

LLMDX

A Framework for Classifying Radiology Reports

Bar Ezra Gafniel¹ · Dr. Oren Feldman² · Dr. Jordanna Koppel² · Prof. Nadav Rappoport¹

¹ Institute for Interdisciplinary Computational Science, Stein Faculty of Computer and Information Science, Ben-Gurion University of the Negev, Israel

² Pediatric Emergency Department, Sheba Medical Center, Tel Hashomer, Israel

BACKGROUND

Clinical free text (e.g. radiology reports) contains critical diagnostic information. Automated classification faces significant barriers:

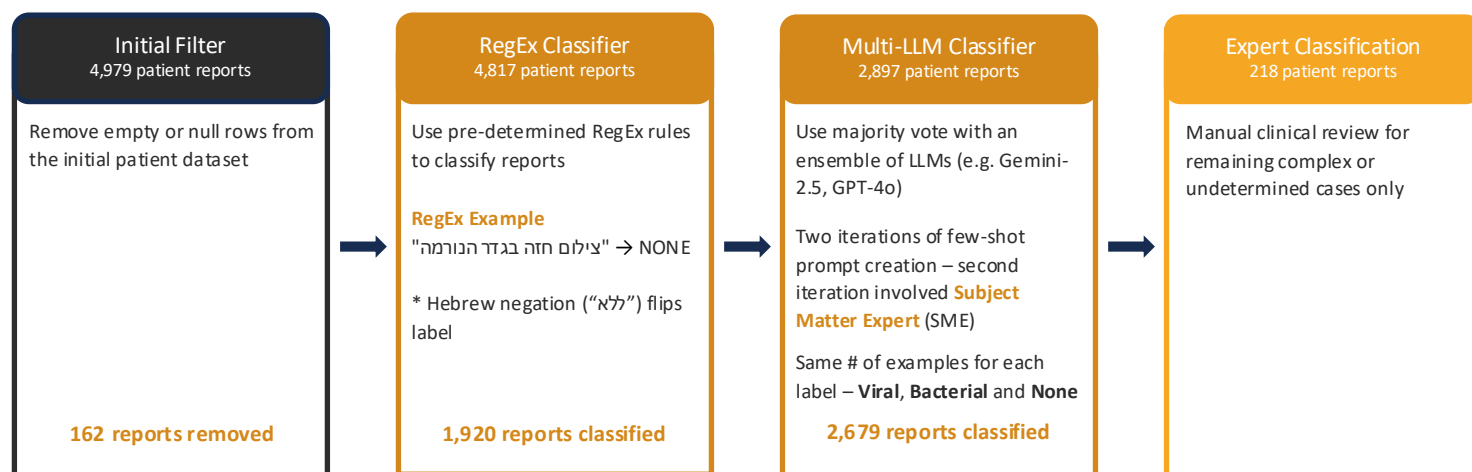
- Data is often **unlabeled** — volume is too large for manual human annotation
- No out-of-the-box solution for classifying **Hebrew medical text**
- High computational costs
- Non-deterministic LLM outputs
- Lack of uncertainty management mechanisms in sensitive medical domains

OBJECTIVES

Develop and evaluate **LLMDX** (Large Language Model Diagnostics) — an open-source framework for **hybrid, tiered classification** of pediatric chest X-ray reports from the ED, combining computational efficiency, clinical accuracy, and expert optimization.

METHODS — Multistage Data Classification and Filtering Flow

The entire pipeline is **fully configurable**: Regular Expression (RegEx) rules & labels, LLM providers, ensemble size, and majority consensus threshold.



RESULTS

Prompt Optimization

v1 – Simple Prompt
79% agreement
21% UNDETERMINED

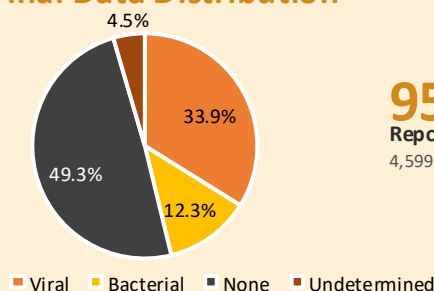
→
SME Curated Examples

v2 – Expert Refined Prompt
92.5% agreement
7.5% UNDETERMINED

K = 0.866

Cohen's Kappa
"Almost Perfect" agreement

Final Data Distribution



95.5%
Reports auto-classified
4,599 of 4,817 total reports

CONCLUSIONS

LLMDX presents a **scalable and reliable methodology** for clinical text classification.

- ✓ Subject Matter Expert involvement in Prompt Engineering is essential for high accuracy
- ✓ Tiered pipeline reduces expert review to only 4.5% of cases
- ✓ Open-source framework available for the clinical NLP community



GitHub: github.com/barezra94/llmdx
Contact: ezrabar@post.bgu.ac.il