Unraveling Clinical Insights: A Lightweight and Interpretable Approach for Multimodal and Multilingual Knowledge Integration

Kanimozhi Uma, Marie-Francine Moens
KU Leuven, Belgium
{kanimozhi.uma, sien.moens}@kuleuven.be

Abstract

In recent years, the analysis of clinical texts has evolved significantly, driven by the emergence of language models like BERT such as PubMedBERT, and ClinicalBERT, which have been tailored for the (bio)medical domain that rely on extensive archives of medical documents. While they boast high accuracy, their lack of interpretability and language transfer limitations restrict their clinical utility. To address this, we propose a new, lightweight graph-based embedding method designed specifically for radiology reports. This approach considers the report's structure and content, connecting medical terms through the multilingual SNOMED Clinical Terms knowledge base. The resulting graph embedding reveals intricate relationships among clinical terms, enhancing both clinician comprehension and clinical accuracy without the need for large pre-training datasets. Demonstrating the versatility of our method, we apply this embedding to two tasks: disease and image classification in X-ray reports. In disease classification, our model competes effectively with BERT-based approaches, yet it is significantly smaller and requires less training data. Additionally, in image classification, we illustrate the efficacy of the graph embedding by leveraging cross-modal knowledge transfer, highlighting its applicability across diverse languages.

Keywords: Multimodal graph learning, Knowledge integration, Clinical reports, Language models

1. Background and Introduction

The advent of transformer-based architectures has revolutionized the field of medical text and image processing. Fine-tuned versions of the Bidirectional Encoder Representation from Transformers (BERT) model, such as ClinicalBERT (Alsentzer et al., 2019) and BioBERT (Lee et al., 2020), have demonstrated remarkable performance (Gu et al., 2021). ClinicalBERT, in particular, excels in tasks related to radiology reports of X-ray scans, including text-based disease classification and report generation. However, the direct application of general natural language processing (NLP) methods to the medical domain presents significant challenges, advocating for the development of domain-specific solutions for processing medical text. Due to the fact that the BERT models primarily focus on the English language, adapting these large language models for multilingual use poses significant challenges, leading to reduced performance. Also, Spanish counterparts of BERT, such as BETO (Cañete et al., 2023) and bio-cli-52k (Carrino et al., 2021) are trained on approximately ten times less data, resulting in inferior performance.

Rather than relying on self-supervision, we can employ structured medical knowledge. The Unified Medical Language System (UMLS) comprises standardized definitions and relationships within medical terminologies and vocabularies across 25 languages (Bodenreider, 2004). The UMLS can be implemented across various national hospitals and even transnationally, and it is particularly beneficial for countries that lack access to large medical datasets due to their smaller population size or limited financial resources. The specific ontologies within UMLS offer additional advantages, as SNOMED CT provides connections between concepts within its respective ontology. The additional information from this knowledge base can be valuable, as expert-level annotation is rare in the medical domain. A particularly useful application of SNOMED CT is in clinical reports, which are widely available in public datasets but are largely unannotated. The structure of UMLS and SNOMED CT makes them suitable for representation with knowledge graphs, which can efficiently represent structured sets of entities (Chang et al., 2020). In specialized domains like medicine, language models must learn directly from domain-specific terminologies to enhance reliability, rather solely relying on corpus-based learning.

This paper introduces a novel self-attention graph embedding method for structuring clinical reports, integrating information from existing medical knowledge by leveraging both the report's structure and its linguistic content. Our experimental results demonstrate that the proposed self-attention graph embedding achieves competitive performance, and the text embeddings from the clinical report offer a more computationally efficient, more interpretable, and more intuitive alternative to existing embedding methods. We demonstrate that the utilization of UMLS and SNOMED CT facilitates effortless translation across languages, and finally, the proposed report graph can be integrated into a multimodal framework for knowledge transfer to images, enabling improved classification accuracy.

Proceedings of CL4Health 2024 @LREC-COLING-2024, pages 197–203
20 May, 2024.

2. Related Work

Two main approaches are currently utilized for embedding medical text: pre-trained models on biomedical datasets and fine-tuned versions of those models. The pre-trained models include BioWordVec (Zhang et al., 2019), and the fine-tuned models include BERT variants such as BioBERT (Lee et al., 2020), ClinicalBERT (Alsentzer et al., 2019), PubMedBERT (Gu et al., 2021) and many more, which are fine-tuned on (bio)medical and clinical datasets. SapBERT (Liu et al., 2021a), which introduced a self-alignment strategy for learning from UMLS synonym pairs via a multi-similarity (MS) loss function to force related concepts closer to one another in BERT's representation space. These embeddings form the basis of the most recent state-of-the-art methods for radiology report generation and outperform previous methods by a large margin. Knowledge graphs have been employed to improve patient record-based diagnosis (Heilig et al., 2022), enhance entity extraction from radiology reports (Jain et al., 2021), and supplement image diagnosis (Prabhakar et al., 2022). The graphs used to generate a clinical report are typically small (approximately containing 15 nodes) consisting of disease labels, proving to be an effective method for capturing the global context (Yan, 2022; Liu et al., 2021c,b). Our approach suggests embedding from a clinical radiology report, rather than for generating one. No prior research has explored structured representations encompassing full clinical reports with knowledge graphs and medical ontologies.

3. Methodology

Our proposed method consists of three steps: extracting entities from the clinical report, constructing a knowledge graph, and encoding the graph, as shown in Figure 1. The graph's nodes represent words in the report that match terms from the clinical database, and the edges represent relationships between these terms and their locations in the report.

3.1. Named Entity Recognition

Clinical concepts C embedded within the plain text of the clinical reports are extracted using Named Entity Recognition (NER) through MetaMap (Aronson and Lang, 2010), on English UMLS concepts; and for Spanish, we utilize UMLSMapper (Perez et al., 2020). From C, we extract UMLS Concept Unique Identifiers $CUIs : \{i_0, i_1, \ldots\} = Concepts(C)$. For each CUI, the corresponding SNOMED-CT concept is extracted, resulting in the final set of clinical concepts from report $C : \{c_0, c_1, \ldots\}$.

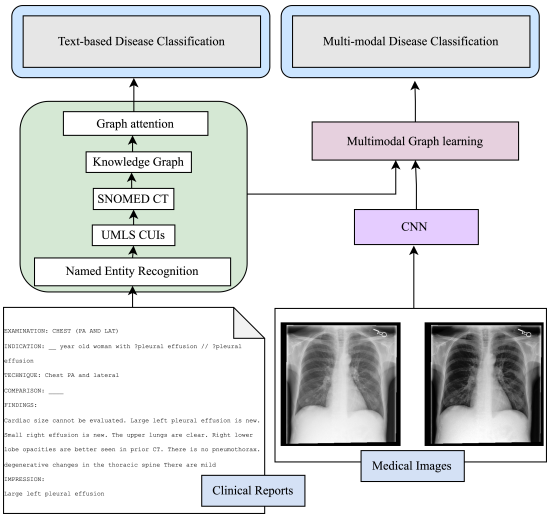

Figure 1: Overview of the architecture showcasing the process of constructing and assessing knowledge graph embeddings and encoding the graph with both textual and image data for disease classification.

3.2. Knowledge Graph Construction

An undirected graph is considered in our graph construction phase where the Graph $G_C = (N_C, E_C)$ defined by a set of nodes $N_C = \{n_0, n_1, \ldots\}$ and edges $E_C = \{e^{l \leftrightarrow m}, \ldots\}$, with $e^{l \leftrightarrow m} = (n_l, n_m)$. The structure of the clinical report is captured by considering each sentence separately with sentence node S^i of sentence i. The SNOMED-CT concepts extracted from the clinical report C gives us a set of concepts per sentence: $\{c^{i,0}, c^{i,1}, \ldots\}$. To capture the context between nodes we have the global connect node g_n with the following set of nodes $N_C = \{s^i, g_n, c\}$. There are three types of edges: edges between the concept nodes, edges from the sentence nodes, and the global connect node, which connects to every concept node $E_C = \{e^{c \leftrightarrow c}, e^{s \leftrightarrow c}, e^{g_n \leftrightarrow s}, e^{g_n \leftrightarrow c}\}$. The types of edges are as follows: (a) Concept nodes c are linked by edges $e^{c \leftrightarrow c}$ if they have a contextual relation based on the SNOMED CT ontology. (b) Sentence nodes s^i are linked to all the c nodes in their sentence by edges $e^{s \leftrightarrow c}$, representing the local structure of the report. (c) A global node g_n connects to every concept node and sentence node through edges $e^{g_n \leftrightarrow c}$ and $e^{g_n \leftrightarrow s}$, allowing for communication across the whole report and thus, $E_C = \{(c^l, c^m)\}, \{(c^l, s^i)\}, \{(s^i, g_n)\}, \{(c^l, g_n)\}$.

3.3. Self-Attention Knowledge Graph Encoding

Graph attention networks use the self-attention mechanism to allow nodes in G_C to focus on their neighbors effectively (Velickovic et al., 2017) and are the preferred method for encoding knowledge graphs. The node N_C^u and neighbor nodes N_C^v

with their weight matrices $W \in R^{L'_f * L_f}$, where L_f is the length of the node feature. The normalized attention score between these nodes can be written as: $att_{uv} = softmax(LRelu(W_a tt^x [W_{N_C^u} || W_{N_C^v}]))$. The node N_C^u is encoded as: $N_C^{u\prime} = \sigma(\sum att_{uv} \cdot W_{N_C^v})$, where σ is a nonlinear function. The whole graph with n attention layers stacked on top of each other is encoded as:$G'_C = f^n_{attention}(N_C, E_C)$.

3.4. Analysis and Inferences

To determine whether the constructed graph embeddings capture information representative of the content of the clinical report, we evaluate them on two classification tasks, as illustrated in Figure 1. The effectiveness of our knowledge graph embedding is assessed for a diagnosis classification based on the clinical report, and we compare our method against biomedical variants of BERT. Disease classification is performed on the entire report through a max pooling operation on encoded node representations N'_C, followed by a classification.

The ability of our graph embedding to transfer the information it contains across modalities is tested by integrating the embeddings into a knowledge distillation with variational inference (KDVI) (Ahn et al., 2019), where graph embeddings are employed to improve image-based disease classifications. A conditional latent variable model was introduced to distill the information from clinical report C to chest X-ray scan through variational inference and we draw inspiration from conditional variational inference through KG reasoning (CVIR) (Chen et al., 2018).

The evidence lower bound objective (ELBO) of CVIR consists of a term for reconstructing the data and a term for measuring the Kullback-Leibler (KL) divergence: L = E$[\log p(y|I, z_I)] - D_{KL}[q(z)||p(z_I|I)]$ where z_I is a hidden representation of I, y are the class labels, and $p(z_I|I)$ is the prior distribution over z. $p(z_I|I) \cdot q(z)$ is the posterior distribution over z, which is usually an isotropic Gaussian distribution $\mathcal{N}(0, I)$. In KDVI, the posterior is $q(z_C|C)$, where z_C is a hidden representation of C. This new posterior allows us to extract information from C to I by minimizing this KL term: $D_{KL}[q(z_C|C)||p(z_I|I)]$. This way, we transfer knowledge and information from text to imaging. During the training phase, we need both as input, but in the testing phase, we only need the imaging data. With this approach, we use our text-based knowledge graph embeddings of the clinical report to enhance the image representations.

4. Experiments and Results

Three datasets were used for training and evaluation: MIMIC-CXR (Johnson et al., 2019), OpenI (Demner-Fushman et al., 2016), and PadChest (Bustos et al., 2020). MIMIC-CXR comprises 377,110 chest X-rays and 227,827 anonymized radiology reports, with disease labels generated using a rule-based labeler. OpenI consists of 7,470 chest X-rays and 3,955 anonymized reports with similar disease labels. PadChest contains 160,000 radiology images and Spanish clinical reports with 174 disease labels. The disease labels from all three datasets can be consolidated into a unified label space. There are no limitations on the number of entities that can be extracted from the clinical report using NER with MetaMap or on the number of edges within a graph. We initialize the nodes using vectorized representations of individual UMLS concepts, which are obtained by pre-training on datasets containing (bio)medical data (Beam et al., 2019). These 200-dimensional non-contextual embeddings are inspired by word2vec and can be directly integrated into our method without requiring additional processing steps. Embedding initializations for s_i and g_n are calculated by averaging the node embeddings n over the sentence and the entire graph, respectively.

Metric	BioBERT	PubMedBERT	ClinicalBERT	Ours
AUC	0.920	0.946	0.974	0.948
Recall	0.840	0.846	0.868	0.850
Precision	0.593	0.600	0.611	0.632
F1	0.688	0.713	0.726	0.726

Table 1: Comparison of our approach with baseline models on MIMIC-CXR dataset

Metric	BioBERT	PubMedBERT	ClinicalBERT	Ours
AUC	0.929	0.929	0.965	0.966
Recall	0.768	0.818	0.818	0.864
Precision	0.560	0.571	0.602	0.625
F1	0.680	0.717	0.700	0.717

Table 2: Comparison of our approach with baseline models on OpenI dataset

Metric	BETO	Bio-cli-52k	Ours
AUC	0.881	0.918	0.966
Recall	0.509	0.678	0.841
Precision	0.219	0.490	0.545
F1	0.555	0.574	0.682

Table 3: Comparison of our approach with baseline models

The graph attention encoder consists of 1, 3, 6, or 12 consecutive graph attention layers with hidden sizes of 512, 1024, 2048, or 4094, and the graph classification is performed using a multilayer perceptron (MLP) with dimensions 512, 256, 14, 8, employing cross-entropy loss. The results are reported using the AUC metric, in line with existing

benchmarks, and the evaluation of VKD is conducted with a latent space size of 2048 and 12 sequential graph attention layers for encoding. Other hyperparameter settings are directly adopted from (Chen et al., 2018), and a dropout rate of 0.5 is applied to all layers of the architecture.

Training is performed on two NVIDIA GeForce RTX 3090 GPUs, utilizing Adam optimizer with early stopping and a tolerance of 1%. Tables 1, 2, and 3 present the disease classification results obtained from our graph embedding approach, and the results are shown for an encoder with a hidden size of 1024 and 3 graph attention layers, which demonstrated the best overall performance. Our method achieves comparable performance to BioBERT and PubMedBERT, and slightly lower performance compared to ClinicalBERT while utilizing significantly fewer parameters (Table 4). On the Spanish PadChest dataset, our method outperforms BERT-based methods. This can be attributed to the smaller training datasets of these models, which are ten times smaller than their English counterparts. Furthermore, our method exhibits faster inference rates (samples per second) on GPU platforms and performs relatively better on the smaller OpenI dataset, highlighting the effectiveness of our embedding for report representation without the need for large datasets.

Dataset	**ClinicalBERT**	**Ours**
MIMIC-CXR (Recall)	0.564	0.528
MIMIC-CXR (Precision)	0.537	0.537
MIMIC-CXR (F1)	0.508	0.469
MIMIC-CXR (AUC)	0.838	0.825
OpenI (Recall)	0.581	0.579
OpenI (Precision)	0.533	0.528
OpenI (F1)	0.509	0.497
OpenI (AUC)	0.884	0.864
PadChest (Recall)	0.528	0.534
PadChest (Precision)	0.510	0.532
PadChest (F1)	0.480	0.474
PadChest (AUC)	0.814	0.821

Table 4: Classification Accuracy

Employing graphs within a multi-modal framework can enhance our understanding of how effectively the graph captures intricate information structures that can span across modalities. Table 4 presents the results of our method for multi-modal knowledge transfer compared to existing methods that utilize ClinicalBERT as a clinical report embedding. Training this framework revealed that the convergence of this model is complex for graphs with shallow encoders and smaller hidden layers. As a result, Table 4 displays results obtained with an encoder comprising 12 graph attention layers and a hidden size of 2048. While not outperforming the existing ClinicalBERT method, graph embeddings demonstrate applicability on both MIMIC-CXR and OpenI. Performance improvements are observed compared to image-only classification, indicating successful multi-modal knowledge transfer with very limited pre-training data.

	MIMIC-CXR	**OpenI**	**PadChest**
Full graph	0.930	0.946	0.950
w/o graph	0.924	0.939	0.940
w/o g_n	0.917	0.931	0.933
w/o C	0.917	0.935	0.925
w/o $g_n \& S \& C$	0.919	0.915	0.920

Table 5: Graph ablations on the classification tasks

Also, we delved into the application of graph encoders for disease classification, investigating the influence of encoder count and hidden size on performance. Parameter count emerges as a crucial factor, with clinicalBERT delivering superior performance but requiring greater computational resources. The performance discrepancy between the smallest, say, 0.4M parameters and the largest, say, 62M parameters models is relatively minor, indicating that graph construction effectively captures medical knowledge irrespective of encoder size. Furthermore, we conducted an ablation study on the graph components in Table 5. The importance of node types is highlighted by removing them from the graph. The global node and edges connecting SNOMED CT concepts stand out as key elements in the graph structure. This underscores how the integration of report composition and the medical knowledge base (SNOMED CT) yields a rich representation of the report and shows how our graph handles repeated terms more efficiently than ClinicalBERT: ours consists of 34 nodes, while tokenization with ClinicalBERT requires as many as 124 tokens. For instance, our knowledge graph captures the entire word such as 'dencities', 'opacities' as an entity, but ClinicalBERT, tokenizes into 'den', 'cities', 'o', 'pa', and 'cities', respectively. The final token in this sequence clearly carries a different contextualized meaning within the general-purpose language BERT model that underlies ClinicalBERT, which demonstrates that graphs can capture medical terminology in a more intuitive and interpretable way.

5. Conclusion

This paper presents a novel knowledge graph-based method for creating structured representations of clinical reports. The knowledge graph embeddings explicitly encode medical knowledge from clinical knowledge bases, facilitating transfer across domains and languages without relying on large datasets. Concurrently, the proposed method maintains a significantly smaller model size compared to existing BERT-based models. By captur-

ing both structural and content relationships embedded within existing knowledge bases, the proposed self-attention-based graph representations achieve comparable performance to current state-of-the-art transformer-based models in English and Spanish, resulting in more informative representations of clinical reports. Our future scope is to extend this architecture to Dutch, German, Estonian, and more languages.

6. Acknowledgements

We thank the anonymous reviewers for the valuable comments.The authors acknowledge the AIDAVA project financed by Horizon Europe (EU HORIZON-HLTH-2021-TOOL-06-03) and the CHIST-ERA ANTIDOTE project financed by FWO G0H6820N for supporting the research.

7. Bibliographical References

Sungsoo Ahn, Shell Xu Hu, Andreas Damianou, Neil D Lawrence, and Zhenwen Dai. 2019. Variational information distillation for knowledge transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9163–9171.

Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. 2019. Publicly available clinical bert embeddings. *arXiv preprint arXiv:1904.03323*.

Alan R Aronson and François-Michel Lang. 2010. An overview of metamap: historical perspective and recent advances. *Journal of the American Medical Informatics Association*, 17(3):229–236.

Andrew L Beam, Benjamin Kompa, Allen Schmaltz, Inbar Fried, Griffin Weber, Nathan Palmer, Xu Shi, Tianxi Cai, and Isaac S Kohane. 2019. Clinical concept embeddings learned from massive sources of multimodal medical data. In *Pacific Symposium on Biocomputing 2020*, pages 295–306. World Scientific.

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

Aurelia Bustos, Antonio Pertusa, Jose-Maria Salinas, and Maria De La Iglesia-Vaya. 2020. Padchest: A large chest x-ray image dataset with multi-label annotated reports. *Medical image analysis*, 66:101797.

José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, and Jorge Pérez. 2023. Spanish pre-trained bert model and evaluation data. *arXiv preprint arXiv:2308.02976*.

Casimiro Pio Carrino, Jordi Armengol-Estapé, Asier Gutiérrez-Fandiño, Joan Llop-Palao, Marc Pàmies, Aitor Gonzalez-Agirre, and Marta Villegas. 2021. Biomedical and clinical language models for spanish: On the benefits of domain-specific pretraining in a mid-resource scenario. *arXiv preprint arXiv:2109.03570*.

Arlene Casey, Emma Davidson, Michael Poon, Hang Dong, Daniel Duma, Andreas Grivas, Claire Grover, Víctor Suárez-Paniagua, Richard Tobin, William Whiteley, et al. 2021. A systematic review of natural language processing applied to radiology reports. *BMC medical informatics and decision making*, 21(1):179.

David Chang, Ivana Balažević, Carl Allen, Daniel Chawla, Cynthia Brandt, and Richard Andrew Taylor. 2020. Benchmark and best practices for biomedical knowledge graph embeddings. In *Proceedings of the conference. Association for Computational Linguistics. Meeting*, volume 2020, page 167. NIH Public Access.

Wenhu Chen, Wenhan Xiong, Xifeng Yan, and William Wang. 2018. Variational knowledge graph reasoning. *arXiv preprint arXiv:1803.06581*.

Dina Demner-Fushman, Marc D Kohli, Marc B Rosenman, Sonya E Shooshan, Laritza Rodriguez, Sameer Antani, George R Thoma, and Clement J McDonald. 2016. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2):304–310.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. 2021. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1):1–23.

Niclas Heilig, Jan Kirchhoff, Florian Stumpe, Joan Plepi, Lucie Flek, and Heiko Paulheim. 2022. Refining diagnosis paths for medical diagnosis based on an augmented knowledge graph. *arXiv preprint arXiv:2204.13329*.

Jinpeng Hu, Jianling Li, Zhihong Chen, Yaling Shen, Yan Song, Xiang Wan, and Tsung-Hui Chang. 2021. Word graph guided summarization for radiology findings. *arXiv preprint arXiv:2112.09925*.

Jinpeng Hu, Zhuo Li, Zhihong Chen, Zhen Li, Xiang Wan, and Tsung-Hui Chang. 2022. Graph enhanced contrastive learning for radiology findings summarization. *arXiv preprint arXiv:2204.00203*.

Saahil Jain, Ashwin Agrawal, Adriel Saporta, Steven QH Truong, Du Nguyen Duong, Tan Bui, Pierre Chambon, Yuhao Zhang, Matthew P Lungren, Andrew Y Ng, et al. 2021. Radgraph: Extracting clinical entities and relations from radiology reports. *arXiv preprint arXiv:2106.14463*.

Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. 2021. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2):494–514.

Alistair EW Johnson, Tom J Pollard, Seth J Berkowitz, Nathaniel R Greenbaum, Matthew P Lungren, Chih-ying Deng, Roger G Mark, and Steven Horng. 2019. Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific data*, 6(1):317.

Kaveri Kale, Pushpak Bhattacharyya, Aditya Shetty, Milind Gune, Kush Shrivastava, Rustom Lawyer, and Spriha Biswas. 2022. Knowledge graph construction and its application in automatic radiology report generation from radiologist's dictation. *arXiv preprint arXiv:2206.06308*.

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.

Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. 2017. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88.

Fangyu Liu, Ehsan Shareghi, Zaiqiao Meng, Marco Basaldella, and Nigel Collier. 2021a. Self-alignment pretraining for biomedical entity representations. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4228–4238, Online. Association for Computational Linguistics.

Fenglin Liu, Xian Wu, Shen Ge, Wei Fan, and Yuexian Zou. 2021b. Exploring and distilling posterior and prior knowledge for radiology report generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13753–13762.

Fenglin Liu, Chenyu You, Xian Wu, Shen Ge, Xu Sun, et al. 2021c. Auto-encoding knowledge graph for unsupervised medical report generation. *Advances in Neural Information Processing Systems*, 34:16266–16279.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Naiara Perez, Pablo Accuosto, Àlex Bravo, Montse Cuadros, Eva Martínez-Garcia, Horacio Saggion, and German Rigau. 2020. Cross-lingual semantic annotation of biomedical literature: experiments in spanish and english. *Bioinformatics*, 36(6):1872–1880.

Chinmay Prabhakar, Anjany Sekuboyina, Hongwei Bran Li, Johannes C Paetzold, Suprosanna Shit, Tamaz Amiranashvili, Jens Kleesiek, and Bjoern Menze. 2022. Structured knowledge graphs for classifying unseen patterns in radiographs. In *Geometric Deep Learning in Medical Image Analysis*, pages 45–60. PMLR.

Ardavan Saeedi, Yuria Utsumi, Li Sun, Kayhan Batmanghelich, and Li-wei Lehman. 2022. Knowledge distillation via constrained variational inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8132–8140.

Kihyuk Sohn, Honglak Lee, and Xinchen Yan. 2015. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems*, 28.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. 2017. Graph attention networks. *stat*, 1050(20):10–48550.

Sixing Yan. 2022. Memory-aligned knowledge graph for clinically accurate radiology image report generation. In *Proceedings of the 21st Workshop on Biomedical Language Processing*, pages 116–122.

Shuxin Yang, Xian Wu, Shen Ge, S Kevin Zhou, and Li Xiao. 2022. Knowledge matters: Chest radiology report generation with general and specific knowledge. *Medical image analysis*, 80:102510.

Dehai Zhang, Anquan Ren, Jiashu Liang, Qing Liu, Haoxing Wang, and Yu Ma. 2022. Improving medical x-ray report generation by using knowledge graph. *Applied Sciences*, 12(21):11111.

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):52.

Yixiao Zhang, Xiaosong Wang, Ziyue Xu, Qihang Yu, Alan Yuille, and Daguang Xu. 2020. When radiology report generation meets knowledge graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12910–12917.

Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2020. Graph neural networks: A review of methods and applications. *AI open*, 1:57–81.