IJWR

MultiCGCN: Multi-Label Text Classification using GCNs and Heterogeneous Graphs

Milad Allahgholi, Hossein Rahmani*, Parinaz Soltanzadeh, Aylin Naebzadeh
School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran;
milad_allahgholi@comp.iust.ac.ir, h_rahmani@iust.ac.ir, p_soltanzadeh77@comp.iust.ac.ir, a_naeb@comp.iust.ac.ir

ABSTRACT

Multi-label text classification is a critical challenge in natural language processing, where the goal is to assign multiple labels to a given document. Recent advances have primarily focused on deep learning approaches, yet many fail to adequately capture the intricate relationships between documents and labels. In this paper, we propose a novel method called MultiCGCN, in which we leverage Graph Convolutional Networks (GCNs) for multi-label text classification by modeling text as a heterogeneous graph. This unified graph incorporates document similarities, label relationships, and document-label associations, enabling the model to effectively capture both document and label dependencies. We transform the multi-label classification problem into a link prediction task, using Term Frequency–Inverse Document Frequency (TF-IDF) for document similarity and applying GCNs to predict label assignments. Our empirical evaluations demonstrate that MultiCGCN achieves a significant performance boost, improving F1 score by 10% over traditional baseline models. This approach opens new avenues for enhancing the accuracy of multi-label classification in various domains.

***Keywords*— Text Classification, Graph Convolutional Neural Networks, Multi-label Text Classification.**

1. Introduction

Multi-label text classification is a fundamental task in natural language processing (NLP) that involves assigning multiple labels to a given text document. This task is essential in various modern applications such as document categorization, where documents need to be classified into multiple categories; tag recommendation, where multiple relevant tags are suggested for a piece of content; and textual recommendations, where multiple recommendations are provided based on the content of the text. The complexity of multi-label text classification arises from the need to handle the interdependencies and correlations between different labels, which is not a concern in single-label classification where each document is assigned only one label [1–3].

Previous approaches to multi-label text classification have faced several challenges and limitations. Traditional methods often rely on simpler models such as Bag of Words (BoW) or Term Frequency-Inverse Document Frequency (TF-IDF), which primarily focus on the frequency of words within the text and do not capture the semantic relationships between words. These models treat each word independently and fail to consider the context in which words appear, leading to a loss of important information about the relationships between words and labels. Additionally, many existing methods predict each label independently, without considering the potential dependencies and correlations between labels. This independent prediction approach can result in suboptimal performance, as it ignores the valuable information that can be gained from understanding how labels are related to each other [4, 5].

Our contribution is the proposal of a novel method for multi-label text classification, named MultiCGCN, which utilizes Graph Convolutional Networks (GCNs) to model the complex interdependencies and correlations between labels. Despite advancements in multi-label text classification, many existing methods treat labels as independent and fail to capture the relationships between them. Recent research has emphasized the importance of modeling label dependencies to improve classification performance. Our approach first discovers the relationships between texts and transforms the multi-label problem into a link

doi http://dx.doi.org/10.22133/ijwr.2024.485064.1243
Citation M. Allahgholi, H. Rahmani, P. Soltanzadeh, A. Naebzadeh, " MultiCGCN: Multi-Label Text Classification using GCNs and Heterogeneous Graphs", *International Journal of Web Research*, vol.7, no.4,pp.29-37, 2024, doi: http://dx.doi.org/10.22133/ijwr.2024.485064.1243.

**Coressponding Author*
Article History: Received: 23 June 2024; Revised: 7 September 2024; Accepted: 14 September 2024.

prediction task. By leveraging GCNs, MultiCGCN captures the structural dependencies between labels, offering a more accurate and comprehensive solution for label prediction.

The structure of this paper is as follows: Section 2 reviews prior research on stance detection. Section 3 describes the proposed method and our main contribution in detail. Section 4 presents the empirical results. Section 5 includes a discussion of the results and proposes promising directions for future research.

2. Related Work

Text classification has gained significant importance due to the increasing volume of textual data. In this context, we categorize previous works into three main categories. These categories include traditional text classification methods, deep learning approaches, and graph neural networks.

2.1. Traditional Text Classification

Traditional text classification research has primarily focused on two areas: feature engineering and classification algorithms. Feature engineering often relies on the bag-of-words approach, which is the most commonly used feature. However, this model, along with other traditional features like n-grams [6] and ontology entities [7], has its limitations. These features are often static, failing to capture the dynamic nature of language, and can miss the context or semantic relationships between words. In addition to these traditional methods, some research has explored transforming texts into graph structures. This involves applying feature engineering to both the graphs and their subgraphs [8–12]. While these methods are innovative, they can be complex and computationally intensive. Despite their potential, they often require significant computational resources and are often challenging to implement effectively. Our method distinguishes itself by learning textual representations through node embeddings, which dynamically capture the nuances of language use. Unlike traditional static features, node embeddings can adapt to the dynamic nature of language, providing a more flexible and accurate representation of text. This approach not only addresses the limitations of traditional methods but also leverages the strengths of graph-based representations to capture complex relationships within the data.

2.2. Deep Learning for Text Classification

Deep learning approaches to text classification can be divided into two main categories. The first category focuses on models based on word embeddings [13–18]. Many studies have shown that the success of deep learning in text classification largely depends on the effectiveness of these word embeddings [19–21]. Some researchers aggregate unsupervised word embeddings into document embeddings before classification, while others learn word/document and document label embeddings concurrently [22, 23]. The second category employs deep neural networks, specifically Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). CNNs, adapted from computer vision applications to one-dimensional convolution, have been used for sentence classification [3]. Character-level CNNs have also shown promise in capturing fine-grained textual features [24, 25]. On the other hand, Long Short-Term Memory (LSTM) networks, a type of RNN, have been utilized to learn text representations by capturing long-term dependencies in sequences [26–28]. To enhance model flexibility and performance, attention mechanisms have been integrated into text classification models [29, 30]. Recent studies have even adopted transformer-based models, which have shown superior performance in various NLP tasks [31–33]. However, these deep learning approaches have their drawbacks. They require large amounts of data to train effectively and can be opaque, making it difficult to interpret how decisions are made. Moreover, they are resource-intensive, often necessitating significant computational power. Transformers, while powerful, can sometimes be overly complex for certain tasks and may not always outperform simpler models [34, 35]. Additionally, they can suffer from attention diffusion, where the attention mechanism becomes less effective as the input sequence length increases. Unlike these methods, our approach learns the relationships between labels and documents as nodes, offering a more dynamic and interpretable framework for text classification. By leveraging graph neural networks, we can capture complex relationships and dependencies within the data, providing a more holistic understanding of the text. This innovative approach not only addresses the limitations of traditional methods but also leverages the strengths of graph-based representations to capture complex relationships within the data.

2.3. Graph Neural Networks

The field of Graph Neural Networks (GNNs) has garnered increasing interest in recent years [36–41]. Several researchers have extended established neural network models, such as CNNs, which are traditionally applied to regular grid structures (like 2D meshes or 1D sequences), to accommodate arbitrarily structured graphs. In a seminal contribution, Kipf and Welling introduced a streamlined GNN variant known as Graph Convolutional Networks (GCNs), which set new benchmarks in classification performance across a range of graph datasets [42, 43]. GCNs have also been applied to various Natural Language Processing

(NLP) tasks, including semantic role labeling [44], relation classification [45], and machine translation [46]. These applications leverage GCNs to capture the syntactic structures within sentences, demonstrating their versatility and effectiveness in handling different NLP tasks. For text classification, GNNs have been previously investigated. These approaches typically represented documents or sentences as graphs composed of word nodes [47], or they depended on the less commonly available document citation relationships for graph construction [42, 48–52]. In contrast, our method for constructing the corpus graph treats both documents and labels as nodes, forming a heterogeneous graph. This approach eliminates the need for inter-document relationship data, which can often be sparse or unavailable. By incorporating labels directly into the graph structure, our method captures the relationships between documents and labels more effectively, providing a richer and more detailed representation of the text data. Our approach not only addresses the limitations of previous methods but also leverages the strengths of GNNs to capture complex relationships within the data. This innovative method paves the way for more accurate and sophisticated text classification models, highlighting the importance of considering both local and global semantic information.

In previous works on multi-label text classification, the relationship between documents and labels has often been overlooked, with many approaches treating the labels as independent of each other. These methods typically rely on shallow models that do not fully exploit the potential interdependencies between labels, which can result in suboptimal performance, especially in complex datasets. In contrast, our proposed approach aims to address this gap by utilizing graph-based methods and Graph Neural Networks (GNNs). By modeling the relationships between documents and their associated labels as a graph, our method captures these dependencies and enhances the prediction accuracy, ultimately improving the performance of multi-label classification tasks.

3. Proposed Method: MultiCGCN

We propose a novel method called MultiCGCN for multi-label text classification. The different steps involved in our proposed method are shown in Figure 1. In the following sections, we will discuss each of these steps in detail.

3.1. Graph Construction

We construct a heterogeneous text graph which contains document nodes and label nodes. The number of nodes in our graph is the number of documents plus the number of labels. We simply set feature matrix X=I as an identity matrix which means every document or label is represented as a one-hot vector as the input to our GCN.

- **Document-Document**: In our heterogeneous graph, we construct document-document edges to reflect the similarity between documents. The weight of an edge connecting two document nodes is determined by the term frequency-inverse document frequency (TF-IDF) of the words shared by the document pair. This method is advantageous as it considers not only the frequency of words but also their importance across the entire corpus. We found that utilizing TF-IDF as a weighting mechanism yields better results than term frequency alone. However, to maintain a high-quality graph structure, we only retain edges where the TF-IDF similarity exceeds a threshold of 0.45. This ensures that only documents with a significant degree of similarity are linked, thereby enhancing the relevance of the connections within the graph.
- **Label-Label**: The second type of edge in our graph is the label-label edge, which is based on the correlation between labels. To quantify this relationship, we employ a correlation coefficient, setting a minimum threshold of 0.05 for the inclusion of an edge. This approach allows us to capture the inherent associations between labels, which can be particularly insightful when labels share common thematic elements or when they frequently co-occur across documents. By establishing these connections, our graph can more accurately model the complex interplay of labels within the corpus.
- **Document-Label**: Lastly, we address the edges that connect documents to their corresponding labels. These edges are pivotal as they directly represent the classification associations we aim to predict. The construction of these edges is straightforward yet critical for the efficacy of our link prediction model. By integrating these edges, our graph encapsulates the fundamental relationships that underpin multi-label text classification.

We combined the three graphs mentioned above—document-document, label-label, and document-label—into a unified composite graph. This integration allows us to capture the complex relationships among documents and labels more effectively, facilitating a comprehensive representation for our multi-label text classification task.

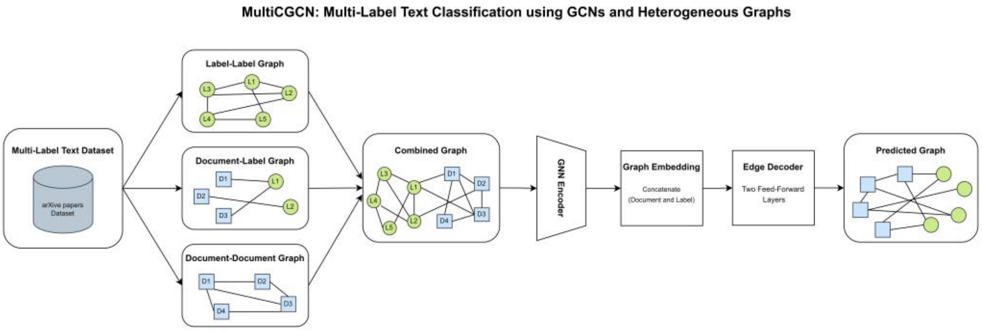

Figure. 1. Overview of our proposed method, MultiCGCN. The method constructs three types of graphs: Label-Label, Document-Label, and Document-Document. These are combined into a unified graph, which is processed by a GNN encoder to generate embeddings. The edge decoder then predicts document-label associations for classification.

3.2. Graph Convolutional Network

After building the text graph, we feed the graph into a simple two-layer GCN as in [7], as shown in Figure 2. The second layer node embeddings are fed into a softmax classifier as shown in Equ(1).

$$\mathbf{Z} = \mathbf{softmax}\left(\tilde{\mathbf{A}}\mathbf{ReLU}\left(\tilde{\mathbf{A}}\mathbf{X}\mathbf{W_0}\right)\mathbf{W_1}\right) \tag{1}$$

where $\tilde{A} = D^{\frac{-1}{2}}AD^{\frac{-1}{2}}$ is the same as in Equ(1), and $softmax(x_i) = \frac{1}{Z}\exp(x_i)$ with $Z = \sum_i \exp(x_i)$. The loss function is defined as the cross-entropy error over all labeled documents, as shown in Equ(2):

$$\boldsymbol{L} = -\sum_{d \in y_D}\sum_{f=1}^{F} \boldsymbol{Y_{df}} \ln \boldsymbol{Z_{df}} \tag{2}$$

Where y_D is the set of document indices that have labels and F is the dimension of the output features, which is equal to the number of classes. Y is the label indicator matrix. The weights parameters W_0 and W_1 can be trained via gradient descent. In equation 3, $E_1 = \tilde{A}XW_0$ contains the first layer document and label embeddings and $E_2 = \tilde{A}ReLU(\tilde{A}XW_0)W_1$ contains the second layer document and labels embeddings.

The final output Z represents the predicted probabilities for each class for the documents, which can be computed by applying the softmax function to the node embeddings obtained from the second layer. This output can be used for classification tasks, where the class with the highest probability is chosen as the predicted label for each document.

For a better understanding, the pseudocode of the proposed method is illustrated in Figure 3.

4. Results

4.1. Dataset

We use "ArXiv CS Papers Multi-Label Classification

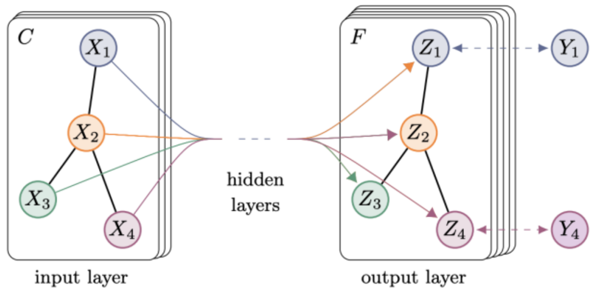

Figure. 2. Graph Convolution Network [7]

Pseudocode

// Graph Construction
// Document-Document Graph Construction
for each pair of documents (d_i, d_j):
  similarity = Compute_TF_IDF_Similarity(d_i, d_j)
  if similarity > threshold:
    AddEdge(DocumentGraph, d_i, d_j, weight=similarity)
// Label-Label Graph Construction
for each pair of labels (l_i, l_j):
  correlation = Compute_Label_Correlation(l_i, l_j)
  if correlation > threshold:
    AddEdge(LabelGraph, l_i, l_j, weight=correlation)
// Document-Label Graph Construction
for each (document d, list of labels L_d):
  for each label l in L_d:
    AddEdge(DocumentLabelGraph, d, l, weight=1)

// Combine Graphs into a Unified Graph
CompositeGraph = CombineGraphs(DocumentGraph, LabelGraph, DocumentLabelGraph)

// GNN Embedding
// Initialize feature matrix X (Identity matrix)
X = InitializeIdentityMatrix(CompositeGraph)
// Encode using Graph Neural Network (GNN)
NodeEmbeddings = GNNEncoder(CompositeGraph, X)
// Edge Prediction

// Pass node embeddings to Edge Decoder
PredictedEdges = EdgeDecoder(NodeEmbeddings)

// Output the Predicted Edges
Return PredictedEdges

Figure. 3. Pseudocode of the proposed method

dataset" in this article. This dataset is a comprehensive collection of research papers from the computer science domain. This dataset is intended for multi-label classification tasks and contains a diverse range of research papers spanning various topics within computer science [53]. The dataset consists of approximately more than 200k research papers and includes the following columns:

- Paper ID
- Title
- Abstract
- Year
- Primary Category
- Categories

We perform some basic preprocessing steps, like checking for duplicate papers based on their "Paper ID". We also apply some filters like "Year" constraint to be between 2016 and 2023. We merge the "Title" and "Abstract" column and save the results in a new column, and apply some preprocessing functions like stemming, lemmatization and removing punctuations and stop words using NLTK [54] and spaCy [55] libraries.

4.2. Evaluation

We split our dataset into train, test and validation sets by 80, 10 and 10 percent of data respectively. We have trained our GCN model for 15 epochs and calculated the loss during training as shown in Figure 4.

We evaluated our proposed method, MultiCGCN, using precision, recall, and F1-score. Due to the novelty of our dataset, there are no existing publications that have worked with it. To provide a more comprehensive evaluation, we implemented and compared our method with three different approaches:

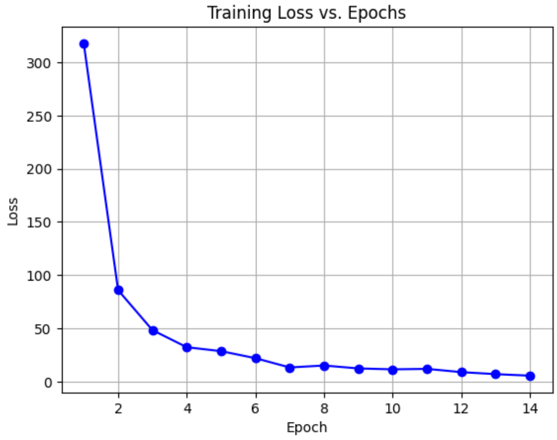

Figure. 4. Training loss of our proposed method

- **Baseline Method (SVM):** In this method, we applied the BERT-large model to extract features from the title and abstract of each sample. These features were then used by an SVM classifier to predict the labels of the test data.
- **Baseline Method (KNN):** In this approach, after feature extraction from the title and abstract of each sample, we used cosine similarity to identify the top K most similar samples in the training data for each test data point. Labels that appeared at least K/2 times among these K samples were assigned to the test sample.
- **LLM-Based Method:** Similar to the KNN-based method, we used cosine similarity to identify the K most similar samples to each test data point as its context. These K samples were formatted as a prompt and provided to ChatGPT-3.5, asking it to predict the corresponding labels for the test data. The prompt used was as follows:

"

Instruction: *Our task is Multi-Label Text Classification. You are provided with the title and abstract of a new paper, along with 20 similar papers that have labels. Based on this data, predict the labels for the new paper.*
Input Title: *{title}*
Input Abstract: *{abstract}*
Similar Papers:
[
{title: $title_1$, abstract: $abstract_1$, labels: [l_{11}, …, l_{1m}]},
…,
{title: $title_k$, abstract: $abstract_k$, labels: [l_{k1}, …, l_{km}]}
]

"

The results are presented in Table 1, while the Receiver Operating Characteristic (ROC) curve is shown in Figure 5. The ROC curve is a graphical representation of a classification model's performance across different decision thresholds. It plots the True Positive Rate against the False Positive Rate. We use ROC curves to evaluate the performance of our classifier. The Area Under the Curve (AUC) quantifies the overall performance of a model. A higher AUC indicates a better model, where 1 being a perfect model and 0.5 representing random chance [56].

In the results presented in Table 1 highlight the superior performance of our proposed method compared to both baseline methods and LLM-based approaches. Our method consistently achieves the best performance in key evaluation metrics, particularly in recall and f1-score, where it outperforms all other methods with a score of 0.6.

Table 1. Comparison of Evaluation Metrics between the Our Model, Baseline Method and LLM-Based Method

Method		***Precision***	***Recall***	***F1-Score***
Baseline	SVM	0.57	0.14	0.22
	KNN (K=10)	0.62	0.27	0.38
	KNN (K=20)	0.51	0.33	0.41
LLM-Based	K=10	**0.67**	0.32	0.43
	K=20	0.63	0.43	0.51
Our Method		0.62	**0.6**	**0.61**

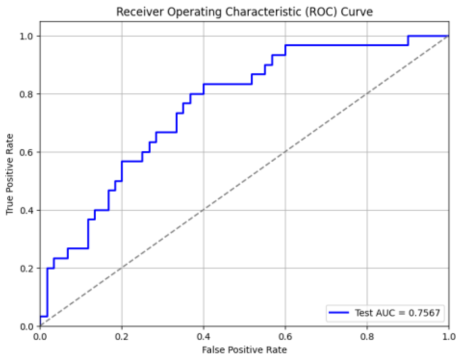

Figure. 5. ROC Curve for the Test Data Demonstrating an AUC of 0.75 in our proposed method

This significant improvement in recall indicates the ability of our model to correctly identify a higher proportion of positive cases, making it more effective for the given classification task. Although the LLM-based method achieves a slightly higher Precision of 0.67 when K=10, its recall remains considerably lower at 0.32, resulting in an imbalanced performance. In contrast, our model delivers a balanced trade-off between Precision and Recall, achieving the highest F1-Score of 0.61. This clearly demonstrates that our model not only identifies more positive cases but also maintains accuracy in its predictions, making it the most reliable approach for this task.

5. Conclusions and Future Work

Multi-label text classification is a key challenge in natural language processing, requiring advanced and innovative techniques to enhance accuracy and efficiency in applications such as document categorization and recommendation systems. In this paper, we proposed a novel method called MultiCGCN that leverages Graph Convolutional Networks (GCNs) to construct a heterogeneous graph encompassing document similarities, label correlations, and document-label associations. This approach not only transforms multi-label classification into a link prediction problem but also effectively captures the intricate interdependencies present in the data. Our empirical results demonstrate that MultiCGCN significantly improves model performance, achieving a 10% increase in the F1 score compared to traditional baseline models. this indicates that considering the relationships among text documents can significantly improve prediction accuracy.

For future work, we aim to extend MultiCGCN to inductive settings for better generalization to unseen data and explore the integration of attention mechanisms to further enhance classification accuracy. Additionally, we will investigate optimizing resource efficiency for our algorithms and adapting our approach to cross-lingual contexts. In summary, our research lays a solid foundation for future investigations in multi-label text classification, and we are eager to explore the potential advancements that can be achieved through these proposed directions.

Declaration

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Data Availability

The data used in this study can be accessed via the following link:
https://www.kaggle.com/datasets/devintheai/arxiv-cs-papers-multi-label-classification-200k-v1/data

Ethical Approval

This study did not involve human or animal data.

Conflict of interest

The authors declare that they have no conflict of interest.

Author Contribution

Milad Allahgholi: Implementation of the proposed method, Drafting the manuscript.

Hossein Rahmani: Interpretation of the results, Revision of the manuscript.

Parinaz Soltanzadeh: Implementation of the proposed method, Drafting the manuscript.

Aylin Naebzadeh: Implementation of the proposed method, Drafting the manuscript.

References

[1] L. Meng, Z. Ye, Y. Yang and H. Zhao, "DeepMCGCN: Multi-channel Deep Graph Neural Networks," *International Journal of Computational Intelligence Systems*, vol. 17, p. 41, 2024. https://doi.org/10.1007/s44196-024-00432-9

[2] J. Xiong, L. Yu, X. Niu and Y. Leng, "XRR: Extreme multi-label text classification with candidate retrieving and deep ranking," *Inf Sci (N Y)*, *622*, 115–132, 2023. https://doi.org/10.1016/j.ins.2022.11.158

[3] A. Rakhlin, "Convolutional neural networks for sentence classification," *GitHub*, 6, 25, 2016.

[4] V. Buchner, L. Cao, J. C. Kalo and V. Von Ehrenheim, "Prompt Tuned Embedding Classification for Industry Sector Allocation," In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 6, Industry Track, 2024, pp. 108–118. https://doi.org/10.18653/v1/2024.naacl-industry.10

[5] D. Li, *et al.*, "Enhancing Extreme Multi-Label Text Classification: Addressing Challenges in Model, Data, and Evaluation," In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, 2023, pp. 313–321. https://doi.org/10.18653/v1/2023.emnlp-industry.30

[6] S. I. Wang and C. D. Manning, "Baselines and bigrams: Simple, good sentiment and topic classification." In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, vol. 2: Short Papers, 2024, pp. 90–94.

[7] V. Chenthamarakshan, P. Melville, V. Sindhwani and R. D. Lawrence, "Concept labeling: Building text classifiers with minimal supervision," In *IJCAI proceedings-international joint conference on artificial intelligence*, 2011, p. 1225.

[8] Y. Luo, Ö. Uzuner and P. Szolovits, "Bridging semantics and syntax with graph algorithms—state-of-the-art of extracting biomedical relations," *Brief Bioinform*, vol. 18, no. 1, pp. 160–178, 2017. https://doi.org/10.1093/bib/bbw001

[9] F. Rousseau, E. Kiagias and M. Vazirgiannis, "Text categorization as a graph classification problem," In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 2015, pp. 1702–1712.

[10] K. Skianis, F. Rousseau and M. Vazirgiannis, "Regularizing text categorization with clusters of words," In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas, 2016, pp. 1827–1837.

[11] Y. Luo, A. R. Sohani, E. P. Hochberg and P. Szolovits, "Automatic lymphoma classification with sentence subgraph mining from pathology reports," *Journal of the American Medical Informatics Association*, vol. 21, pp. 824–832, 2014. https://doi.org/10.1136/amiajnl-2013-002443

[12] Y. Luo, Y. Xin, E. Hochberg, R. Joshi, O. Uzuner and P. Szolovits, "Subgraph augmented non-negative tensor factorization (SANTF) for modeling clinical narrative text," *Journal of the American Medical Informatics Association*, vol. 22, no. 5, pp. 1009–1019, 2015. https://doi.org/10.1093/jamia/ocv016

[13] Y. Yan, F. Liu, X. Zhuang and J. Ju, "An R-transformer_BiLSTM model based on attention for multi-label text classification, *Neural Process Lett*, vol. 55, pp. 1293–1316, 2023. https://doi.org/10.1007/s11063-022-10938-y

[14] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean, "Distributed representations of words and phrases and their compositionality," *Adv Neural Inf Process Syst*., vol. 26, 2013.

[15] J. Pennington, R. Socher and C. D. Manning, "Glove: Global vectors for word representation," In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, Doha, Qatar, 2014, pp. 1532–1543. https://doi.org/10.3115/v1/D14-1162

[16] H. Yu, F. Xiong and Z. Chen, "Text Classification Based on Natural Language Processing and Machine Learning in Multi-Label Corpus," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 8, pp. 1–14, 2024. https://doi.org/10.1145/3617831

[17] J. Wang, H. Xie, F. L. Wang and L. K. Lee, "Improving text classification via a soft dynamical label strategy," *International Journal of Machine Learning and Cybernetics*, vol. 14, pp. 2395–2405, 2023. https://doi.org/10.1007/s13042-022-01770-w

[18] W. Liu, J. Pang, N. Li, X. Zhou and F. Yue, "Research on multi-label text classification method based on tALBERT-CNN," *International Journal of Computational Intelligence Systems*, vol. 14, p. 201, 2021. https://doi.org/10.1007/s44196-021-00055-4

[19] A. Joulin, E. Grave and P. B. T. Mikolov, "Bag of Tricks for Efficient Text Classification," arXiv preprint *arXiv:1607.01759*, 2017. https://doi.org/10.48550/arXiv.1607.01759

[20] D. Shen *et al.*, "Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms," In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Volume 1: Long Papers, 2018, pp. 440–450. https://doi.org/10.18653/v1/P18-1041

[21] G. Wang *et al.*, "Joint Embedding of Words and Labels for Text Classification," In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Volume 1: Long Papers, 2018, pp. 2321–2331. https://doi.org/10.18653/v1/P18-1216

[22] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," *Proceedings of the 31st International Conference on Machine Learning, PMLR* , 2014, pp. 1188-1196.

[23] J. Tang, M. Qu and Q. Mei, "Pte: Predictive text embedding through large-scale heterogeneous text networks," In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 2015, pp. 1165–1174. https://doi.org/10.1145/2783258.2783307

[24] X. Zhang, J. Zhao and Y. LeCun, "Character-level convolutional networks for text classification," *Adv. Neural Inf Process Syst.*, vol. 28, 2015.

[25] A. Conneau, H. Schwenk, L. Barrault and Y. Lecun, "Very Deep Convolutional Networks for Text Classification," In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Volume 1: Long Papers, 2017, pp. 1107–1116. https://aclanthology.org/E17-1104

[26] P. Liu, X. Qiu and X. Huang, "Recurrent neural network for text classification with multi-task learning," *In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016, pp. 2873–2879. https://www.ijcai.org/Proceedings/16/Papers/408.pdf

[27] Y. Luo, "Recurrent neural networks for classifying relations in clinical notes," *J. Biomed Inform*., vol. 72, pp. 85–95, 2017. https://doi.org/10.1016/j.jbi.2017.07.006

[28] K. S. Tai, R. Socher and C. D. Manning, "Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks," In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Volume 1: Long Papers, 2015. pp. 1556–1566. https://doi.org/10.3115/v1/P15-1150

[29] Y. Wang, M. Huang, X. Zhu and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," In *Proceedings of the 2016 conference on empirical methods in natural language processing*. 2016, pp. 606–615. https://doi.org/10.18653/v1/D16-1058

[30] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola and E. Hovy, "Hierarchical attention networks for document classification." In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, 2016, pp. 1480–1489. https://doi.org/10.18653/v1/N16-1174

[31] T. Lin, Y. Wang, X. Liu and X. Qiu, "A survey of transformers," *AI open*, vol. 3, pp. 111–132, 2022. https://doi.org/10.1016/j.aiopen.2022.10.001

[32] F. Zhao, Q. Ai, X. Li, W. Wang, Q. Gao and Y. Liu, "TLC-XML: Transformer with Label Correlation for Extreme Multi-label Text Classification," *Neural Process Lett*., vol. 56, p. 25, 2024. https://doi.org/10.1007/s11063-024-11460-z

[33] W. Cunha, F. Viegas, C. França, T. Rosa, L. Rocha and M. A. Gonçalves, "A Comparative Survey of Instance Selection Methods applied to Non-Neural and Transformer-Based Text Classification," *ACM Comput Surv.*, vol. 55, pp. 1–52, 2023. https://doi.org/10.1145/3582000

[34] Q. Li *et al*., "A survey on text classification: From traditional to deep learning," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 2, pp. 1–41, 2022. https://doi.org/10.1145/3495162

[35] A. Palanivinayagam and C. Z., El-Bayeh and R. Damaševičius, "Twenty years of machine-learning-based text classification: A systematic review," Algorithms, vol. 16, no. 5, p. 236, 2023. https://doi.org/10.3390/a16050236

[36] H. T. Vu, M. T. Nguyen, V. C. Nguyen, M. H. Pham, V. Q. Nguyen and V. H. Nguyen, "Label-representative graph convolutional network for multi-label text classification," *Applied Intelligence*, vol. 53, pp. 14759–14774, 2023. https://doi.org/10.1007/s10489-022-04106-x

[37] H. Cai, V. W. Zheng and K. C. C. Chang, "A comprehensive survey of graph embedding: Problems, techniques, and applications," *IEEE Trans Knowl Data Eng*., vol. 30, no. 2, pp. 1616–1637, 2018. https://doi.org/10.1109/TKDE.2018.2807452

[38] K. Wang, Y. Ding and S. C. Han, "Graph neural networks for text classification: A survey," *Artif Intell Rev*., vol. 57, p. 190, 2024. https://doi.org/10.1007/s10462-024-10808-0

[39] D. Zeng, E. Zha, J. Kuang and Y. Shen, "Multi-label text classification based on semantic-sensitive graph convolutional network," *Knowl Based Syst*., vol. 284, p. 111303, 2024. https://doi.org/10.1016/j.knosys.2023.111303

[40] X. Li, B. You, Q. Peng and S. Feng, "Dual-view graph convolutional network for multi-label text classification," *Applied Intelligence*, vol. 54, pp. 9363–9380, 2024. https://doi.org/10.1007/s10489-024-05666-w

[41] Y. Ma, N. Yan, J. Li, M. Mortazavi and N. V. Chawla, "HetGPT: Harnessing the power of prompt tuning in pre-trained heterogeneous graph neural networks," In *Proceedings of the ACM on Web Conference 2024*, 2024, pp. 1015–1023. https://doi.org/10.1145/3589334.3645685

[42] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," In *International Conference on Learning Representations*, 2017. https://doi.org/10.48550/arXiv.1609.02907

[43] J. Bruna, W. Zaremba, A. Szlam and Y. Lecun, "Spectral networks and locally connected networks on graphs," In *International Conference on Learning Representations (ICLR2014)*, CBLS, April 2014. https://doi.org/10.48550/arXiv.1312.6203

[44] D. Marcheggiani and I. Titov, "Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling," In: Palmer, M., Hwa, R., and Riedel, S. (eds.) *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Copenhagen, Denmark, 2017, pp. 1506–1515. https://doi.org/10.18653/v1/D17-1159

[45] J. Bastings, I. Titov, W. Aziz, D. Marcheggiani and K. Sima'an, "Graph Convolutional Encoders for Syntax-aware Neural Machine Translation. In: Palmer, M., Hwa, R., and Riedel, S. (eds.) *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Copenhagen, Denmark, 2017, pp. 1957–1967. https://doi.org/10.18653/v1/D17-1209

[46] Y. Li, R. Jin and Y. Luo, "Classifying relations in clinical narratives using segment graph convolutional and recurrent neural networks (Seg-GCRNs)," *Journal of the American Medical Informatics Association*, vol. 26, no. 3, pp. 262–268, 2019. https://doi.org/10.1093/jamia/ocy157

[47] M. Defferrard, X. Bresson and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, Curran Associates Inc., Red Hook, NY, USA, 2016, pp. 3844–3852.

[48] Z. Cao, X. Deng, S. Yue, P. Jiang, J. Ren and J. Gui, "Dependent Task Offloading in Edge Computing Using GNN and Deep Reinforcement Learning," *IEEE Internet Things J*., vol. 11, no. 12, pp. 21632-21646, 2024, https://doi.org10.1109/JIOT.2024.3374969.

[49] X. Li, B. Wang, Y. Wang and M. Wang, "Graph-based text classification by contrastive learning with text-level graph augmentation," *ACM Trans Knowl Discov Data*., vol. 18, pp. 1–21, 2024. https://doi.org/10.1145/3638353

[50] S. S. Ziaee, H. Rahmani, M. Tabatabaei, A. H. C. Vlot and A. Bender, "DCGG: drug combination prediction using GNN and GAE. *Progress in Artificial Intelligence*, vol. 13, pp. 17–30, 2024. https://doi.org/10.1007/s13748-024-00314-3

[51] A. Sharma, S. Singh and S. Ratna, "Graph neural network operators: a review," *Multimed Tools Appl*., vol. 83, pp. 23413–23436, 2024. https://doi.org/10.1007/s11042-023-16440-4

[52] P. C. Kuo, Y. T. Chou, K. Y. Li, W. T. Chang, Y. N. Huang and C. S. Chen, "GNN-LSTM-based fusion model for

structural dynamic responses prediction," *Eng Struct*., vol. 306, p. 117733, 2024. https://doi.org/10.1016/j.engstruct.2024.117733

[53] https://www.kaggle.com/datasets/devintheai/arxiv-cs-papers-multi-label-classification-200k-v1/data

[54] www.nltk.org

[55] www.spacy.io

[56] Z. H. Hoo, J. Candlish and D. Teare, "What is an ROC curve?," *Emergency Medicine Journal*, vol. 34, no. 6, pp. 357-359, 2017. https://doi.org/10.1136/emermed-2017-206735

Milad Allahgholi received the M.Sc. degree from Iran University of Science and Technology (IUST), Tehran, in 2020. He is currently a Ph.D. candidate in Computer Engineering (Software) at Iran University of Science and Technology. His research interests include machine learning, complex networks, and text mining.

Hossein Rahmani received his Ph.D. from Leiden Institute of Advanced Computer Science (LIACS), the Netherlands, in 2012. Following his Ph.D., he completed a postdoctoral fellowship in Text Mining at Maastricht University, the Netherlands, from 2012 to 2014. He has been an assistant professor at Iran University of Science and Technology (IUST) from 2015. His research interests mainly include data analysis, graph mining, text mining, and complex networks.

Parinaz Soltanzadeh received the B.Sc. degree from University of Science and Culture, Tehran, in 2021. She is currently a M.Sc. student in Computer Engineering (Software) at Iran University of Science and Technology.

Aylin Naebzadeh received the B.Sc. degree from Iran University of Science and Technology, Tehran, in 2024. Her research interests lie at the intersection of Data Science, Machine and Deep Learning, NLP, and Generative AI.