Enhancing Radiology Report Interpretation through Modality-Specific RadGraph Fine-Tuning

Haoyue Guan[1], Yuwei Dai[2], Shadi Afyouni[2], Alec Kain[1], Wen-Chi Hsu[1], Jiashu Cheng[1], Sophie Yao[1], Yuli Wang[1], Jing Wu[3], Rishitha Pulakhandam[1], Lin-mei Zhao[2], Chengzhang Zhu[4,*], Zhicheng Jiao[5], Craig Jones[1], and Harrison Bai[2,*]

[1] Johns Hopkins University, MD, 21218, USA
hguan16@jh.edu
[2] Johns Hopkins University School of Medicine, MD, 21205, USA
hbai7@jh.edu
[3] Second Xiangya Hospital, Central South University, China
[4] Central South University, Changsha, China
chzhzhu@csu.edu.cn
[5] Brown University, RI, 02912, USA

Abstract. Radiology reports contain free-form text that conveys critical clinical information derived from imaging studies and patient history. However, the unstructured nature of these reports, coupled with the complexity and ambiguity of natural language, poses significant challenges for automated information extraction, particularly in domains with limited labeled data. To address this, we introduce a novel expert-annotated dataset encompassing four new imaging modalities: cardiac magnetic resonance imaging (MRI), abdominal ultrasound, head computerized tomography (CT), and CT pulmonary angiography (CTPA). Leveraging this dataset, we developed transformer-based models optimized for entity recognition and relation extraction within specific modalities, enabling the generation of high-quality radiology annotations. Our evaluation of fine-tuning methods demonstrate that modality-specific models achieve a 12.5% macro F1 score improvement in entity recognition and a 28.3% improvement on relation extraction tasks compared to prior approaches. These findings highlight the potential of fine-tuned, modality-specific models in enhancing automated radiology text processing and downstream applications. By releasing the model and datasets, we aim to foster research on wider modalities in medical natural language processing across a broader range of imaging modalities. The code is available at https://github.com/tonikroos7/RadGraph-Multimodality.

Keywords: Radiology Report · Deep Learning · Natrual Language Processing.

1 Introduction

Radiology reports are essential for assessing a patient's clinical condition based on imaging studies. They serve a key role in medical natural language processing (NLP) applications, including information extraction, sentiment analysis, and predictive modeling [2]. However, nearly 80% of electronic health records (EHRs) consist of unstructured, free-form texts, making automated processing for secondary applications challenging and complicating the extraction of valuable clinical insights [17]. Supervised deep learning approaches, which dominate current NLP methods, require large amounts of high-quality labeled data. This need is particularly evident in radiology, where imaging modalities such as head CT and cardiac MRI rely on precise, high-fidelity annotations. However, high-quality datasets remain scarce, and even when available, annotation inconsistencies across datasets can introduce discrepancies that hinder model performance [20]. To address these challenges, developing a high-quality radiology report dataset with a standardized annotation framework is crucial for improving the accessibility and usability of medical data. Existing datasets like MIMIC-CXR [15] and CheXpert [12] link radiographic studies to free-text radiology reports but often fail to capture fine-grained clinical details. More advanced approaches integrate entity extraction [4] and prioritize factual accuracy [6], though they remain heavily dependent on expert annotations.

In medical information extraction, traditional machine learning methods, such as sequential classifiers, have been widely used to extract medical concepts [19]. More recent approaches leverage deep learning architectures, including long short-term memory (LSTM) networks, recurrent neural networks (RNNs), and bi-directional LSTMs [5, 9, 10, 13]. Transformer-based architectures [21] have further advanced the field, enabling large language models (LLMs) to achieve state-of-the-art performance across a range of medical NLP tasks. Within LLMs, named entity recognition (NER) serves as a fundamental component for applications such as knowledge graph construction, sentiment analysis, and question answering [18, 25].

RadGraph utilizes language models to identify clinical entities and inter-entity relationships in chest X-ray reports [14]. While RadGraph has been instrumental in downstream tasks, such as automatic radiology report generation [7], knowledge graph integration [24], style-aware radiology report generation [23], and long-time disease progression tracking [16], its applicability remains limited to the X-ray modality. To expand its utility, RadGraph-XL [8] extends the framework to a broader range of imaging modalities. However, the availability of high-quality, annotated radiology reports across multiple modalities remains insufficient, limiting the effectiveness of current models. To address this gap, we extend the RadGraph dataset to support a more diverse set of radiology modalities. Our contributions can be summarized as follows: 1) We introduce a new dataset covering four commonly used but previously underrepresented imaging modalities, annotated by expert radiologists. 2) We develop a fine-tuning approach leveraging a BERT-based language model, achieving high accuracy in both NER and relation extraction tasks. 3) We demonstrate the effectiveness of modality-specific RadGraph extension, highlighting its potential for broader applications in clinical practice and research.

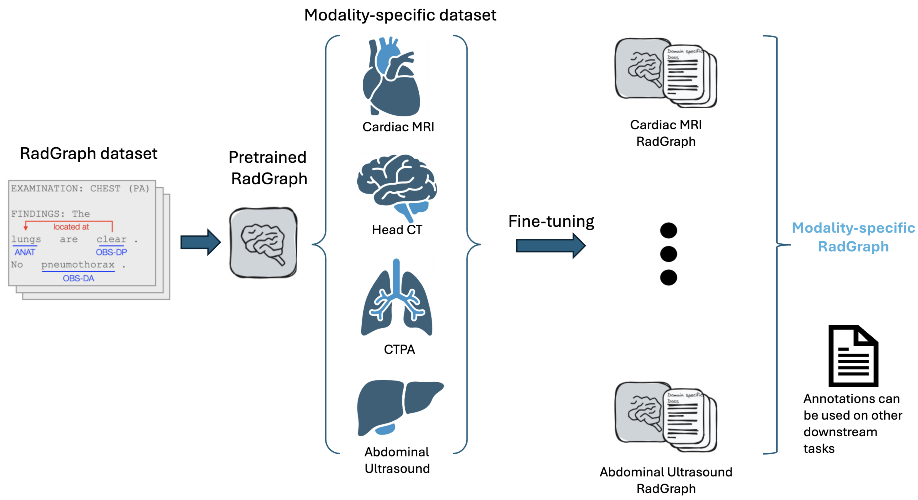

Fig. 1. Overview of the fine-tuning pipeline. The model is initially trained on the RadGraph dataset [14] and then fine-tuned on different modalities. The resulting high-quality, modality-specific model can then be applied to information extraction and other downstream tasks.

2 Proposed Framework

We adopt the annotation scheme introduced in RadGraph [14] to label entities and relations in radiology reports. To extend the model's multi-modality capabilities, we developed a new expert-annotated dataset in collaboration with radiologists. To ensure high-quality standards while expanding upon prior work, annotators were trained using the same guidelines as RadGraph and evaluated on a subset of its dataset. This approach maintains compatibility between our dataset and RadGraph, enabling robust multi-modality information extraction and evaluation.

New Modality We extend RadGraph by incorporating additional modalities, expanding its applicability beyond chest X-rays. The newly included modalities are cardiac MRI for detailed tissue characterization and cardiac function assessment, CT pulmonary angiography for pulmonary vasculature evaluation, including pulmonary embolism detection, head CT for rapid assessment of acute conditions such as stroke, hemorrhage, and trauma, and abdominal ultrasound for non-invasive evaluation of abdominal organs, including hepatobiliary pathology. These imaging modalities serve as critical first-line diagnostic and screening tools for cardiovascular, digestive, and neurological disorders, conditions with high morbidity and mortality rates that often necessitate urgent intervention. By incorporating these underrepresented modalities, our dataset introduces clinically significant scenarios ranging from routine imaging to emergency conditions that are currently lacking in existing medical NLP research.

New Anatomical and Pathological Annotations Our dataset introduces annotations for previously unrepresented anatomical regions and diseases, enhancing its clinical relevance for downstream applications. These additions include cardiac abnormalities including pulmonary vein anomalities, atrial dilation, and tricuspid regurgitation, abdominal conditions such as cholelithiasis and acute cholecystitis, pulmonary vasculature pathologies including pulmonary embolism, and head and brain pathologies such as acute hemorrhage, ischemic stroke, and brain masses. By incorporating a diverse range of anatomical structures and pathological findings, our dataset enables more comprehensive information extraction, fostering advancements in multi-modality medical NLP and decision support systems.

2.1 Annotation Scheme

Entity Entities are defined as spans of text or individual words within the radiology reports. These entities are categorized into two main types: Anatomy and Observation. Anatomy is the text that is related to anatomical body parts that exist in the radiology report. Observation refers to words associate with visual features, identifiable pathophysiological processes, or diagnostic disease classifications. There are three categories for observation: Observation: Definitely Present, Observation: Definitely Absent, and Observation: Uncertain, which denotes the level of certainty regarding the observation's presence. In total, the schema defines four entity categories.

Relation Relations are defined as directed edges between two entities, representing their interactions. The schema includes three types of relations: Located At, Modify, and Suggestive Of. Located At links an anatomy entity to an observation, indicating either the location of the observation or a relationship between the anatomy and the observation. Modify denotes the relationship between two observations or two anatomy entities, where the first entity alters or impacts the second. Suggestive Of connects two observations if the first observation infers the presence of the second observation. Compared to previous annotation techniques, this schema captures relationships more precisely while simplifying the annotation process.

2.2 Dataset

To explore the multi-modality capabilities of the RadGraph model, we extend its dataset by incorporating four additional imaging modalities while closely following the annotation procedures outlined in the original RadGraph study [14]. Each dataset was annotated by two radiologists or trainees who independently labeled the reports before reviewing their annotations together to reach a consensus. To ensure consistency, all annotators followed the same structured annotation guidelines as in the original RadGraph dataset and were trained using the previous RadGraph annotations. The annotation process was conducted using

Datasaur [1], a widely used text annotation platform that allow radiologists to annotate reports directly based on a predefined schema. A detailed breakdown of our dataset is provided in Table 1. In total we collected 800 radiology reports, comprising of 100 cardiac MRI reports, 200 CTPA reports, 300 head CT reports, and 200 abdominal ultrasound reports. We aimed to maintain a balanced report distribution across modalities, ensuring sufficient annotations for training and fine-tuning. Our dataset contains 14,813 unique entity pairs (entity, label) and 11,179 unique relation tuples (entity 1, entity 2, relation). Overall, the dataset size has increased by 60% compared to the RadGraph, which was limited to chest X-ray reports and different from RadGraph-XL which contains chest CT, abdomen/pelvis CT, brain MR, and chest X-rays [8]. The entity distribution is relatively balanced, with anatomical annotations comprising approximately 40% of total annotations and observation annotations accounting for 60%. The relation distribution follows a similar trend observed in previous work [14, 8], reflecting the inherent structure of radiology reports.

Table 1. Category distribution statistics of our dataset.

	MRI (%)	CTPA (%)	HeadCT (%)	Ultrasound(%)
Anatomy	979 (42.1)	2671 (45.3)	1765 (40.5)	734 (32.9)
Observation: Definitely Present	1175 (50.5)	2546 (43.2)	1550 (35.5)	916 (41.0)
Observation: Uncertain	57 (2.5)	247 (4.2)	74 (1.7)	152 (6.8)
Observation: Definitely Absent	115 (4.9)	429 (7.3)	973 (22.3)	430 (32.9)
Total Entities	2,326 (100)	5893 (100)	4362 (100)	2232 (100)
Modify	1232 (69.4)	3256 (72.0)	2404 (71.9)	986 (64.1)
Located at	454 (25.6)	962 (21.3)	829 (24.8)	389 (25.3)
Suggestive of	90 (5.1)	302 (6.7)	112 (3.3)	163 (10.6)
Total Relations	1,776 (100)	4,520 (100)	3,345 (100)	1,538 (100)

3 Experiments and Results

3.1 Approaches

We developed a modality-specific model using our newly annotated dataset in conjunction with the RadGraph dataset [14]. The model leverages RadGraph annotation to achieve performance comparable to previous methods while fine-tuning pre-trained models for adaptation across multiple imaging modalities. For each modality in our dataset, we fine-tuned a modality-specific version of RadGraph for performance evaluation. To ensure compatibility with transformer-based language models, we preprocessed radiology reports by tokenizing text, separating punctuation from words, and structuring them into sequences suitable for training and inference. As a baseline, we employed a BERT-base-uncased model, providing a reference point for evaluating improvements gained from transitioning from a general-purpose model to a domain-specific one. To comprehensively evaluate performance, we integrated several biomedical language models pretrained on domain-specific corpora. These models were incorporated into the DYGIE++ [22] framework, which has demonstrated state-of-the-art results in NER and relation extraction tasks, as evidenced by the RadGraph [14] and RadGraph-XL [8] models. For the biomedVLP-CXR-BERT [3] and BiomedNLP-BiomedBERT [11] models, we used the following hyperparameters: a learning rate of 2e-3, a batch size of 1, the AdamW optimizer with an embedder learning rate of 1e-5, and weight decay of 0.01. A 7:1.5:1.5 split was used for training, validation, and testing. Training was performed on an NVIDIA GeForce RTX 3080 GPU for up to 50 epochs with early stopping to prevent overfitting.

3.2 Pre-trained model selection

We first evaluate the performance of various base models to determine the most suitable pre-trained model for our task. For entity recognition, a prediction is considered correct if the predicted span and entity type exactly match the ground truth. For relation extraction, correctness requires that the span boundaries, entity types, and relation type between entities are all accurately identified. We select the model with the highest overall performance as the base model. Our results indicate that all Biomed-BERT models achieved higher macro F1 scores than the baseline BERT model. This finding reinforces the notion that general-purpose models, which are not fine-tuned on biomedical-specific corpora, struggle to capture the domain-specific nuances of radiology reports, leading to lower performance in clinical NLP tasks. We observe that BiomedNLP-BiomedBERT achieves the highest macro F1 scores on the test set. Based on these results, we select BiomedNLP-BiomedBERT as our base model for further fine-tuning. The performance results for all approaches are summarized in Table 2.

Table 2. Aggregated results of the selected base model on RadGraph dataset

Approach	**Entity**			**Relations**		
	Precision	**Recall**	**Macro F1**	**Precision**	**Recall**	**Macro F1**
Baseline BERT	0.867	0.919	0.916	0.786	0.758	0.772
BiomedVLP-CXR-BERT	**0.916**	0.924	0.920	0.805	0.772	0.788
BlueBERT	0.898	0.910	0.903	0.793	0.748	0.769
BiomedNLP-BiomedBERT*	0.913	**0.937**	**0.924**	**0.827**	**0.825**	**0.826**

3.3 Performance analysis

We fine-tuned four different RadGraph-specific models, each tailored to a single imaging modality. As shown in Table 3, our RadGraph-specific model outperforms both RadGraph and RadGraph-XL across nearly all entity and relation categories for all modalities. For entity recognition, our fine-tuned model

Table 3. Comparative analysis between our fine-tuning method with previous works. "O-DP" stands for "Observation: Definitely Present", "O-DA" stands for "Observation: Definitely Absent", and "O-U" stands for "Observation: Uncertain". The best performance score is shown in bold.

Models	**ANAT**	**O-DP**	**O-DA**	**O-U**	**Modify**	**LocatedAt**	**SuggestiveOf**
Cardiac MRI							
RadGraph	0.827	0.701	0.659	0.207	0.503	0.472	0.261
RadGraph-XL	0.826	0.732	0.720	0.343	0.643	0.550	0.440
RadGraph+fine-tuning	**0.955**	**0.871**	**0.757**	**0.424**	**0.847**	**0.775**	**0.444**
CTPA							
RadGraph	0.900	0.785	0.795	0.595	0.580	0.464	0.590
RadGraph-XL	0.923	0.753	**0.862**	0.549	0.626	0.466	0.526
RadGraph+fine-tuning	**0.925**	**0.822**	0.860	**0.656**	**0.664**	**0.468**	**0.608**
Head CT							
RadGraph	0.886	0.671	0.772	0.444	0.555	0.415	0.370
RadGraph-XL	0.900	0.653	0.793	0.400	0.675	0.488	**0.514**
RadGraph+fine-tuning	**0.912**	**0.833**	**0.896**	**0.667**	**0.797**	**0.602**	0.400
Ultrasound							
RadGraph	0.872	0.810	0.876	0.595	0.616	0.555	0.426
RadGraph-XL	**0.889**	0.815	0.891	0.546	0.685	0.629	0.377
RadGraph+fine-tuning	0.876	**0.843**	**0.897**	**0.615**	**0.730**	**0.655**	**0.576**

achieves an macro F1 score of 0.801, which is 14.3% higher than RadGraph-XL (0.712) and 10.5% higher than RadGraph (0.725). In the major entity category Anatomy: Definitely Present, our model achieves an F1 score of 0.92, which is 5.6% higher than RadGraph and 4% higher than RadGraph-XL. For relation extraction, our model demonstrates even more significant improvements, achieving an F1 score of 0.621, which is 14.8% higher than RadGraph-XL (0.541) and 28.3% higher than RadGraph (0.484). The performance of Suggestive Of category in head CT is affected by the scarcity of corresponding labels, which account for only 3.3% of the dataset. Due to this imbalance, minor variations in predictions can lead to substantial fluctuations in F1 scores. Notably, RadGraph-XL achieved only three additional true positives compared to our model, suggesting a minimal practical difference.

We hypothesize that the performance of Observation: Definitely Absent and Anatomy in the CTPA and ultrasound datasets may be influenced by annotation inconsistencies among different annotators. Despite these minor misalignments, our fine-tuned model remains competitive, with only 0.002 and 0.013 differences in F1 scores, respectively, when compared to RadGraph-XL. Overall, our findings confirm that modality-specific pre-training enhances model performance. Fine-tuning models on domain-specific datasets improves their ability to identify clinical entities and relationships, as they become better adapted to the specialized language and context of radiology reports. These results highlight the robustness of our model and establish it as a strong candidate for downstream tasks in multi-modality radiology NLP.

We evaluated the multi-modality capabilities of our model by training it on the full dataset, which includes cardiac MRI, CTPA, ultrasound, and head CT reports. As shown in Table 4, our multi-modality model achieved and F1 score of 0.854 for entity recognition, representing a 14.6% improvement over RadGraph-XL and 7.4% over RadGraph. Additionally, for relation extraction, our model outperformed RadGraph by 31.9% and RadGraph-XL by 28.3%. To further examine the impact of training data proportions on modality-specific model performance, we selected the Head CT dataset as a case study. As shown in Table 5, increasing the proportion of Head CT data in the training set generally led to performance improvements, with 70% training data yielding the highest F1 scores for both tasks. Interestingly, at 30% training data, entity recognition performance reached a sub-optimum result. This suggests that, under a fine-tuning framework, the model can achieve strong performance even with limited modality-specific data. As the dataset size increases, incorporating a higher proportions of reliable modality-specific data further enhances model performance for specialized tasks.

Table 4. Comparison of previous models with our multi-modality model on our test set, evaluated using Macro F1 scores for both tasks.

Approach	**Entity**	**Relation**
RadGraph	0.795	0.526
RadGraph-XL	0.745	0.541
Multi-Modality RadGraph	**0.854**	**0.694**

Table 5. Pre-trained model performance comparison based on different proportions of training data for Head CT.

Dataset Portion (%)	10	20	30	40	50	60	70
Entity	0.777	0.762	0.830	0.827	0.819	0.792	**0.849**
Relation	0.628	0.675	0.684	0.690	0.691	0.659	**0.733**

4 Conclusion

We expanded the RadGraph framework by introducing a new multi-modality dataset encompassing cardiac MRI, CTPA, head CT, and abdominal ultrasound reports, thereby broadening the model's applicability across diverse imaging modalities. Our experiments demonstrate that fine-tuning modality-specific models is highly effective, enabling strong performance with less training data while providing insights into optimal training strategies. We further emphasize the importance of modality-specific RadGraph models, showing that they consistently outperform out-of-domain models in entity recognition and relation extraction tasks. These findings reinforce the necessity of tailored models for specialized radiology datasets. Future work will focus on enhancing relation extraction performance and evaluating modality-specific models on additional domain-specific downstream tasks. By advancing the RadGraph framework, we aim to facilitate its broader adoption in multi-modal clinical applications, ultimately improving automated radiology information extraction and medical NLP research.

Acknowledgments. This work was funded by the National Natural Science Foundation of China (No. 62376287), the International Science and Technology Innovation Joint Base of Machine Vision and Medical Image Processing in Hunan Province (No. 2021CB1013), the Natural Science Foundation of Hunan Province (Nos. 2022JJ30762, 2023JJ70016), and the 111 Project (No. B18059).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Datasaur, inc., sunnyvale, california, united states. datasaur, 2019
2. Assale, M., Dui, L.G., Cina, A., Seveso, A., Cabitza, F.: The revival of the notes field: leveraging the unstructured content in electronic health records. Frontiers in medicine **6**, 66 (2019)
3. Boecking, B., Usuyama, N., Bannur, S., Castro, D.C., Schwaighofer, A., Hyland, S., Wetscherek, M., Naumann, T., Nori, A., Alvarez-Valle, J., Poon, H., Oktay, O.: Making the most of text semantics to improve biomedical vision-language processing (2022). https://doi.org/10.48550/ARXIV.2204.09817, https://arxiv.org/abs/2204.09817
4. Bustos, A., Pertusa, A., Salinas, J.M., De La Iglesia-Vaya, M.: Padchest: A large chest x-ray image dataset with multi-label annotated reports. Medical image analysis **66**, 101797 (2020)
5. Cho, H., Lee, H.: Biomedical named entity recognition using deep neural networks with contextual information. BMC bioinformatics **20**, 1–11 (2019)
6. Datta, S., Si, Y., Rodriguez, L., Shooshan, S.E., Demner-Fushman, D., Roberts, K.: Understanding spatial language in radiology: Representation framework, annotation, and spatial relation extraction from chest x-ray reports using deep learning. Journal of biomedical informatics **108**, 103473 (2020)
7. Delbrouck, J.B., Chambon, P., Bluethgen, C., Tsai, E., Almusa, O., Langlotz, C.: Improving the factual correctness of radiology report generation with semantic rewards. In: Goldberg, Y., Kozareva, Z., Zhang, Y. (eds.) Findings of the Association for Computational Linguistics: EMNLP 2022. pp. 4348–4360. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022). https://doi.org/10.18653/v1/2022.findings-emnlp.319, https://aclanthology.org/2022.findings-emnlp.319/

8. Delbrouck, J.B., Chambon, P., Chen, Z., Varma, M., Johnston, A., Blankemeier, L., Van Veen, D., Bui, T., Truong, S., Langlotz, C.: Radgraph-xl: A large-scale expert-annotated dataset for entity and relation extraction from radiology reports. In: Findings of the Association for Computational Linguistics ACL 2024. pp. 12902–12915 (2024)
9. Du, N., Chen, K., Kannan, A., Tran, L., Chen, Y., Shafran, I.: Extracting symptoms and their status from clinical conversations. arXiv preprint arXiv:1906.02239 (2019)
10. Gligic, L., Kormilitzin, A., Goldberg, P., Nevado-Holgado, A.: Named entity recognition in electronic health records using transfer learning bootstrapped neural networks. Neural Networks **121**, 132–139 (2020)
11. Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J., Poon, H.: Domain-specific language model pretraining for biomedical natural language processing (2020)
12. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., et al.: Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 590–597 (2019)
13. Jagannatha, A.N., Yu, H.: Bidirectional rnn for medical event detection in electronic health records. In: Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting. vol. 2016, p. 473. NIH Public Access (2016)
14. Jain, S., Agrawal, A., Saporta, A., Truong, S.Q., Duong, D.N., Bui, T., Chambon, P., Zhang, Y., Lungren, M.P., Ng, A.Y., Langlotz, C.P., Rajpurkar, P.: Radgraph: Extracting clinical entities and relations from radiology reports (2021), https://arxiv.org/abs/2106.14463
15. Johnson, A.E., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M.P., Deng, C.y., Mark, R.G., Horng, S.: Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports. Scientific data **6**(1), 317 (2019)
16. Khanna, S., Dejl, A., Yoon, K., Truong, Q.H., Duong, H., Saenz, A., Rajpurkar, P.: Radgraph2: Modeling disease progression in radiology reports via hierarchical information extraction (2023), https://arxiv.org/abs/2308.05046
17. Li, I., Pan, J., Goldwasser, J., Verma, N., Wong, W.P., Nuzumlalı, M.Y., Rosand, B., Li, Y., Zhang, M., Chang, D., et al.: Neural natural language processing for unstructured data in electronic health records: a review. Computer Science Review **46**, 100511 (2022)
18. Lin, Y., Liu, Z., Sun, M.: Neural relation extraction with multi-lingual attention. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 34–43 (2017)
19. Pyae Khin, N.P., Lynn, K.T.: Medical concept extraction: A comparison of statistical and semantic methods. In: 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). pp. 35–38 (2017). https://doi.org/10.1109/SNPD.2017.8022697
20. Tripathi, S., Gabriel, K., Dheer, S., Parajuli, A., Augustin, A.I., Elahi, A., Awan, O., Dako, F.: Understanding biases and disparities in radiology ai datasets: a review. Journal of the American College of Radiology **20**(9), 836–841 (2023)
21. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)

22. Wadden, D., Wennberg, U., Luan, Y., Hajishirzi, H.: Entity, relation, and event extraction with contextualized span representations. CoRR **abs/1909.03546** (2019), http://arxiv.org/abs/1909.03546
23. Yan, B., Liu, R., Kuo, D., Adithan, S., Reis, E., Kwak, S., Venugopal, V., O'Connell, C., Saenz, A., Rajpurkar, P., Moor, M.: Style-aware radiology report generation with RadGraph and few-shot prompting. In: Bouamor, H., Pino, J., Bali, K. (eds.) Findings of the Association for Computational Linguistics: EMNLP 2023. pp. 14676–14688. Association for Computational Linguistics, Singapore (Dec 2023). https://doi.org/10.18653/v1/2023.findings-emnlp.977, https://aclanthology.org/2023.findings-emnlp.977/
24. Yang, S., Wu, X., Ge, S., Zhou, S.K., Xiao, L.: Knowledge matters: Chest radiology report generation with general and specific knowledge. Medical Image Analysis **80**, 102510 (Aug 2022). https://doi.org/10.1016/j.media.2022.102510, http://dx.doi.org/10.1016/j.media.2022.102510
25. Zheng, S., Wang, F., Bao, H., Hao, Y., Zhou, P., Xu, B.: Joint extraction of entities and relations based on a novel tagging scheme. CoRR **abs/1706.05075** (2017), http://arxiv.org/abs/1706.05075