NLP_project 2

Group F

November 2025

1 Introduction

1. **Investigate which standard and potential new NER types are most prominent in your dataset (i.e., manual data inspection)**

   As our dataset consists of sentence-level descriptions extracted from MIMIC-CXR, most samples describe findings within the thoracic cavity on chest X-rays. From these sentences, we can typically identify the following entity types: – Findings (e.g., atelectasis, nodule) – Location (the spatial description of a finding) – Anatomy (e.g., heart, left atrium, ribs) – Descriptive terms (e.g., enlarged) – Other

2. **Apply the standard NER classifier of spaCy to your data**

   For the majority of sentences, no entity is predicted. In some cases, numerical values are identified, while certain nouns are incorrectly labeled as proper names (e.g., "pleural fluid" → "PERSON").

3. **Find out how to evaluate the quality of the NER classification and evaluate it automatically and manually with 100 randomly sampled entities**

   Standard metrics such as Precision, Recall, F1-score, and Exact/Partial match can be used for evaluation. Manual assessment can also be performed using predefined criteria, such as: – Correct span? YES / PARTIAL / NO – Correct entity type? YES / NO

4. **Extend the standard NER types using the NER Annotator and re-run the NER classification as well as its evaluation**

   We used the NER Annotator to label five entity types—Finding, Anatomy, Location, Description, and Other—and subsequently performed training and validation. The evaluation results are shown below:

5. **Investigate using an LLM-based NER classifier**

   We conducted an evaluation using Groq as the large language model. The results are presented below:

Entity Type	Precision	Recall	F1-score
LOCATION	0.0000	0.0000	0.0000
ANATOMY	0.2083	0.2381	0.2222
FINDING	0.4130	0.4634	0.4368
DESCRIPTION	0.2632	0.3448	0.2985
OTHER	0.2683	0.3056	0.2857

Table 1: NER classification performance with training from scratch NER model.

Label	TP	FP	FN	Precision	Recall	F1
DESCRIPTION	1	10	36	0.091	0.027	0.042
ANATOMY	1	33	35	0.029	0.028	0.029
OTHER	0	17	49	0.000	0.000	0.000
FINDING	0	37	40	0.000	0.000	0.000
LOCATION	0	26	13	0.000	0.000	0.000

Table 2: NER performance per entity type using LLM.

6. **Describe how NER type information could help in other NLP use cases**

By identifying NER types, one can determine the presence or absence of specific findings, enabling the standardisation and structuring of raw radiology reports. After applying NER, raw sentences may be converted into structured representations such as: FINDING: pleural effusion LOCATION: left.