

MedMap – The NLP-Driven Healthcare Visualization Tool

Team 050: Stefan Lehman, Kelly Lewis, Ruby Truong, Ying Wang, Luke Williams

Introduction/ Motivation

Physicians dedicate over half their workday managing electronic health records (EHRs), reducing time spent with patients. This leads to **physician dissatisfaction** and **harm to patient health outcomes**. Our tool, called MedMap, aims to address this by using natural language processing (NLP) to transform raw clinical data into interactive visual reports, simplifying EHR interactions and improving clinical productivity.

Approaches

The team used industry standard NLP packages, including SpaCy, NLTK, in Python for **entity recognition and extraction of key medical terms**. This provides more accessible annotations and visualizations and expands the scope of proactive management of more chronic diseases to facilitate an **effective patient care model**. The **Patient Matcher** feature was developed using Term Frequency Inverse Document Frequency (TF-IDF) **vectorization** and an unsupervised **clustering algorithm** to identify similar subgroups of patients. **Cosine similarities** are then computed between the currently displayed patient and patients within that subgroup, pointing to the top similar patients for further interactive exploration.

Data

The data was downloaded from the **MIMIC-IV clinical database**, with 331,794 deidentified discharge summaries for 145,915 patients.

- Most discharge data was in one long format text column, so key sections were parsed out into separate columns.
- Based on the scope of MedMap, the team removed all prior office visits to focus on the latest record, creating a new variable to retain the number.
- After performing data cleaning and parsing, the final patient-level dataset is split into 70/30 across training and testing.

Innovations

- Instant Access to the Latest Doctor Visits:** Providing doctors with quick lookup access to the latest doctor visits enhances the efficiency of healthcare delivery.
- Streamline Patient Matching for Chronic Diseases:** A logic system that allows for easy application to multiple chronic diseases enables doctors to efficiently match patients with similar conditions and formulate targeted treatment plans.
- Enhanced User Interface (UI):** MedMap offers a better experience for users to explore key medical identifiers and comparisons across patient profiles.

Experiment and Results

Table 1. Clinical SpaCy Model Performance

Model Name	Average Precision	Average Recall	Average F1 Score
en_core_sci_sm	0.488	0.876	0.573
en_ner_bc5cdr_md	0.483	0.881	0.571
medspacy_md	0.796	0.909	0.783

Table 2. Top Performing Model per Condition

Condition	Model Name	F1 Score
ALZHEIMERS	medspacy_md	0.783
CANCER	medspacy_md	0.948
CHRONIC_KIDNEY_DISEASE	en_core_sci_sm	0.573
DIABETES	medspacy_md	0.898
HEART_DISEASE	medspacy_md	0.961
LUNG_DISEASE	medspacy_md	1.00

Approach: The NLP models were applied to sections such as 'History of present illness', 'Past medical history', 'Discharge medications', and 'Discharge diagnosis'.

Analysis:

- The team evaluated these models using recall, precision, and F1 scores. The top performing model was the medspacy_md with an Average F1 Score of 0.783, Average Recall of 0.909, and Average Precision of 0.796, scoring the highest in five out of the six chronic diseases.
- The ScispaCy en_ner_bc5cdr_md model performed the lowest with an Average F1 Score of 0.571, Average Recall of 0.881, and Average Precision of 0.483.
- When considering the context of the medical domain, usually medical practitioners require models to score 0.95 or higher in precision, recall, and F1 Scores, meaning the medspacy_md model performed reasonably well in identifying heart and lung disease as well as cancer.
- The model has room for improvement in recognizing chronic kidney disease, diabetes, and Alzheimer's.

MedMap 

The NLP-Driven Healthcare Visualization Tool

Team 050: Stefan Lehman, Kelly Lewis, Ruby Truong, Ying Wang, Luke Williams

Report Patient Matcher Visualization About

The NLP-Driven Healthcare Visualization Tool

Select a subject ID

10245250

Below are the details about this patient:

10245250-discharge summary-8

Report tab:

Provides a summary of patient information in easily viewable categories and medical terms, with two of many such fields shown below.

History of present illness

Select entity labels

___ history of alcoholic cirrhosis c/b hepatocellular **carcinoma CANCER** s/p ddlt (___) with post-op course c/b coagulopathy leading to washout/exploration, and roux en y hepaticojejunostomy, **aortic stenosis HEART_DISEASE** well **aortic stenosis HEART_DISEASE** portal vein stenosis s/p ptbd, cytomegalovirus viremia who now presents with one day of cough, congestion and elevated lfts

Medications on admission

Select entity labels

the preadmission medication list **incentive spirometry LUNG_DISEASE** accurate and complete. 1. **metformin DIABETES** (**metformin DIABETES**) 500 magnesium by mouth twice a day 2. ferrous sulfate 325 magnesium by mouth daily 3. multivitamins 1 tablet by mouth daily 4. fluticasone propionate 110mcg 2 puff intrahospital twice a day 5. omeprazole 20 magnesium by mouth daily 6.

MedMap Patient Matcher

Using the selected ID, you can look up the most similar patients to enable physicians to quickly identify treatment plans per cohort.

Select a subject ID

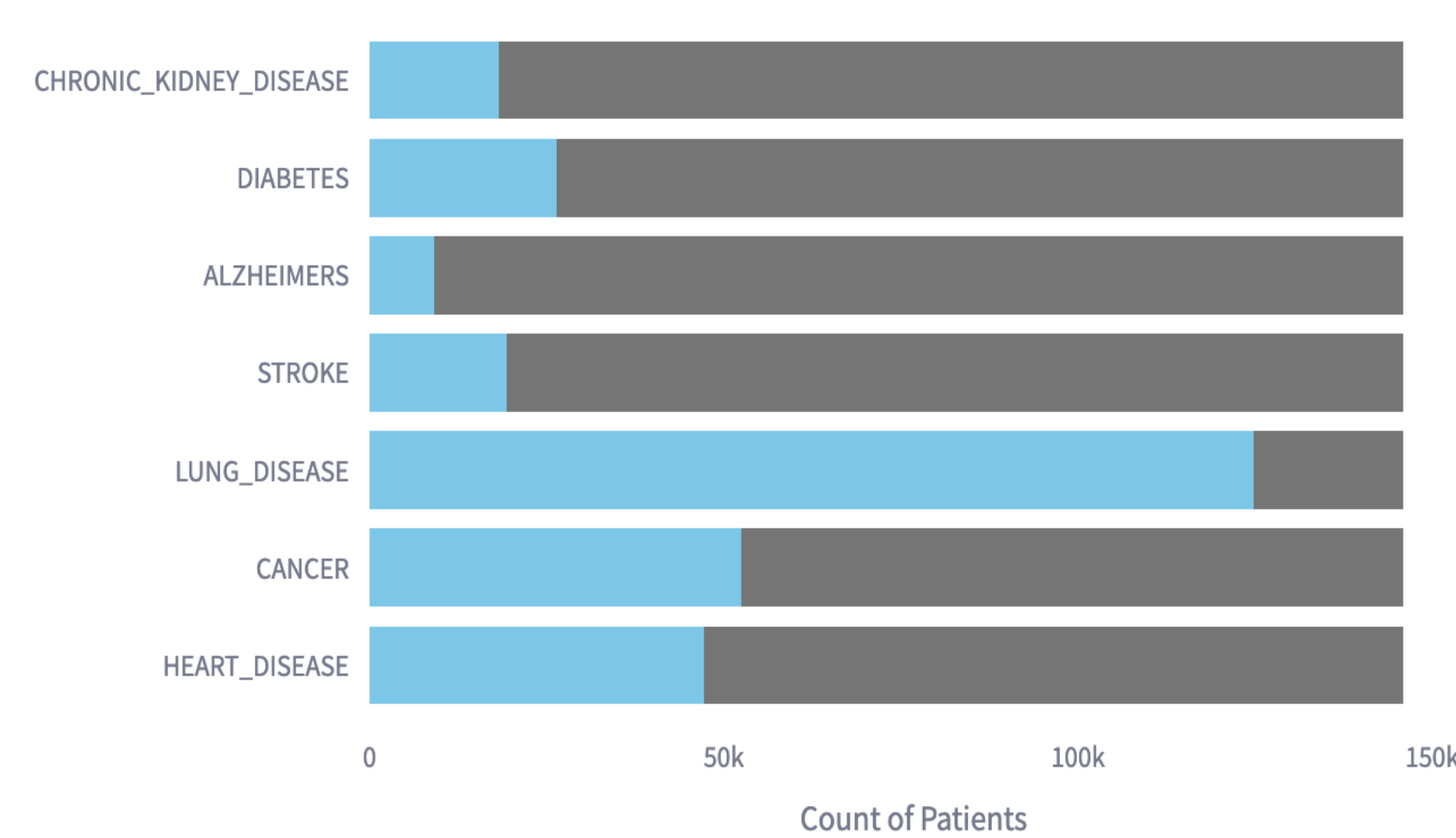
10368348

	similar_patient_id	similarity_score
0	10285646	0.9894
1	10148333	0.9867
2	10423496	0.9863

Patient Matcher tab:

Displays the most similar patients to the selected patient based on the similarity algorithm. It also provides a new dropdown box to explore a highlighted report for the chosen similar patient to compare to the originally selected patient.

Breakdown of Chronic Disease



Visualizations tab:

Includes a bar graph of patients with and without chronic conditions and a histogram of the number of prior office visits by gender with an interactive option for the user to select different Service types.

Prior Office Visits

