Summarizing Clinical Text: An NLP Approach to Simplifying Doctor Notes

Presented by: **Aditya Jadhav, Rashmi Kulkarni, Jay Sarode**

Overview

- Introduction
- Motivation
- Dataset Information
- Synthetic summary generation
- Abstractive Summarization
- Extractive Summarization
- Hybrid
- Multi-Document Summarization

Introduction

Need for Summary Generation in Healthcare:

- Doctor notes contain extensive details like patient history, diagnoses, and treatment plans.
- Reviewing lengthy documentation is time-consuming in fast-paced clinical environments.
- Healthcare providers require quick access to critical patient information.
- Streamlining information is essential for managing complex cases effectively.

Motivation

- Interpreting medical documentation is time-intensive, especially in dynamic clinical settings.
- Summarized notes help clinicians quickly overview essential patient details.
- Focused summaries improve communication among healthcare providers.

The Dataset

OPEN PATIENTS DATASET:

- 180,142 patient descriptions from 4 sources (TREC CDS/CT, MedQA-USMLE, PMC-Patients).
- Includes synthetic and real patient notes, clinical questions, and case reports.

MEDICAL TRANSCRIPTIONS DATASET:

- Consists of medical keywords later used in Extractive summarisation.

Synthetic Summary Generation

- Absence of Ground Truth
- Training data for the models

Steps taken

Abstractive Summarisation + Extractive Summarisation + Hybrid

Abstractive Summarisation

Bleu & Rouge Scores?

- **Generate summaries with paraphrased or new content not directly present in the original text.**
- **Models Used:**

PEGASUS-XSum: 'The summarization specialist'

T5: The 'jack-of-all-trades'

Abstractive Summarisation of Original Text

Model	Average Bleu Score	Average ROUGE1 Score	Average ROUGE2 Score	Average ROUGEL Score
PEGASUS	0.0011	0.1189	0.0530	0.0902
T5	0.0493	0.1468	0.0465	0.1131

Fig. 5: Abstractive Summarisation of Original Text

T5 or Pegasus

Choice of hybrid approach

Original Text - example

A 58-year-old African-American woman presents to the ER with episodic pressing/burning anterior chest pain that began two days earlier for the first time in her life. The pain started while she was walking, radiates to the back, and is accompanied by nausea, diaphoresis, and mild dyspnea, but is not increased on inspiration. The latest episode of pain ended half an hour before her arrival. She is known to have hypertension and obesity. She denies smoking, diabetes, hypercholesterolemia, or a family history of heart disease. She currently takes no medications. Physical examination is normal. The EKG shows nonspecific changes.

Abstractive Summary - example

A 58-year-old African-American woman presents to the ER with episodic pressing/burning anterior chest pain that began two days earlier for the first time in her life. The American College of Emergency Physicians would like to hear from you if you have a patient who presents to the emergency department with chest pain that has not gone away for more than a day or is accompanied by nausea, vomiting, or dyspnea.

Extractive Summarisation

- **What is Extractive Summarisation.?**
- **Dynamic Keyword Extraction**
- **Maximal Marginal Relevance in action.!!**

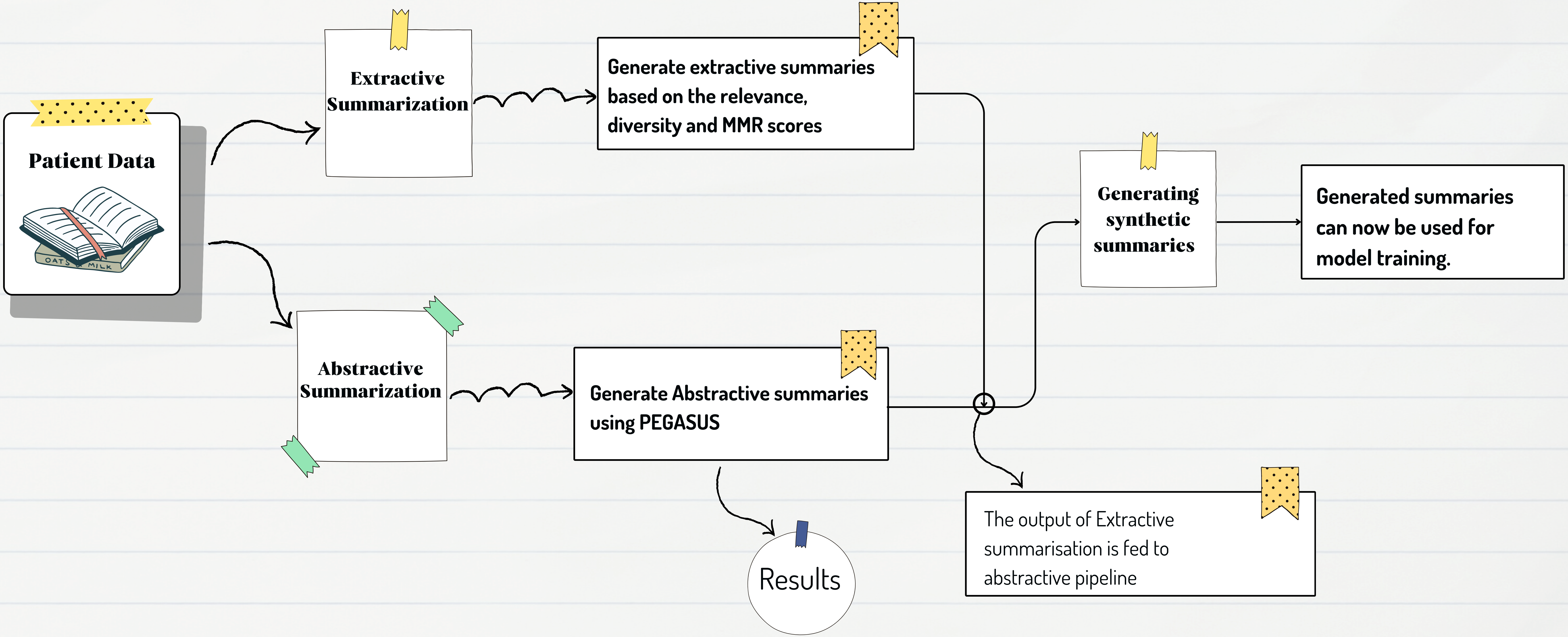
1. What is MMR.?
2. Striking a balance between Relevance and Diversity.!
3. Significance of having a dynamic lambda(λ).??

```python
def dynamic_lambda(num_sentences):
    if num_sentences < 5:
        return 0.86
    elif num_sentences < 10:
        return 0.82
    else:
        return 0.72
```

$$MMR \stackrel{\text{def}}{=} Arg \max_{D_i \in R \setminus S} \left[\lambda (Sim_1(D_i, Q) - (1-\lambda) \max_{D_j \in S} Sim_2(D_i, D_j)) \right]$$

Why MMR

Summarization Work Flow



Hybrid Summarisation

- Combines extractive summarization for factual accuracy with abstractive summarization for fluency and coherence.
- Models Used:
  PEGASUS-XSum

Extractive + Abstractive using only PEGASUS for abstractive

Average Bleu Score	Average ROUGE1 Score	Average ROUGE2 Score	Average ROUGEL Score
0.0823	0.3937	0.2814	0.3234

Fig. 6: Hybrid Summarisation of Original Text

Why choose a Hybrid Approach.?

Model training and testing

- How does preprocessing differ from training.???
- Train-test split.

PEGASUS vs T5 model

Pegasus model	T5 model
Transformer-based encoder-decoder architecture.	Transformer-based encoder-decoder architecture.
Focused primarily on **abstractive summarization**.	General-purpose, can handle **multiple NLP tasks** (summarization, translation, QA, etc.).
Gap-Sentence Generation (GSG)—masks whole sentences and predicts them.	**Text-to-Text Transformation**: Converts all tasks (e.g., summarization, translation) into text generation problems.
Highly specialized for summarization tasks.	General-purpose, adaptable to many NLP tasks.
PEGASUS-Large (~350M parameters).	**Scalable options from T5-Small (60M) to T5-11B (11 billion) parameters.**

Results

Original Text: A 29-year-old male was brought to the emergency department following a car accident. It was reported that the wheel of a 5-ton water tanker rolled over his lower body as he was stepping into a car. Workup revealed a Glasgow Coma Scale (GCS) score of 15 with a pulse rate of 80 beats per minute and blood pressure of 90/60 mmHg. There was profuse bleeding due to a perineum injury. No fractures of the skull, cervical spine, and pelvis were detected in the X-rays. Urinary catheterization was performed without difficulty and there was no hematuria. No internal bleeding was seen during the abdominal ultrasound. The patient was immediately taken to the operating theatre for injury assessment after the replacement of fluid and blood. While the patient was placed in the lithotomy position, the anterior displacement of the anus could be seen with a circular skin defect surrounding the region. The anus looked normal......
Summary generated using T5: A 29-year-old male was brought to the emergency department following a car accident. The patient was immediately taken to the operating theatre for injury assessment after the replacement of fluid and blood. While the patient was placed in the lithotomy position, the anterior displacement of the anus could be seen with a circular skin defect surrounding the region. It was reported that the wheel of a 5-ton water tanker rolled over his lower body as he was stepping into a car. Workup revealed a Glasgow Coma Scale (GCS) score of 15 with a pulse rate of 80 beats per minute and blood......
Summary generated using T5: A 29-year-old male was brought to the emergency department following a car accident. Workup revealed a Glasgow Coma Scale (GCS) score of 15 with a pulse rate of 80 beats per minute and blood pressure of 90/60 mmHg. The patient was immediately taken to the operating theatre for injury assessment after the replacement of fluid and blood. Operative Findings and Procedure While the patient was placed in the lithotomy position, the anterior displacement of the anus could be seen with a circular skin defect surrounding the region. Digital and proctoscopy examination showed an intact anorectum and its surrounding sphincters......

Results

PEGASUS outperformed T5 based on the **ROUGE** scores making it the better choice for our summarization task that involved training the data on hybrid summaries.

Summarisation of text trained on synthetically generated hybrid summaries

Model	Average Bleu Score	Average ROUGE1 Score	Average ROUGE2 Score	Average ROUGEL Score
PEGASUS	0.2576	0.4034	0.2897	0.3319
T5	0.1705	0.3840	0.2732	0.3150

Fig. 7: Predicted Summarization results

Multi-Document Summarisation

Multi-Document Summarization

Intro

A technique that condenses key information from multiple related documents into a concise and coherent summary.

Experimental

It involves labeling the unstructured data, grouping them based on the labels, and summarizing it. Both of these processes are challenging to implement.

Motivation

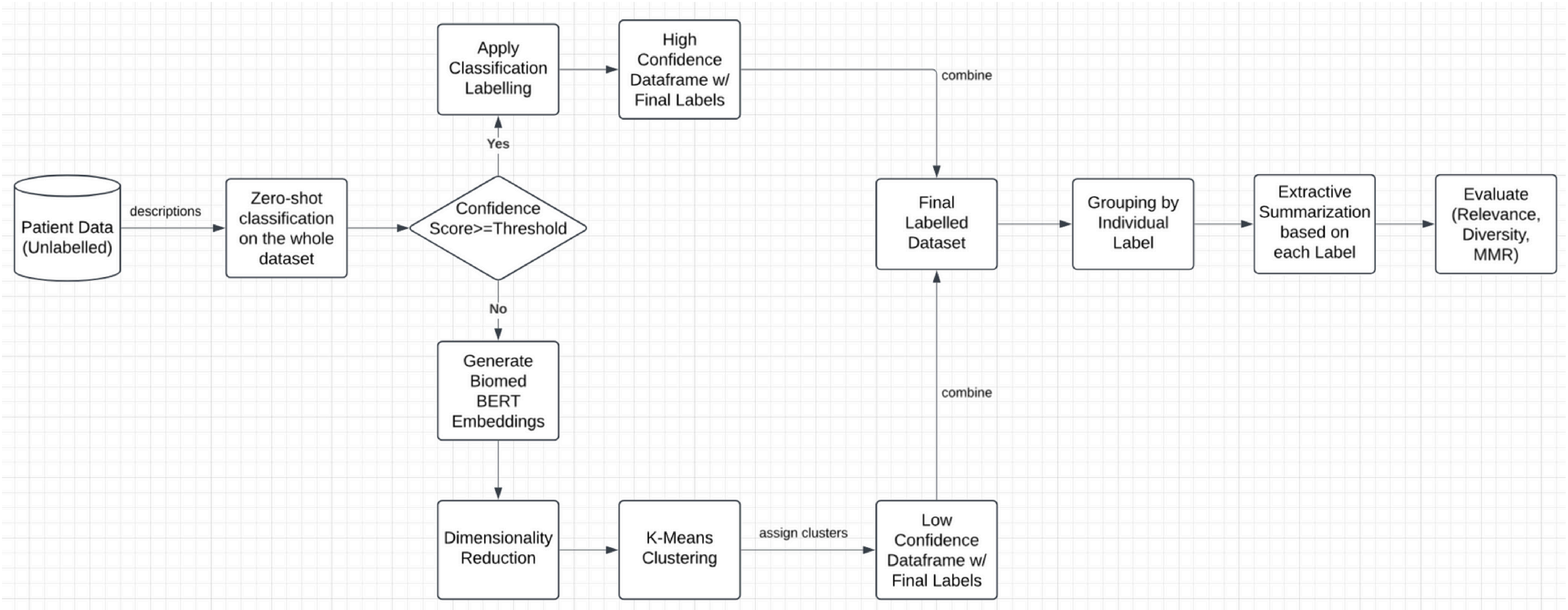
Since medical professionals deal with vast patient records and diagnostic reports, summarizing can save time and improve decision-making.

Dataset Labeling Process

+

Kmeans Clustering

Multi-Document Summarization Pipeline



Results

A small snapshot of the outputs:

01 Zero-shot Classification

Text	Final Label	Confidence Score
An 8-year-old male presents in March to the ER...	Respiratory Issues	0.880696
A 66yo female with significant smoking history...	Respiratory Issues	0.911800
A 62 yo male presents with four days of non-pr...	Respiratory Issues	0.851462

02 K-Means Clustering

Text	Final Label
A 37-year old female presented to the neurology...	Neurological Disorders
The patient was a 9-year-old boy with a history...	Infectious Diseases
A 52-year-old female with no past medical or s...	Cardiac Conditions

Results (Contd.)

Label	Relevance	Diversity	MMR
Cardiac Conditions	0.734872	0.627412	0.107459
Endocrine/Diabetes-Related Issues	0.736672	0.483604	0.253068
Infectious Diseases	0.739921	0.629245	0.110675
Neurological Disorders	0.738828	0.627022	0.111806
Other Issues	0.737954	0.494882	0.243072
Post-Surgical Complications	0.737590	0.631152	0.106439
Psychiatric/Mental Health	0.737409	0.627838	0.109571
Respiratory Issues	0.740512	0.628432	0.112080
Skin Conditions	0.737728	0.634669	0.103059

Questions?