Predicting Patient Readmission Risk from Medical Text via Knowledge Graph Enhanced Multiview Graph Convolution

Qiuhao Lu
University of Oregon
Eugene, OR, USA
luqh@cs.uoregon.edu

Thien Huu Nguyen
University of Oregon
Eugene, OR, USA
thien@cs.uoregon.edu

Dejing Dou
University of Oregon
Baidu Research
Eugene, OR, USA
dou@cs.uoregon.edu
doudejing@baidu.com

ABSTRACT

Unplanned intensive care unit (ICU) readmission rate is an important metric for evaluating the quality of hospital care. Efficient and accurate prediction of ICU readmission risk can not only help prevent patients from inappropriate discharge and potential dangers, but also reduce associated costs of healthcare. In this paper, we propose a new method that uses medical text of Electronic Health Records (EHRs) for prediction, which provides an alternative perspective to previous studies that heavily depend on numerical and time-series features of patients. More specifically, we extract discharge summaries of patients from their EHRs, and represent them with multiview graphs enhanced by an external knowledge graph. Graph convolutional networks are then used for representation learning. Experimental results prove the effectiveness of our method, yielding state-of-the-art performance for this task.

CCS CONCEPTS

• **Applied computing → Health informatics**; • **Computing methodologies → Natural language processing**.

KEYWORDS

patient readmission prediction; graph convolutional networks; knowledge graph

ACM Reference Format:
Qiuhao Lu, Thien Huu Nguyen, and Dejing Dou. 2021. Predicting Patient Readmission Risk from Medical Text via Knowledge Graph Enhanced Multiview Graph Convolution. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3404835.3463062

1 INTRODUCTION

Patients who are readmitted to intensive care units (ICUs) after transfer or discharge usually have a greater chance of developing dangerous symptoms that can result in life-threatening situations. Readmissions also put families at higher financial burden and increase healthcare providers' costs. Therefore, it is beneficial for both patients and hospitals to identify patients that are inappropriately or prematurely discharged from ICU.

Over the past few years, there has been a surge of interest in applying machine learning techniques to clinical forecasting tasks, such as readmission prediction [12], mortality prediction [6], length of stay prediction [14], etc. Earlier studies generally select statistically significant features from patients' Electronic Health Records (EHRs), and feed them into traditional machine learning models like logistic regression [19]. Deep learning models have also been gaining more and more attention in recent years, and have shown superior performance in medical prediction tasks. For example, Lin *et al.* select 17 types of chart events (diastolic blood pressure, capillary refill rate, etc.) over a 48-hour time window and put them into a LSTM-CNN model [12] and achieve much better performance than previous work in readmission prediction.

A common theme among these studies is that they all rely on numerical and time-series features of patients, while neglecting rich information in the clinical notes of EHRs. This motivates us to tackle this task from a pure natural language processing perspective, which is not well explored in literature. Essentially, in this work, we consider the task of ICU readmission prediction as binary text classification, i.e., for a given clinical note, the model aims to predict whether or not the patient will be readmitted to ICU within 30 days after discharge.

Although it is possible to directly apply existing text classification methods to the readmission prediction task, two major challenges need to be addressed: (1) clinical notes, e.g., discharge summaries, are generally long and noisy, which makes it difficult to capture the inherent semantics to support classification; (2) general methods do not consider domain knowledge in the medical area, which is critical as medical concepts are hard to interpret with limited training for downstream tasks.

Recently, a useful strategy is proposed to tackle the first challenge, where it encodes documents with graphs-of-words to enhance the interactions of context, and to capture the global semantics of the document. The strategy has been applied to different NLP tasks, including document-level relation extraction [3, 4, 15], question answering [5, 18], and text classification [17, 20, 23]. But constructing graphs of clinical notes for patient outcome prediction, to our knowledge, is underexplored.

Motivated by this, we propose a novel graph-based model that represents clinical notes as document-level graphs to predict patient readmission risk. Moreover, to address the second challenge, we

---

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8037-9/21/07…$15.00
https://doi.org/10.1145/3404835.3463062

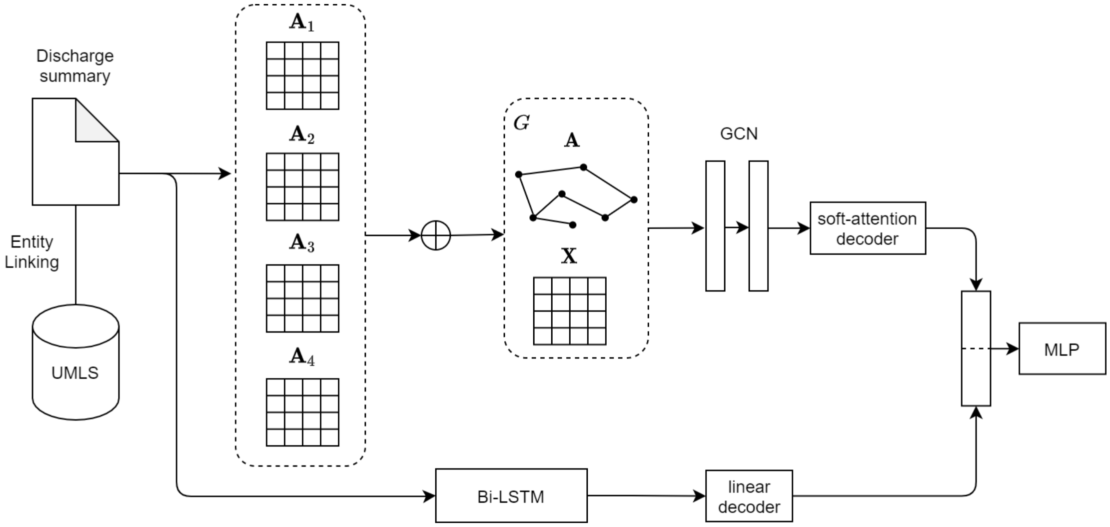

Figure 1: Architecture of MedText.

incorporate an external knowledge graph, i.e., the Unified Medical Language System (UMLS) [2] Metathesaurus, to construct a four-view graph for each input clinical note. The four views correspond to intra-document, intra-UMLS, and document-UMLS interactions, respectively. By constructing such a enhanced graph representation for clinical notes, we inject medical domain knowledge to improve representation learning for the model. Our contribution can thus be summarized as follows:

- We propose a novel graph-based text classification model, i.e., MedText, to predict ICU patient readmission risk from clinical notes in patients' EHRs. Unlike previous studies that rely on numerical and time-series features, we only use clinical notes to make predictions, which provides some insights on utilizing medical text for clinical predictive tasks.
- We construct a specifically designed multiview graph for each clinical note to capture the interactions among words and medical concepts. In this way we inject domain-specific information from an external knowledge graph, i.e., UMLS, into the model. The experimental studies demonstrate the superb performance of this method, by updating the state-of-the-art results on readmission prediction.

2 METHODOLOGY

2.1 Graph Construction

For each document (e.g., clinical note), we construct a weighted and undirected four-view graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ with an associated adjacency matrix $\mathbf{A}$, where $\mathcal{N}$ and $\mathcal{E}$ refer to the vertex set and edge set respectively. We also denote the representation of vertices by $\mathbf{X}$. Instead of using unique words in the document as vertices, we first conduct entity linking over the text and link the entity mentions to UMLS[1]. Consequently, we consider two types of vertices in the document-level graph $\mathcal{G}$, i.e., the unique words $\mathcal{N}_w$ and the linked UMLS entities $\mathcal{N}_e$. The vertex set $\mathcal{N}$ is thus formed as the union of $\mathcal{N}_w$ and $\mathcal{N}_e$: $\mathcal{N} = \mathcal{N}_w \cup \mathcal{N}_e$. Four views are then designed to exploit intra-document, intra-UMLS, and document-UMLS interactions that will be combined to form the adjacency matrix as follows.

2.1.1 Intra-Document: $\mathcal{V}_1$. $\mathcal{V}_1$ is designed to capture the intra-document interactions among words and entities. Essentially, we expect the edge weights between vertices to estimate the level of interaction, so that vertices can directly interact during message passing even if they are sequentially far away from each other in the document. In this work, we generate the adjacency matrix $\mathbf{A}_1$ for $\mathcal{V}_1$ by counting the co-occurrences of vertices within a fixed-size sliding window (size 3 in this work) over the text.

2.1.2 Intra-UMLS: $\mathcal{V}_2$,$\mathcal{V}_3$. In this work, we aim to inject external knowledge from UMLS to the document-level graph representation. To this end, we consider two types of information, i.e., the internal structure of UMLS and the semantic similarities between medical concepts. Specifically, we construct $\mathcal{V}_2$ by computing the shortest path lengths between entity vertices as edge weights in $\mathbf{A}_2$, where a shorter path indicates a higher relevance. We further construct $\mathcal{V}_3$ by computing the string similarities based on the word overlap ratios of entity descriptions for $\mathbf{A}_3$.

2.1.3 Document-UMLS: $\mathcal{V}_4$. $\mathcal{V}_4$ is constructed by calculating the cosine similarities between initial representations of all vertices, including words and entities, which aims to capture the interactions between the information sources, i.e., the document itself and the knowledge base. The similaries are used for edge weights $\mathbf{A}_4$.

2.1.4 View Combination. By combining the four views, we expect to leverage three levels of interactions, i.e., intra-document, intra-UMLS, and document-UMLS, to generate rich interaction structures for documents to aid representation learning. Intuitively, the four views are combined via a weighted sum of the four adjacency matrices as the final adjacency matrix $\mathbf{A}$:

---

[1] We use ScispaCy [16] as the entity linker in this work.

$$\mathbf{A} = \text{MASK}(\sum_{i=1}^{4} \alpha_i \mathbf{A}_i) \tag{1}$$

where $\mathbf{A}_i$ refer to each view's normalized adjacency matrix and α_i are the balancing factors that are determined by cross-validation. The adjacency matrix is then masked with a threshold, i.e., $\gamma = 0.5$, where only edges with larger weights are kept for further message passing. The motivation for the masking is to improve robustness and efficiency by decreasing some density.

The representation of vertices, i.e., $\mathbf{X}$, are initialized with a pre-trained word embedding BioWordVec [22]. For entity vertices, we take the average values of word embeddings of the entity names as the representation for the entity.

2.2 Encoding and Decoding

In this work, we incorporate a two-layer graph convolutional network (GCN) [9] to encode the graph representation of clinical notes, as depicted in Figure 1. We include an attention layer after GCN, which serves as a decoder to decode the document-level representation $\mathbf{D}_G$ from node embeddings. The encoding process can be described as:

$$\mathbf{X}^{(l+1)} = \text{LeakyReLU}(\hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{A}} \hat{\mathbf{D}}^{\frac{1}{2}} \mathbf{X}^{(l)} \mathbf{W}^{(l)}) \tag{2}$$

where $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, and $\mathbf{I}$ is the identity matrix of $\mathbf{A}$. $\hat{\mathbf{D}}$ is the diagonal degree matrix of $\hat{\mathbf{A}}$, and $\mathbf{W}^{(l)}$ is the weight matrix for the l-th layer where $l = 0, 1, 2$ in this work.

We incorporate a graph summation module [11, 23] to decode the document-level representation $\mathbf{D}_G$ from the constructed graph, by assigning different attention weights to the nodes. The decoding process can be described as:

$$\mathbf{X}_G = f_1(\mathbf{X}^{(2)}) \odot f_2(\mathbf{X}^{(2)}) \tag{3}$$

$$\mathbf{D}_G = \text{mean}(\mathbf{X}_G) + \max(\mathbf{X}_G) \tag{4}$$

where $\mathbf{X}^{(2)}$ is the output of the GCN encoder and f_1, f_2 are two feed-forward networks with sigmoid and leakyrelu activation, respectively. The f_1 network acts as a soft attention mechanism that indicates the relative importance of nodes, while f_2 serves as feature transformation. The operator $\odot$ denotes element-wise multiplication. Then the document-level representation $\mathbf{D}_G$ is summarized as the addition of the mean and maximum values of the attentive node embeddings.

We also use a two-layer bidirectional LSTM to directly encode the document and decode the document-level representation $\mathbf{D}_T$ with a linear decoder, where linear transformation and max-pooling are applied. Then the two document-level representations, i.e., $\mathbf{D}_G$ and $\mathbf{D}_T$, are concatenated and fed into a MLP classifier. The model is optimized with cross-entropy loss.

3 EXPERIMENT

3.1 Dataset

The experiment is conducted based on the MIMIC-III Critical Care (Medical Information Mart for Intensive Care III) Database, which is a large, freely-available database composed of de-identified EHR data [8]. For a fair comparison, we use the same data split with the baseline [21], where the discharge summaries are extracted from EHRs and the generated 48, 393 documents are split into training (80%), validation (10%), and testing (10%).

3.2 Evaluation Metrics

We use three metrics for evaluation, i.e., the area under the receiver operating characteristics curve (AUROC), the area under the precision recall curve (AUPRC), and the recall at precision of 80% (RP80). AUROC and AUPRC are widely used for evaluating patient outcome prediction tasks, including readmission prediction [12, 13, 21]. RP80 is a clinically-relevant metric that helps minimize the risk of alarm fatigue, as introduced in ClinicalBERT [7], where we fix the precision at 80% and calculate the recall rate.

3.3 Baselines

The following baselines are used for comparison.

- BioBERT. BioBERT is a domain-specific BERT variant pre-trained on large biomedical corpora, e.g., PubMed abstracts and PMC full-text articles [10]. In the experiment, we use the latest version, i.e., BioBERT v1.1, with a classification head as the baseline. The last 512 tokens of each note are used as input to the model.
- ClinicalBERT. ClinicalBERT is initialized from BioBERT v1.0 and pre-trained on MIMIC notes [1]. Note that there is another ClinicalBERT [7] model which presents a similar idea.
- CC-LSTM. Zhang *et al.* propose CC-LSTM that encodes UMLS knowledge into text representations and report state-of-the-art performance on readmission prediction on the MIMIC-III dataset [21]. For a fair comparison, we use the same pre-trained word embeddings, i.e., BioWordVec [22], in our model.
- MedText-x. Specifically, we replace the Bi-LSTM encoder with ClinicalBERT and BioBERT to demonstrate the effectiveness of the proposed graph-based knowledge injection strategy. The last two baselines are denoted by MedText-ClinicalBERT and MedText-BioBERT, respectively.

3.4 Results

The experimental results are presented in Table 1. Generally, the proposed method, i.e., MedText, compares favorably with all the other baselines and outperforms the state-of-the-art method. Besides, directly applying pre-trained language models, such as BioBERT and ClinicalBERT, to readmission prediction does not work well. It is most likely due to the long and noisy nature of clinical notes, and only the last 512 tokens are taken as input in the experiment. However, by combining with MedText, the performance gets improved greatly, indicating the effectiveness of the proposed graph-based knowledge injection method.

Additionally, Lin *et al.* propose a readmission prediction model that takes numerical features, e.g., chart events, of patients as input, and claim a state-of-the-art AUROC of 0.791 with AUPRC of 0.513 on the same dataset [12]. This is essentially not comparable as they are using numerical features instead of text, but it highlights the value of clinical notes in EHRs.

Table 1: Performance on 30-day unplanned ICU patient readmission prediction.

Method	AUROC	AUPRC	RP80
BioBERT	0.775	0.538	0.200
MedText-BioBERT	0.811	0.610	0.278
ClinicalBERT	0.781	0.536	0.189
MedText-ClinicalBERT	0.812	0.615	0.277
CC-LSTM [21]	0.804	0.613	N/A
MedText	**0.825**	**0.632**	**0.319**

Table 2: Ablation analysis of MedText.

Method	AUROC	AUPRC	RP80
w/o $\mathcal{V}_1$	0.803	0.605	0.300
w/o $\mathcal{V}_{1,2}$	0.809	0.615	0.296
w/o $\mathcal{V}_{1,2,3}$	0.801	0.607	0.290
w/o $\mathcal{V}_{1,2,3,4}$	0.799	0.601	0.288
w/o $\mathbf{D}_T$	0.808	0.601	0.275
Full	0.825	0.632	0.319

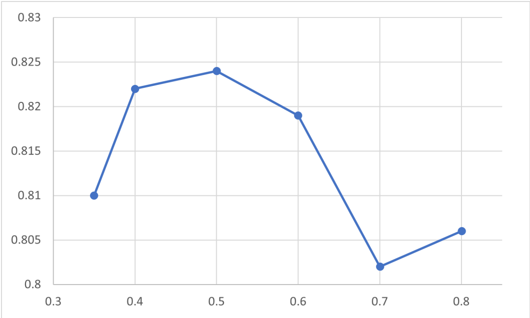

Figure 2: Sensitivity of masking threshold γ.

3.5 Ablation and Sensitivity Study

We present the ablation study in Table 2. As shown in the table, removal of the four views will cause the performance to drop greatly, indicating the effectiveness and necessity of the four views. It is also worth noticing that the model still performs on par with CC-LSTM if the Bi-LSTM module is removed, i.e., w/o $\mathbf{D}_T$, and it would be more efficient in training. We also show the AUROC score with different masking threshold in Figure 2, where AUROC reaches the peak when $\gamma = 0.5$. To further assess the performance of the model in terms of precision and recall, we show the P-R curve in Figure 3.

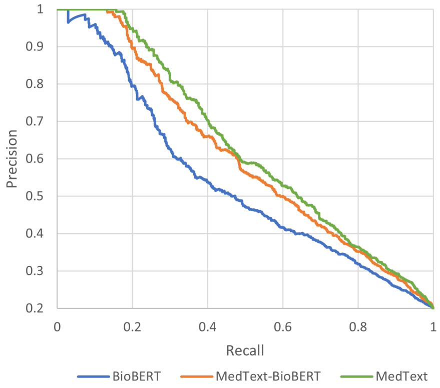

Figure 3: Precision-recall curve of MedText.

3.6 Error Analysis

Entity linking plays an important role in this method as it is the first step of graph construction and all four views either directly or indirectly depend on the linked entities. Since a relatively high linking precision can be achieved by setting appropriate parameters of the ScispaCy linker, we mainly focus on the missed entities in the text. After manually examining a subset of notes, we roughly estimate that 15% to 25% of entities are not recognized or linked, which may have negatively influenced the prediction model. Some example snippets of clinical notes include:

“this is a 69 year old man with a history of end stage cardiomyopathy (nyha class 4) and severe chf with an ef of 15 (ef of 20 on milrinone drip) as well as severe mr p/w sob , doe , pnd , weight gain of 6lbs in a week , likely due to chf **exacerbation** . ”

“he has a history of v-tach which responded to amiodarone . patient also has icd in place . respiratory : sob and increased o2 requirement were likely secondary to chf **exacerbation** and resultant **pulmonary edema**”

“you were admitted for increasing **shortness of breath** and oxygen requirements on increasing doses of lasix”

The texts in bold refer to unrecognized entity mentions. Essentially they should be linked to UMLS entities C4086268 (Exacerbation), C0034063 (pulmonary edema) and C0013404 (Dyspnea), respectively. These uncovered entities might indicate the severity of patients’ conditions and thus are critical for predicting the readmission risk.

4 CONCLUSION

In this study, we propose a novel graph-based text classification model, i.e., MedText, to predict ICU patient readmission risk, using clinical notes from patients’ EHRs. The experiments demonstrate the effectiveness of the method and an updated state-of-the-art performance is observed on the benchmark.

ACKNOWLEDGMENTS

This research has been supported by the Army Research Office (ARO) grant W911NF-21-1-0112 and the NSF grant CNS-1747798 to the IU-CRC Center for Big Learning. This research is also based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2019-19051600006 under the Better Extraction from Text Towards Enhanced Retrieval (BETTER) Program. We also would like to thank the IBM-Almaden research group for their support in this work.

REFERENCES

[1] Emily Alsentzer, John Murphy, William Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. 2019. Publicly Available Clinical BERT Embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. Association for Computational Linguistics, Minneapolis, Minnesota, USA, 72–78. https://doi.org/10.18653/v1/W19-1909

[2] Olivier Bodenreider. 2004. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research* 32, suppl_1 (2004), D267–D270.

[3] Hui Chen, Pengfei Hong, Wei Han, Navonil Majumder, and Soujanya Poria. 2020. Dialogue Relation Extraction with Document-level Heterogeneous Graph Attention Networks. *arXiv preprint arXiv:2009.05092* (2020).

[4] Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Connecting the Dots: Document-level Neural Relation Extraction with Edge-oriented Graphs. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, 4927–4938.

[5] Nicola De Cao, Wilker Aziz, and Ivan Titov. 2019. Question Answering by Reasoning Across Documents with Graph Convolutional Networks. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 2306–2317.

[6] Hrayr Harutyunyan, Hrant Khachatrian, David C. Kale, and Aram Galstyan. 2017. Multitask Learning and Benchmarking with Clinical Time Series Data. *CoRR* abs/1703.07771 (2017).

[7] Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. 2019. Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342* (2019).

[8] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data* 3, 1 (2016), 1–9.

[9] Thomas N Kipf and Max Welling. 2016. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv preprint arXiv:1609.02907* (2016).

[10] 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 (2020), 1234–1240.

[11] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. 2015. Gated graph sequence neural networks. *arXiv preprint arXiv:1511.05493* (2015).

[12] Yu-Wei Lin, Yuqian Zhou, Faraz Faghri, Michael J Shaw, and Roy H Campbell. 2019. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PloS one* 14, 7 (2019), e0218942.

[13] Qiuhao Lu, Nisansa de Silva, Sabin Kafle, Jiazhen Cao, Dejing Dou, Thien Huu Nguyen, Prithviraj Sen, Brent Hailpern, Berthold Reinwald, and Yunyao Li. 2019. Learning Electronic Health Records through Hyperbolic Embedding of Medical Ontologies. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. 338–346.

[14] Xin Ma, Yabin Si, Zifan Wang, and Youqing Wang. 2020. Length of stay prediction for ICU patients using individualized single classification algorithm. *Computer methods and programs in biomedicine* 186 (2020), 105224.

[15] Guoshun Nan, Zhijiang Guo, Ivan Sekulić, and Wei Lu. 2020. Reasoning with Latent Structure Refinement for Document-Level Relation Extraction. *arXiv preprint arXiv:2005.06312* (2020).

[16] Mark Neumann, Daniel King, Iz Beltagy, and Waleed Ammar. 2019. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. In *Proceedings of the 18th BioNLP Workshop and Shared Task*. Association for Computational Linguistics, Florence, Italy, 319–327. https://doi.org/10.18653/v1/W19-5034 arXiv:arXiv:1902.07669

[17] Giannis Nikolentzos, Antoine Tixier, and Michalis Vazirgiannis. 2020. Message passing attention networks for document understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 8544–8551.

[18] Lin Qiu, Yunxuan Xiao, Yanru Qu, Hao Zhou, Lei Li, Weinan Zhang, and Yong Yu. 2019. Dynamically fused graph network for multi-hop reasoning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 6140–6150.

[19] Y Xue, D Klabjan, and Luo Yuan. 2018. Predicting ICU readmission using grouped physiological and medication trends. *Artificial intelligence in medicine* (2018), 4.

[20] Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7370–7377.

[21] Xiao Zhang, Dejing Dou, and Ji Wu. 2020. Learning Conceptual-Contextual Embeddings for Medical Text.. In *AAAI*. 9579–9586.

[22] Yijia Zhang, Qingyu Chen, Zhihao Yang, Hongfei Lin, and Zhiyong Lu. 2019. BioWordVec, improving biomedical word embeddings with subword information and MeSH. *Scientific data* 6, 1 (2019), 1–9.

[23] Yufeng Zhang, Xueli Yu, Zeyu Cui, Shu Wu, Zhongzhen Wen, and Liang Wang. 2020. Every Document Owns Its Structure: Inductive Text Classification via Graph Neural Networks. *arXiv preprint arXiv:2004.13826* (2020).