A Novel Corpus of Annotated Medical Imaging Reports and Information Extraction Results Using BERT-based Language Models

Namu Park[1*], Kevin Lybarger[2*], Giridhar Kaushik Ramachandran[2], Spencer Lewis[3], Aashka Damani[4], Özlem Uzuner[2], Martin Gunn[5] and Meliha Yetisgen[1]

[1]Department of Biomedical Informatics & Medical Education, University of Washington
[2]Department of Information Sciences and Technology, George Mason University
[3]Department of Radiology, School of Medicine, Stanford University
[4]School of Medicine, University of Washington
[5]Department of Radiology, School of Medicine, University of Washington
[*]Equal contribution

Introduction

Problem

- Improved image resolution has led to the rapid rise in demand for medical imaging (e.g. CT, MRI, PET-CT)
- Radiologists interpret important medical concepts identified in imaging tests through narrative reports that remain largely unstructured (Figure1)
- **Conversion into semantic representation** is required to facilitate secondary use of radiology reports

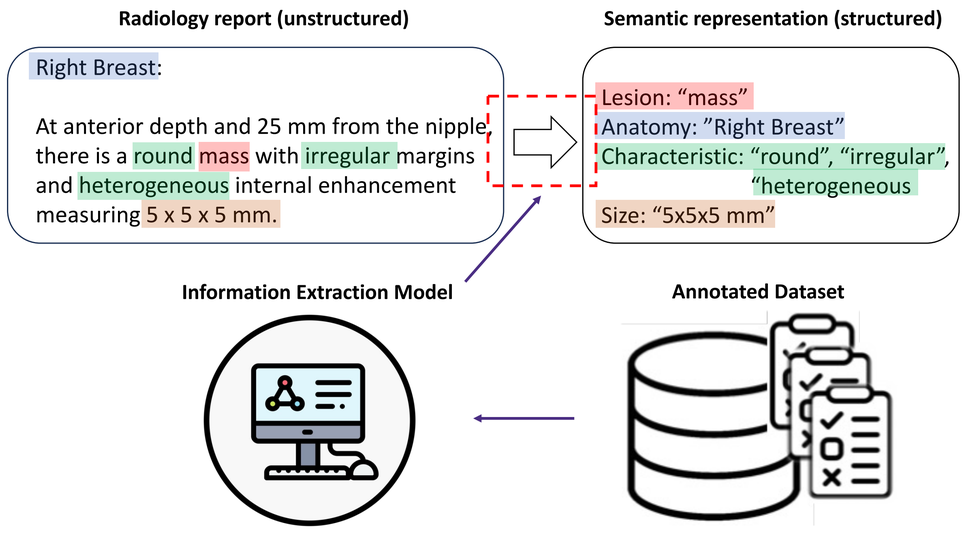

Figure 1. Semantic representation of radiology reports

Goals

- Introduce the **Corpus of Annotated Medical Imaging Reports (CAMIR)** with granular clinical annotations
- Explore information extraction results using two BERT-based models trained on CAMIR, that can be widely used as a foundational information extraction model in radiology

Corpus Development

Data

- 609 de-identified radiology reports from University of Washington Medical System hospitals
- Imaging modality types: CT (n=203), MRI (n=202), PET-CT (n=204)

Event-based annotation schema (Figure 2)

- Events: Indication, Medical Problem, Lesion
- Each event includes Trigger-Argument relations
  - Trigger: Text span identifying the event
  - Argument: Text span that characterizes the event

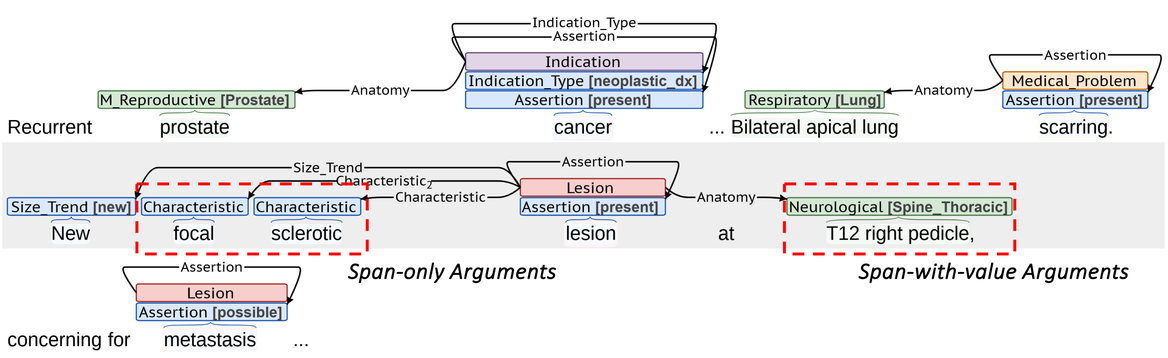

Figure 2. Annotation example using event-based annotation schema

Annotation Process

- Annotation guidelines designed by a senior radiology resident and an experienced board-certified radiologist with 20+ years of experience
- Annotators: 4 medical students from UW Medical School
- 5 double-annotated rounds with high inter-annotator agreement (IAA)
  - Triggers: Indication 0.856 F1, Medical Problem 0.854 F1, Lesion 0.805 F1
  - Overall (triggers and arguments): 0.762 F1
- 4 single-annotated rounds

Table 1. Event distribution based on imaging modality type

Modality	# Notes	# Events (Avg. per report)		
		Indication	Medical Problem	Lesion
CT	203	507 (2.5)	2063 (10.2)	1855 (9.2)
MRI	202	496 (2.4)	2111 (10.4)	1967 (9.7)
PET-CT	204	491 (2.4)	2080 (10.2)	1887 (9.3)
OVERALL	609	1494 (2.5)	6254 (10.3)	5709 (9.4)

Information Extraction

Task

- Events extracted as relations
- Relation mapping: (Trigger, Argument, Role) → (Head, Tail, Relation)

BERT-based information extraction models

- mSpERT[1]
  - Information extraction leveraging span-level representations
  - Modified to generate CAMIR event structure
- PL-Marker++
  - Augmented version of Packed-Levitated Marker[2] (PL-Marker)
  - Addition of subtype markers allow the extraction of all CAMIR events (Figure 3)

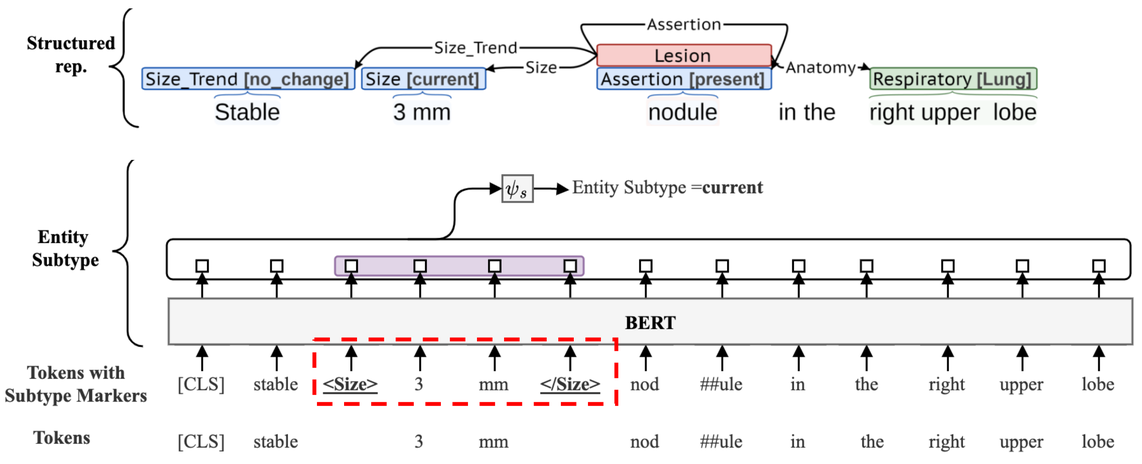

Figure 3. PL-Marker++ subtype classification

Prediction performance

- Table 2 summarizes the prediction performance of mSpERT and PL-Marker++
- PL-Marker++ achieved higher overall performance (0.759 F1 vs. 0.736 F1)
- PL-Marker++ performed significantly better in extracting Lesion triggers and all but one argument type

Table 2. Information extraction performance using mSpERT and PL-Marker++

Event	Argument	F1 (Overlap Criterion)	
		mSpERT	PL-Marker++
Indication	Trigger	**0.787**	0.782
	Assertion	**0.770**	0.759
	Anatomy Parent	0.617	**0.639**
	Anatomy Child	0.593	**0.639**
	Indication Type	**0.732**	0.729
Lesion	Trigger	0.853	**0.884**†
	Assertion	0.825	**0.866**†
	Anatomy Parent	0.680	**0.718**†
	Anatomy Child	0.646	**0.684**†
	Characteristic	0.512	**0.591**†
	Count	**0.816**	0.809
	Size	0.713	**0.778**†
	Size Trend	0.647	**0.752**†
Medical Problem	Trigger	0.863	**0.875**
	Assertion	0.839	**0.844**
	Anatomy Parent	**0.697**	0.688
	Anatomy Child	**0.633**	0.628
OVERALL		0.736	**0.759**†

† indicates statistical significance ($p<0.05$)

Conclusions

- CAMIR is a novel corpus containing clinical information from radiology reports
- PL-Marker++ can serve as a foundational model for extracting clinical findings from radiology reports
- We plan to explore the generalizability of our research, as well as assess the capability of generative LLMs (e.g. Llama3, GPT-4)

References

1. Lybarger, K., Dobbins, N. J., Long, R., Singh, A., Wedgeworth, P., Uzuner, Ö., & Yetisgen, M. (2023). Leveraging natural language processing to augment structured social determinants of health data in the electronic health record. Journal of the American Medical Informatics Association, 30(8), 1389-1397.
2. Ye, D., Lin, Y., Li, P., & Sun, M. (2021). Packed levitated marker for entity and relation extraction. arXiv preprint arXiv:2109.06067.