CaseMirror

Finding Similar Medical Cases from Interviews Using LLMs

By: Nicole Poliak and Naveh Nissan

The Problem

Use Case

Clinicians often rely on memory to recall similar cases, which can be slow and unreliable—especially for rare or vague conditions. This project proposes a semantic similarity engine that embeds patient-doctor interviews and retrieves the most clinically relevant past cases based on semantic closeness. The system can support faster, more consistent clinical decisions and improve diagnostic confidence.

The Problem

Problem Definition

- Input: A transcript of a doctor-patient interview
- Output: Top-N semantically similar past interviews, extracted features (symptoms, risk factors, etc.), potential weak labels (condition groups or clusters)
- NLP Tasks: Semantic similarity search, keyword extraction, cluster-Based Labeling

Key challenges

- Informal, unstructured medical dialogues
- Varied symptom phrasing
- Ensuring clinical relevance in similarity-matching

Training and Test Data

Public Dataset

217 simulated patient-doctor interviews (focused on respiratory complaints).

Synthetic Data

ChatGPT generated for broader symptom diversity and for a bigger combined dataset.

Data Type

Unlabeled, unstructured, dialogue-based free-text clinical interviews.

Input/Output Example

Input

"Patient: I've had a dry cough for six months and feel short of breath walking."

Output

Top-N similar interviews with matching symptoms.
Suggested label: Specific suspected respiratory condition.

Evaluation ✓

Evaluation Metrics

- Recall and Precision
- Jaccard or cosine similarity

Evaluation Strategy

- Cluster interviews using LLM-based semantic similarity
- Measure symptom overlap within clusters using Jaccard or Cosine similarity
- Validation by checking if top shared features support the cluster label.

Baseline

- TF-IDF + Cosine Similarity
- Compare with advanced LLM embeddings (e.g., BioClinicalBERT)