity. To overcome this, alternative approaches involve integrating additional representations, such as motif graphs for drugs or structural domains for proteins, and employing interaction models capable of handling multiple inputs. For example, using a Two-Side Multi-Head Attention Mechanism allows the model to consider multiple aspects of the drug and target simultaneously, capturing intricate patterns of interaction and leading to more accurate predictions.

To address these challenges, this study introduces a computational approach that integrates GAT for drugs molecular graph embedding with motif graph embedding as additional drug representation and BiLSTM for target protein sequences. This integration aims to enhance the DTA prediction accuracy by capturing both the overall molecular structure and substructures (motifs) of drugs, as well as improving the contextual understanding of protein sequences. To improve the prediction phase, this study also incorporates a Two-Side Multi-Head Attention Mechanism inspired by AttentionDTA. This mechanism dynamically models the interdependencies between drug molecular graphs embedding, drug motif graphs embedding, and protein embeddings to predict the binding affinities value. This research contribution is to enhance the model's ability to predict drug-target binding affinity values by providing a more comprehensive drug representation and improved interaction modeling between drugs and proteins.

2. METHODS

This study focuses on developing a robust deep-learning model for DTA prediction by leveraging advanced computational techniques. The research workflow involves several key steps, including dataset selection, pre-processing of data to create suitable representations for drugs and proteins, model development, and evaluation of predictive performance using established metrics. Two widely recognized datasets, Davis and KIBA, were employed for model training and validation due to their rich and diverse information on drug-protein interactions. Pre-processing steps were designed to extract meaningful features from drugs and targets, including molecular graphs, motif graphs, and protein sequences. The proposed model integrates sequence- and graph-based approaches with advanced attention mechanisms to enhance predictive accuracy. Finally, rigorous evaluation metrics were applied to assess the model's effectiveness in predicting binding affinities, ensuring robust and reliable performance. This study workflow can be seen in Fig. 1.

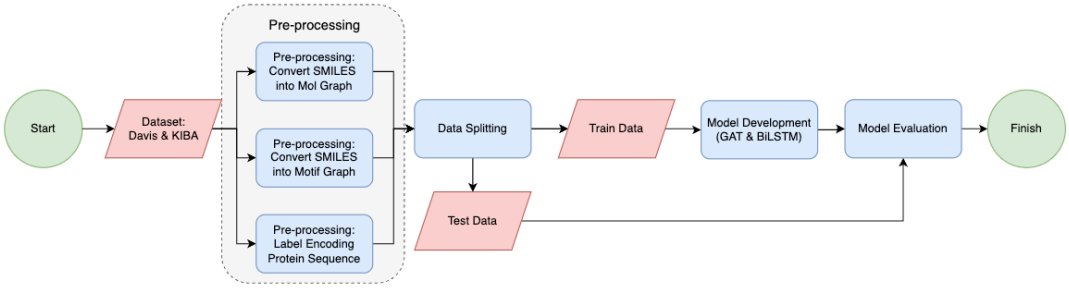

Fig. 1. Proposed Method study workflow

2.1. Dataset

In this study, the Davis and KIBA datasets were utilized [40], [41]. Both datasets consist of binding affinity values, treated as the dependent variable, while drugs and targets/proteins were considered independent variables. The datasets were checked for duplicates and missing values to ensure the validity of the data set. The Davis dataset consists of selectivity tests for relevant protein kinase and inhibitor families, along with their corresponding dissociation constant (K_d) values, measured in nanomolar units. To get a better representation of the relationship between K_d and binding affinity, we calculated pK_d the variable as formulated in (1). Higher pK_d values signify stronger binding affinities and vice versa. The distribution of value shows that the pK_d value ranges from 5.0 to 10.8.

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

The variable provided in the KIBA dataset is derived by using the KIBA score, which integrates kinase inhibitor bioactivity data from various sources, including K_i, K_d, and IC_{50}, to compute a unified metric known as the KIBA score. These scores range from 0.0 to 17.2, where lower values indicate stronger binding affinities. A summary of the drugs, targets, and their interaction counts can be seen in Table 1.

Table 1. Dataset Description

Description	Drug	Protein	Total Interaction	Train set (80%)	Test set (20%)
Davis	62	379	30,056	25,046	5,010
KIBA	2,068	229	118,254	94,603	23,651

2.2. Pre-processing

SMILES (Simplified Molecular Input Line Entry System) is widely used to represent the three-dimensional structure of drug molecules in textual format. SMILES strings capture key characteristics of drug molecules, such as atomic weights and valence electrons [32]. In this study, each drug's SMILES representation was converted into a molecular graph where the nodes represent the collection of atoms in the drug, and the edges represent the chemical bonds between atoms [2]. To better describe the node feature in graphs, we adapted a set of atomic features from DeepChem [42]. Here, Each node in the molecular graph encodes the chemical properties of its corresponding atom using a 78-dimensional feature vector, where each dimension corresponds to a specific chemical attribute. Detailed descriptions of these atomic features are provided in Table 2.

To further enhance the representation of drug structure information, a motif-level graph was constructed alongside the molecular graph. Motifs in drugs, such as the benzene ring, are closely tied to molecular properties. For instance, a benzene ring retains its significance as a whole structure but loses meaning when its bonds are considered in isolation. Several layers of graph neural networks (GNNs) struggle to capture the complete information within these cyclic structures, leading to an incomplete feature extraction [39]. An example of transformation from SMILES into a graph can be seen in Fig. 2.

Table 2. Molecular Graph Node Feature Details

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

The motif-level graph was formed by identifying fundamental building blocks consisting of both cyclic arrangements of atoms and bonds, and individual bonds that are not embedded in any ring structure together with their connected atom pairs [39]. Each of these building blocks is represented as a node in the motif graph. Nodes corresponding to cyclic structures capture groups of atoms and bonds arranged in rings, while the remaining nodes denote single chemical bonds and their associated pairs of atoms. Edges in the motif graph represent chemical bonds connecting these nodes. Similar to molecular graphs, the features of motif graph nodes were encoded into a 92-dimensional vector based on the DeepChem [42]. The details is described in Table 3.

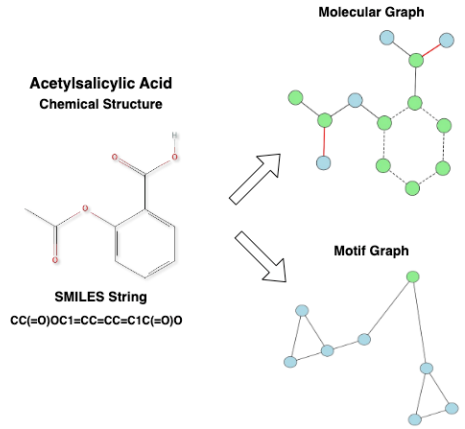

Fig. 2. Illustration of drug graph construction

For target processing, the amino acid sequence of the protein's primary structure was utilized. Protein sequences consist of 25 types of amino acids, each represented by a specific ASCII character denoting its properties. Following the approach used in the GraphDTA study, these sequences were label-encoded with a maximum length of 1,000 characters. Sequences shorter than this length were zero-padded, while longer sequences were truncated [32]. This is done to ensure the same input dimension size for training convenience. Additionally, this length also accommodates the majority of proteins as the length varies from 200 to 2000 with a median length of 700 characters.

Table 3. Motif Graph Node Feature Details

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

2.3. Model Training

This study proposes DTA prediction by integrating graph-based and sequence-based models. The model processes drugs as molecular and motif graphs using Graph Attention Networks (GAT) [43] and proteins as sequences with Bidirectional Long Short-Term Memory (BiLSTM) [44] and combines these embeddings through a two-side multi-head attention mechanism for prediction [31]. The final embeddings will be passed through a fully connected layer for the model to predict the affinity value. Fig. 3 provides an overview of the proposed architecture.

The drug graphs are represented as $G = (V, E)$, where V (nodes) denotes atoms encoded as 78-dimensional feature vectors for molecular graph and 92-dimensional feature vectors for motif graph, and E (edges) represent chemical bonds between the atoms. Both graphs pass through five GAT layers to learn node embeddings, respectively. Each GAT layer computes the updated representation for a node i as in equation (3).

$$a_{ij} = Softmax\left(LeakyReLU\left(a^T\left[Wh_i \parallel Wh_j\right]\right)\right) \quad (2)$$

$$h_i' = \sigma\left(\sum_{j \in N(i)} a_{ij} W h_j\right) \tag{3}$$

Here h_j represent the feature vector corresponding to the neighboring node j, W represent a learnable weight matrix, a_{ij} is the attention coefficient, and σ is a non-linear activation (ReLU in this implementation). The attention coefficient a_{ij} is calculated as in equation (2) where a^T which is a learnable weight vector and $\|$ denotes concatenation [43]. After each GAT layer, a gated skip connection is applied to combine features from adjacent layers while mitigating gradient vanishing and feature degradation [39]. For a node i in the layer l, the skip connection updates the representation in equation (4) and (5).

$$z_i = sigmoid\left(U_1 H_i^{(l+1)} + U_2 H_i^{(l)} + b\right) \tag{4}$$

$$H_i^{(l+1)} = z_i H_i^{(l+1)} + (1 - z_i) H_i^l \tag{5}$$

Here U_1 and U_2 are trainable parameters, b is bias, $H_i^{(l)}$ and $H_i^{(l+1)}$ represent the feature vectors of node i at the lth and l+1th layer, respectively. The coefficient z_i represent the learned proportion coefficient that retains the information of the preceding hidden layer. A sigmoid activation function is employed to ensure that this learned proportion coefficient remains within the interval [0,1]. This process is repeated across the three GAT layers and passed into mean global pooling and max global pooling aggregate information across all nodes. The pooled features are then passed through two fully connected layers to obtain final drug embedding molecular graphs (E_{mol}) and motif graphs (E_{motif}).

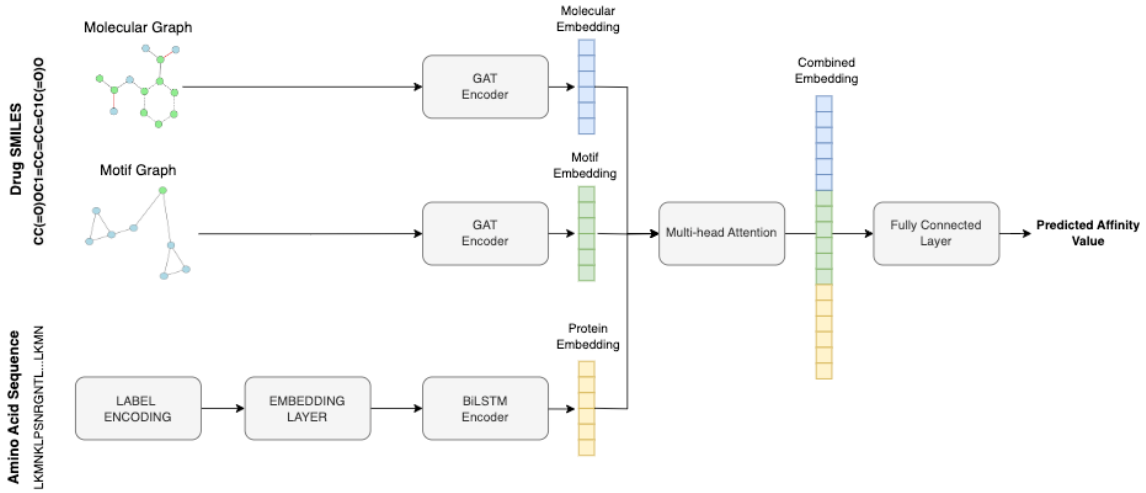

Fig. 3. Proposed Model Architecture Overview

Protein targets are represented by their amino acid sequences. Each amino acid is encoded as a 128-dimensional vector using an embedding layer. The sequence is processed by a BiLSTM, which captures contextual relationships in both forward and backward directions. The final BiLSTM output is pooled and passed through two fully connected layers to generate the protein embedding (E_{prot}).

The embeddings E_{mol}, E_{motif}, and E_{prot} are combined using a two-side multi-head attention mechanism to capture the interactions between drugs and proteins across modalities [31]. For each attention head k, attention scores are computed as in equation (6) where W_k^{mol}, W_k^{motif}, W_k^{prot} are trainable weight matrices. The weighted embedding is computed as in equation (7). The final combined embedding is the concatenation of outputs from all attention heads, which can be seen in equation (8). The combined embedding H is passed through two fully connected layers to predict the DTA value.

$$a_k = softmax\left(W_k^{mol} E_{mol} + W_k^{motif} E_{motif} + W_k^{prot} E_{prot}\right) \tag{6}$$

$$H_k = a_k \odot [E_{mol} \parallel E_{motif} \parallel E_{prot}] \quad (7)$$

$$H = Concat(H_1, H_2, \dots, H_k) \quad (8)$$

The combined embedding H is passed through two fully connected layers to predict the DTA value. The final prediction is computed in equation (9).

$$y = \sigma\,(WH + b) \quad (9)$$

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

2.4. Evaluation Metrics

The performance of the proposed model was evaluated by comparing it with several state-of-the-art models using standard evaluation metrics. Since Drug-Target Affinity (DTA) prediction is a regression task, the three most commonly used metrics in DTA studies are Mean Squared Error (MSE), Concordance Index (CI) as proposed by Gönen and Heller [45], and regression towards the mean (r_m^2) introduced by Roy *et al.* [46]. MSE quantifies the average squared difference between the predicted and actual values. A smaller MSE value indicates better model performance, as it reflects a lower degree of error in the predictions. The formula for calculating MSE is shown in equation (10).

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(y_i - y_i)^2 \quad (10)$$

CI is a metric that measures predictive accuracy by checking whether, for two randomly selected drug–target pairs, the predicted rankings mirror their actual binding affinities. Specifically, if $y_i - y_j$, then the predicted binding affinity $b_i$should be greater than b_j. A higher CI value indicates better predictive performance. The CI is calculated in equation (11).

$$CI = \frac{1}{z}\sum_{y_i - y_j} h(b_i - b_j) \quad (11)$$

Here z is a normalization constant and $h(u)$ is the step function. CI evaluates the consistency of prediction rankings with the actual dataset [47].

The r_m^2 metric assesses the external predictive performance of a regression model. It evaluates how closely a variable approaches the mean in subsequent measurements, even for variables with extreme values. This r_m^2 is defined as in equation (12).

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (12)$$

where r^2 is the squared correlation coefficient with an intercept, and r_0^2 is the squared correlation coefficient without an intercept [22]. This metric is handy for assessing the model's generalization performance on unseen data.

2.5. Experimental Design

We used PyTorch and PyTorch Geometric libraries to develop the proposed model. For drug preprocessing, which involved converting drug SMILES into graphs, we used the RDKit library. Table 4 provides an overview of the hyperparameter settings used in our experimental design. These hyperparameters value were chosen based on current knowledge from previous studies and experiments. This values were further validated when we conducted a manual experiment with various learning rate (0.001, 0.0001, 0.0005), batch size values (32, 128, 512), and dropout rate values (0.1, 0.2, 0.3). Specifically, values for the learning rate (0.0005), batch size (512), dropout rate (0.2), and optimizer (Adam) were adopted because they consistently demonstrated strong performance in models for similar drug-target prediction tasks. Due to computational limitations, the model was trained for 250 epochs. Additionally, 20% of the training data was used as the validation set.

Several experimental scenarios were conducted to evaluate the model's performance. In the first experiment, we performed a parameter search to determine the best configuration for achieving optimal performance. Three key parameters were explored: the hidden size for the BiLSTM, the number of heads in

the GAT layers, and the number of heads in the two-side multi-head attention mechanism. All parameter explorations were conducted sequentially. First, we determined the best hidden size value for the BiLSTM while keeping the number of heads in the GAT and the attention mechanism fixed at two. Once the best hidden size was established, we identified the optimal number of heads for the GAT. Lastly, we determined the best number of heads for the attention mechanism with BiLSTM hidden size and GAT heads best value. The details of the parameter settings can be seen in Table 5.

Table 4. Hyperparameter Settings for Training

Hyperparameter	Value
Learning rate	0.0005
Epoch	250
Batch size	512
Dropout rate	0.2
Loss Function	(MSE)
Optimizer	Adam

Table 5. Parameter Exploration Details

Parameter	Value	Description
BiLSTM Hidden Size	16, 32, 64	Number of units in the hidden layer of the BiLSTM to increase the model's capacity to learn complex sequential dependencies
GAT Head	2, 4, 8	Number of attention heads in the GAT layers. Higher values allow the model to capture diverse relationships among nodes.
Attention Mechanism Head	2, 4, 8	Number of concatenated attention heads in the two-side multi-head attention mechanism. Controls the richness of modality interactions.

For the second experiment, after the best key parameters were all obtained, we compared out the proposed model with baseline models and state-of-the-art benchmark models from previous studies. For drug embedding, the Graph Convolutional Network (GCN) [48] was chosen due to its foundational role in graph-based representation learning and its widespread adoption for aggregating information from neighbouring nodes to capture the structural properties of molecular graphs. This comparison allows us to assess whether the proposed Graph Attention Network (GAT)-based approach, with its ability to incorporate attention mechanisms, provides a meaningful improvement in modeling complex molecular structures. For protein embedding, Recurrent Neural Networks (RNN) [49] and Gated Recurrent Units (GRU) [50] were used as baselines, as they are commonly employed in sequential data. RNNs effectively model sequential dependencies, while GRUs enhance efficiency by addressing the vanishing gradient problem, making them suitable for long sequence processing. These models serve as robust benchmarks for evaluating the proposed BiLSTM, which introduces bidirectional processing to capture contextual relationships within protein sequences more effectively. Additionally, we compared our proposed model with two recent state-of-the-art models: GraphDTA [32] and MSGNN-DTA [39]. These studies is chosen because their code is publicly available, and therefore we were able to reproduce it. To ensure a fair comparison, we re-ran these models in the same environment using the same dataset and experimental setup as in the paper but limited the training epoch to 250 epochs, which is the same as our experiment setup. This comprehensive comparison enables us to assess whether our model not only outperforms baseline models but also advances the current state-of-the-art in drug-target affinity prediction.

3. RESULTS AND DISCUSSION

3.1. Parameters Exploration

In the first experiment, we conducted parameter exploration for three key parameter for out proposed model which using Graph Attention Networks (GAT) for drug embeddings and Bidirectional Long Short-Term Memory (BiLSTM) for proteins embeddings combines these embeddings through a two-side multi-head attention mechanism for prediction. Therefore, the three parameters we explore were BiLSTM hidden size, number of GAT heads, and number of attention mechanism heads for both dataset, Davis and KIBA. For this experiment, we consider the value of CI as the overall metric measurement to define the model with the best parameters. For the Davis Dataset, the results are summarized in Table 6. The BiLSTM hidden size, representing the number of units in its hidden layer, was varied among 16, 32, and 64. The results show that increasing the hidden size improved performance, with the best results obtained at a hidden size of 64, achieving an MSE of 0.3269, a CI of 0.8632, and r_m^2 of 0.4820. These improvements highlight the importance of larger hidden sizes in capturing complex sequential dependencies in protein sequences. For the number of

GAT heads, values of 4, 8, and 10 were tested. The optimal performance was achieved with 8 heads, produce an MSE of 0.3361, a CI of 0.8621, and r_m^2 of 0.4907. Increasing the number of heads enhances the model's expressiveness to capture important information on molecular and motif graph. using fewer heads, such as two, lacked the diversity needed for effective graph representation learning. However, a further increase to 10 heads resulted in a performance decline, likely due to noise and overfitting. The number of heads in the two-side multi-head attention mechanism was also tested between 4, 8, and 10. The best performance was observed with 8 heads, achieving the lowest MSE of 0.3209, the highest CI of 0.8646, and r_m^2 of 0.5046. This indicates that the richness of the interaction modeling is maximized at this configuration, effectively aggregating and fusing multi-scale features from molecules, motifs, and proteins to capture complex interactions. The decline in performance at 10 heads suggests that excessive attention heads can introduce noise or lead to overfitting. The optimal configuration identified through these experiments utilized a BiLSTM Hidden Size of 64, 8 GAT heads, and 8 attention heads in the two-side multi-head attention mechanism. This configuration allowed the model to capture richer relationships among molecular, motif, and protein features.

Table 6. Parameter Exploration Results for Davis Dataset

Parameter	Value	MSE ↓	CI ↑	r_m^2 ↑
BiLSTM Hidden Size	16	0.3550	0.8546	0.4542
	32	0.3524	0.8591	0.4310
	64	**0.3269**	**0.8632**	**0.4820**
GAT Head	4	0.3394	0.8591	0.4463
	8	**0.3361**	**0.8621**	**0.4907**
	10	0.3410	0.8546	0.4698
Attention Mechanism Head	4	0.3354	0.8600	0.4348
	8	**0.3209**	**0.8646**	**0.5046**
	10	0.3671	0.8553	0.4475

For the KIBA dataset, as shown in Table 7, The best BiLSTM hidden size was also 64, achieving the lowest MSE of 0.1939, the highest CI of 0.8541, and r_m^2 of 0.4804. These results confirm that larger hidden sizes enhance the ability to capture long-range dependencies in protein sequences. For the GAT heads, the best configuration was 8 heads, resulting in an MSE of 0.2002, a CI of 0.8555, and r_m^2 of 0.4343. Similar to the Davis dataset, this configuration effectively balances complexity and representation capacity. Increasing to 10 heads led to a slight degradation in performance, mirroring the results observed in the Davis dataset. Interestingly, the optimal configuration for the attention mechanism was 4 heads, yielding the lowest MSE of 0.1864, the highest CI of 0.8616, and r_m^2 of 0.6672. This differs from the Davis dataset, where 8 heads were optimal, striking a balance between representational capacity and computational efficiency. The better performance with fewer attention heads may be attributed to the size of the dataset and dataset-specific characteristics such as different affinity value calculation. Overall, the optimal configuration for KIBA dataset is a BiLSTM Hidden Size of 64, 8 GAT heads, and 4 attention heads in the two-side multi-head attention mechanism.

Table 7. Parameter Exploration Results for KIBA Dataset

Parameter	Value	MSE ↓	CI ↑	r_m^2 ↑
BiLSTM Hidden Size	16	0.2046	0.8486	0.4621
	32	0.2115	0.8434	0.4853
	64	**0.1939**	**0.8541**	**0.4804**
GAT Head	4	0.1934	0.8536	0.7016
	8	**0.2002**	**0.8555**	**0.4343**
	10	0.1992	0.8521	0.4699
Attention Mechanism Head	**4**	**0.1864**	**0.8616**	**0.6672**
	8	0.1978	0.8517	0.5057
	10	0.1997	0.8507	0.5665

These results underscore the importance of carefully selecting model parameters tailored to each dataset, as seen in the differences between the Davis and KIBA datasets. However, due to current computational power, we are only able to do a parameter exploration with small sample values. further exploration of the parameter space is warranted to provide deeper insights into the sensitivity of the model to specific changes. For example, expanding the range and granularity of hidden sizes (e.g., using increments of 16, 32, and 64) could reveal more nuanced trends in model performance. Conducting and experiment on how variations in network

structure, such as employing different GAT layers or alternative attention mechanisms, might influence the model's predictive accuracy across diverse contexts.

3.2. Overall Comparison

In the second experiment, we compared our proposed model with baseline models. For our proposed model, the learning curves for both the Davis and KIBA datasets, depicted in Fig. 4, show rapid convergence of the training loss during the initial epochs, followed by a declining trend. Although the regularization technique (L1 and L2) and early stopping were not utilized, the risk of overfitting was partially mitigated through careful dataset splitting (80% training, 20% testing) and consistent monitoring of validation loss during training. These measures helped ensure that the model maintained a balance between training performance and generalization. For both datasets, the validation loss closely mirrors the training loss, indicating that the models generalize well to unseen data and avoid overfitting. These results also reveal the effectiveness of the gated skip-connection mechanism in stabilizing the training process and avoiding gradient degradation even when using five GAT layers for each drug molecule embedding and drug motif embedding, respectively. This is evident from the smooth convergence of training and testing losses in both datasets.

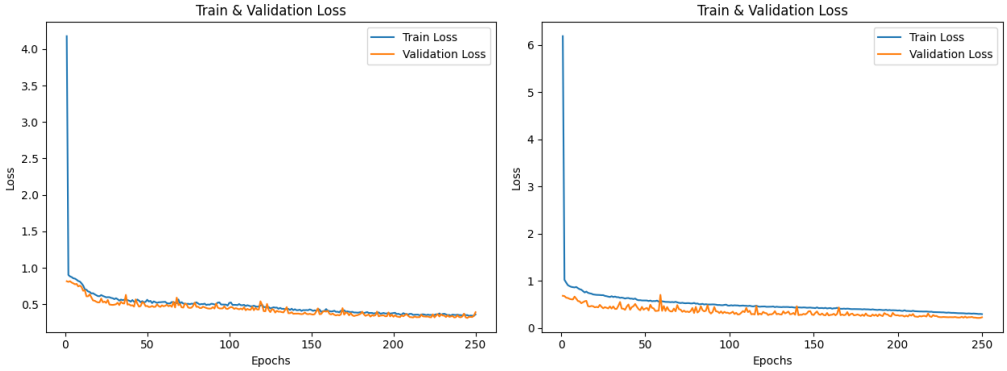

Fig. 4. Proposed Model Training and Validation Loss Curves of Davis (left) and KIBA (right) Dataset

The performance results for the Davis dataset, shown in Table 8, highlight the strengths of the proposed model (GATBILSTM). This model achieved the lowest MSE (0.2369 for the training set and 0.3209 for the testing set), effectively minimizing prediction errors compared to baseline architectures. Additionally, it achieved the highest CI values (0.8874 for training and 0.8646 for testing), demonstrating its superior capability to rank drug-target binding affinities correctly. The model's r_m^2 values (0.5432 for training and 0.5046 for testing) further demonstrate its robustness in capturing the variance in binding affinity data. Notably, the GCN-BiLSTM model emerged as the second-best performer across all baselines, suggesting that the BiLSTM component better captures sequential protein features compared to RNN and GRU. Furthermore, GCN-based models, while competitive, displayed slightly lower performance metrics than their GAT counterparts. This finding suggests that the inclusion of attention mechanisms (GAT) enables more effective learning of complex molecular and motif representations, which are critical for accurate drug-target affinity prediction. Overall, the results from the Davis dataset demonstrate that the proposed GATBILSTM model not only enhances prediction accuracy but also maintains strong generalization capability, as evidenced by the close alignment between training and testing metrics.

Table 8. Model Performance Comparison for Davis Dataset

Model		MSE ↓		CI ↑		r_m^2 ↑	
Drug	**Protein**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
GCN	RNN	0.2848	0.3505	0.8700	0.8537	0.4598	0.4580
GCN	GRU	0.2520	0.3338	0.8819	0.8599	0.4921	0.4406
GCN	BILSTM	0.2511	0.3370	0.8865	0.8599	0.4236	0.4596
GAT	RNN	0.3161	0.3781	0.8629	0.8451	0.4372	0.3963
GAT	GRU	0.2829	0.3527	0.8707	0.8545	0.3981	0.3777
GAT	BILSTM	**0.2369**	**0.3209**	**0.8874**	**0.8646**	**0.5432**	**0.5046**

For the KIBA dataset, as shown in Table 9, the proposed GAT-BiLSTM model demonstrated substantial improvements over all baseline models. It achieved an MSE of 0.1096 for the training set and 0.1864 for the

testing set, surpassing all other architectures in prediction accuracy. Its CI values, 0.8928 for training and 0.8616 for testing, indicate exceptional ranking ability, outperforming the baseline models. Furthermore, the r_m^2 values, 0.8127 for training and 0.6672 for testing, highlight the model's strong predictive power and its ability to capture variance in binding affinity data, even in a more complex and larger dataset like KIBA. Compared to the Davis dataset, the improvement in predictive performance for the KIBA dataset can be attributed to its richer feature representations and larger sample size, which provide more robust training opportunities for the model. The two-sided multi-head attention mechanism in the proposed model likely played a crucial role in modeling the interdependencies among drug molecular graph embeddings, motif graph embeddings, and protein sequence embeddings. This advanced attention mechanism enabled better context aggregation, resulting in higher prediction performance. While GCN-based models were competitive, the GAT-based models consistently outperformed them, indicating the superiority of attention mechanisms in learning complex molecular and motif representations. Overall, the results for the KIBA dataset emphasize the effectiveness of the proposed GATBILSTM model in capturing intricate drug-target interactions and delivering superior prediction accuracy across a larger and more challenging dataset.

Table 9. Model Performance Comparison for KIBA Dataset

Model		MSE ↓		CI ↑		r_m^2 ↑	
Drug	**Protein**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
GCN	RNN	0.1522	0.1990	0.8680	0.8489	0.7787	0.5090
GCN	GRU	0.1416	0.1933	0.8768	0.8559	0.7689	0.5644
GCN	BILSTM	0.1382	0.1937	0.8781	0.8571	0.7627	0.5493
GAT	RNN	0.1587	0.2148	0.8650	0.8426	0.7589	0.4622
GAT	GRU	0.1520	0.2065	0.8664	0.8439	0.7004	0.5818
GAT	BILSTM	**0.1096**	**0.1864**	**0.8928**	**0.8616**	**0.8127**	**0.6672**

Unlike conventional GCN-based models, which cannot focus on specific nodes or connections, the attention mechanism in GAT ensures that critical features within the drug structure are prioritized during feature extraction. This leads to improved molecular graph embeddings, as reflected in the superior performance metrics. Similarly, the use of BiLSTM for protein sequence processing capitalizes on its ability to capture long-term dependencies and contextual relationships within sequential data. This makes it particularly suited for encoding the amino acid sequences of proteins, where local and global sequence contexts are critical for understanding binding mechanisms. The results indicate that BiLSTM outperforms RNN and GRU models, which lack sufficient capacity to model such dependencies.

We also compared our prosed model with benchmark model from previous study, GraphDTA and MSGNN-DTA, as shown in Table 10. On the Davis dataset, our proposed model demonstrated the highest Concordance Index (CI) and r_m^2, showcasing its superior capability to rank drug-target binding affinities correctly and explain variance in binding affinity data. While the MSE of the proposed model (0.3209) was slightly higher than GraphDTA (0.3079), the substantial improvements in CI and r_m^2 underscore its enhanced predictive power in capturing complex relationships between drug and protein features. Similarly, on the KIBA dataset, the proposed model achieved the lowest MSE, the highest CI, and the highest r_m^2, further demonstrating its robustness and scalability in more diverse and complex datasets compared to the baseline models.

The superior performance of the proposed model can be attributed to several key factors. First, the use of GAT as the drug encoder allows the model to effectively capture complex molecular interactions by assigning adaptive attention weights to critical nodes and edges within molecular and motif-level graphs. While MSGNN-DTA combines GAT and GCN for feature learning, it is not enough to fully exploit the molecular structure and motif-level features due to the static nature of GCN, which limits its ability to capture dynamic relationships and complex dependencies within the graph. In contrast, GAT's adaptive attention mechanism provides a more flexible and expressive representation of these features, leading to enhanced performance. Second, the incorporation of BiLSTM as the protein encoder enables the model to learn long-term dependencies and contextual relationships within protein sequences, outperforming CNN-based protein encoding used in GraphDTA. For MSGNN-DTA, the use of a weighted protein graph adds additional layers of complexity without significantly improving the ability to capture sequential relationships. This increases computational overhead while providing only marginal gains in predictive accuracy. BiLSTM, on the other hand, effectively models both local and global dependencies in amino acid sequences with a lower computational cost compared to graph-based methods for protein representation. Third, the two-sided multi-head attention mechanism serves as a critical component in fusing drug and protein features, effectively modeling their interactions. By dynamically weighting the contributions of molecular graphs, motif graphs, and protein sequences, the mechanism captures the complex dependencies between drugs and their targets. This not only enhances prediction accuracy but also addresses the limitations of simple concatenation methods used in previous

studies, such as the fully connected layers employed by GraphDTA and the self-attention mechanisms used in MSGNN-DTA. The two-sided multi-head attention mechanism leverages the complementary strengths of the fused features, enabling the model to better learn the intricate relationships required for accurate drug-target affinity predictions.

Table 10. state-of-the-art benchmark comparison for Davis & KIBA dataset

Dataset	Model	Drug Encoder	Protein Encoder	Interaction Modelling	MSE ↓	CI ↑	r_m^2 ↑
Davis	GraphDTA	GCN	CNN	FC	**0.3079**	0.8498	0.4700
	MSGNN-DTA	GATGCN		Self AM	0.3301	0.8628	0.4444
	Proposed Model	GAT	BILSTM	TS-MH AM	0.3209	**0.8646**	**0.5046**
KIBA	GraphDTA	GCN	CNN	FC	0.2529	0.8176	0.5995
	MSGNN-DTA	GATGCN		**Self AM**	0.1916	0.8504	0.6185
	Proposed Model	GAT	BILSTM	TS-MH AT	**0.1864**	**0.8616**	**0.6672**

The prediction accuracy is further illustrated in Fig. 5. When the predicted value closely matches the actual value, it indicates better model performance, meaning the sample points should align near the diagonal line (green line). For the Davis dataset, the dense region of pK_d values fall between 5 and 6 on the X-axis, aligning with the actual data distribution. Similarly, the X-axis density of KIBA scores ranges from 10 to 14. In both datasets, the sample points are distributed close to the straight line (p = y. This further demonstrates the strong predictive performance of the integration of GAT and BiLSTM, coupled with the two-side multi-head attention mechanism, which facilitates effective feature fusion, allowing the model to capture complex interaction patterns between drugs and proteins with high precision.

However, these advantages come with the trade-off of increased computational complexity. The GAT layers, with their quadratic scaling relative to the number of graph nodes, and the two-side multi-head attention mechanism, with its demand for parallel processing across multiple attention heads, present significant challenges in terms of scalability. These challenges could impact the practical application of the model in large-scale drug discovery projects. Future work should explore optimization strategies, such as sparse attention mechanisms, graph sampling techniques, and hardware accelerations, to balance the computational demands while maintaining the model's superior predictive performance. Despite these limitations, the proposed model demonstrates its potential as a robust and effective tool for drug-target affinity prediction, particularly in scenarios where accuracy outweighs computational constraints.

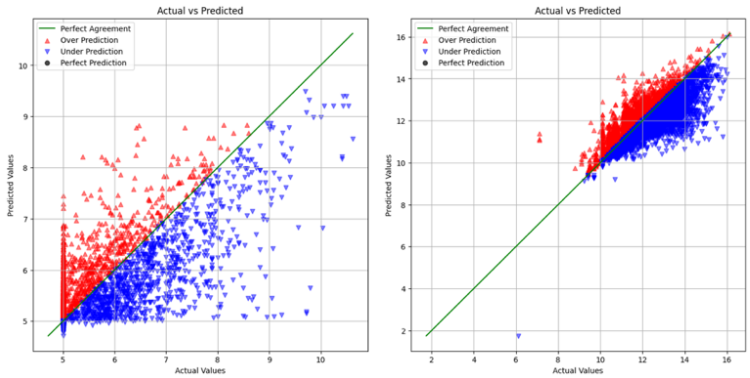

Fig. 5. Actual vs. Predicted Binding Affinities Values of Davis (left) and KIBA (right) Dataset

4. CONCLUSION

This study proposed a novel framework for predicting drug-target binding affinities using Graph Attention Networks (GAT), BiLSTM, with two-side multi-head attention mechanism. The proposed model achieved the lowest MSE, highest CI, and highest r_m^2 among all tested architectures for both datasets. These results confirm the model's effectiveness in capturing structural and sequential features of drugs and proteins, as well as their interactions. The integration of molecular and motif-level drug graphs with BiLSTM-encoded protein sequences provides comprehensive feature representations, while the multi-head attention mechanism enables robust feature fusion and improved predictive performance. The findings of this study demonstrate the potential of the proposed approach as a reliable computational tool for DTA prediction, reducing reliance on time-consuming and costly laboratory experiments.

Despite these promising results, this work acknowledges certain limitations. The evaluation was limited to the Davis and KIBA datasets, raising questions about the model's generalizability to other datasets or types of drug-target interactions. Additionally, the computational complexity of the GAT layers and multi-head attention mechanisms, while enabling high performance, poses challenges for scalability in large-scale drug discovery projects. Strategies such as sparse attention mechanisms, graph sampling, and hardware optimization should be explored to address these challenges. Future research should test the model across additional datasets with more diverse drug and targets representation, such as molecular fingerprints, protein subsequence's, and protein 3D structures, to assess its robustness and broader applicability. By extending this framework to incorporate additional features and optimize computational efficiency, future improvements could enhance the model's utility in real-world drug discovery processes, enabling faster and more accurate identification of potential therapeutic candidates.

ACKNOWLEDGMENT

We extend our gratitude to Telkom University for facilitating this research and providing administrative support and resource throughout this project. This research was funded by the Directorate of Research, Technology, and Community Service (DRTPM), under the Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology, as part of the DRTPM Academic Research Grant 2024 (Grant Number 106/E5/PG.02.00.PL/2024, 043/SP2H/RT-MONO/LL4/2024, 081/LIT07/PPM-LIT/2024)

REFERENCES

[1] K. Sachdev dan M. K. Gupta, "A comprehensive review of feature based methods for drug target interaction prediction," *J. Biomed. Inform.*, vol. 93, pp. 103159, Mei 2019, https://doi.org/10.1016/j.jbi.2019.103159.

[2] Y. Zhang, Y. Hu, N. Han, A. Yang, X. Liu, dan H. Cai, "A survey of drug-target interaction and affinity prediction methods via graph neural networks," *Comput. Biol. Med.*, vol. 163, pp. 107136, Sep 2023, https://doi.org/10.1016/j.compbiomed.2023.107136.

[3] A. Suruliandi, T. Idhaya, dan S. P. Raja, "Drug Target Interaction Prediction Using Machine Learning Techniques – A Review," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 8, no. 6, pp. 86, 2024, https://doi.org/10.9781/ijimai.2022.11.002.

[4] D. Butina, M. D. Segall, dan K. Frankcombe, "Predicting ADME properties in silico: methods and models," *Drug Discov. Today*, vol. 7, no. 11, pp. S83–S88, Mei 2002, https://doi.org/10.1016/S1359-6446(02)02288-2.

[5] E. Byvatov, U. Fechner, J. Sadowski, dan G. Schneider, "Comparison of Support Vector Machine and Artificial Neural Network Systems for Drug/Nondrug Classification," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 1882–1889, Nov 2003, https://doi.org/10.1021/ci0341161.

[6] H. Bhargava, A. Sharma, dan P. Suravajhala, "Chemogenomic Approaches for Revealing Drug Target Interactions inDrug Discovery," *Curr. Genomics*, vol. 22, no. 5, pp. 328–338, Des 2021, https://doi.org/10.2174/1389202922666210920125800.

[7] H. Li *et al.*, "TarFisDock: a web server for identifying drug targets with docking approach," *Nucleic Acids Res.*, vol. 34, pp. W219–W224, Jul 2006, https://doi.org/10.1093/nar/gkl114.

[8] A. C. Cheng *et al.*, "Structure-based maximal affinity model predicts small-molecule druggability," *Nat. Biotechnol.*, vol. 25, no. 1, pp. 71–75, Jan 2007, https://doi.org/10.1038/nbt1273.

[9] G. Pujadas *et al.*, "Protein-ligand Docking: A Review of Recent Advances and Future Perspectives," *Curr. Pharm. Anal.*, vol. 4, no. 1, pp. 1–19, Feb 2008, https://doi.org/10.2174/157341208783497597.

[10] W. Zhang, Y. Chen, dan D. Li, "Drug-Target Interaction Prediction through Label Propagation with Linear Neighborhood Information," *Molecules*, vol. 22, no. 12, pp. 2056, Nov 2017, https://doi.org/10.3390/molecules22122056.

[11] X. Zhang, L. Li, M. K. Ng, dan S. Zhang, "Drug–target interaction prediction by integrating multiview network data," *Comput. Biol. Chem.*, vol. 69, pp. 185–193, Agu 2017, https://doi.org/10.1016/j.compbiolchem.2017.03.011.

[12] Z. Shi dan J. Li, "Drug-Target Interaction Prediction with Weighted Bayesian Ranking," dalam *Proceedings of the 2nd International Conference on Biomedical Engineering and Bioinformatics*, pp. 19–24, Sep 2018, https://doi.org/10.1145/3278198.3278210.

[13] A. C. A. Nascimento, R. B. C. Prudêncio, dan I. G. Costa, "A multiple kernel learning algorithm for drug-target interaction prediction," *BMC Bioinformatics*, vol. 17, no. 1, pp. 46, Jan 2016, https://doi.org/10.1186/s12859-016-0890-3.

[14] 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, pp. 11174, Sep 2017, https://doi.org/10.1038/s41598-017-10724-0.

[15] M. Ohue, T. Yamazaki, T. Ban, dan Y. Akiyama, "Link Mining for Kernel-Based Compound-Protein Interaction Predictions Using a Chemogenomics Approach," *Intelligent Computing Theories and Application*, vol. 10362, pp. 549–558, 2017, https://doi.org/10.1007/978-3-319-63312-1_48.

[16] Y.-B. Wang, Z.-H. You, S. Yang, H.-C. Yi, Z.-H. Chen, dan 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, pp. 49, Mar 2020, https://doi.org/10.1186/s12911-020-1052-0.

[17] L. Xu, X. Ru, dan R. Song, "Application of Machine Learning for Drug–Target Interaction Prediction," *Front. Genet.*, vol. 12, pp. 680117, Jun 2021, https://doi.org/10.3389/fgene.2021.680117.
[18] N. R. C. Monteiro, B. Ribeiro, dan J. P. Arrais, "Drug-Target Interaction Prediction: End-to-End Deep Learning Approach," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 18, no. 6, pp. 2364–2374, Nov 2021, https://doi.org/10.1109/TCBB.2020.2977335.
[19] C. Chen, H. Shi, Y. Han, Z. Jiang, X. Cui, dan B. Yu, "DNN-DTIs: improved drug-target interactions prediction using XGBoost feature selection and deep neural network," *Computers in Biology and Medicine*, vol. 136, p. 104676. 2021, https://doi.org/10.1101/2020.08.11.247437.
[20] S. M. Hasan Mahmud, W. Chen, H. Jahan, B. Dai, S. U. Din, dan A. M. Dzisoo, "DeepACTION: A deep learning-based method for predicting novel drug-target interactions," *Anal. Biochem.*, vol. 610, pp. 113978, Des 2020, https://doi.org/10.1016/j.ab.2020.113978.
[21] G. Liu *et al.*, "GraphDTI: A robust deep learning predictor of drug-target interactions from multiple heterogeneous data," *J. Cheminformatics*, vol. 13, no. 1, pp. 58, Agu 2021, https://doi.org/10.1186/s13321-021-00540-0.
[22] H. Öztürk, A. Özgür, dan E. Ozkirimli, "DeepDTA: deep drug–target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, Sep 2018, https://doi.org/10.1093/bioinformatics/bty593.
[23] H. Öztürk, E. Ozkirimli, dan A. Özgür, "WideDTA: prediction of drug-target binding affinity," *arXiv*: arXiv:1902.04166, 2019, https://doi.org/10.48550/arXiv.1902.04166.
[24] Y. Pu, J. Li, J. Tang, dan F. Guo, "DeepFusionDTA: Drug-Target Binding Affinity Prediction With Information Fusion and Hybrid Deep-Learning Ensemble Model," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 19, no. 5, pp. 2760–2769, Sep 2022, https://doi.org/10.1109/TCBB.2021.3103966.
[25] W. Yuan, G. Chen, dan C. Y.-C. Chen, "FusionDTA: attention-based feature polymerizer and knowledge distillation for drug-target binding affinity prediction," *Brief. Bioinform.*, vol. 23, no. 1, pp. bbab506, Jan 2022, https://doi.org/10.1093/bib/bbab506.
[26] A. Ghimire, H. Tayara, Z. Xuan, dan K. T. Chong, "CSatDTA: Prediction of Drug–Target Binding Affinity Using Convolution Model with Self-Attention," *Int. J. Mol. Sci.*, vol. 23, no. 15, pp. 8453, Jul 2022, https://doi.org/10.3390/ijms23158453.
[27] L. Deng, Y. Zeng, H. Liu, Z. Liu, dan X. Liu, "DeepMHADTA: Prediction of Drug-Target Binding Affinity Using Multi-Head Self-Attention and Convolutional Neural Network," *Curr. Issues Mol. Biol.*, vol. 44, no. 5, pp. 2287–2299, Mei 2022, https://doi.org/10.3390/cimb44050155.
[28] Q. Zhao, F. Xiao, M. Yang, Y. Li, dan J. Wang, "AttentionDTA: prediction of drug–target binding affinity using attention model," dalam *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 64–69, 2019, https://doi.org/10.1109/BIBM47256.2019.8983125.
[29] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J. B. Ghasemi, dan 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, Nov 2020, https://doi.org/10.1093/bioinformatics/btaa544.
[30] A. Mahdaddi, S. Meshoul, dan M. Belguidoum, "EA-based hyperparameter optimization of hybrid deep learning models for effective drug-target interactions prediction," *Expert Syst. Appl.*, vol. 185, pp. 115525, Des 2021, https://doi.org/10.1016/j.eswa.2021.115525.
[31] Q. Zhao, G. Duan, M. Yang, Z. Cheng, Y. Li, dan J. Wang, "AttentionDTA: Drug–Target Binding Affinity Prediction by Sequence-Based Deep Learning With Attention Mechanism," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 20, no. 2, pp. 852–863, Mar 2023, https://doi.org/10.1109/TCBB.2022.3170365.
[32] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, dan S. Venkatesh, "GraphDTA: predicting drug–target binding affinity with graph neural networks," *Bioinformatics*, vol. 37, no. 8, pp. 1140–1147, Mei 2021, https://doi.org/10.1093/bioinformatics/btaa921.
[33] S. Zhang, M. Jiang, S. Wang, X. Wang, Z. Wei, dan Z. Li, "SAG-DTA: Prediction of Drug–Target Affinity Using Self-Attention Graph Network," *Int. J. Mol. Sci.*, vol. 22, no. 16, pp. 8993, Agu 2021, https://doi.org/10.3390/ijms22168993.
[34] Y. Liang, S. Jiang, M. Gao, F. Jia, Z. Wu, dan Z. Lyu, "GLSTM-DTA: Application of Prediction Improvement Model Based on GNN and LSTM," *J. Phys. Conf. Ser.*, vol. 2219, no. 1, pp. 012008, Apr 2022, https://doi.org/10.1088/1742-6596/2219/1/012008.
[35] A. Banerjee, Z.-H. Zhou, E. E. Papalexakis, dan M. Riondato, Ed., *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2022. https://doi.org/10.1137/1.9781611977172.
[36] P. Chen, H. Shen, Y. Zhang, B. Wang, dan P. Gu, "SGNet: Sequence-Based Convolution and Ligand Graph Network for Protein Binding Affinity Prediction," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 20, no. 5, pp. 3257–3266, Sep 2023, https://doi.org/10.1109/TCBB.2023.3262821.
[37] W. Qiu, Q. Liang, L. Yu, X. Xiao, W. Qiu, dan W. Lin, "LSTM-SAGDTA: Predicting Drug-target Binding Affinity with an Attention GraphNeural Network and LSTM Approach," *Curr. Pharm. Des.*, vol. 30, no. 6, pp. 468–476, Feb 2024, https://doi.org/10.2174/0113816128282837240130102817.
[38] H. N. T. Tran, J. J. Thomas, dan N. H. Ahamed Hassain Malim, "DeepNC: a framework for drug-target interaction prediction with graph neural networks," *PeerJ*, vol. 10, pp. e13163, Mei 2022, https://doi.org/10.7717/peerj.13163.

[39] S. Wang *et al.*, "MSGNN-DTA: Multi-Scale Topological Feature Fusion Based on Graph Neural Networks for Drug–Target Binding Affinity Prediction," *Int. J. Mol. Sci.*, vol. 24, no. 9, pp. 8326, Mei 2023, https://doi.org/10.3390/ijms24098326.
[40] M. I. Davis *et al.*, "Comprehensive analysis of kinase inhibitor selectivity," *Nat. Biotechnol.*, vol. 29, no. 11, pp. 1046–1051, Nov 2011, https://doi.org/10.1038/nbt.1990.
[41] 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, Mar 2014, https://doi.org/10.1021/ci400709d.
[42] B. Ramsundar, P. Eastman, P. Walters, dan 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. https://books.google.co.id/books?hl=id&lr=&id=6OiRDwAAQBAJ.
[43] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, dan Y. Bengio, "Graph Attention Networks," *arXiv*: arXiv:1710.10903, 2024, http://arxiv.org/abs/1710.10903.
[44] S. Hochreiter dan J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov 1997, https://doi.org/10.1162/neco.1997.9.8.1735.
[45] M. Gönen dan G. Heller, "Concordance probability and discriminatory power in proportional hazards regression," *Biometrika*, vol. 92, no. 4, pp. 965–970, Des 2005, https://doi.org/10.1093/biomet/92.4.965.
[46] K. Roy, P. Chakraborty, I. Mitra, P. K. Ojha, S. Kar, dan 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, Mei 2013, https://doi.org/10.1002/jcc.23231.
[47] T. Pahikkala *et al.*, "Toward more realistic drug-target interaction predictions," *Brief. Bioinform.*, vol. 16, no. 2, pp. 325–337, Mar 2015, https://doi.org/10.1093/bib/bbu010.
[48] T. N. Kipf dan M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," *arXiv*: arXiv:1609.02907, 2017, http://arxiv.org/abs/1609.02907.
[49] D. E. Rumelhart, G. E. Hinton, dan R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Okt 1986, https://doi.org/10.1038/323533a0.
[50] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," *arXiv*: arXiv:1406.1078, 2024. [Daring], http://arxiv.org/abs/1406.1078.

BIOGRAPHY OF AUTHORS

Muhammad Rizky Yusfian Yusuf, is currently pursuing informatics master's degree in school of computing, Telkom university, Bandung, Indonesia. He received his bachelor's degree at Surabaya university, Surabaya, Indonesia. He is interested in the fields related to topics such as artificial intelligence, machine learning, and software engineering. Email: rzkysfian@student.telkomuniversity.ac.id

Isman Kurniawan holds a master's degree in computational science from Kanazawa University and Institut Teknologi Bandung, and a Doctoral degree in Computational Science from Kanazawa University. He is currently a researcher and academic staff member at the School of Computing, Telkom University. His research focuses on machine learning, computational chemistry, and modeling and simulation, with a particular interest in applying artificial intelligence to drug discovery and disease detection. Email: ismankrn@telkomuniversity.ac.id