AI-powered Extraction of Relevant Information from Prior Radiology Reports

Yiting Xie, PhD, Senior Data Scientist, Merge by Merative
Linda Bagley; Marwan Sati, PhD

Introduction

With increasing image volume and staffing shortages, radiologists have less time to find and adequately read patient history. This represents a significant risk to patients' health and increases radiologist's liability. Automatically finding relevant history is extremely challenging because it requires deep medical knowledge to identify terms that are not semantically similar. For example, there are over 100 conditions that could be related to a chief complaint of "abdominal pain". Lung nodule mentions may be relevant for knee surgery prep due to their potential impact on surgery risks.

Hypothesis

Relevant information can be automatically extracted from radiology reports using machine learning (ML) and large language models (LLMs).

Methods

We algorithmically determined relevant priors based on the patient's "Reason for Exam" in the imaging order. The framework uses a weighted combination of ChatGPT 4o and o1 LLMs and our proprietary ML model trained on extensive medical data from sources including PubMed and Wikipedia. Intelligent prompt engineering was employed to enhance LLM performance. The framework extracts relevant sentences and associated probabilities, with the highest probability sentence indicating the overall relevancy of each report.
We evaluated the framework on 1366 radiology reports from 300 patients using binary ground truth provided by 3 radiologists (single reader per report). Report level probability was compared to the ground truth. Performance of the framework was compared against a baseline model using ChatGPT 4o alone with a simple prompt.

Results

Out of the box ChatGPT 4o model achieved accuracy of only 74%. The combined framework achieved an AUC of 0.87 (see Figure 1) and an optimal accuracy was 85%. The confusion matrix at this accuracy is shown in Table 1.

Conclusion

This complex problem cannot be solved with the latest out-of-the-box LLMs however combining state of the art LLMs with medical-specific ML is promising for this critical yet very challenging problem.

2005 | Transforming Radiology Reporting: AI-Driven Tools for Education, Communication, and Reliability
Scientific Research Abstracts

Figure(s)

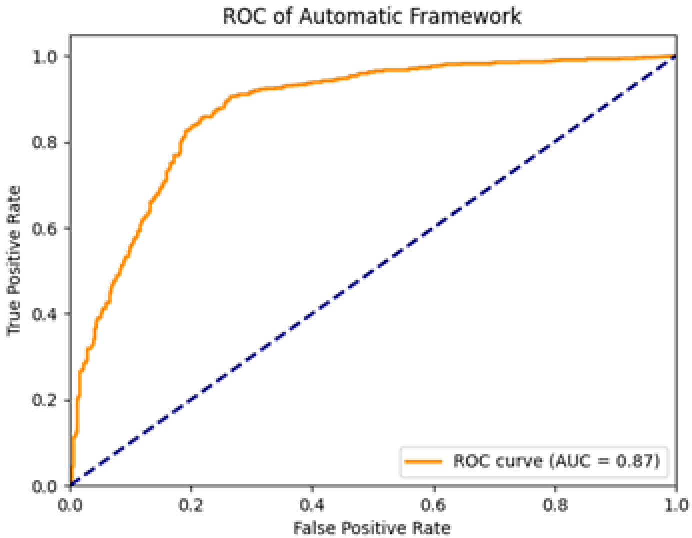

Figure 1. ROC and AUC of the automatic framework

CM	Prediction=0	Prediction=1
Ground truth=0	354	128
Ground truth=1	83	801

Table 1. Confusion Matrix (CM) at optimal accuracy

Keywords

Artificial Intelligence/Machine Learning