Predicting Emergency Department Disposition from Radiology Reports

Yue Andy Zhang, Karen Garcia
{yuezha,karengar}@stanford.edu

Abstract

Radiology reports are a key input in deciding the disposition of emergency department patients. We have two objectives: unsupervised multi-label classification to extract 39 medical condition from radiology reports, followed by supervised classification to predict patient's disposition.

We evaluate unsupervised label extraction methods by considering CheXbert labels as a ground truth. Our work demonstrate superior performance of using a pretrained NLI model for zero-shot label extraction. We find that models trained using the medical annotations, together with the 39 extracted medical conditions outperforms the other methods to predict patients' disposition.

Introduction

Radiology reports are a key input used by physicians to decide the disposition of emergency department patients.

There are two tasks relevant to disposition prediction from radiology reports:

1. Extract human-readable labels from radiology reports. These labels indicate the presence or absence of medical conditions, eg. pneumonia. There are 39 conditions, and each report may discuss multiple conditions.
2. Predict the patient's disposition based on the text report and/or extracted condition labels.

We propose and compare systems to perform zero-shot/unsupervised label extraction and supervised disposition prediction.

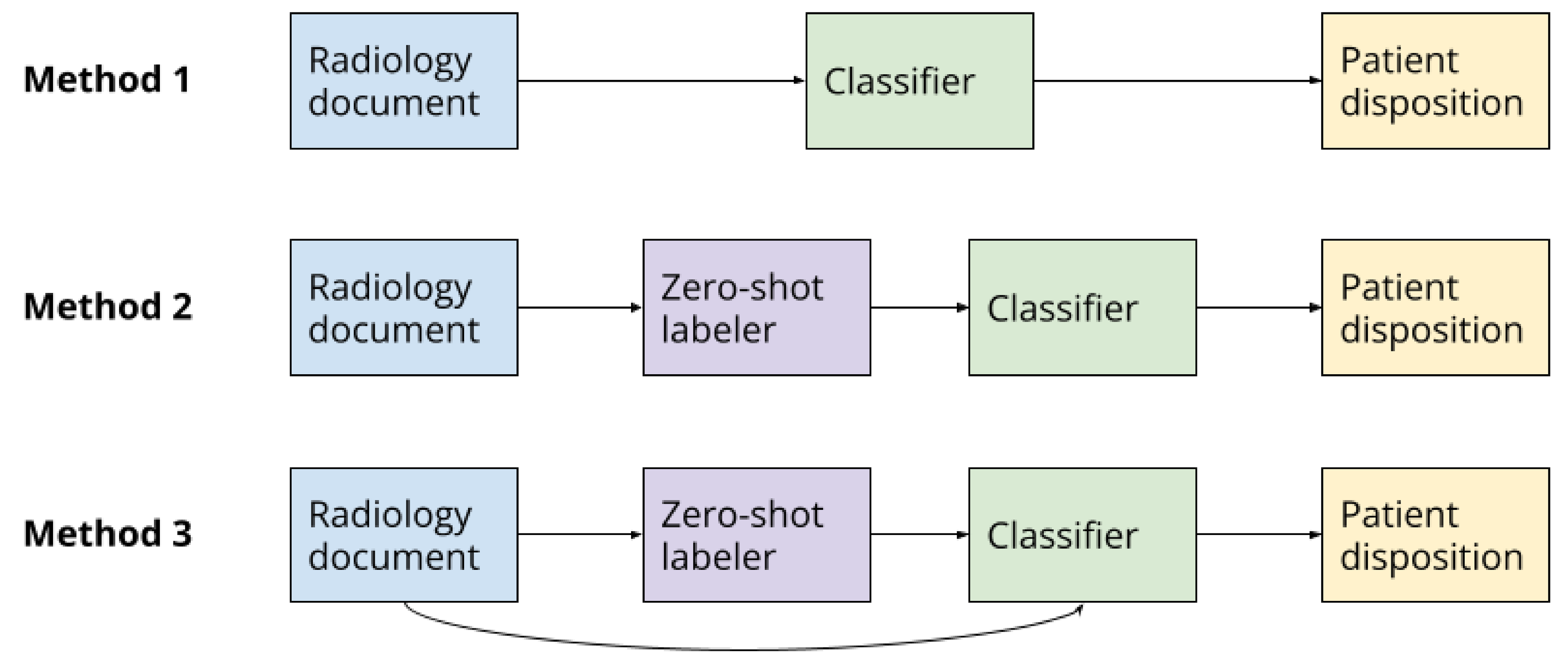
Dataset

The dataset consists of 150,000 emergency department records from Stanford Hospital. The key components of each record are:

1. **Report:** a document describing the radiology results, up to a few thousand words
2. **Disposition:** the patient's outcome, determined by a physician. The four categories are Admit to Inpatient, Discharge, Observation and CDU Observation.

Each record also has some additional metadata, such as which radiology procedure performed. We did not use the metadata, so that the results here can be solely attributed the methods' ability to understand report documents. This metadata likely has predictive value and would be used in a real-world system.

Method



Zero-shot label extraction methods

1. Lbl2vec [1]: Embed documents and labels using transformer. Cluster documents with labels by cosine similarity. Apply Naive Bayes decision rule.
2. Pretrained NLI model: Use BART-large pretrained on NLI as zero-shot classifier.
3. Finetuned NLI model: Same as 2, but with additional fine tuning.

We use the CheXbert labeler [2] to extract 7 of 39 labels. We consider the 7 CheXbert labels as ground truth to evaluate the two proposed methods.

Disposition prediction methods

For methods 1 and 3, we fine-tuned several transformer models:

1. BERT
2. BioBERT
3. ClinicalBERT

For method 2, we trained multilayer perceptron classifiers, performing grid search over hyperparameters such as the number of hidden units and learning rate.

References

[1] Tim Schopf, Daniel Braun, and Florian Matthes.
Evaluating unsupervised text classification: Zero-shot and similarity-based approaches.
In *2022 6th International Conference on Natural Language Processing and Information Retrieval (NLPIR)*, NLPIR 2022, New York, NY, USA, 2023. Association for Computing Machinery.

[2] Akshay Smit, Saahil Jain, Pranav Rajpurkar, Anuj Pareek, Andrew Y. Ng, and Matthew P. Lungren.
Chexbert: Combining automatic labelers and expert annotations for accurate radiology report labeling using bert, 2020.

Results

Result 1: Zero-shot labeler to predict CheXbert conditions

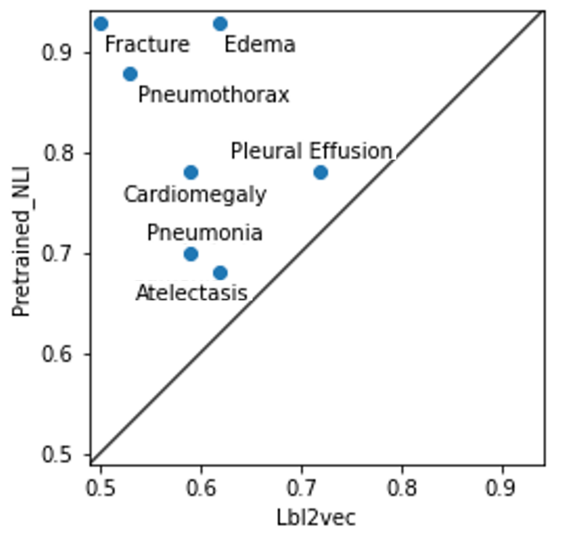


Figure 1. Scatter plot showing the AUC obtained for Pretrained NLI and Lbl2vec.

1. For all classes, pretrained NLI model outperforms Lbl2vec in terms of AUC.
2. The best performance is associated with the labels Fracture and Edema both with an AUC of 0.93.
3. Label extraction with the NLI model takes roughly twice as long as Lbl2vec.

Result 2: Model performance of disposition prediction

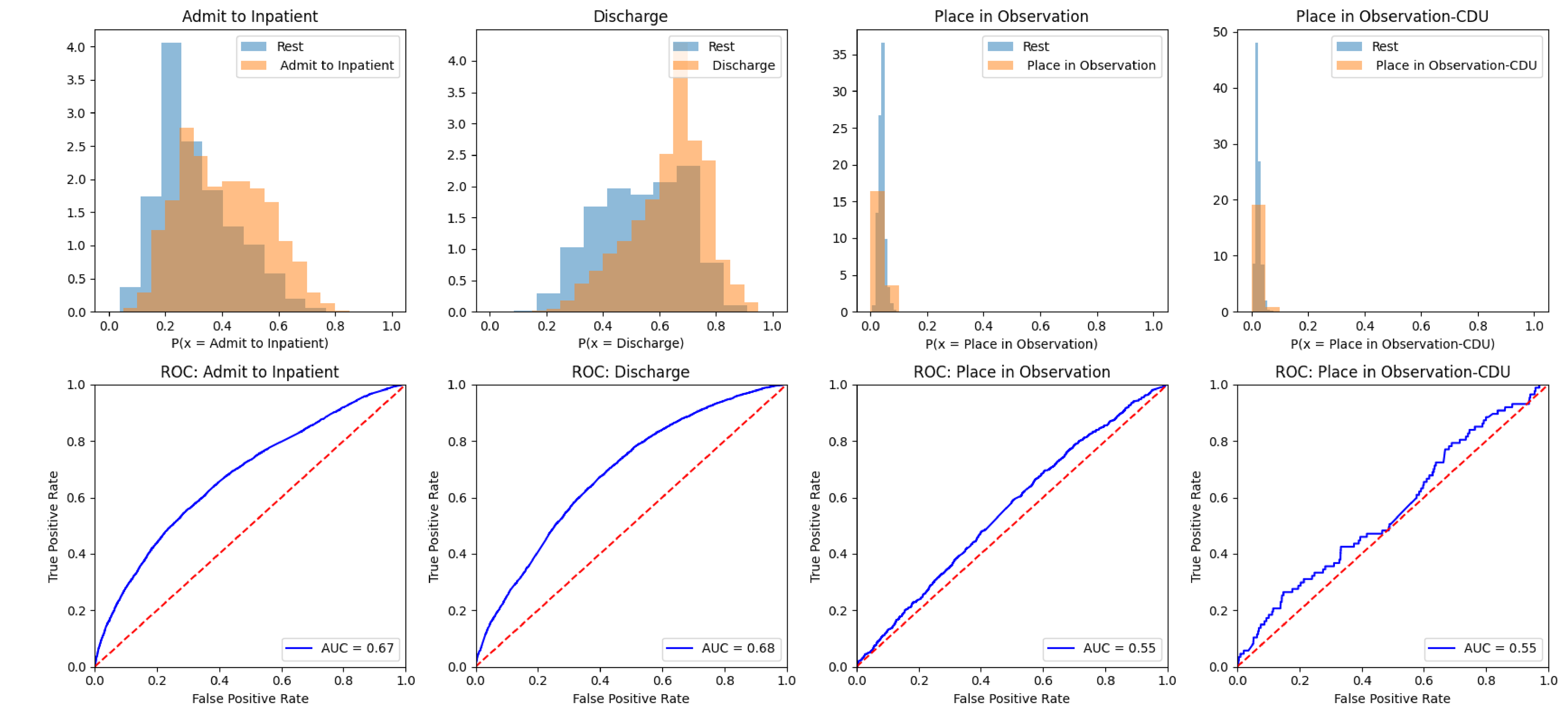


Figure 2. Histograms and ROC curves of each disposition class. Classification results after fine-tuning pretrained model ClinicalBERT.

Analysis and Conclusions

- The best overall method is extracting labels using the pre-trained NLI model, and then fine-tuning ClinicalBERT.
- Augmenting the disposition classifier input with zero-shot labels is effective in improving performance of disposition classifiers. Prediction disposition based on just labels also yields decent results.
- For both tasks, rare classes have much weaker performance.
- In future work, disposition prediction can be further improved by augmenting input data beyond radiology documents.

Special thanks to Abhinav Garg and Prof. David Kim