Power System Technology
ISSN:1000-3673

Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

Drug-Target Interaction Prediction Using Relative Multi-Head Self-Attention and Graph Attention Exchange Network

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

[1] Research Scholar (Registration No-22121242282005),Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous),
Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India
Email: jane77johnson@gmail.com

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

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

[4] Pharm D, Affiliated to Annamalai University, Chidambaram, Tamilnadu, India
Email: serinahingis6@gmail.com

Abstract

Finding drug-target interactions will drastically limit the number of candidate medications that need to be looked for, making it the crucial initial stage in the drug discovery process. The number of biochemical experiments that can be undertaken is significantly constrained by the use of DTIs in the choice of potential drugs. They can also provide genuine insight on the health consequences and operating principles of drugs. Even now, research designs for finding drug-target interactions are nevertheless cost prohibitive, time-consuming, and challenging despite the availability of various biological techniques and elevated testing. As a result, numerous digital patterns have been created to broadly forecast possible drug-target relationships. This work proposes an Attention Exchange network of graph and encoder model for DTI (RMHSA_GAEN). First, a GNN receives pre-processed molecular graphs from SMILES from RDKit of the drug compounds. Similarly encoder relying on a relative multi-head self-attention is used extract the feature vectors of the protein sequences in the form of n-grams. The proposed concept RMHSA_GAEN exchange of its individual attention weights to improve the feature correlation and fine-tune the output from the graph network and Relative Multi Head Attention Transformer blocks. This model achieved accuracy 95.16% for the Human Dataset and 96.55% for the C.Elegans Dataset during testing on those datasets.

Keywords- Multi-Head Self-Attention, Drug-Target, Transformer, GNN, Encoder

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

1. Introduction

The current drug discovery techniques depend on Drug Target Interactions (DTIs) and it is vital in drug research and development (R&D). Finding drug indications, drug–drug interactions, adverse drug responses, and drug mechanism of action can all be inferred by identifying the drug's interaction with the target. Predicting DTIs has been widely used for drug repositioning and anticipating adverse reactions [1]. The concept of repurposing medications to novel indications may not only lower drug research costs but also diminish the threat associated with drug safety because pharmaceutical drugs have obvious accessibility and known security assessments [2]. Despite the availability of numerous biological test tools, huge drug–target interaction investigations still have their drawbacks. Additionally, the expensive nature of the test and the dearth of publicly accessible drug reposition assay data need the creation of adequate analytical models that could properly identify drug-target interactions. There have been lot of works on drug-target interactions introduced, involving machine learning and deep learning. Several silico approaches to discovering new drug-target interactions have been proposed (DTI). The two most common approaches are ligand- and structure-based [3]. Quantitative structure activity relationship (QSAR) is the commonly used ligand technique for predicting a molecule's biocompatibility on a goal. According to the QSAR hypothesis, compounds with comparable structural characteristics have equivalent bioactivity [4]. Recently, a number of network techniques for predicting drug - related problems have been put forth. In a drug-target interaction network, medications and targets are symbolized by nodes, and known interactions between them are represented by the lines connecting the nodes. The current network serves as the foundation for the future DTIs. For instance, Cheng et al. [5] developed an inference model based on network (NBI) to estimate fresh DTIs. Drug-target bipartite network topology similarity is the sole basis for NBI. Using a score function, the relationship between a medicine and a target was graded. NBI has the drawback of not being applicable to new medications unless the training set includes known target information.

2. Literature Review

An ANN with numerous hidden units and a more intricate parameter training process is referred to as deep learning. Due to its comparatively improved performance and capacity to generate visualisations with several levels of abstraction for a range of regression problems, deep learning is becoming more and more popular [6].

Deep learning (DL) models recently exhibit promising DTI prediction performance. The ability of deep learning approaches to predict DTIs has also been demonstrated [7, 8, 9, 10, 11, 12], often outperforming more conventional machine learning techniques. However, both computer scientists who are new to the biomedical sector and bioinformaticians with hardly DL knowledge may find it challenging to use these models. Predicting how

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

medications will interact with their target proteins is crucial for drug research and development.

Traditional experimental paradigms, on the other hand, are expensive, and earlier in silico prediction paradigms have been constrained by the diversity of data platforms and the dearth of data[13]. Although there have been significant developments in deep learning technology, it is still interesting to look into how to create effective sequence based networks for DTI prediction.

Numerous biological and chemical sectors have made use of deep learning. To anticipate the sequence particularities of DNA- and RNA-binding proteins, Frey et al. [14], for instance, modified the Deep Bind deep-learning technique. Even when taught on in vitro data and tested on in vivo data, deep learning beats conventional techniques. A Deep-learning Network technique (DN-Fold) developed by Cheng et al. [15] greatly enhanced protein fold analysis.

To find new DTIs, Nascimento et al. [16] used a variety of heterogeneous sources of news. There has recently been a tendency toward integrating many resources to learn more about DTIs, thanks to the advent of numerous experimental tools and technologies, including high-throughput studies and next-generation sequencing. Without identifying the targets, the Deep Belief Network (DBN) was employed as an efficient deep-learning technique to predict novel DTIs between FDA-approved medicines and targets. Wen et al. created the innovative predictor DeepDTIs to foresee potential drug-target interactions [17].

Farshid et al. suggested FRnet-Encode to retrieve 4096 features and FRnet-Predict to forecast drug-target interactions relying on a deep CNN [18]. Hu et al. created a CNN method to identify drug and target interactions. To train the CNNs model, the concatenated descriptors from drugs and target proteins were randomly projected into a lower-dimensional subspace. They also presented a reasonable methodology for selecting negative instances, which reduces the possibility of false negative instances and thus significantly improves predictive performance [19]The amount of hyperparameter modifications is the main reason limiting deep learning applications. The growing dimensionality of features in DTIs, even with deep learning, is another difficulty.

Yamanishi et al. [20] employed the kernels technique to forecast DITs by fusing together several pieces of data and gauging how similar medications and targets are. The benefit of integrating several resources is that it allows for the detection of data that either other resources do not retain or that was not represented by the calculated characteristics. However, this can lead to a mistake if the resources weren't reliable. Furthermore, to extract the features from all these resources, further domain knowledge is needed. It can be challenging to choose the best descriptor for a particular task when there are numerous chemical composition and sequence - based descriptors accessible, for instance [21].

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

Due to its capacity to organise and extract information from unstructured data as well as to acquire data with different levels of abstraction, deep learning is a particularly effective tool for DTI prediction. A DTI forecasting approach on drug-target networks, drug structural similarity, and protein sequence similarity was presented by Eslamiet et al. [22]. It has the benefit of automatically picking up topological aspects from the DTI network and obtaining more beneficial relational information. Additionally, complicated topological properties that heuristics cannot represent are learned using neural networks.

In order to recognize localized residue characteristics of protein classifications by convolution on varying lengths of amino acid sub-sequences, Lee et al. [23] utilised a CNN on unprocessed protein sequences. They used a vast amount of DTI data to train the model, then they used a different dataset to show how well it performed. For large-scale DTI prediction, it performs better than recently established deep learning algorithms and earlier protein descriptor-based approaches. It uses pooled convolution data to find protein binding locations for DTIs. In order to identify commonly produced antiviral medications that may be able to interfere with the viral components of SARS-CoV-2, Beck et al. [24] employed a MT-DTI model and it quickly add target proteins, such as the SARS-CoV-2 viral proteins, to the model. Chu et al. [25] offer the cascade deep forest (CDF) model DTI-CDF, which incorporates a number of similarity-based features between medicines and target proteins that are retrieved from a heterogeneous graph comprising known DTIs.

The construction of DPP network with DPPs as nodes and relationships among DPPs as network edges, based on various medications and proteins were employed [26]. A framework based on learning is called "Graph Convolutional Network-DTI." In order to determine the ultimate label that use the visual features as input, it first uses a GNN followed by a dnn. [27] proposes DeepFusion, a multi-scale feature fusion approach for DTI prediction. In order to create localized chemical sub-structure semantic characteristics for both drugs and proteins, researchers used transformer networks along with similarity theory, and CNN. In a case study on forecasting potential DTIs, DeepFusion produces encouraging prediction results.

The MINN-DTI approach enhances DTI forecasting by extracting the reciprocal effects among targets and drugs, that are expressed by proteins length maps and molecular graph [28]. A newly developed Interformer and an enhanced CMPNN were put together to create a MINN, which was used to capture the context of how medicines and targets interact (dubbed Inter-CMPNN). Zhang et al. [29] present a DeepMGT-DTI, transform network combining multilayer graph knowledge for Drug-Target interaction prediction. This model includes a transformer network that incorporates multilayer graph knowledge to capture the

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

properties of a medication's molecular structure, enabling a more thorough investigation of the interactions between the atoms of pharmacological molecules. Mehdi, et al. [30] suggested GCNN-based system on a self-attention framework for trying to capture any relationship between binding of a specific protein and the drug in a series comparable to a sentence with relationship definition among its biochemical entities.

3. Proposed Attention Exchange Network for Drug Target Interaction (DTI) Prediction

This work proposes a deep learning cross attention architecture to precisely predict the Drug-Target Interactions (DTIs). Our approach moves beyond simple concatenation by establishing a dedicated, bidirectional information exchange between the heterogeneous inputs. We utilize a specialized Graph Neural Network (GNN) to capture the complex, non-sequential features of the drug molecule, and a Multi-Head Self-Attention (MHSA) network to encode the long-range dependencies within the protein's amino acid sequence. The exchange attention mechanism allows contextually-aware representations that explicitly model their binding compatibility. Finally, these enriched features are aggregated to deliver a highly accurate and interpretable DTI prediction. The following diagram (Figure 1.) shows the overall architecture of the proposed model. The protein and drug subsequences of the proposed work are explained in the following sub sections.

3.1 Protein Sequence Processing Network

3.1.1 Pre-processing and Position Encoding

The "words" of a "protein sequence language" have been examined as a variety of fundamental protein sequence building blocks, such as patterns, motifs, and n-grams The term "n-grams" refers to all possible amino acid sub-sequences of a certain length. To describe the "word" in sequences of amino acids in this work, we followed [31]'s advice and employed the n-gram. By breaking up sequence data into overlaying n-gram amino acids, we may get the vector representation of an amino acid sequence, $P = \{p_1, p_2,...,p_{|P|}\}$, where $pi \in R^d$ = embedding vector of i-th word and d = embedding dimension. The model lacking both recurrence and convolution cannot fully utilise the sequence's order in the input portion of the data. This issue is resolved by using a "position encoding" to learn the relative positions of each item in the sequence. The final step is to combine the position encoding and "word" embedding vector as the input to the encoder for subsequent operations.

Power System Technology
ISSN:1000-3673

Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

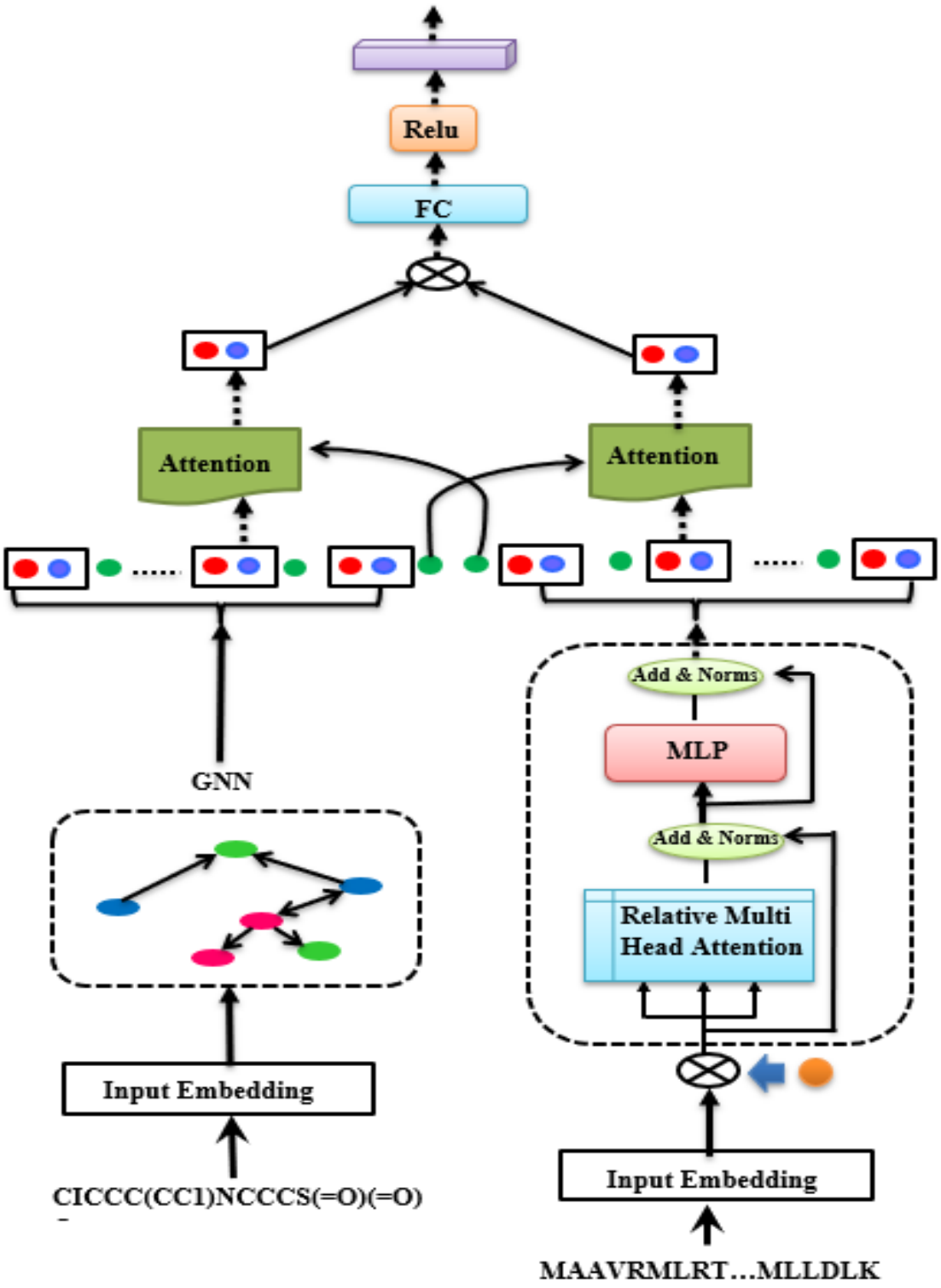

Figure 1. Architecture of Proposed RMHSA_GAEN DTI Prediction Model

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

3.1.2 Relative Multi-head Attention

Deep learning tasks involving contextual interactions, such as machine translation and semantic understanding, have frequently exploited the self-attention mechanism [32, 33]. The ability to simultaneously respond to data from several representational subsets at various places makes multi-head attention interesting [34]. A significant advancement in machine translation is made possible by using self-attention rather than sequence-aligned RNNs or CNNs to determine how its inputs and outputs are expressed. To produce an utterance level representation, this technique functions as a pooling layer that selects the most discriminative features from the sequence [35].

In this part, we employ the Transformer's encoder structure to extract the features and distinctive information of the protein sequences. Along Relative multi-head self-attention mechanism's effectiveness and resilience, this structure also addresses certain crucial issues with CNN's inability to gather contextual information in the sequences. The visualization outcomes taught by the attention mechanism are displayed in the experimental section to demonstrate the efficiency of RMHSA_GAEN.

3.1.3 Encoder based on the Multi-head Self-Attention Mechanism

In order to make the reasons in our proofs simpler, the transformer model with Relative Multi Head Attention model is adapted for protein sequence analysis. The input vectors pre-processed in Section 2.2.1 is given as input to the transformer model. From the embedding vectors, Query, Key, and Value representations are obtained through mapping, each set of amino acids, $Q = \{ q^T_1 , q^T_2 , ..., q^T_L \} \{ , K = k^T_1 , k^T_2 , ..., k^T_L \} , V = \{ v^T_1 , v^T_2 , ..., v^T_L \}$, where L stands for the sequence's length and $q^T_i \in Rd$, $k^T_i \in Rd$, $v^T_i \in Rd$. Following processing for self-attention, the result of the i-th position is finally available.

$$Attention (q_i, K, V) = softmax (q_i K^T \sqrt{d}) V \quad (6)$$

The model must continue to be stable and valid in order to incorporate the relative multi-attention learning outcomes as well. It has the capacity to learn the diverse semantic data of the elements employing diverse self-attention in various spaces scale matrices. Finally, as output, the relative multi-attention mechanism's findings are concatenated.

$$head\__i = Att (LWL_i , MWM_i , NWN_i) \quad (7)$$

$$MultiHead (L, M, N) = Concatination(head\__1, ..., head\__h) Wo \quad (8)$$

Where $WL_i \in Rd \times dk$, $WM_i \in Rd \times dk$, $WN_i \in Rd \times dv$ and $Wo \in Rh \bullet dv \times d$ was utilized for feature space transformation. The vector concatenated operation is (•). A fully linked forward-looking network featuring two linear transformation structures that are applied

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

evenly and individually to every position in the sequence makes up the second part of the encoder.

$$MLP\ (x) = W_2\ (relu\ (W1_x + b_1)) + b_2 \quad (9)$$

W1 = dinner where dinner, W2, and Rd are transformations parameters, the activation function is Relu, and the output is x from the encoder's 1st segment. In along with the two components of the encoder, the residual connection structure [36] and the layer normalisation structure [37] are implemented as the input & output for the majority of network topologies. The characteristics of every amino acid sequence can then be determined by understanding the k-layer encoder architecture, wherein k is a hyper-parameter.

3.1.4 Output: Protein Vector Representation

We apply the average of C to produce the ultimate output $y_{protein} \in R^d$ deriving out from the set of concealed vectors $C = \{C_1^{(t)}, C_2^{(t)}, \ldots, C_{|c|}^{(t)}\}$ as follows.

$$y_{protein} = \frac{1}{|C|}\sum_{i=1}^{|c|} ci(t) \quad (10)$$

3.2 Drug Sequence Processing

3.2.1 Graph Neural Network for Molecular Graph

According to Scarselli et al. [43], a GNN converts a graph G into a vector $y\ \epsilon\ R^d$ by employing 2 purposes, specifically transitional and yielded functions. An acronym for a graph is $G = (Vs, Es)$, where Vs stands for the group of vertices and Es for the group of edges. $U_i \in Vs$ is the ith atom of a drug, while $e_{ij} \in Es$ is the chemical bond connecting the ith and jth atoms.In more detail, collection of every neighbouring vertex indices inside radius ra from the ith vertex as $Neg\ (i, ra)$ for given a graph $G = (Vs, Es)$ is defined. Consider the fact that $Neg\ (i,0) = \{i\}$. The ra-radius subgraph for vertex U_i is therefore defined as,

$$U_i^{(ra)} = (Vs_i^{(ra)}, Es_i^{(ra)}) \quad (11)$$

Where,

$$Vs_i^{(ra)} = \{U_j \mid j \in Neg\ (i, ra)\} \quad (12)$$

$$Es^{(ra)}{}_i = \{e_{mn} \in \mid (m,n) \in Neg\ (i, ra) \times Neg(i, ra\text{-}1)\} \quad (13)$$

U_i^{ra} is referred to as the r-radius vertex in this work. The ra-radius subgraph for edge e_{ij} is defined as follows:

$$e_{ij}^{(ra)} = (Vs_i^{(ra-1)} \cup Vs_j^{(ra-1)}, Es_i^{(ra)}, \cap\ Es_j^{(ra)}) \quad (14)$$

The edge transition function, also referred to as the edge transition, is also taken into consideration because edges are also characterised by vector embeddings.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

The *i*th vertex immersing at timing *ti* is described as $v_i^{(ti)} \in R^d$ given a graph *G* and the randomized generated implants of vertices and edges. Then, adjusting the $v_i^{(t)}$ just utilizing the transition function shown below:

$$v_i^{(ti+1)} = \sigma\left(v_i^{(ti)} + \sum_{j \in N(i)} h_{ij}^{(ti)}\right) \quad (15)$$

Where the sigmoid function is represented by σ: *Neg (i),* $\sigma(x) = 1/(1+e^{-x})$ is the collection of *i*'s nearby indices, $h_{ij}^{(ti)} \in R^d$ is the concealed neighbourhood vector. The preceding neural network can be used to compute this hidden vector by taking into account the surrounding vertex u_j and edge e_{ij}.

$$h_{ij}^{(ti)} = f\left(Weight_{neighbor}\begin{bmatrix} V_j^{(ti)} \\ e_{ij}^{(ti)} \end{bmatrix} + bios_{neighbour}\right) \quad (16)$$

where *f* is an activation function that is not linear, like *ReLU*. The weight matrix is $Weight_{neighbor} \in R^{d \times 2d}$, the bios vector is $bios_{neighbor} \in R^d$ and $e_{ij}^{(ti)} \in R^d$ is the edge wrapping among both the *i*th and *j*th vertices at time *ti*.

Edge embeddings can likewise be processed using the aforementioned iterative method in a similar way. Here, considering the two side vertex embeddings $v_i^{(ti)}$ and $v_j^{(ti)}$, we update $e_{ij}^{(ti)}$ as follows:

$$e_{ij}^{(ti+1)} = \sigma(e_{ij}^{(ti)} + g_{ij}^{(ti)}) \quad (17)$$

$$g_{ij}^{(ti)} = f(Weight_{side}\,(v_i^{(ti)} + vj^{(ti)}) + bios_{side}) \quad (18)$$

where $Weight_{side} \in R^{d \times d}$ *is the weight matrix and* $bios_{side} \in R^d$ is the bios Vector.

3.2.2 Output: Molecular Vector Representation

We apply the average of the vertex vectors to get the last output, $Y_{drug} \in 2\ R^d$, deriving out of the collection of vertex vectors given by the transition function, i.e. $V = \{V_1^{(ti)}, V_2^{(ti)}, \ldots, V_{|ti|}^{(ti)}\}$.

$$Y_{drug} = \frac{1}{|v|}\sum_{i=1}^{|v|} vi(t) \quad (19)$$

Where |v| is the no. of vertices in the molecular graph.

3.3 Interaction between Drug and Protein with Attention Exchange Interaction

To weight for $c_i^{(ti)}$ and $d_i^{(ti)}$ while taking into account a drug vector Y_{drug} and protein vector $Y_{protein}$. The given collection of hidden vectors of sub-sequences in a protein is C = $\{c_1^{(ti)}, c_2^{(ti)} \ldots, c_{|c|}^{(ti)}\}$ and collection of hidden vectors of sub-sequences in a drug is D = $\{d_1^{(ti)}, d_2^{(ti)} \ldots, d_{|d|}^{(ti)}\}$. To put it another way, to calculate which protein sub-sequences are more crucial for the drug by giving the sub-sequences heavier weights and vice-versa.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

Specifically, the dot product scalar values are calculated as the weights given y_{drug} and $c_i^{(ti)}$ as follows,

$$h_{drug} = f(Weight_{inter}\, Y_{drug} + b_{inter}), \quad (20)$$

$$h_{ic} = f(Weight_{inter}\, c_i + b_{inter}), \quad (21)$$

$$\alpha_{ic} = \sigma(h_{drug}\, h_{ic}), \quad (22)$$

where b_{inter} is the bias vector and $Weight_{inter}$ is the weight matrix. One can presume that the weight value α_i, called attention, represents the degree of connection between a drug and a protein subsequence. The weighted sum of h_{ic} using the attention weights is obtained as follow,

$$Y_{protein} = \sum_{i=1}^{|c|} \alpha_{ic}\ h_{ic} \quad (23)$$

Similarly, the dot product scalar values are calculated as the weights given $y_{protein}$ and $d_i^{(t)}$ as follows:

$$h_{protein} = f(Weight_{inter}\, Y_{protein} + b_{inter}), \quad (24)$$

$$h_{id} = f(Weight_{inter}\, d_i + b_{inter}), \quad (25)$$

$$\alpha_{id} = \sigma(h_{protein}\, h_{id}), \quad (26)$$

The weighted sum of h_{id} using the attention weights is obtained as follow,

$$Y_{drug} = \sum_{i=1}^{|d|} \alpha_{id}\ h_{id} \quad (27)$$

3.4 The Pseudo code

Input: Drug Dataset and Protein Dataset

Output: Predicted Class

Begin

// Convert SMILES into molecular graph to preprocess the drug dataset

For each drug in Drug Dataset:

Initialize by converting SMILES to molecular graph (G)

// Convert sequence into n-gram tokens to preprocess the protein dataset

Generate numerical embedding for each n-gram

// Drug feature extraction using GNN (Graph Neural Network)

For each Molecular Graph G:

Generate feature vector D using low dimension

Apply message passing layers

// Protein feature extraction using Transformer encoder with RMHSA

For each protein sequence embedding:

Generate feature vector P using low dimension

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

// Perform feature fusion with attention mechanism

Combine D and P

Fine-tune combined representation using weights

// Prediction model for drug-target classification

Input: Combined feature vector of Drug and Protein dataset

Output: Predicted Target Class either having interaction or not.

// Performance evaluation

Evaluate the model using standard performance metrics discussed in Section 4.2

End

4. EXPERIMENTAL RESULT AND DISCUSSION

The experimental analysis and its findings are discussed in this section. Two open datasets were analysed to demonstrate that our proposed model performs superior to conventional approaches.

4.1 Dataset Description

4.1.1. Human Dataset[42]

A total of 33984 sequences, 3,369 positive interactions between 1,052 distinct chemicals & 852 distinct proteins, and 33984 sequences make up the Human Dataset. There are 27187 and 6797 sequences in the training and testing sets, respectively.

4.1.2. C.Elegans Dataset[42]

Totalling 4800 sequences, the C.elegans Dataset also includes 4,000 positive interactions between 1,434 distinct chemicals and 2,504 distinct proteins. There are 3840 and 960 sequences in the training and testing sets, respectively.

4.2 Evaluation Protocol and Metrics

In this proposed DTI prediction work, the confusion matrix plays a critical role in evaluating the model's classification performance. It provides a detailed breakdown of the model's ability to correctly and incorrectly predict drug-target interactions. The matrix consists of four key components: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). True Positives represent the number of instances where the model correctly predicts an actual drug-target interaction, while True Negatives indicate the correct identification of non-interacting pairs. False Positives occur when the model predicts an interaction that does not actually exist, and False Negatives represent missed interactions where the model fails to identify true interacting pairs. Table 1 shows the confusion matrix for Human and C.Elegans Dataset.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

Actual \ Predicted	Predicted 0	Predicted 1	Total	Actual \ Predicted	Predicted 0	Predicted 1	Total
Actual 0	353	9	362	**Actual 0**	274	11	285
Actual 1	17	372	389	**Actual 1**	19	317	336

Table 1 Confusion matrix for Human & C.Elegans Dataset

By analyzing these four elements, important performance metrics such as Accuracy, Precision, Recall, and F1-Score are calculated. Particularly in DTI prediction, Recall is crucial because failing to predict true interactions may result in overlooking potentially valuable drug candidates.

Accuracy

This shows how well the classified images performed overall. The effectiveness of the model is assessed by comparing the proportion of accurate predictions to the total cases.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (28)$$

Precision

A total of positive samples is correctly or wrongly identified as positive depending on the proportion of actually positive patterns

$$Precision = \frac{TP}{TP + FP} \quad (29)$$

Recall

Calculating the percentage of positive patterns that are correctly categorised requires the measurement of recall.

$$Recall = \frac{TP}{TP + FN} \quad (30)$$

F1_Score

The term used to identify the harmonic mean values between recall and precision values is F1 score.

$$F1_{score} = \frac{2 * Precision * Recall}{Precision + Recall} \quad (31)$$

AUC

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

Area Under the Curve, which summarises the ROC curve, examines how well a classifier can distinguish between classes. The higher the AUC, the better the model does at separating the positive and negative groups.

AUPR

The area under the precision-recall curve is the likelihood that, if a "positive" edge is chosen from the method's ranked list, an edge on the list immediately above it will also be "positive." Alternatively, it is the average precision across all recall values. The following Table 2 displays the outcomes of the proposed RMHSA_GAEN and a comparison with earlier studies.

Methods	Accuracy	AUC	Precision (std)	Recall (std)	AUPR	F1_Score (std)
RWR [38]	-	83.75	77.07%	72.43%	81.65	74.66%
DrugE-Rank [39]	-	85.62	71.81%	86.68%	82.57	78.51%
DeepConv-DTI [40]	-	97.38	92.95%	91.75%	94.37	92.04%
DeepCPI [41]	-	96.92	91.87%	92.10%	93.99	90.96%
MHSADTI [42]	94.52%	98.22	94.72%	93.65%	95.68	93.46%
RMHSA_GAEN	**95.17%**	**98.73**	**95.08%**	**95.24%**	**96.37**	**95.16%**

Table 2. Performance Analysis of RMHSA_GAEN on Human Database

The outcomes demonstrate that RMHSA GAEN outperforms the other algorithms, indicating that it is effective in detecting images. In terms of Accuracy, AUC, Precision, Recall, and AUPR, F1 Score, our technique performs at 95.17%, 98.73%, 95.08%, 95.24%, 96.37%, and 95.16%, respectively.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

In terms of Precision, RMHSA_GAEN gives 0.36% greater than MHSADTI [42], 3.21% greater than DeepCPI [41], 2.13% greater than Deepconv-DTI [40], 23.27% greater than DrugE-Rank [39], and 18.01% greater than RWR [38].

In terms of Recall, RMHSA_GAEN gives 1.59% greater than MHSADTI [42], 3.14% greater than DeepCPI [41], 3.49% greater than Deepconv-DTI [40], 8.56% greater than DrugE-Rank [39], and 22.81% greater than RWR [38].

In terms of F1_Score, RMHSA_GAEN gives 1.70% greater than MHSADTI [42], 4.2% greater than DeepCPI [41], 3.12% greater than Deepconv-DTI [40], 16.65% greater than DrugE-Rank [39], and 20.5% greater than RWR [38]. Figure 2 illustrates the comparative result analysis of the existing and proposed methodologies using the Human Dataset.

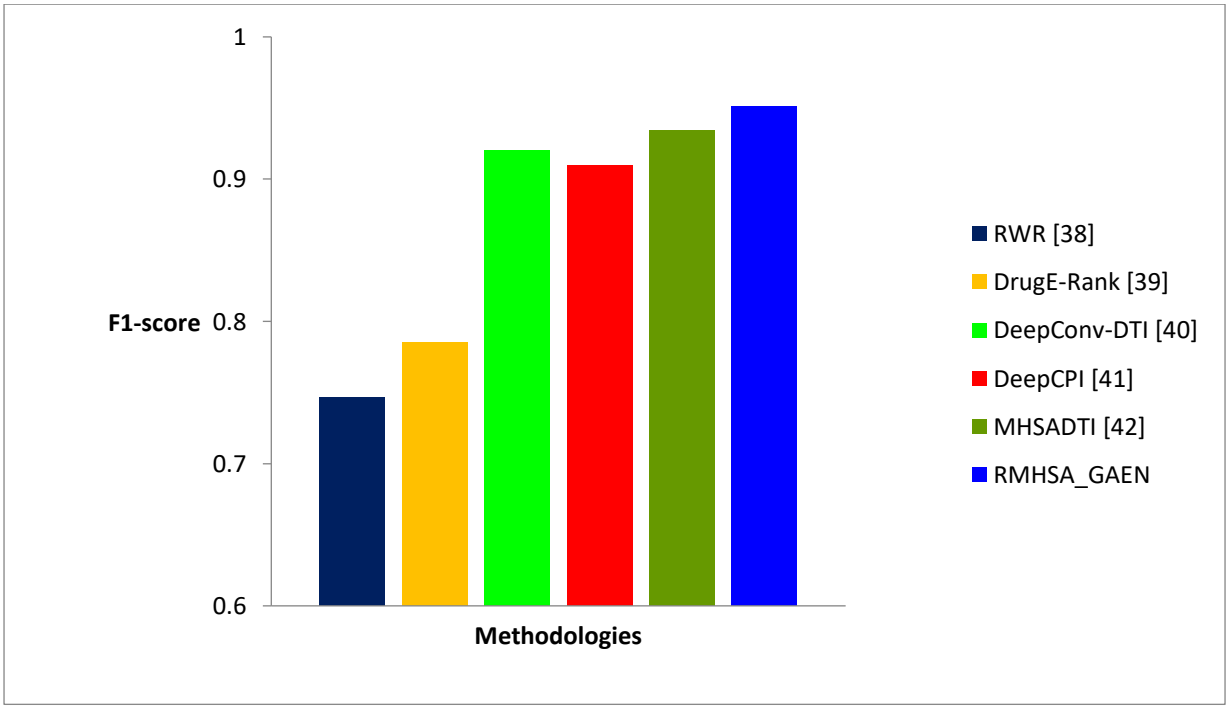

Figure 2. Performance analysis of RMHSA_GAEN with previous works in terms of F1_Score for Human Dataset

In terms of AUPR, RMHSA_GAEN gives 0.69 greater than MHSADTI [42], 2.38 greater than DeepCPI [41], 2 greater than Deepconv-DTI [40], 13.8 greater than DrugE-Rank [39], and 14.72 greater than RWR [38]. In terms of AUC, RMHSA_GAEN gives 0.51 greater than MHSADTI [42], 1.81 greater than DeepCPI [41], 1.35 greater than Deepconv-DTI [40], 13.11 greater than DrugE-Rank [39], and 14.98 greater than RWR [38]. That is clearly shown in the following figure 3.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

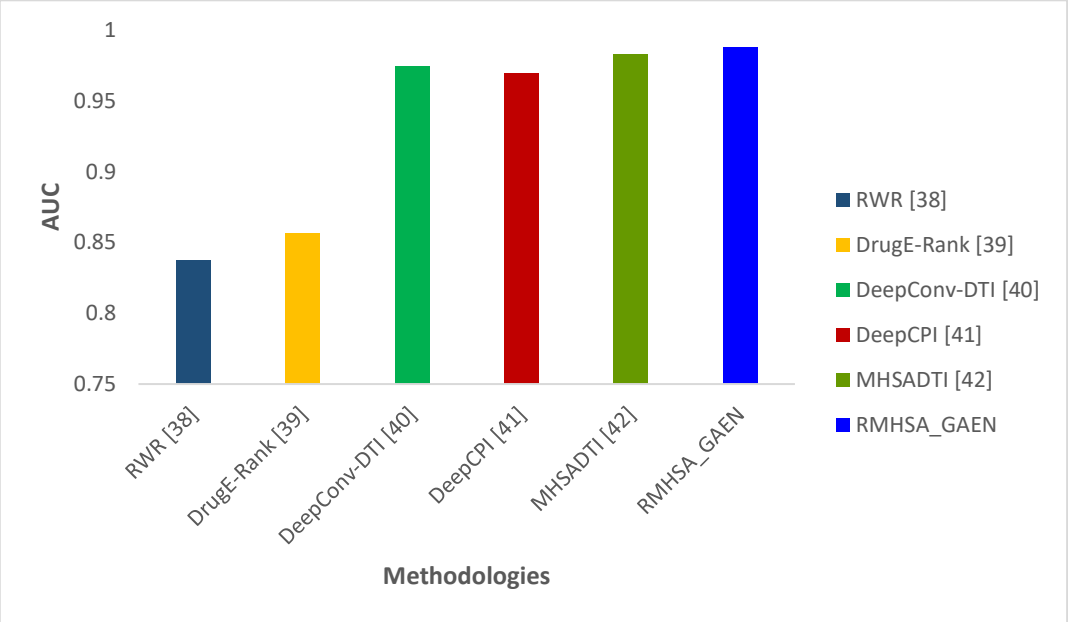

Figure 3. Performance analysis of RMHSA_GAEN with previous works in terms of Area Under the Curve for Human Dataset

Methods	Accuracy	AUC	Precision (std)	Recall (std)	AUPR	F1_Score (std)
RWR [38]	-	84.93	78.60%	71.28%	82.12	74.75%
DrugE-Rank [39]	-	82.21	79.06%	74.74%	83.22	76.84%
DeepConv-DTI [40]	-	97.82	94.35%	94.23%	97.11	95.79%
DeepCPI [41]	-	97.58	93.93%	92.71%	95.71	93.94%
MHSADTI [42]	94.54%	98.38	94.65%	94.51%	98.32	97.63%
RMHSA_GAEN	**96.54%**	**98.67**	**96.52%**	**96.57%**	**98.87**	**96.55%**

Table 3. Performance Analysis of RMHSA_GAEN on C.Elegans Dataset

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

Table 3 displays the outcomes of our approach and a comparison with earlier research. The outcomes demonstrate that RMHSA GAEN outperforms the other algorithms, indicating that it is effective in detecting images. In terms of Accuracy, AUC, Precision, Recall, and AUPR. F1 Score, our technique performs at 96.54%, 98.67%, 96.52%, 96.57%, 98.87%, and 96.55%, respectively.

In terms of Precision, RMHSA_GAEN gives 1.87% greater than MHSADTI [42], 2.59% greater than DeepCPI [41], 2.17% greater than Deepconv-DTI [40], 17.46% greater than DrugE-Rank [39], 17.92% greater than RWR [38].

In terms of Recall, RMHSA_GAEN gives 2.06% greater than MHSADTI [42], 3.86% greater than DeepCPI [41], 2.34% greater than Deepconv-DTI [40], 21.83% greater than DrugE-Rank [39], 25.29% greater than RWR [38].

In terms of F1_Score, RMHSA_GAEN gives 1.08% greater than MHSADTI [42], 2.61% greater than DeepCPI [41], 0.76% greater than Deepconv-DTI [40], 19.71% greater than DrugE-Rank [39], 21.80% greater than RWR [38]. Figure 4 presents a pictorial representation of the F1-Score result analysis using both the existing and proposed methodologies on the C.Elegans Dataset.

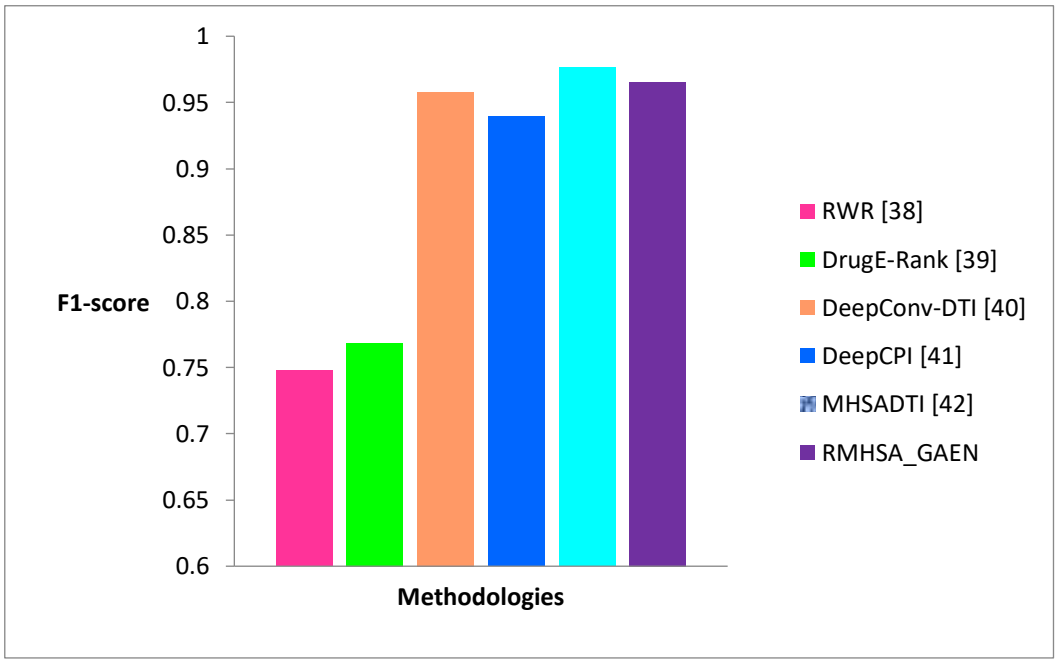

Figure 4. Performance analysis of RMHSA_GAEN with previous works in terms of F1_Score for C.Elegans Dataset

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

In terms of AUPR, RMHSA_GAEN gives 0.55 greater than MHSADTI [42], 3.16 greater than DeepCPI [41], 1.76 greater than Deepconv-DTI [40], 15.65 greater than DrugE-Rank [39], 16.75 greater than RWR [38].

RMHSA GAEN provides an AUC that is 0.29 higher than MHSADTI [42], 1.09 higher than DeepCPI [41], 0.85 higher than Deepconv-DTI [40], 16.46 higher than DrugE-Rank [39], and 13.74 higher than RWR [38]. This is clearly shown in the following figure 5.

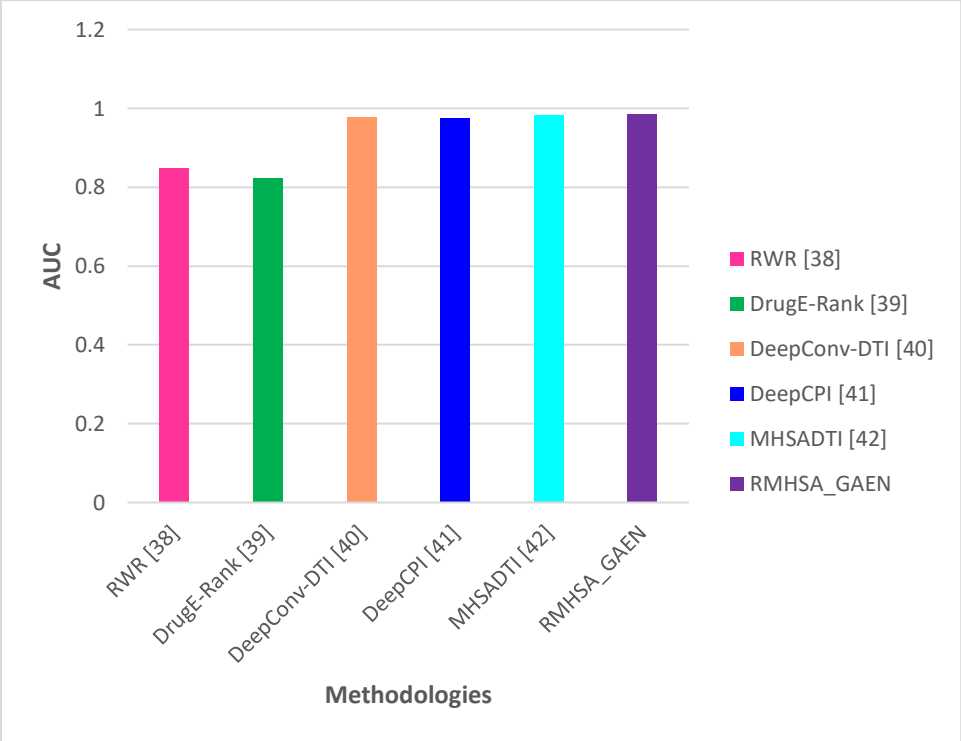

Figure 5. Performance analysis of RMHSA_GAEN with previous works in terms of AUC for C.Elegans Dataset

Conclusion

Powerful computational prediction techniques could be used as prospective approaches for DTI prediction, given that in vitro tests are very expensive and time-consuming. There is, however, still a lot of room for development. In order to predict drug target interaction and offer prompt assistance, this invention thereby produces instantaneous and extremely accurate results. The proposed method RMHSA_GAEN helps transformer for protein and Graph Neural Network for drug. We used attention exchange mechanism to calculate which protein sub-sequences are more crucial for the drug by giving the sub-sequences heavier weights and vice-versa. This cross and exchange attention mechanism delivers best accuracy level of 95.17% for Human dataset and 96.54% for C.Elegans dataset comparatively higher than the previous models.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

In the future, this work can be further improved by using more advanced deep learning techniques and larger, more diverse datasets. One possible direction is to apply pre-trained models like Protein Language Models (PLMs) for better protein feature extraction. Similarly, Graph Neural Networks (GNNs) can be enhanced using graph attention mechanisms or 3D molecular structure data to capture more detailed chemical information from drugs. Finally, applying explainable AI (XAI) methods can help researchers understand why the model predicts certain drug-target pairs, making it more trustworthy for use in real-world drug discovery.

References

[1] Novac, Nicolae. "Challenges and Opportunities of Drug Repositioning." Trends in Pharmacological Sciences, vol. 34, no. 5, 2013, pp. 267–272.

[2] Lounkine, E., et al. "Large-Scale Prediction and Testing of Drug Activity on Side-Effect Targets." Nature, vol. 486, no. 7403, 2012, pp. 361–367.

[3] Yamanishi, Y., et al. "Drug-Target Interaction Prediction from Chemical, Genomic and Pharmacological Data in an Integrated Framework." Bioinformatics, vol. 26, no. 12, 2010, pp. i246–i254.

[4] Zanni, R., et al. "QSAR Multi-Target in Drug Discovery: A Review." Current Computer-Aided Drug Design, vol. 10, no. 2, 2014, pp. 129–136.

[5] Cheng, F., et al. "Prediction of Drug-Target Interactions and Drug Repositioning via Network-Based Inference." PLoS Computational Biology, vol. 8, no. 5, 2012, e1002503.

[6] LeCun, Y., et al. "Deep Learning." Nature, vol. 521, no. 7553, 2015, pp. 436–444.

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

[8] Gao, K. Y., et al. "Interpretable Drug Target Prediction Using Deep Neural Representation." Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI), 2018, pp. 3371–3377.

[9] Zheng, X., et al. "DTIRCNN: New Efficient Hybrid Neural Network Model to Predict Drug–Target Interactions." International Conference on Artificial Neural Networks. Springer, 2018, pp. 104–114.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

[10] Zheng, S., et al. "Predicting Drug Protein Interaction Using Quasi-Visual Question Answering System." 2019, https://doi.org/10.1101/588178.

[11] Lin, C., et al. "Learning to Predict Drug Target Interaction from Missing Not at Random Labels." IEEE Transactions on Nanobioscience, 2019.

[12] Feng, Qingyuan, et al. "PADME: A Deep Learning-Based Framework for Drug-Target Interaction Prediction." preprint arXiv:1807.09741, 2018.

[13] Xie, Lingwei, et al. "Deep Learning-Based Transcriptome Data Classification for Drug-Target Interaction Prediction." BMC Genomics, vol. 19, 2018, https://doi.org/10.1186/s12864-018-5031-0.

[14] Alipanahi, B., et al. "Predicting the Sequence Specificities of DNA- and RNA-Binding Proteins by Deep Learning." Nature Biotechnology, vol. 33, no. 8, 2015, pp. 831–838.

[15] Jo, T., et al. "Improving Protein Fold Recognition by Deep Learning Networks." Scientific Reports, vol. 5, 2015, Article no. 17573.

[16] Nascimento, A. C., et al. "A Multiple Kernel Learning Algorithm for Drug-Target Interaction Prediction." BMC Bioinformatics, vol. 17, 2016, pp. 1–12.

[17] Wen, M., et al. "Deep-Learning-Based Drug-Target Interaction Prediction." Journal of Proteome Research, vol. 16, 2017, pp. 1401–1409.

[18] Rayhan, F., et al. "FRnet-DTI: Deep Convolutional Neural Networks with Evolutionary and Structural Features for Drug-Target Interaction." arXiv preprint arXiv:1806.07174, 2018.

[19] Hu, S., et al. "A Convolutional Neural Network System to Discriminate Drug-Target Interactions." IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 99, 2019, pp. 1–12.

[20] Yamanishi, Y., et al. "Prediction of Drug–Target Interaction Networks from the Integration of Chemical and Genomic Spaces." Bioinformatics, vol. 24, no. 13, 2008, pp. i232–i240.

[21] Sawada, R., et al. "Benchmarking a Wide Range of Chemical Descriptors for Drug-Target Interaction Prediction Using a Chemogenomic Approach." Molecular Informatics, vol. 33, no. 11–12, 2014, pp. 719–731.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

[22] Manoochehri, Hafez Eslami, and Mehrdad Nourani. "Drug-Target Interaction Prediction Using Semi-Bipartite Graph Model and Deep Learning." BMC Bioinformatics, vol. 21, no. 4, 2020, pp. 1–16.

[23] Lee, Ingoo, et al. "DeepConv-DTI: Prediction of Drug-Target Interactions via Deep Learning with Convolution on Protein Sequences." PLoS Computational Biology, vol. 15, no. 6, 2019, e1007129.

[24] Beck, Bo Ram, et al. "Predicting Commercially Available Antiviral Drugs That May Act on the Novel Coronavirus (SARS-CoV-2) through a Drug-Target Interaction Deep Learning Model." Computational and Structural Biotechnology Journal, vol. 18, 2020, pp. 784–790.

[25] Chu, Yanyi, et al. "DTI-CDF: A Cascade Deep Forest Model towards the Prediction of Drug-Target Interactions Based on Hybrid Features." Briefings in Bioinformatics, vol. 22, no. 1, 2021, pp. 451–462.

[26] Zhao, Tianyi, et al. "Identifying Drug–Target Interactions Based on Graph Convolutional Network and Deep Neural Network." Briefings in Bioinformatics, vol. 22, no. 2, 2021, pp. 2141–2150.

[27] Song, Tao, et al. "DeepFusion: A Deep Learning-Based Multi-Scale Feature Fusion Method for Predicting Drug-Target Interactions." Methods, vol. 204, 2022, pp. 269–277.

[28] Li, Fei, et al. "Effective Drug-Target Interaction Prediction with Mutual Interaction Neural Network." Bioinformatics, 2022.

[29] Zhang, Peiliang, et al. "DeepMGT-DTI: Transformer Network Incorporating Multilayer Graph Information for Drug–Target Interaction Prediction." Computers in Biology and Medicine, vol. 142, 2022, 105214.

[30] Yazdani-Jahromi, Mehdi, et al. "AttentionSiteDTI: An Interpretable Graph-Based Model for Drug-Target Interaction Prediction Using NLP Sentence-Level Relation Classification." Briefings in Bioinformatics, vol. 23, no. 4, 2022, bbac272.

[31] Dong, Q.-W., et al. "Application of Latent Semantic Analysis to Protein Remote Homology Detection." Bioinformatics, vol. 22, no. 3, 2005, pp. 285–290.

[32] Paulus, R., et al. "A Deep Reinforced Model for Abstractive Summarization." arXiv preprint arXiv:1705.04304, 2017.

Power System Technology
ISSN:1000-3673
Received: 16-03-2025 Revised: 05-04-2025 Accepted: 02-05-2025

[33] Lin, Z., et al. "A Structured Self-Attentive Sentence Embedding." arXiv preprint arXiv:1703.03130, 2017.

[34] Li, Jian, et al. "On the Diversity of Multi-Head Attention." Neurocomputing, vol. 454, 2021, pp. 14–24.

[35] India, Miquel, et al. "Self Multi-Head Attention for Speaker Recognition." arXiv preprint arXiv:1906.09890, 2019.

[36] He, K., et al. "Deep Residual Learning for Image Recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[37] Ba, J. L., et al. "Layer Normalization." arXiv preprint arXiv:1607.06450, 2016.

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

[39] Yuan, Q., et al. "DrugE-Rank: Improving Drug–Target Interaction Prediction of New Candidate Drugs or Targets by Ensemble Learning." Bioinformatics, forthcoming.

[40] Lee, I., et al. "DeepConv-DTI: Prediction of Drug-Target Interactions via Deep Learning with Convolution on Protein Sequences." PLoS Computational Biology, vol. 15, no. 6, 2019, e1007129.

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

[42] Cheng, Z., et al. "Drug-Target Interaction Prediction Using Multi-Head Self-Attention and Graph Attention Network." IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 4, 2022, pp. 2208–2218.

[43] Scarselli, F., et al. "The Graph Neural Network Model." IEEE Transactions on Neural Networks, vol. 20, 2009, pp. 61–80.