Ingénierie des Systèmes d'Information
Vol. 31, No. 6, June, 2026, pp. 1927-1938
Journal homepage: http://iieta.org/journals/isi

A Deep Residual Graph Attention Network Integrating Protein Language Models and Structural Features for Residue-Level Protein–Protein Interaction Site Prediction

Oras A. Hussein*, Eman Salih Al-Shamery

Department of Software, College of Information Technology, University of Babylon, Hillah 51001, Iraq

Corresponding Author Email: orasabdulkhudhurh.sw@student.uobabylon.edu.iq

https://doi.org/10.18280/isi.310615

Received: 21 April 2026
Revised: 15 June 2026
Accepted: 23 June 2026
Available online: 30 June 2026

Keywords:
protein-protein interaction site prediction, graph attention network, protein language model, ESM-2 embeddings, residue-level graph learning, structural bioinformatics, deep learning

ABSTRACT

Protein-protein interaction site (PPIS) prediction is essential for understanding molecular recognition mechanisms and supporting applications such as drug discovery and protein engineering. However, existing computational approaches often suffer from limited structural representation capability or insufficient integration of sequence and three-dimensional information. This study proposes a deep residual graph attention network, termed DR-GATv2, for residue-level PPIS prediction by integrating protein language model representations with structural and physicochemical features. In the proposed framework, proteins are represented as residue-level graphs, where residues serve as nodes and spatial relationships are modeled as graph edges. The model combines ESM-2 embeddings with DSSP-derived structural descriptors and physicochemical characteristics within residual GATv2 blocks. Furthermore, jumping knowledge aggregation is introduced to preserve multi-scale residue representations across different graph layers. Experiments were conducted on a benchmark of protein complex datasets with strict protein-level separation to avoid data leakage. The proposed model achieved an AUROC of 0.9625, an AUPRC of 0.9149, an F1-score of 0.8192, and an MCC of 0.7674 on the independent test set. Comparative experiments and ablation studies demonstrate that the integration of sequence-derived embeddings, structural information, and hierarchical graph representation contributes to improved PPIS prediction performance. The proposed framework provides an effective computational strategy for residue-level protein interaction analysis.

1. INTRODUCTION

Protein-protein interactions (PPIs) are key to biological systems and regulate a range of activities, including signal transduction, metabolic regulation, transcriptional control, and immune response. The formation of protein complexes, where some residues on the surfaces of proteins come together to form stable interfaces, allows a wide range of biological activities. Identification of protein-protein interaction sites (PPIS) at the residue level is important for understanding molecular recognition processes, and it has wide implications in drug development, protein engineering, and disease mechanism study. Experimental techniques such as X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy provide accurate structural information on protein interfaces. However, the current computational algorithms for PPIS prediction can be categorized into sequence-based and structure-based methods. Sequence-based approaches depend on amino acid sequences only and often include evolutionary profiles, sequence conservation, physicochemical attributes, and contextual residue information. Examples include SCRIBER [1] and DELPHI [2], which use machine learning and deep learning approaches to predict interaction propensity from sequence-based information. However, the predictive power of these approaches is sometimes hampered by the lack of specific structural information, while these methods are generally relevant. Structure-based methods use 3D protein structures to describe spatial contexts of residues and interactions between residues. Methods such as IntPred [3], PPIcons [4], and GraphPPIS [5] have shown that the addition of structural information considerably increases the prediction accuracy. Graph-based representations have proven especially popular, as proteins can be represented as residue-level graphs, where residues are nodes and spatial or sequential interactions are edges. In recent years, four important research lines have generated considerable progress: protein language models, graph transformers, geometric deep learning, and attention-based graph neural networks. Protein language models trained on millions of protein sequences, such as ESM-based methods, have proven the potential to acquire complex contextual and evolutionary representations directly from raw sequences. ISPRED-SEQ [6] has shown that embeddings derived from protein language models may greatly enhance the effectiveness of interaction site prediction without the need for extensive handcrafted sequence characteristics. At the same time, geometric deep learning approaches have developed as useful tools to utilize three-dimensional structure information. PeSTo [7] proposed a parameter-free geometric approach that learns surface representations crucial to interactions directly from protein structure. Spatom [8] further proved the power of graph neural networks for modeling residue-level spatial

connections. On the other hand, graph transformer topologies have enhanced classic graph neural networks with improved long-range dependency modeling and multi-scale feature integration; it has allowed more effective learning from complicated protein structures. Attention-based graph neural networks have also become a popular way to predict PPIS. Techniques like EGRET [9], GrASP [10], and GACT-PPIS [11] use attention processes to assign weights to surrounding residues in real-time, based on their structural importance. This allows the model to focus on physiologically relevant interaction patterns. These investigations all show that adaptive attention mechanisms may greatly enhance the interface detection compared with the traditional graph convolutional methods. Despite these gains, there are significant difficulties. First, a significant number of graph-based PPIS predictors have a somewhat shallow design and hence are limited in capturing higher-level structural connections. Second, the over-smoothing issue [12] happens with the increase in network depth, when the node representations become too similar and lose their discriminative strength. Third, current techniques mostly rely on either structural features or learned sequence embeddings, and the successful combination of protein language model's embeddings [13] with rich structural descriptors remains largely unexplored. Finally, it is still a difficulty to preserve the multi-scale information across graph layers in the deep graph topologies. To overcome these limitations, we propose a Deep Residual Graph Attention Network with GATv2 [14] (DR-GATv2) for residue-level PPIS prediction. The proposed system blends contextualized ESM-2 protein language model embeddings with extensive structural data such as secondary structure, solvent accessibility, geometric descriptors, and residue coordinates. We model proteins as residue-level graphs and transmit the information using stacked residual GATv2 layers to enhance feature learning and alleviate over-smoothing. In addition, the model uses jumping knowledge aggregation to maintain and merge multi-scale representations at multiple graph levels, which allows it to use both local and global structural information concurrently. Unlike existing graph-based methods such as GraphPPIS [5] and EGRET [9] that primarily leverage shallow graph architectures and limited feature aggregation, the proposed DR-GATv2 exploits deep residual attention-based message passing, expressive GATv2 attention mechanisms, protein language model representations, and multi-scale feature aggregation within a single unified framework. This approach allows better modeling of complicated residue interactions and patterns of cooperative structural features that characterize protein interaction surfaces. The key contributions of this study are the following: (1) to solve over-smoothing and achieve deeper graph representation learning, we provide a deep residual GATv2 architecture for residue-level PPIS prediction. (2) We fuse the ESM-2 protein language model's embeddings with rich structural descriptors to collect complementary information at the sequence and structure levels. (3) We utilize jumping knowledge aggregation to retain and integrate hierarchical representations gained across different graph levels. (4) We show the proposed framework is successful on benchmark Protein Data Bank (PDB) datasets with good prediction performance and better discrimination of protein interaction sites. Accurate predictions are shown on several PDB [15] protein complex datasets, with an Area Under the Receiver Operating Characteristic Curve (AUROC) [16] and an Area Under the Precision-Recall Curve (AUPRC) [17] on an independent test set. The proposed graph-based PPIS prediction method is interpretable and robust. This area is a promising field for the study of structural bioinformatics and protein interaction modeling.

2. MATERIALS AND METHODS

This section presents the proposed framework to apply deep residual GAT to predict PPIS at residue level. The approach includes dataset construction, feature extraction, residual graph creation, deep residual GATv2 architecture, training and evaluation of the model.

2.1 Dataset construction

The dataset used in this study was developed using a procedure similar to established benchmarks like Train_335 [18], but expanded to include a more extensive and comprehensive human PPI collection. Initially, experimentally derived protein complex structures were sourced from the PDB [15] and refined to include only Homo sapiens entries with at least two interacting protein chains, excluding nucleic acids, ligands, or branching entities to ensure the purity of PPI. Each compound was further processed at the chain level, where amino acid sequences were removed and only standard residues retained. Residues were categorized as interfaces (binding sites) based on structural closeness if their atoms were within a certain distance (e.g., < 8 Å) from other protein chains in the complex. Alternatively, they were classified as non-binding sites. This labeling technique produces binary sequences similar to Train_335. Binding residues are denoted by the binary string of each protein chain, derived from its sequence. Geometric descriptors, physicochemical properties, and protein language model embeddings were obtained for each residue. A graphical representation of residue-level annotated protein chains was created for model training. This construction technique surpasses conventional datasets in quality and structure-focused supervision, enabling robust learning for binding site prediction and protein interaction modeling. We partition the data into intricate training, validation, and test sets to avert data leakage. Proteins were grouped using CD-HIT [19] at a 40% sequence identity threshold. Partitions for training, validation, and testing were executed using protein clusters instead of PDB entries to prevent the presence of analogous proteins inside the data divisions. The Euclidean distance between atoms was computed as:

$$d_{\{ij\}} = \sqrt{\left\{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2 + \left(z_i - z_j\right)^2\right\}} \qquad (1)$$

where, (x_i, y_i, z_i) and $\left(x_j, y_j, z_j\right)$ denote the 3D coordinates of atoms *i* and *j* respectively. Residues satisfying $d_{\{ij\}} < \delta$ were assigned label 1 (interface residues), while others were labeled 0. The dataset consists of 6,782 protein chains from 2,242 different PDB complexes covering 1,242,811 residues. Among them, 295,858 residues (23.81%) are labelled as binding sites and 946,953 residues (76.19%) are non-binding, which presents a large class imbalance. The dataset was divided into three parts: training (5,466 chains), validation (658 chains), and test (658 chains). At the PDB level, there was no overlap. All dataset statistics are shown in Table 1. The average chain length is about 183 residues, which is sufficient

for modeling graphs at the residue level with structural context.

The collection consists of 2,242 unique PDB structures and covers more than 2.4 million residues, far above typical standards, which allows for capturing complex and generalizable interaction patterns.

Table 1. Dataset statistics for residue-level protein-protein interaction (PPI) site prediction

Dataset	Chains	Unique PDBs	Residues	Non-Binding (0)	Binding (1)	Positive Ratio	Negative Ratio	Avg Residues/Chain
All	6782	2242	1,242,811	946,953	295,858	0.2381	0.7619	183.25
Train	5466	1793	1,003,604	765,150	238,454	0.2376	0.7624	183.61
Validation	658	224	124,859	95,51	29,340	0.2350	0.7650	189.76
Test	658	225	114,348	86,284	28,064	0.2454	0.7546	173.78

Note: Only residues from chains with valid interaction annotations used for training and evaluation, resulting in a labeled subset of 2.15 million residues.

2.2 Feature extraction and preprocessing

In order to obtain sequence and structural information, we removed most residue-level features. For each residue, a 537-dimensional feature vector is created with structural, physicochemical, and sequence-derived features. DSSP is used to compute secondary structure and solvent accessibility [20]. Physicochemical descriptors assess the charge, hydrophobicity, polarity, and molecular weight of amino acids. The ESM-2 protein language model generates high-dimensional contextual embeddings (512-d) that provide insight into evolution and structure [6]. ESM-2 embeddings provide extensive contextual sequence information obtained from large-scale protein datasets, including evolutionary and biological markers that cannot be hand-crafted. Conversely, hand-crafted structural descriptors, including secondary structure, solvent accessibility, geometric properties, and spatial coordinates, unambiguously define local three-dimensional settings. Collectively, these feature groups provide complementary sequence-level and structural representations that enable the model to more accurately define protein interaction sites. Table 2 illustrates feature groups used in our DR-GATv2 model.

Table 2. Feature sets used in the DR-GATv2 model

Feature Type	Description	Dimension
	Structural Features	
Secondary Structure	Encodes backbone conformation (helix, sheet, coil) derived from DSSP	8
Solvent Accessibility (RSA/ASA)	Measures residue exposure to solvent, indicating surface accessibility	2
	Physicochemical Properties	
Charge	Indicates residue charge (positive, negative, neutral)	1
Hydrophobicity	Quantifies residue affinity to water (hydrophobic/hydrophilic)	1
Polarity	Represents polarity class of amino acids	1
Molecular Weight	Mass of each amino acid residue	1
Additional amino acid properties	Includes flexibility, volume, and other biochemical descriptors	11
	Sequence/Contextual Features	
Protein Language Model (ESM-2) Embeddings	High-dimensional contextual embeddings capturing evolutionary and structural patterns from sequence	**512**
	Total Feature Dimension	
	Combined residue representation used as node features in the graph	**537**

Before model training, residues without structural information or incorrect coordinate values were removed. When proper, default values were used in the absence of missing structural features. Data consistency was achieved by removing duplicate residue records. Due to heterogeneous sources and different numerical scales of recovered features, feature normalization was done before graph building. Z-score normalization was applied to continuous features, including ASA, RSA, torsion angles, geometric descriptors, coordinate-based characteristics, and neighborhood statistics.

$$x' = \frac{x - \mu}{\sigma} \tag{2}$$

The mean and standard deviation from the training set are represented by μ and σ. Standardization improves numerical stability and avoids large-scale influences from obscuring optimization. After preprocessing, all feature groups were combined into one residue representation. We concatenated features [21], integrated ESM-2 embeddings, structural descriptors, and physicochemical attributes into a residue feature vector to fuse features:

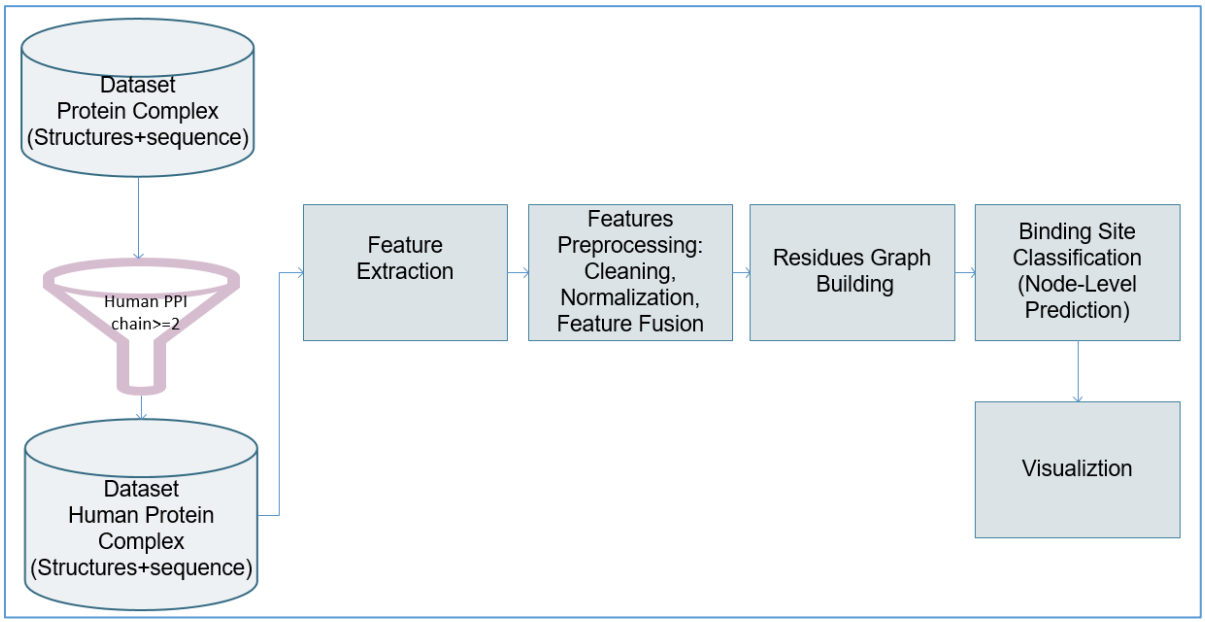
$$h_i = [h_i^{(ESM2)} || h_i^{(DSSP)} || h_i^{(Geo)} || h_i^{(Phys)}] \tag{3}$$

The subsets $h_i^{(ESM2)} || h_i^{(DSSP)} || h_i^{(Geo)} || h_i^{(Phys)}$ represent sequence embedding, structural, geometric, and physicochemical features, respectively. The final fused representation comprises 537 characteristics per residue and feeds the DR-GATv2 architecture. The integrated feature vector combines evolutionary, structural, geometric, and biochemical data to predict PPIS accurately.

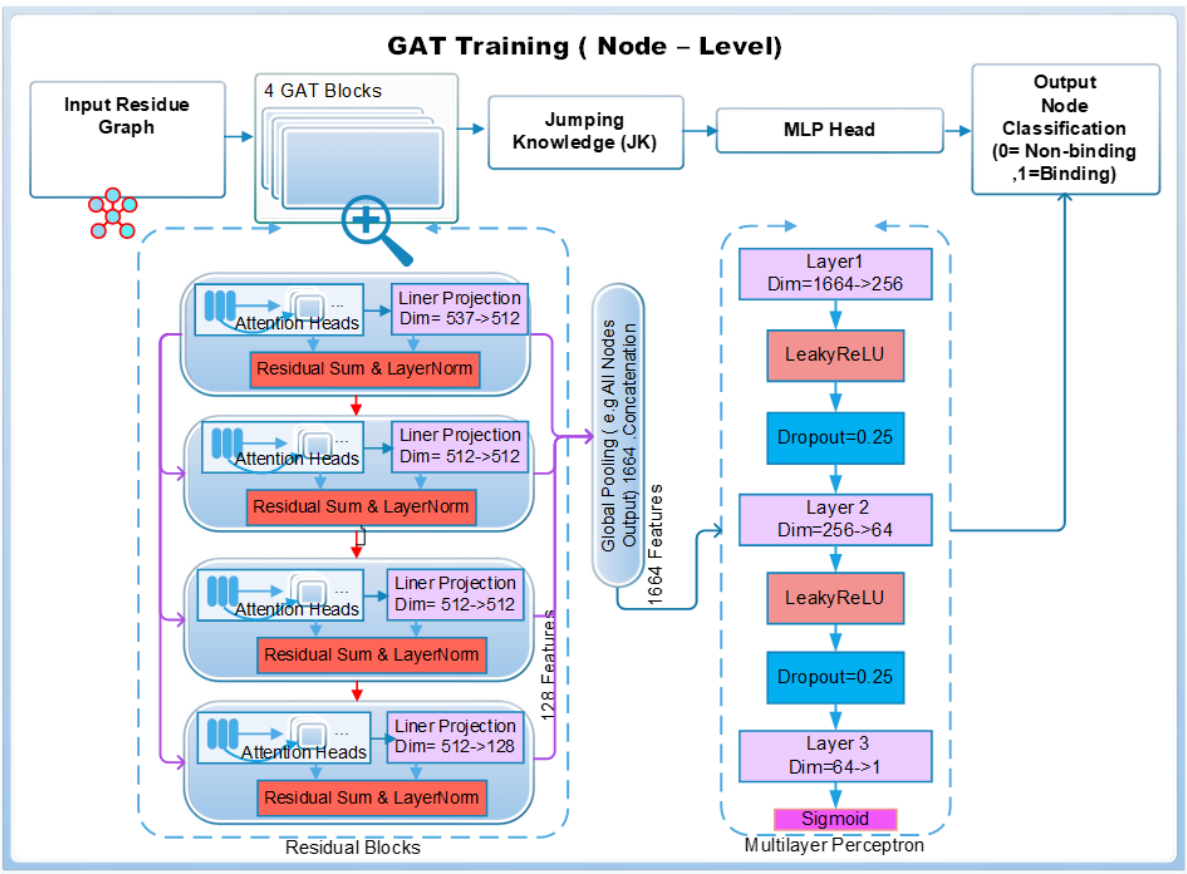
2.3 Protein representation

Each protein chain is depicted as a residue-level graph. In this graph, amino acid residues are nodes and edges represent the spatial relationships among the residues. If the distance between the Cα atoms of two residues was smaller than a certain distance (<8 $\AA$), they were considered connected. The spatial connectivity lets the graph show local structural areas

that are significant for protein interactions. We represent a protein graph as an undirected graph $G = (V, E)$ [22] in which nodes V represent amino acid residues, and edges E are formed based on a distance-based threshold between C_α atoms [23]. Each residue is represented as a node $v_i \in V$ where V is the set of all residues in the protein graph. A comprehensive feature vector is generated for each residue by amalgamating sequence-derived and structure-derived descriptors [24]. The residue representation comprises ESM-2 protein language model embeddings, secondary structure annotation (DSSP), solvent accessibility (ASA and RSA), backbone geometric features (φ and ψ angles along with their sine and cosine transformations) [25], structural position descriptors, local neighborhood features, and sequence position information. The resulting feature vector provides a comprehensive characterization of the physicochemical, evolutionary, and structural attributes of each residue.



(a) Main steps for the DR-GATv2 framework



(b) The proposed DR-GATv2 architecture

Figure 1. The overall architecture of the proposed framework

2.4 Deep residual GATv2 architecture

We propose a model using a Deep Residual Graph Attention Network based on GATv2 [14] (DR-GATv2) to understand complex structural patterns in residue-level protein graphs. The design is to model the short and long-range interactions among residues respectively and the stability of deep Graph Neural Network (GNN) during the optimization process. We employed a four-layer GATv2 architecture to expand the receptive field without oversmoothing. Shallow designs (two or three layers) gathered local neighborhood data and validated poorly. In deeper designs (five or more layers), residual connections did not reduce feature homogeneity. Empirically, four layers balanced representational capacity, training stability, and predictive performance well. This enables us to learn deep representations without suffering from gradient decay or oversmoothing. Each residual block contains four principal components: (1) GATv2 convolutional layer, (2) layer normalization, (3) nonlinear activation function (Exponential Linear Unit), and (4) dropout regularization. The GATv2 attention mechanism calculates the adaptive feature-dependent attention coefficients for neighboring residues in each block. This lets the model concentrate on interactions relevant to structure and function. This dynamic attention mechanism provides a more detailed representation of residue-residue interactions than static aggregation approaches. It is important to find interaction interfaces. Residual connections are critical for allowing the flow of gradients through the layers. This results in improved training stability and alleviates the over-smoothing problem that is common in deep GNN. This design enables efficient expansion of the network to deeper levels, while preserving important distinctions between nodes. As the number of GATv2 layers increases, the receptive field of each residual component of the four GATv2 layers gradually expands. In this way, information can go up to four hops across communities. Every residue embedding captures local structural context (e.g., neighboring spatial residues) and global dependencies (e.g., distant residues for interface development). This hierarchical feature aggregation is key for identifying cooperative residue patterns, which signal protein interactions. Figure 1 shows the overall architecture of the proposed framework for data collection, feature extraction, construction of a residue graph, and a deep residual GATv2 network for interface prediction.

The complete framework of the proposed deep residual GATv2 model for residue-level PPI site prediction is demonstrated in Figures 1(a) and (b). The model consists of stacked ResidualGATBlocks with multi-head attention and residual connections. On the right, a close-up of the inside of a Residual GAT block. It shows the GATv2 convolution, normalization, activation, dropout, and residual connection. In summary, the proposed DR-GATv2 architecture provides a powerful and scalable framework for modeling residue-level interactions, combining deep attention-based message passing with residual learning to generate robust and expressive protein representations.

2.5 Training and evaluation

The DR-GATv2 was trained in a supervised framework of residue-level classification, which assigned to each residue a binary label indicating whether it was an interaction site or not. Training was performed to discriminate residue representations that predict participation of the interface well. We improved the model by binary cross-entropy (BCE) loss [26], which is suitable for probabilistic binary classification. The loss function penalizes the difference between expected and actual class assignment according to the likelihood of the residue and the ground-truth label. This allows for the distinction of interface and non-interface residues. The interface residues make up a tiny fraction of the protein structural residues, which makes the PPI site prediction difficult due to the class imbalance [27] Otherwise the model may prefer the dominant non-interface class and become less sensitive to true binding residues. To address this issue, the loss function was weighted class-wise [28] giving a higher weight to positive samples during optimization. This strategy improves the detection of interface residues and the stability of the model when labels are not equally distributed. The Adam optimizer [21] found optimal parameters for the model. It uses adaptive estimates of first and second-order moments to speed up the learning process. Adaptive learning helped the model converge faster and maximize each training cycle. This optimization parameter facilitates deep graph topology learning, especially for large residue-level graphs with many features. We used a set of standard metrics for binary classification under class imbalance to evaluate the performance of the proposed model in predicting residue level PPI sites. These measures are calculated from the confusion matrix [29] with true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). These include: The AUROC [30] quantifies the model's ability to discriminate between interface and non-interface residues across all classification levels.

$$\mathrm{AUROC} = \int_{\{0\}}^{\{1\}} \mathrm{TPRate(t)} d\mathrm{FPRate(t)} \tag{4}$$

Unbalanced datasets benefit from the AUPRC [5], which illustrates the favorable class accuracy-recall trade-off.

$$\mathrm{AUPRC} = \int_{\{0\}}^{\{1\}} \mathrm{Precision(R)dR} \tag{5}$$

The balanced Matthews Correlation Coefficient (MCC) [27] considers all four confusion matrix outputs. Not uniformly distributed categorization issues benefit from it.

$$\mathrm{MCC} = \frac{(\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN})}{\sqrt{[(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})]}} \tag{6}$$

The harmonic mean of recall and accuracy is the F1-score.

$$\mathrm{F1-score} = \frac{(2 \times \mathrm{Precision} \times \mathrm{Recall})}{(\mathrm{Precision} + \mathrm{Recall})} \tag{7}$$

Precision and recall evaluate positive prediction reliability and interface residue detection sensitivity. Together, these measurements give a complete picture of the model performance. In the case of residue-level PPI site prediction, AUPRC and MCC are particularly useful because the positive class is rare and standard accuracy metrics can be misleading. The assessment approach thus provides a comprehensive view of the model's discrimination ability, reliability, and utility by combining threshold-independent metrics (AUROC and AUPRC) with threshold-dependent metrics (MCC, F1-score, precision, and recall) [26].

2.6 Model interpretation

Multi-hop graph attention and residual propagation are used to transform each residue into a context-aware embedding that captures its protein structure and function. For the final prediction layer, a residue representation captures the biochemical properties and spatial arrangement of neighboring residues. The number of layers above determines the receptive fields of GNN nodes. This receptive field characterizes the propagation of structural information in residue-level protein graphs. Its four GATv2 layers allow residues to gather information from their four-hop neighbors. This allows the model to describe the structural interactions of proteins within short and long range. Four-hop neighborhoods spread a residue's structural context over the graph. Such cascade aggregation enables the network to learn hierarchical structural features of PPI sites. The first layer receives information from neighboring residues within 8 $\AA$ distance to the central residue. Structural factors: solvent accessibility and contextual sequence embeddings ESM-2. This stage represents the immediate physicochemical environment of the residue. The second layer uses the neighbor-of-neighbors data. The network captures local structural motifs like twists of α-helices or ß-sheets around the residue. The receptive field of the protein structure increases with the third layer. The residue representation also includes environmental information, e.g., whether a flexible loop region is supported by hard secondary structures. After four propagation cycles, the residue representation shows a large protein structural neighborhood. This larger context allows the model to reason about whether the surrounding surface patch is an interaction interface. Multi-hop graph attention and residual propagation convert each residue into a context-aware embedding that reveals its structural and functional neighbors across the protein. The residue representation of the last prediction layer encodes the biological properties of the residue and the spatial arrangement of its neighbors. This contextualized representation helps the model to find interaction interfaces, as protein binding sites are often surface patches of many cooperating residues and not a single amino acid. This is why the proposed model performs well on AUPRC, which matters for tasks with imbalanced interface prediction. Surface patches of many cooperative residues, not isolated amino acids, normally characterize protein binding sites, and this contextualized representation helps the model to find interaction interfaces. Thus, the proposed model performs well on measures such as AUPRC, which are important for unbalanced interface prediction. Unlike conventional graph convolutional networks that aggregate input from neighboring nodes uniformly, our proposed architecture utilizes the attention mechanism of GATv2 to dynamically weight contributions from individual residues. The attention mechanism allows the model to pay attention to residues that are important for the structure during aggregation. For instance, residues that are highly conserved in evolution or have unique physicochemical features may be more attended to during message passing. The GATv2 formulation improves upon the original graph attention network with a flexible attention scoring function. This enables the model to derive attention scores that are contingent on both the main residue and its neighboring residues. The influence of neighboring residues thus varies with the protein's structure. Deeper GNNs are usually too smooth to be trained, as the representations of the nodes become indistinguishable after a few layers. To remedy this, each GATv2 block has residual connections. These skip connections maintain the identity of the original residue representation, while enabling the network to learn more and more complicated structural links across many propagation steps.

3. RESULTS AND DISCUSSION

This section generally evaluates the performance of the suggested DR-GATv2 model in prediction of the residue level PPI sites. We evaluate the model performance using standard evaluation metrics and compare it with other state-of-the-art approaches.

3.1 Experimental setup

The model presented here was developed into a deep GAT, DR-GATv2, for residue-level PPI site prediction. Each protein is represented as a graph, where residues are nodes and links are edges. The input feature dimensions d_{in} is dynamically generated from the dataset, which is the residue-level feature vector, which is derived from sequence, structural, and embedding attributes. The dataset consisted of protein complexes from the Protein Data Bank for all experiments. To enable a fair evaluation, the data set was split into training, validation, and test subsets. All parameters and hyperparameters were selected on the training dataset and early stopping with the validation set was used to avoid overfitting [31]. In particular, training was stopped if there was no improvement in validation performance for a certain number of epochs. This method helps the model to work well on a new data and not to overfit to the training set. The model was improved to maximize as much as possible the AUPRC. This is particularly helpful for tasks such as the classification of residues at an imbalanced level. We tuned the learning rate, weight decay, dropout rate, and focus loss parameters for steady convergence and better discrimination of binding residues. For training, we used the AdamW optimizer with an initial learning rate of 6×10^{-4} and a weight decay of 1×10^{-4} [32] The learning rate was tuned using a learning rate scheduler (ReduceLROnPlateau) based on the validation AUPRC [33]. Gradient clipping was used to keep training stable and dropout was used throughout the network to minimize overfitting. To address the class imbalance, a focal loss function was used with a dynamically calculated positive class weighting to focus on hard samples during training. After training, the MCC [11] was used to search for the best classification threshold on the validation set. Then the test set was tested using this threshold. All reported results are on the best model chosen based on the validation AUPRC. The full list of parameters and hyperparameters used in this study are shown in Table 3. All experiments were performed on Google Colab with an NVIDIA A100 GPU with 40 GB RAM using PyTorch and PyTorch Geometric [34]. The GPU acceleration made the DR-GATv2 model training fast and it converged in ~ 4 hours. Table 3 shows the parameters used in our DR-GATv2 Model.

3.2 Evaluation scheme

We examine the proposed model using several parameters, as in the previous studies. Examples are precision, recall, F1-score (F1), MCC, AUROC, and AUPRC. ROC and PR curves obtained by using the prediction probability for each residue

of the model. Scikit-learn [35] calculates AUROC and AUPRC. These are not threshold-dependent metrics and are representative of the performance of the model across various classification levels. Other metrics such as accuracy, recall, F1-score, and MCC are threshold-dependent. The MCC is used to find the best classification threshold for the validation set in this research. Again, a threshold was applied to the independent test set. This approach is a fair and consistent comparison of performance. Because there are more non-binding residues than binding residues, protein interaction prediction is imbalanced. AUPRC and MCC are higher-performing metrics as they highlight minority class prediction. However, AUROC measures the performance of both classes and is not sensitive to the class imbalance [36]. The confusion matrix provides the formulas for these measurements. TP and TN correctly predict binding and non-binding residues, but FP and FN are errors. The performance of the proposed DR-GATv2 model on the test dataset is summarized in Table 4, and the training history of the DR-GATv2 model is shown in Figure 2.

The results show that DR-GATv2 is able to effectively identify structural patterns related to protein interaction interfaces. The graph attention allows the model to weigh residues that are structurally important for binding regions more heavily. The model learns better representations of residues when sequence embeddings are combined with structural properties. To evaluate the efficiency of the DR-GATv2 model, we compared it with many common PPI site prediction techniques, as shown in Table 5. The models represent recent advances in graph-based protein interaction prediction.

Table 3. Parameters used in DR-GATv2

Item	Value
Model	DR-GATv2 classifier
Task	Residue-level binary node classification (PPI sites)
Input feature dimension	537
Number of layers	4 Residual GATv2 Blocks
Hidden dimension (H)	128
Epochs	100
Attention heads	4 (first 3 layers), 1 (final layer)
Block 1	GATv2Conv((d), 128, heads=4) → output = 512
Block 2	GATv2Conv(512, 128, heads=4) → output = 512
Block 3	GATv2Conv(512, 128, heads=4) → output = 512
Block 4	GATv2Conv(512, 128, heads=1, concat=False) → output = 128
Residual connections	Linear projection when dimensions differ, identity otherwise
Normalization	LayerNorm after each GAT layer
Activation (GAT blocks)	ELU
Dropout (GAT blocks)	0.25
Jumping knowledge (JK)	Concatenation of all layer outputs
JK dimension	(128 * 4 * 3) + 128 = 1664
MLP head	1664 → 256 → 64 → 1
Activation (MLP)	ReLU
Dropout (MLP)	0.25
Output	Single logit per residue
Loss function	BCEWithLogitsLoss
Class imbalance handling	Pos-weight (computed from training set)
Optimizer	AdamW
Learning rate	$(8 * 10^{-4})$
Learning rate scheduler	ReduceLROnPlateau
Loss	BCE Focal ($\gamma = 2.0$, $\alpha = 0.35$)
Weight decay	$(1*10^{-4})$
Positive class weight	3.2917
Regularization	Dropout + Residual connections + LayerNorm
Total parameters	1,663,265

Table 4. Performance of DR-GATv2 on test set

Metric	Value
AUPRC	0.9149
AUROC	0.9625
F1-score	0.8192
Precision	0.8715
Recall	0.7729
MCC	0.7674

Note: AUROC = Receiver Operating Characteristic Curve; AUPRC = Area Under the accuracy-Recall Curve; MCC = Matthews Correlation Coefficient.

Table 5 presents a comparative analysis of the proposed DR-GATv2 model against various sampling methods for predicting PPIS. The results indicate a significant and consistent improvement by DR-GATv2 across evaluation metrics. The first approach, DeepPPISP [37], primarily employs PSSM and structural descriptors and is executed using a conventional deep neural network. Although successful upon publication, its performance remains limited, shown by an AUROC of 0.653 and an MCC of just 0.167. This indicates that traditional brain architectures struggle to reflect complex spatial interactions in protein structures. GraphPPIS [5] directly modeled residue interactions from protein structures with graph convolutional networks, resulting in substantial improvements compared to DeepPPISP. The AUROC increased from 0.653 to 0.786, and the MCC almost doubled from 0.167 to 0.333. These results underscore the need of using graph-based models to simulate three-dimensional residue interactions. Contemporary graph attention-based methodologies, such EDG-PPIS [38] and

AGAT-PPIS [39], enhance prediction accuracy by including attention mechanisms and more sophisticated structural representations. AGAT-PPIS had an AUROC of 0.867 and an MCC of 0.484. EDG-PPIS outperformed AGAT-PPIS slightly, with an AUROC of 0.871 and an MCC of 0.487. The improvements in performance from these methods underscore the efficacy of attention-based message forwarding and geometric-aware protein modeling. The proposed DR-GATv2 model achieves superior performance across all specified metrics, with an AUROC of 0.8893, an AUPRC of 0.6717, and an MCC of 0.5655. In comparison to the most robust baseline EDG-PPIS, DR-GATv2 improves AUROC by 1.83 percentage points, AUPRC by 9.47 percentage points, and MCC by 7.85 percentage points. The enhancement of AUPRC is particularly significant due to the high imbalance in datasets for PPIS prediction, making precision-recall performance more valuable than AUROC alone. The superior performance of DR-GATv2 may be attributed to many factors. Initially, ESM-2 embeddings provide comprehensive contextual and evolutionary representations derived from extensive protein sequence data, allowing the model to identify functional patterns that are difficult to get from manually crafted descriptions alone. Secondly, structural attributes obtained from DSSP provide explicit insights into secondary structure and solvent accessibility, both of which are significantly associated with interface formation. Third, the physical attributes of residues provide additional biological insights essential for molecular recognition mechanisms. The Residual GATv2 design, together with Jumping Knowledge aggregation, enhances information transmission in protein graphs while preserving local neighborhood characteristics and higher-level structural linkages. In conclusion, our findings indicate that integrating protein language model embedding with structural and physicochemical descriptors within a residual graph attention framework provides a highly effective approach to residue-level protein-protein interaction site prediction, surpassing current state-of-the-art performance on the Test_60 benchmark dataset.

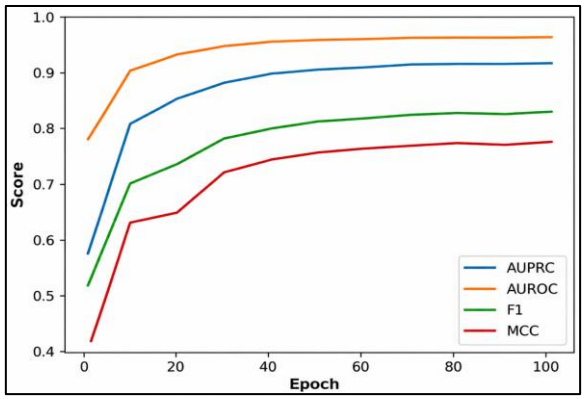


Figure 2. DR-GATv2 performance over 100 epochs

Table 5. DR-GATv2 Performance in comparison with other recent PPI site prediction models on the Test_60 benchmark dataset

Model	Year	Input Features	Architecture	AUROC	AUPRC	MCC
DeepPPISP	2020 [37]	PSSM + structure	Deep NN	0.653	0.276	0.167
GraphPPIS	2022 [5]	PSSM + DSSP	GCN	0.786	0.429	0.333
EDG-PPIS	2025 [38]	Evolutionary + structural features	Equivariant Dual-Scale Graph Network	0.871	0.577	0.487
AGAT-PPIS	2023 [39]	geometric features	Augmented GAT	0.867	0.574	0.484
DR-GATv2	**2026**	**ESM-2+DSSP+Physiochemical**	**Residual GATv2+JK**	**0.8893**	**0.6717**	**0.5655**

Note: AUROC = Receiver Operating Characteristic Curve; AUPRC = Area Under the accuracy-Recall Curve; MCC = Matthews Correlation Coefficient; PPI = protein-protein interactions; PPIS = protein-protein interaction sites.

3.3 Ablation studies

Ablation studies to validate the contribution of each component in the Deep Residual GATv2 architecture. The model performance was evaluated using AUROC and AUPRC on validation and independent test datasets. The study investigates how feature groups and architectural aspects impact performance and resilience.

3.4 Feature ablation experiments

To ablate the effects of input factors, we performed experiments where we excluded ESM-2 embeddings, DSSP structural features, and physicochemical descriptors from the model. The full model combines sequence-derived and structural features to include both local and global residue information. When ESM-2 embeddings were not used, the performance dropped significantly, suggesting the need for contextual sequence representation for interaction site detection. Removing DSSP structural features leads to a drop in performance, demonstrating the importance of secondary structure and solvent accessibility. Removing both types of features causes the biggest drop in performance of the model, suggesting that they are complementary features. Enhanced prediction accuracy and robustness with ESM-2 and DSSP features.

3.5 Experimental ablation on model architecture

We tested the proposed concept on many model configurations by modifying some important network components. First, we replaced GATv2 layers with GCN and GAT. GATv2 outperforms GCN and GAT in general, which demonstrates their ability to model complex interactions through dynamic attention mechanisms. Second, we cut the remaining links from the network to see what effect they had. The performance degrades significantly, which shows that residual learning benefits deeper structures and avoids oversmoothing. Third, we examined Jumping Knowledge (JK) concatenation using just the last layer representation. The performance drop without JK aggregation suggests that multi-scale features within layers are crucial for capturing both local and global structural context. The GATv2 layers, residual connections, and JK aggregation model achieve the best performance in terms of accuracy, stability, and generalization. Tables 6–8 present the ablation study on the contribution of input features and architectural components. The effectiveness of the proposed Deep Residual GATv2 architecture is proved by the fact that the performance always degrades when important parts of the model or features are removed.

Table 6. Features of ablation study

Feature Group	AUROC	AUPRC	MCC
Full Model (ESM-2 + DSSP + Physicochemical)	**0.9625**	**0.9149**	**0.767**
Without (ESM-2)	0.9180	0.8425	0.701
Without DSSP features	0.9342	0.8651	0.724
Without Physicochemical features	0.9456	0.8823	0.741
Without (ESM-2+DSSP)	0.8894	0.8017	0.652

Note: AUROC = Receiver Operating Characteristic Curve; AUPRC = Area Under the Accuracy-Recall Curve; MCC = Matthews Correlation Coefficient.

Table 7. Architecture of DR-GATv2 ablation study

Model Variant	AUROC	AUPRC	MCC
GCN (no attention)	0.8812	0.7924	0.641
GAT (standard attention)	0.9125	0.8356	0.689
GATv2 (no residual)	0.9354	0.8712	0.726
GATv2+Residual (no JK)	0.9487	0.8926	0.751
Full Model (Residual GATv2 + JK)	**0.9625**	**0.9149**	**0.767**

Note: AUROC = Receiver Operating Characteristic Curve; AUPRC = Area Under the Accuracy-Recall Curve; MCC = Matthews Correlation Coefficient.

Table 8. Architecture of DR-GATv2 ablation study

Configuration	Resi-Dual	JK	GA-Tv2	AUROC	AUPRC	MCC
Baseline GAT	✗	✗	✗	0.912	0.835	0.689
With GATv2	✗	✗	✓	0.935	0.871	0.726
With Residual	✓	✗	✓	0.948	0.892	0.75
Full Model	✓	✓	✓	**0.962**	**0.914**	**0.767**

Note: AUROC = Receiver Operating Characteristic Curve; AUPRC = Area Under the Accuracy-Recall Curve; MCC = Matthews Correlation Coefficient.

We conducted a series of ablation experiments to evaluate the impact of different feature groups and architectural elements in the proposed DR-GAT framework. Our results demonstrate that both the feature engineering technique and the network architecture significantly influence overall predictive performance. The feature ablation results indicate that the whole model, including ESM-2 embeddings, DSSP structural descriptors, and physicochemical features, achieves optimal performance with an AUROC of 0.9625, an AUPRC of 0.9149, and an MCC of 0.7674. The elimination of any feature category leads to a substantial decline in performance, indicating that the three feature categories provide complementary information for predicting PPIS. The ESM-2 embeddings are the most significant among all feature groupings. The most pronounced decline in performance occurs in the absence of ESM-2, with the AUROC decreasing from 0.9625 to 0.9180 and the MCC diminishing from 0.7674 to 0.7012. This result underscores the significance of evolutionary and contextual sequence representations derived from extensive protein language models. These embeddings may include long-range sequence correlations and functional motifs that are difficult to derive from only handcrafted characteristics. Structural features derived from DSSP are also significant. Eliminating DSSP data reduces the AUROC to 0.9342 and the MCC to 0.7248, indicating that secondary structure and solvent accessibility provide valuable structural context for identifying interface residues. Protein-protein interactions occur on protein surfaces, where solvent exposure and local structural conformation are particularly significant. The exclusion of physicochemical descriptors results in a little decrease in performance (AUROC = 0.9456, MCC = 0.7415), suggesting that residue-specific characteristics such as hydrophobicity, polarity, charge, and volume provide additional discriminative information. These features are intrinsically associated with molecule recognition and interface formation. The most significant decline in performance occurs when both ESM-2 and DSSP features are removed simultaneously (AUROC = 0.8894, MCC = 0.6523). This significant decline illustrates the strong complementarity between structure-derived descriptors and sequence-derived contextual representations. These results together underscore the need to integrate evolutionary, structural, and physicochemical information for accurate prediction of PPIS. The proposed Residual GATv2 with Jumping Knowledge (JK) aggregation is shown to be effective via architectural ablation analysis. The baseline GCN model has the worst performance (AUROC = 0.8812, MCC = 0.6417), indicating that fundamental neighborhood aggregation fails to encapsulate the complex structural connections inside the protein graphs. Upon substituting the GCN with the traditional GAT, the AUROC is 0.9125, and the MCC is 0.6893. The enhancement confirms the advantage of attention processes, enabling the model to assign differing relevance weights to adjacent residues rather than treating all neighbors uniformly. The further enhancement is achieved by the use of GATv2, resulting in an AUROC of 0.9354 and an MCC of 0.7265. GATv2 provides a more nuanced attention formulation capable of learning dynamic attention patterns, facilitating a more accurate representation of residue interactions within protein structures. Residual connections enhance performance (AUROC = 0.9487, MCC = 0.7512). The residual pathways facilitate the gradient's descent down the lower levels of the graph, preserving low-level residual information and reducing the oversmoothing phenomenon often seen in graph neural networks. The most effective model is the whole DR-GAT model, which incorporates both residual connections and Jumping Knowledge aggregation (AUROC = 0.9625, AUPRC = 0.9149, MCC = 0.7674). The JK mechanism facilitates the integration of representations from various graph levels, thus providing the model with access to both local and global structural contexts. This is particularly pertinent in the context of proteins, where contact determinants may be encoded in the immediate proximity of residues and in long-range structural interactions. The ablation studies indicate that the superior performance of DR-GAT arises from the synergistic integration of diverse multi-source residue representations and an advanced graph attention framework. ESM-2 embeddings provide the most substantial singular data, but DSSP and physicochemical descriptors furnish further structural and biological insights. The architectural elements of GATv2, including residual connections and Jumping Knowledge aggregation, consistently improve performance, with the model consistently achieving the highest scores across all evaluation metrics.

3.6 Visualization of predicted interaction sites

The attention weights derived from the GATv2 layers were analyzed and mapped into the corresponding protein structures to understand the model's predictions. The results of the visualization show that the residues with high prediction scores are frequently in agreement with the experimentally verified interaction sites. Moreover, clusters of predicted binding residues are mainly located on regions of the protein surface, indicating that the model captures the spatial

arrangement of the interaction interfaces well. The results indicate that the proposed strategy provides accurate predictions and gives physiologically relevant insights into protein interactions.

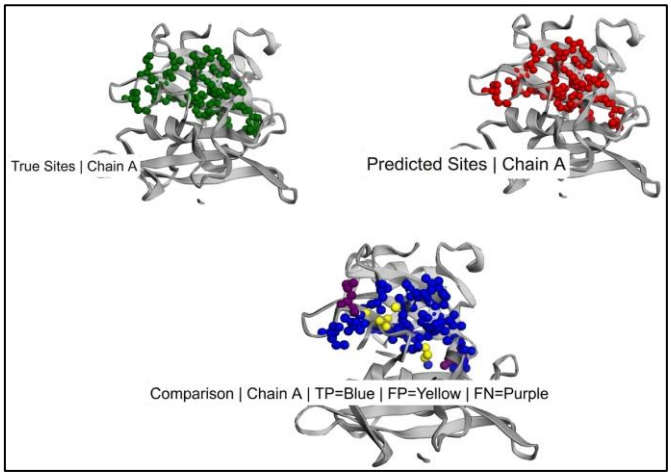


Figure 3. Comparison of protein 1A07 chain A
Note: TP = true positives; FP = false positives; FN = false negatives.

Figure 3 shows the predicted and experimental binding sites for a sample protein chain(1A07) at the residue level. The predicted interface residues match the true binding sites, demonstrating that DR-GATv2 may capture protein interaction interface structural properties. Most anticipated residues (blue true positives) are in surface-exposed locations, generating continuous contact patches. This work shows that the model learns protein interface structural properties such as solvent accessibility, local residue density, and spatial neighborhood relationships. Aggregating true positives shows that the model recognizes interface areas as cohesive structural units rather than separate residues. Purple false negative residues are rare and predominantly near the interface periphery. These residues may have less distinct structural signatures or evolutionary conservation than core interface residues, making them harder to distinguish from non-interface surface residues. Few false negatives suggest that the model can account for much of the physiologically relevant interaction surface. The result shows that DR-GATv2 represents protein interactions structurally. True positives are common in the central interface region, and most errors occur near interface boundaries, suggesting that the model identifies the core interaction surface but struggles with residues at interface-non-interface areas.

4. CONCLUSION AND FUTURE WORKS

This study presents DR-GATv2, a deep residual graph attention framework for predicting PPIS at the residue level. The proposed model integrates ESM-2 contextual embeddings, structural descriptors derived from DSSP, and physicochemical characteristics of residues inside a Residual GATv2 architecture using Jumping Knowledge aggregation. We represent proteins as residue-level graphs that accurately encapsulate local structural contexts and long-range interactions among residues. The experimental evaluation on the Test60 benchmark demonstrated that DR-GATv2 surpassed many leading methodologies, achieving an AUROC of 0.8893, an AUPRC of 0.6717, and an MCC of 0.5655. The results indicate that integrating protein language model embedding with graph-based structural learning enhances the precision of PPI site prediction. Future research will explore cross-protein interaction graphs, geometry-informed graph attention mechanisms, and multi-task learning frameworks for the simultaneous prediction of interaction sites and partners. Moreover, the incorporation of extensive protein data and high-quality projected structures may enhance the model's generalizability and practical use in structural biology, protein engineering, and drug development.

REFERENCES

[1] Zhang, J., Kurgan, L. (2019). SCRIBER: Accurate and partner type-specific prediction of protein-binding residues from proteins sequences. Bioinformatics, 34(14): i343-i353. https://doi.org/10.1093/bioinformatics/btz324

[2] Li, Y., Golding, G.B., Ilie, L. (2020). DELPHI: Accurate deep ensemble model for protein interaction sites prediction. Bioinformatics, 37(7): 896-904. https://doi.org/10.1093/bioinformatics/btaa750

[3] Northey, T.C., Barešić, A., Martin, A.C.R. (2017). IntPred: A structure-based predictor of protein-protein interaction sites. Bioinformatics, 34(2): 223-229. https://doi.org/10.1093/bioinformatics/btx585

[4] Sriwastava, B.K., Basu, S., Maulik, U., Plewczynski, D. (2013). PPIcons: Identification of protein-protein interaction sites in selected organisms. Journal of Molecular Modeling, 19(9): 4059-4070. https://doi.org/10.1007/s00894-013-1886-9

[5] Yuan, Q., Chen, J., Zhao, H., Zhou, Y., Yang, Y. (2021). Structure-aware protein-protein interaction site prediction using deep graph convolutional network. Bioinformatics, 38(1): 125-132. https://doi.org/10.1093/bioinformatics/btab643

[6] Manfredi, M., Savojardo, C., Martelli, P.L., Casadio, R. (2023). ISPRED-SEQ: Deep neural networks and embeddings for predicting interaction sites in protein sequences. Journal of Molecular Biology, 435(14): 167963. https://doi.org/10.1016/j.jmb.2023.167963

[7] Krapp, L.F., Abriata, L.A., Rodriguez, F.C., Peraro, M.D. (2023). PeSTo: Parameter-free geometric deep learning for accurate prediction of protein binding interfaces. Nature Communications, 14(1): 1-11. https://doi.org/10.1038/s41467-023-37701-8

[8] Wu, H., Han, J., Zhang, S., Xin, G., Mou, C., Liu, J. (2023). Spatom: A graph neural network for structure-based protein-protein interaction site prediction. Briefings in Bioinformatics, 24(6): bbad345. https://doi.org/10.1093/bib/bbad345

[9] Mahbub, S., Bayzid, S. (2022). EGRET: Edge aggregated graph attention networks and transfer learning improve protein-protein interaction site prediction. Briefings in Bioinformatics, 23(2): bbab578. https://doi.org/10.1093/bib/bbab578

[10] Smith, Z., Strobel, M., Vani, B.P., Tiwary, P. (2024). Graph attention site prediction (GrASP): Identifying druggable binding sites using graph neural networks with attention. Journal of Chemical Information and Modeling, 64(7): 2637-2644. https://doi.org/10.1021/acs.jcim.3c01698

[11] Meng, L., Zhang, H. (2024). GACT-PPIS: Prediction of

protein-protein interaction sites based on graph structure and transformer network. International Journal of Biological Macromolecules, 283: 137272. https://doi.org/10.1016/j.ijbiomac.2024.137272

[12] Hoseinnia, F., Ghatee, M., Chehreghani, M.H. (2025). Mitigating over-smoothing in graph neural networks for node classification through adaptive early embedding and biased DropEdge procedures. Knowledge-Based Systems, 320: 113615. https://doi.org/10.1016/j.knosys.2025.113615

[13] Albu, A., Bocicor, M., Czibula, G. (2023). MM-StackEns: A new deep multimodal stacked generalization approach for protein-protein interaction prediction. Computers in Biology and Medicine, 153: 106526. https://doi.org/10.1016/j.compbiomed.2022.106526

[14] Feng, Y., Yang, X., Guan, Y., Zhang, J., Yang, H., Wang, Z., Cheng, Q. (2025). KG-MACNF: A nonlinear cross-modal fusion model for predicting drug-target interactions via multi-relational embedding and fine-grained structure. PLOS ONE, 20(9): e0331037. https://doi.org/10.1371/journal.pone.0331037

[15] Westbrook, J., Feng, Z., Jain, S., et al. (2002). The Protein Data Bank: Unifying the archive. Nucleic Acids Research, 30(1): 245-248. https://doi.org/10.1093/nar/30.1.245

[16] Sun, H., Wang, J., Wu, H., Lin, S., Chen, J., Wei, J., Lv, S., Xiong, Y., Wei, D. (2023). A multimodal deep learning framework for predicting PPI-modulator interactions. Journal of Chemical Information and Modeling, 63(23): 7363-7372. https://doi.org/10.1021/acs.jcim.3c01527

[17] Cong, H., Liu, H., Cao, Y., Liang, C., Chen, Y. (2023). Protein-protein interaction site prediction by model ensembling with hybrid feature and self-attention. BMC Bioinformatics, 24(1): 1-21. https://doi.org/10.1186/s12859-023-05592-7

[18] Roche, R., Moussad, B., Shuvo, H., Bhattacharya, D. (2023). E(3) equivariant graph neural networks for robust and accurate protein-protein interaction site prediction. PLOS Computational Biology, 19(8): e1011435. https://doi.org/10.1371/journal.pcbi.1011435

[19] Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics, 28(23): 3150-3152. https://doi.org/10.1093/bioinformatics/bts565

[20] Kabsch, W., Sander, C. (1983). How good are predictions of protein secondary structure? FEBS Letters, 155(2): 179-182. https://doi.org/10.1016/0014-5793(82)80597-8

[21] Li, H., Gong, X., Yu, H., Zhou, C. (2018). Deep neural network based predictions of protein interactions using primary sequences. Molecules, 23(8): 1923. https://doi.org/10.3390/molecules23081923

[22] Yuen, H.Y., Jansson, J. (2023). Normalized L3-based link prediction in protein-protein interaction networks. BMC Bioinformatics, 24(1): 1-28. https://doi.org/10.1186/s12859-023-05178-3

[23] Jha, K., Saha, S., Singh, H. (2022). Prediction of protein-protein interaction using graph neural networks. Scientific Reports, 12(1): 1-12. https://doi.org/10.1038/s41598-022-12201-9

[24] Jha, K., Karmakar, S., Saha, S. (2023). Graph-BERT and language model-based framework for protein-protein interaction identification. Scientific Reports, 13(1): 1-10. https://doi.org/10.1038/s41598-023-31612-w

[25] McPartlon, M., Xu, J. (2023). Deep learning for flexible and site-specific protein docking and design. bioRxiv. https://doi.org/10.1101/2023.04.01.535079

[26] Gu, Z., Luo, X., Chen, J., Deng, M., Lai, L. (2023). Hierarchical graph transformer with contrastive learning for protein function prediction. Bioinformatics, 39(7): btad410. https://doi.org/10.1093/bioinformatics/btad410

[27] Gao, X., Cao, H., Li, J., Qiu, J., Chen, G., Heng, P. (2025). Gated-GPS: Enhancing protein-protein interaction site prediction with scalable learning and imbalance-aware optimization. Briefings in Bioinformatics, 26(3): bbaf248. https://doi.org/10.1093/bib/bbaf248

[28] Sinha, S., Ohashi, H., Nakamura, K. (2020). Class-wise difficulty-balanced loss for solving class-imbalance. In Asian Conference on Computer Vision, Kyoto, Japan.

[29] Zhao, S., Cui, Z., Zhang, G., Gong, Y., Su, L. (2024). MGPPI: Multiscale graph neural networks for explainable protein-protein interaction prediction. Frontiers in Genetics, 15: 1440448. https://doi.org/10.3389/fgene.2024.1440448

[30] Balogh, O.M., Benczik, B., Horváth, A., Pétervári, M., Csermely, P., Ferdinandy, P., Ágg, B. (2022). Efficient link prediction in the protein-protein interaction network using topological information in a generative adversarial network machine learning model. BMC Bioinformatics, 23(1): 1-19. https://doi.org/10.1186/s12859-022-04598-x

[31] Bai, Y., Yang, E., Han, B., et al. (2021). Understanding and improving early stopping for learning with noisy labels. In NIPS'21: Proceedings of the 35th International Conference on Neural Information Processing Systems, New York, United States, pp. 24392-24403. https://doi.org/10.48550/arXiv.2106.15853

[32] Yan, Y., Huang, S. (2021). Accurate prediction of inter-protein residue-residue contacts for homo-oligomeric protein complexes. Briefings in Bioinformatics, 22(5): bbab038. https://doi.org/10.1093/bib/bbab038

[33] Volzhenin, K., Bittner, L., Carbone, A. (2024). SENSE-PPI reconstructs interactomes within, across, and between species at the genome scale. iScience, 27(7): 110371. https://doi.org/10.1016/j.isci.2024.110371

[34] Reim, T., Hartebrodt, A., Blumenthal, D.B., Bernett, J., List, M. (2025). Deep learning models for unbiased sequence-based PPI prediction plateau at an accuracy of 0.65. Bioinformatics, 41: i590-i598. https://doi.org/10.1093/bioinformatics/btaf192

[35] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12: 2825-2830. https://doi.org/10.48550/arXiv.1201.0490

[36] Saito, T., Rehmsmeier, M. (2015). The Precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3): e0118432. https://doi.org/10.1371/journal.pone.0118432

[37] Zeng, M., Zhang, F., Wu, F., Li, Y., Wang, J., Li, M. (2019). Protein-protein interaction site prediction through combining local and global features with deep neural networks. Bioinformatics, 36(4): 1114-1120. https://doi.org/10.1093/bioinformatics/btz699

[38] Zhang, Z., Li, Z., Li, W., Zhang, Q., Xiao, J., Ding, S., Han, Y. (2025). EDG-PPIS: An equivariant and dual-scale graph network for protein-protein interaction site prediction. BMC Genomics, 26(1): 1-15. https://doi.org/10.1186/s12864-025-12084-w

[39] Zhou, Y., Jiang, Y., Yang, Y. (2023). AGAT-PPIS: A novel protein-protein interaction site predictor based on augmented graph attention network with initial residual and identity mapping. Briefings in Bioinformatics, 24(3): bbad122. https://doi.org/10.1093/bib/bbad122