



CaseMirror

Finding Similar Patient Cases from Discharge Summaries 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 discharge summaries 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.



Problem Definition

Input: A new discharge summary in free-text format.

Output: Top-N most similar historical discharge notes, ranked by semantic similarity

Relevant NLP Tasks: Semantic text embedding, information retrieval, similarity search

Key challenges

- Unstructured and noisy medical language.
- Varied symptom phrasing.
- Large dataset.



Training and Test Data

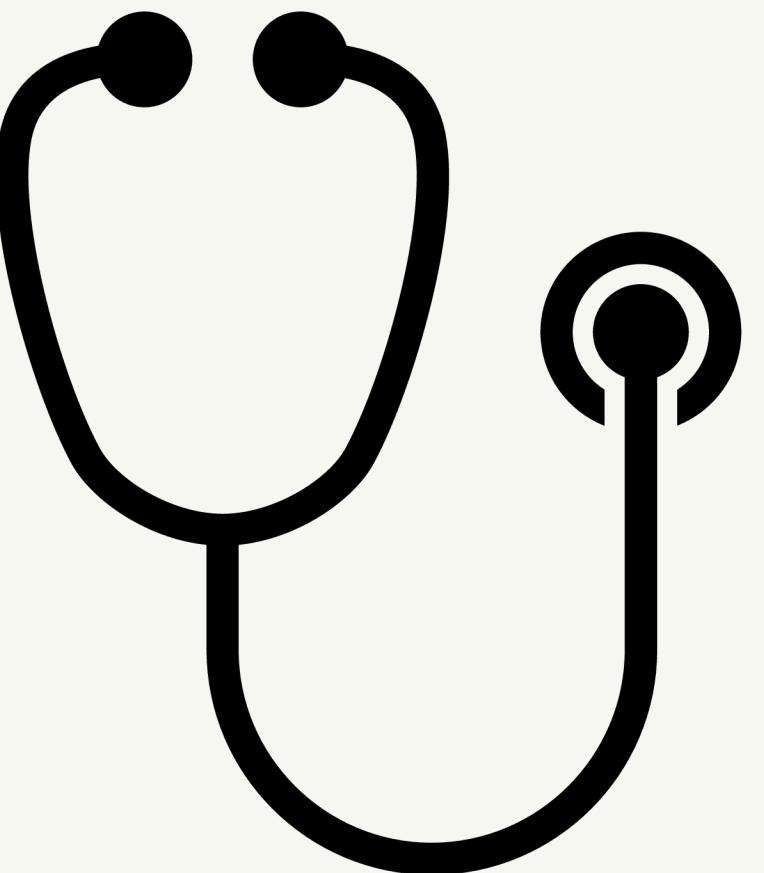
Dataset

MIMIC-IV-Note: Deidentified free-text clinical notes (PhysioNet).
Publicly available, deidentified clinical notes (requires DUA).
Focused on 50k discharge summaries sampled from 331K total.

Data Type

Unlabeled, unstructured free-text discharge summaries.
Similarity can be evaluated using diagnosis codes/medical condition names.

Input/Output Example



Input:

"78-year-old male with history of CHF admitted for worsening shortness of breath, elevated BNP, pulmonary congestion."

Output:

Top-N most semantically similar discharge summaries describing patients with similar symptoms or diagnoses (e.g., congestive heart failure, fluid overload).

Evaluation:

Match between retrieved notes and original based on diagnosis labels like “Heart Failure” or ICD codes such as I50.9.

Evaluation ✓

Evaluation Metrics:

- Recall: Proportion of relevant discharge summaries retrieved in top-N results.
- Precision: Proportion of top-k retrieved summaries that are clinically relevant.
- Cosine similarity of embeddings as a proxy for semantic distance.



Evaluation Strategy:

- Data split: Train/test split of MIMIC-IV discharge summaries.
- Ground truth: Similarity defined via diagnosis labels (e.g., ICD codes).
- Example: Input case about "CHF and shortness of breath" → relevant if output includes other heart failure summaries.

Baseline:

- TF-IDF + Cosine Similarity
- Compare results with advanced LLM-based embeddings (e.g., BioClinicalBERT or Med-PaLM).