

Multimodal Graph-based Transformer Framework for Biomedical Relation Extraction

Sriram Pingali, Shweta Yadav*, Pratik Dutta*, Sriparna Saha

Indian Institute of Technology Patna, India

{1801cs37, shweta.pcs14, pratik.pcs16, sriparna}@iitp.ac.in

Abstract

The recent advancement of pre-trained Transformer models has propelled the development of effective text mining models across various biomedical tasks. However, these models are primarily learned on the textual data and often lack the domain knowledge of the entities to capture the context beyond the sentence. In this study, we introduced a novel framework that enables the model to learn multi-omics biological information about entities (proteins) with the help of additional multi-modal cues like molecular structure. Towards this, rather developing modality-specific architectures, we devise a generalized and optimized graph based multi-modal learning mechanism that utilizes the GraphBERT model to encode the textual and molecular structure information and exploit the underlying features of various modalities to enable the end-to-end learning. We evaluated our proposed method on Protein-Protein Interaction task from the biomedical corpus, where our proposed generalized approach is observed to be benefited by the additional domain-specific modality.

1 Introduction

The biomedical scientific articles hold the valuable knowledge of biomedical entities (such as protein, drug, gene) and their relationships. However, with the exponential increase in the volume of biomedical articles (Lu, 2011), it is imperative to advance the development of an accurate biomedical text mining tool to extract and curate meaningful information from huge unstructured texts automatically.

One of the cardinal tasks in biomedical document processing is Protein-protein interaction (PPI), where the relation (*'interaction'* or *'non-interaction'*) between two protein mentions is identified from the given biomedical text. The knowl-

edge about protein interactions is critical in understanding the biological processes, such as signaling cascades, translations and metabolism, that are regulated by the interactions of proteins that alter proteins to modulate their stability (Elangovan et al., 2020).

Majority of the existing works on PPI in the literature primarily focused only on the textual information present in the biomedical article. However, these approaches lack in capturing (1) multi-omics biological information regarding protein interactions, and (2) genetic and structure information of the proteins. A few works (Dutta and Saha, 2020; Asada et al., 2018; Jha et al., 2020; Jha and Saha) have been reported in the literature where the researchers have considered different modalities of the biomedical corpus. However, these multi-modal architectures are modality-specific and thus are very complex. Hence, there is a surge to develop a generalized and optimized model that can understand all the modalities rather than developing various architectures for different modalities.

Towards this, we explore Graph-based Transformer model (GraphBERT) (Zhang et al., 2020) to learn the modality independent graph representation. This enables the model to acquire the joint knowledge of both the modalities (textual and protein structure) under a single learning network. The main contributions of this work are:

1. Besides the textual information of the biomedical corpus, we have also utilized protein atomic structural information while identifying the protein interactions.
2. Developed a generalized modality-agnostic approach that is able to learn the feature representations of both the textual and the protein-structural modality.
3. Our analysis reveals that addition of protein-structure modality increases the efficiency of

*These authors contributed equally to this work.

model in identifying the interacted protein mentions.

Related Work: Existing studies have adopted traditional statistical and graphical methods (Miyao et al., 2008; Chang et al., 2016) to identify the protein interactions from the textual content. Later, with the success of deep learning, several techniques based on Convolutional Neural Network (Choi, 2018; Peng and Lu, 2017; Ekbal et al., 2016), Recurrent Neural Network (Hsieh et al., 2017; Ahmed et al., 2019), Long Short Term Memory network (Yadav et al., 2019; Ningthoujam et al., 2019; Yadav et al., 2020), and language models (Yadav et al., 2021) based methods are proposed for extracting the relationships from biomedical literature and clinical records. Fei et al. (2020) proposed a span-graph neural model for jointly extracting overlapping entity relationships from biomedical text. The recent advancement of the Transformer model (Lee et al., 2020; Beltagy et al., 2019) in the biomedical domain has also led to significant performance improvement in biomedical relation extraction task (Giles et al., 2020). Recently, the use of multi-modal dataset in BioNLP domain (Dutta and Saha, 2020; Asada et al., 2018) draws the attention of the researchers due to its better performance than the traditional approaches. In contrast, our model is independent of handling multiple modalities without relying on modality-specific architectures.

2 Proposed Method

In this section, we introduce our proposed method and its detailed implementation. The proposed deep multi-modal architecture is illustrated in Figure-1, that consists of four main components: **(1) Multi-modal Graph Constructor**, **(2) Multi-modal Graph Fusion**, **(3) Multi-modal Graph Encoder**, **(4) PPI Predictor**. Below we briefly describe each of the model components.

Problem Statement: Given a biomedical input text $S = \{w_1, w_2, \dots, w_n\}$ having n words, and a pair of protein mentions $p_1, p_2 \in S$, we aim to predict, whether the protein mentions will ‘interact’ or ‘non-interact’.

2.1 Multi-modal Graph Constructor

This component consists of two distinct graph constructors for two different modalities, which are Textual Graph Constructor and Protein Structure Graph Constructor. The former, constructs

the graph by considering the textual content that aims to capture the lexical and contextual information present in the input. The later, exploits the atomic structure (3D PDB structure) of the protein molecules to build the graph.

Textual Graph Constructor: To generate the textual graph, we begin by first constructing the vocabulary from the training corpus. For each input text S , we use one-hot-encoding mechanism to encode them as a vector representation $R_S \in \mathcal{R}^{|V|}$. However, the representation R_S suffers from the data sparsity as the vocabulary size can become very large for the entire training corpus. To deal with this, we utilized the Principal Component Analysis (PCA) (Wold et al., 1987) to reduce the vector dimensionality. The textual graph $\mathcal{G}_T = \{\mathcal{V}_T, \mathcal{E}_T\}$ is formulated by the nodes $\mathcal{V}_T = \{\hat{R}_{S_1}, \hat{R}_{S_2}, \dots, \hat{R}_{S_{|N|}}\}$, where $|N|$ is the number of input sentences in the training dataset and $\hat{R}_{S_i} \in \mathcal{R}^{|\hat{V}|}$ is reduced vector representation of size $|\hat{V}|$ for sentence S_i . The link $e_{i,j}$ between nodes \hat{R}_{S_i} and \hat{R}_{S_j} is determined by the common entities (*protein*) present in both the sentences S_i and S_j , if there is no common entity, then link does not exist between the nodes. The edges $\mathcal{E}_T = \{e_{i,j} \mid i, j \in \mathcal{V}_T, \text{ and } \textit{protein} \in i, j\}$ are the set of all the links that exist between any two nodes in the graph, \mathcal{G}_T .

Protein Structure Graph Constructor: For the protein structural modality, we created a graph where each node represents an atom and the edge represents the connection between the atoms. To obtain the atomic information about the proteins, first we have mapped the proteins into genes and utilized the PDB (Protein Data Bank)¹ for each associated protein mention. Each protein information obtained from PDB consists of set of atoms $\{a_1, a_2, \dots, a_A\}$, and a node feature matrix, $N_p \in \mathcal{R}^{A \times d_p}$. The node feature matrix for each protein k undergoes the convolutional operation $CNN(\cdot)$ followed by the max-pooling operation, $pool(\cdot)$. Formally, $P_k = pool(relu(CNN(N_{p_k})))$. The final protein representation, P_{S_i} , for both the proteins present in the given input sentence S_i is computed as follows: $P_{S_i} = P_1 \oplus P_2$. Following this, the protein structure graph $\mathcal{G}_P = \{\mathcal{V}_P, \mathcal{E}_P\}$ is formulated by the nodes $\mathcal{V}_P = \{P_{S_1}, P_{S_2}, \dots, P_{S_{|N|}}\}$, where $|N|$ is the number of input sentences in the training dataset and $P_{S_i} \in \mathcal{R}^{d_s}$ is the protein structure

¹<https://www.rcsb.org/>

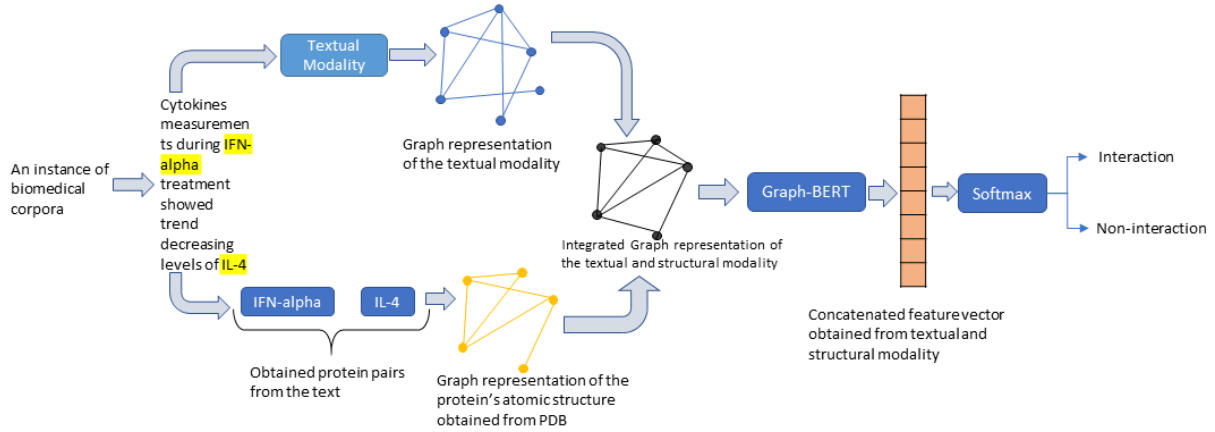


Figure 1: An outline of the proposed deep multi-modal architecture for PPI.

representation of size d_s for sentence S_i .

2.2 Multi-modal Graph Fusion

In this component, we fused the textual graph \mathcal{G}_T and protein structure graph \mathcal{G}_P with the aim of generating a joint representation that is capable of capturing the contextual, lexical, and multi-omics information. Towards this, we expanded the node information of textual graph with the node information obtained in the protein-structure graph. Specifically, we created a multi-modal graph \mathcal{G} with the nodes \mathcal{V} having concatenated vector representations from the respective nodes of textual graph and protein structure graph. Formally,

$$\mathcal{V}_i = \hat{R}_{S_i} \oplus P_{S_i} \quad (1)$$

The link information remains intact in the multi-modal graph fusion, thus, $\mathcal{E} = \mathcal{E}_T$.

2.3 Multi-modal Graph Encoder

Majority of the existing works on multi-modal relation extraction have treated multiple modalities separately and exploited the modality-specific architectures to learn the respective feature representations. However, these strategies inhibit the learning of inherent shared complementary features, that are often present across the modalities. To address this, we present an end-to-end multi-modality learning mechanism that exploits the single expanded multi-modal graph (obtained from the Multi-modal Graph Expansion component) with the Graph-based Transformer encoder. Specifically, we utilized the Graph-BERT (Zhang et al., 2020) encoder over the other dominants graph neural networks (GNNs) primarily due to its capability to avoid the (a) *suspended animation problem* (Zhang

and Meng, 2019), and (b) *over-smoothing problem* (Li et al., 2018) that hinders the applications of GNNs for deep graph representation learning tasks. For a given multi-modal graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the set of nodes (\mathcal{V}) and edges (\mathcal{E}), Graph-BERT sampled set of graph batches for all the nodes as set $\mathcal{G} = \{g_1, g_2, \dots, g_{|\mathcal{V}|}\}$. For all the nodes v_j in sub-graph g_i , the Graph-BERT computes raw feature vector embedding e_j^x , role embedding e_j^r , position embedding e_j^p and distance embedding e_j^d . The initial input vector for node v_j is computed as follows: $h_j^{(0)} = e_j^x + e_j^r + e_j^p + e_j^d$. Furthermore, the initial input vectors for all the nodes in g_i can be organized into a matrix $H^{(0)} = [h_{i,1}^{(0)}, h_{i,1}^{(0)}, \dots, h_{i,k}^{(0)}]^\top$, where k is a hyper-parameter. The Graph-Transformer (Zhang et al., 2020) computes the vector representation of D layers of transformers. The final feature (z_i) for node v_j is computed as follows:

$$\begin{aligned} H^{(0)} &= [h_{i,1}^{(0)}, h_{i,1}^{(0)}, \dots, h_{i,k}^{(0)}]^\top \\ H^{(l)} &= G\text{-Transformer} \left(H^{(l-1)} \right) \\ z_i &= \sum_{m=0}^{m=D} H^{(m)} \end{aligned} \quad (2)$$

2.4 PPI Predictor

The final feature (z_i) of each node i is used to predict the PPI category. Towards this, we employed a feed-forward network with *softmax* activation layer to predict the input text into one of the two classes *interaction* or *non-interaction*. Formally,

$$\text{prob}(c = \text{interact} | \mathcal{G}, S, \theta) = \frac{\exp(\mathbf{W}^T z_i^{\text{interact}} + \mathbf{b})}{\sum_{k=1}^K \exp(\mathbf{W}^T z_i^k + \mathbf{b})} \quad (3)$$

where, \mathbf{W} and \mathbf{b} are the weight matrix and bias vector, respectively. K denotes total number of

	Precision	Recall	F-score
Proposed Model	80.84	80.87	80.86
Dutta and Saha (2020)	69.04	88.49	77.54
Yadav et al. (2019)	80.81	82.57	81.68
Hua and Quan (2016)	73.40	77.00	75.20
Choi (2018)	72.05	77.51	74.68
Qian and Zhou (2012)	63.61	61.24	62.40
Peng and Lu (2017)	62.70	68.2	65.30
Zhao et al. (2016)	53.90	72.9	61.60
Tikk et al. (2010)	53.30	70.10	60.00
Li et al. (2015)	72.33	74.94	73.61
Choi and Myaeng (2010)	74.50	70.90	72.60

Table 1: Comparative analysis of the proposed multi-modal approach with state-of-the-art techniques for BioInfer dataset.

		Precision	Recall	F-score
HPRD50	Textual Modality	90.44	92.18	91.28
	Proposed Model	95.47	94.69	95.06
BioInfer	Textual Modality	78.49	79.78	79.06
	Proposed Model	80.84	80.87	80.86

Table 2: Results by uni-modal and multimodal approaches

distinct classes, which are ‘*interaction*’ and ‘*non-interaction*’ in our case.

3 Datasets and Experimental Analysis

Datasets: In this work, we have collected two exemplified multi-modal protein protein interaction datasets (Dutta and Saha, 2020). In these datasets, the authors exemplified two popular benchmark PPI corpora, namely BioInfer² and HPRD50³.

Experimental Setup We have utilized the pre-trained Graph-BERT⁴ in our experiment. The initial vocabulary for BioInfer and HPRD50 datasets are 6561 and 1277, respectively. We have projected them into 1000 and 1185 dimension vectors using PCA, respectively. We have kept maximum of 5052 and 1185 number of words in both the datasets, respectively. The filter-size of CNN is set to 3, 4. We have obtained 1185 length node feature representation for protein structure graph. The nodes of multi-modal graph received the 2185 sized feature representation. We have obtained 2500 and 25859 number of nodes and edges from HPRD50 dataset and 13675 and 15930214 number of nodes and edges from BioInfer dataset for the Graph-BERT training, respectively. We have used all the hyper-parameters of Graph-BERT model in our proposed

²<http://corpora.informatik.hu-berlin.de/>

³<https://goo.gl/M5tEJj>

⁴<https://github.com/jwzhanggy/Graph-Bert>

	Precision	Recall	F-score
Proposed Model	95.47	94.69	95.06
Dutta and Saha (2020)	94.79	75.21	83.87
Yadav et al. (2019)	79.92	77.58	78.73
Tikk et al. (2010)	68.20	69.80	67.80
Tikk et al. (2010)(with SVM)	68.20	69.80	67.80
Palaga (2009)	66.70	80.20	70.90
Airola et al. (2008)(APG)	64.30	65.80	63.40
Van Landeghem et al. (2008)	60.00	51.00	55.00
Miwa et al. (2009)	68.50	76.10	70.90
Airola et al. (2008)(Co-occ)	38.90	100	55.40
Pyysalo et al. (2008)	76.00	64.00	69.00

Table 3: Comparative analysis of the proposed multi-modal approach with other state-of-the-art approaches for HPRD50 dataset.

model. We have kept following hyper-parameters values: subgraph size = 5, hidden size = 32, attention head number = 2, Transformer layers, D = 2, learning rate = 0.01, weight decay = 5e -4, hidden dropout rate = 0.5, attention dropout rate = 0.3, loss = cross entropy, optimizer = adam (Kingma and Ba, 2014). The hyper-parameters are chosen based on the 5-fold cross-validation experiments on both the datasets.

Results and Analysis: We have compared the performance (*c.f.* Table-1,3) of our proposed model with the existing state-of-the-art methods on PPI for both the datasets. These existing methods are based on different techniques like kernel-based (Choi and Myaeng, 2010; Tikk et al., 2010; Qian and Zhou, 2012; Li et al., 2015), deep neural network-based (Zhao et al., 2016; Yadav et al., 2019), multi-channel dependency-based convolutional neural network model (Peng and Lu, 2017), semantic feature embedding (Choi, 2018), shortest dependency path (Hua and Quan, 2016) and a recent deep multi-modal approach (Dutta and Saha, 2020). It is to be noted that our results on BioInfer and HPRD50 are not directly comparable with the existing approaches as other methods have utilized different test sets for evaluation. From the above comparative study, it is evident that our proposed multi-modal approach identifies the protein interactions in an efficient way and can be further improved in different ways.

Discussion: To analyze the role of each modality, we conducted ablation study as shown in Table 2. We performed the experiments with the textual modality. Here, we could not consider the protein-structural modality alone as it would bring the conflicting labeling relation. For example, consider two sentences that contain same pair of proteins

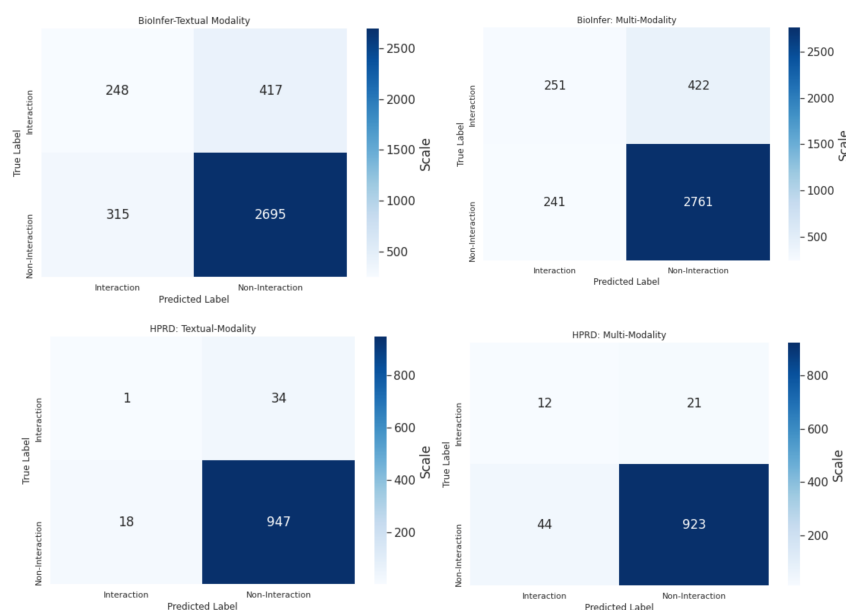


Figure 2: Confusion matrices of our proposed approach on both HPRD and BioInfer datasets with only textual modality and text+structure modality.

but these proteins can have conflicting relations (interacting or non-interacting) depending on the context of sentences in which they appear. Hence, we could not consider the protein-structural modality alone. Though the structural modality is unable to draw any conclusion alone, however the integration of both the modalities demonstrates the improvements (3.78% and 1.8%, in terms of F-score for HPRD50 and BioInfer, respectively) over the textual modality alone.

4 Error Analysis

The comparative confusion matrices with only textual-modality and multi-modality for both the datasets are shown in Figure-2. We have performed error analysis to postulate possible reasons and areas with scope of improvement in our experiments. After careful study on false positive and false negative classes, following observations can be made.

- 1) Instances with a large number of protein mentions in a single sentence can cause misclassification. For example, the maximum number of proteins in any instances of BioInfer and HPRD50 datasets are 26 and 24, respectively. These large number of proteins present in a single instance may lead the network to misclassification.
- 2) Few samples contain repeated mentions of the same protein. This adds noise and might lead to losing useful contextual information.
- 3) To get a consistent graph from molecular structure, the nodes were required to be of the same

length. This is done by padding the vectors with zeros, and when the PDB is not available, a null vector is used for consistency. A better handling of missing data will help in learning the proposed model.

5 Conclusion

This work presents a novel modality-agnostic Graph-based framework to identify the interactions between the proteins. Specifically, we explored two modalities: *textual*, and *molecular structure* that enable the model to learn the domain-specific multi-omics information complementary with the task-specific contextual information. A detailed comparative results and analysis proves that our proposed multi-modal approach can capture underlying molecular structure information without relying on sophisticated modality-specific architectures. Future work aims at extending this study to the other related tasks like drug-drug interactions.

Acknowledgment

Sriparna Saha gratefully acknowledges the Young Faculty Research Fellowship (YFRF) Award, supported by Visvesvaraya Ph.D. Scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India, being implemented by Digital India Corporation (Formerly Media Lab Asia) for carrying out this research.

References

- Mahtab Ahmed, Jumayel Islam, Muhammad Rifayat Samee, and Robert E Mercer. 2019. Identifying protein-protein interaction using tree lstm and structured attention. In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pages 224–231. IEEE.
- Antti Airola, Sampo Pyysalo, Jari Björne, Tapio Pahikkala, Filip Ginter, and Tapio Salakoski. 2008. All-paths graph kernel for protein-protein interaction extraction with evaluation of cross-corpus learning. *BMC bioinformatics*, 9(11):S2.
- Masaki Asada, Makoto Miwa, and Yutaka Sasaki. 2018. Enhancing drug-drug interaction extraction from texts by molecular structure information. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 680–685, Melbourne, Australia. Association for Computational Linguistics.
- Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*.
- Yung-Chun Chang, Chun-Han Chu, Yu-Chen Su, Chien Chin Chen, and Wen-Lian Hsu. 2016. PIPE: a protein-protein interaction passage extraction module for BioCreative challenge. *Database*, 2016. Baw101.
- Sung-Pil Choi. 2018. Extraction of protein-protein interactions (ppis) from the literature by deep convolutional neural networks with various feature embeddings. *Journal of Information Science*, 44(1):60–73.
- Sung-Pil Choi and Sung-Hyon Myaeng. 2010. Simplicity is better: revisiting single kernel ppi extraction. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 206–214. Association for Computational Linguistics.
- Pratik Dutta and Sriparna Saha. 2020. Amalgamation of protein sequence, structure and textual information for improving protein-protein interaction identification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6396–6407, Online. Association for Computational Linguistics.
- Asif Ekbal, Sriparna Saha, Pushpak Bhattacharyya, et al. 2016. A deep learning architecture for protein-protein interaction article identification. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 3128–3133. IEEE.
- Aparna Elangovan, Melissa Davis, and Karin Verspoor. 2020. Assigning function to protein-protein interactions: a weakly supervised biobert based approach using pubmed abstracts. *arXiv preprint arXiv:2008.08727*.
- Hao Fei, Yue Zhang, Yafeng Ren, and Donghong Ji. 2020. A span-graph neural model for overlapping entity relation extraction in biomedical texts. *Bioinformatics*.
- Oliver Giles, Anneli Karlsson, Spyroula Masiala, Simon White, Gianni Cesareni, Livia Perfetto, Joseph Mullen, Michael Hughes, Lee Harland, and James Malone. 2020. Optimising biomedical relationship extraction with biobert: Best practices for data creation. *bioRxiv*.
- Yu-Lun Hsieh, Yung-Chun Chang, Nai-Wen Chang, and Wen-Lian Hsu. 2017. Identifying protein-protein interactions in biomedical literature using recurrent neural networks with long short-term memory. In *Proceedings of the eighth international joint conference on natural language processing (volume 2: short papers)*, pages 240–245.
- Lei Hua and Chanqin Quan. 2016. A shortest dependency path based convolutional neural network for protein-protein relation extraction. *BioMed Research International*, 2016.
- K. Jha and S. Saha. Amalgamation of 3d structure and sequence information for protein-protein interaction prediction. *Sci Rep*, 10.
- Kanchan Jha, Sriparna Saha, and Matloob Khushi. 2020. Protein-protein interactions prediction based on bi-directional gated recurrent unit and multi-modal representation. In *Neural Information Processing - 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 18-22, 2020, Proceedings, Part V*, volume 1333 of *Communications in Computer and Information Science*, pages 164–171. Springer.
- Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.
- Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.
- Lishuang Li, Rui Guo, Zhenchao Jiang, and Degen Huang. 2015. An approach to improve kernel-based protein-protein interaction extraction by learning from large-scale network data. *Methods*, 83:44–50.
- Qimai Li, Zhichao Han, and Xiao-Ming Wu. 2018. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.
- Zhiyong Lu. 2011. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database*, 2011. Baq036.
- Makoto Miwa, Rune Sætre, Yusuke Miyao, and Jun'ichi Tsujii. 2009. Protein-protein interaction extraction by leveraging multiple kernels and parsers. *International journal of medical informatics*, 78(12):e39–e46.

- Yusuke Miyao, Kenji Sagae, Rune Sætre, Takuya Matsuzaki, and Jun'ichi Tsujii. 2008. Evaluating contributions of natural language parsers to protein-protein interaction extraction. *Bioinformatics*, 25(3):394–400.
- Dhanachandra Ningthoujam, Shweta Yadav, Pushpak Bhattacharyya, and Asif Ekbal. 2019. Relation extraction between the clinical entities based on the shortest dependency path based lstm. *arXiv preprint arXiv:1903.09941*.
- Peter Palaga. 2009. Extracting relations from biomedical texts using syntactic information. *Mémoire de DEA, Technische Universität Berlin*, 138.
- Yifan Peng and Zhiyong Lu. 2017. [Deep learning for extracting protein-protein interactions from biomedical literature](#). In *BioNLP 2017*, pages 29–38, Vancouver, Canada., Association for Computational Linguistics.
- Sampo Pyysalo, Antti Airola, Juho Heimonen, Jari Björne, Filip Ginter, and Tapio Salakoski. 2008. Comparative analysis of five protein-protein interaction corpora. In *BMC bioinformatics*, volume 9, page S6. BioMed Central.
- Longhua Qian and Guodong Zhou. 2012. Tree kernel-based protein-protein interaction extraction from biomedical literature. *Journal of biomedical informatics*, 45(3):535–543.
- Domonkos Tikk, Philippe Thomas, Peter Palaga, Jörg Hakenberg, and Ulf Leser. 2010. A comprehensive benchmark of kernel methods to extract protein-protein interactions from literature. *PLoS computational biology*, 6(7):e1000837.
- Sofie Van Landeghem, Yvan Saeys, Bernard De Baets, and Yves Van de Peer. 2008. Extracting protein-protein interactions from text using rich feature vectors and feature selection. In *3rd International symposium on Semantic Mining in Biomedicine (SMBM 2008)*, pages 77–84. Turku Centre for Computer Sciences (TUCS).
- Svante Wold, Kim Esbensen, and Paul Geladi. 1987. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52.
- Shweta Yadav, Asif Ekbal, Sriparna Saha, Ankit Kumar, and Pushpak Bhattacharyya. 2019. Feature assisted stacked attentive shortest dependency path based bi-lstm model for protein-protein interaction. *Knowledge-Based Systems*, 166:18–29.
- Shweta Yadav, Usha Lokala, Raminta Daniulaityte, Krishnaprasad Thirunarayan, Francois Lamy, and Amit Sheth. 2021. “when they say weed causes depression, but it’s your fav antidepressant”: Knowledge-aware attention framework for relationship extraction. *PLoS one*, 16(3):e0248299.
- Shweta Yadav, Srivastva Ramesh, Sriparna Saha, and Asif Ekbal. 2020. Relation extraction from biomedical and clinical text: Unified multitask learning framework. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.
- Jiawei Zhang and Lin Meng. 2019. Gresnet: Graph residual network for reviving deep gnns from suspended animation. *arXiv preprint arXiv:1909.05729*.
- Jiawei Zhang, Haopeng Zhang, Congying Xia, and Li Sun. 2020. Graph-bert: Only attention is needed for learning graph representations. *arXiv preprint arXiv:2001.05140*.
- Zhehuan Zhao, Zhihao Yang, Hongfei Lin, Jian Wang, and Song Gao. 2016. A protein-protein interaction extraction approach based on deep neural network. *International Journal of Data Mining and Bioinformatics*, 15(2):145–164.