Enhancing Patient Care through Data Visualization and Interactive Chatbot Solutions

Aditya Saxena (saxena11), Aryamaan Sen (as202), Aditya Raju (raju7), Nischay Singh (nischay2)

Project Summary

This project aims to enhance patient care by integrating data visualization and chatbot using two healthcare-related datasets. The first component uses the Healthcare Documentation Database to create an interactive visualization tool that maps common patient symptoms to diagnosis. This tool will help patients and healthcare professionals identify patterns and relationships, providing valuable insights into common symptom-diagnosis connections. We will also use this dataset to create a search engine to allow users to search for all patient records containing information for their related topics. However, it does not provide a conclusive diagnosis, just a correlation of symptoms.

The second component involves developing an AI-powered chatbot augmented with the Healthcare NLP dataset, which contains symptom-related questions and corresponding answers. This chatbot will assist patients by providing accurate and contextually relevant responses to their health-related queries. By integrating language models, the chatbot will improve engagement and accessibility, ensuring users receive reliable information about their symptoms. Together, these tools will bridge the gap between medical data and patient understanding, making healthcare insights more accessible and actionable.

Project Description

The primary problem we aim to solve is the lack of easily interpretable symptom-diagnosis relationships and the challenge of obtaining accurate health-related information outside clinical settings. Many patients struggle to understand medical data and find reliable information about their symptoms. By leveraging real-world healthcare datasets, our project aims to make medical insights more accessible and actionable.

Our project focuses on improving patient access to healthcare information through two key applications: an interactive symptom-diagnosis visualization tool and a chatbot for answering health-related questions. The visualization tool will help users explore how different symptoms correlate with diagnoses, revealing patterns that could improve patient awareness and aid medical professionals in decision-making. We will use third-party software like the Gephi API along with other data manipulations and graphing tools on Python. Meanwhile, the chatbot will provide instant responses to common symptom-related queries, offering a user-friendly interface for patients seeking guidance before consulting a healthcare

professional. To achieve this, we will create a Retrieval Augmented Generation (RAG) system using a vector database and an LLM for the output generation. Lastly, we will create a search engine to allow users to see patient documentation for related topics using the Elasticsearch API. These solutions will enhance patient education, engagement, and proactive health management.

Creative Components

This project uniquely integrates graph visualization, a chatbot, and a search engine to improve healthcare insights. The graph visualization, built using Gephi, maps symptom-diagnosis relationships, allowing doctors and patients to intuitively identify co-occurring conditions. The RAG-based chatbot provides instant AI-driven responses, offering quick explanations and guidance on symptoms. Meanwhile, the Elasticsearch-powered search engine enables patients to explore anonymized medical records, helping them gain insights from similar cases. By combining visual, conversational, and search-based tools, this project makes symptom analysis more intuitive, accessible, and informative.
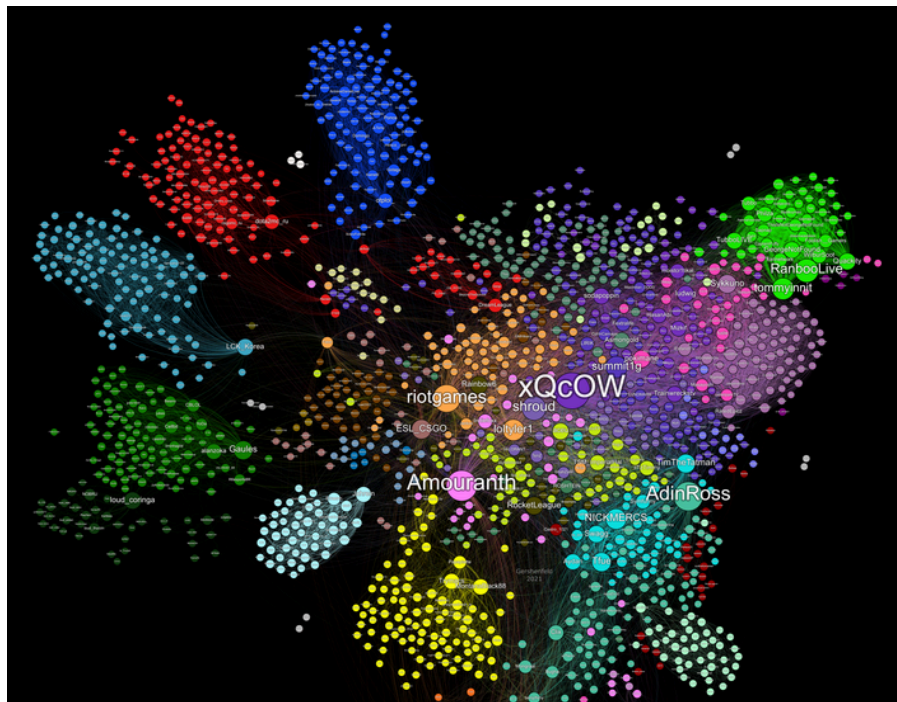
Example Visualization Style

Usefulness

Our project will serve as a real-time application as a diagnosis recommendation tool based on previous medical histories that are publicly available. Essentially, the tool will allow users to

find symptom-diagnosis. Additionally, it could be used as a medical learning tool, proving to be crucial in the education industry.

Secondly, the healthcare industry is expensive for normal lightweight diagnosis. The search engine along with a concentrated chatbot will allow users to find trends in symptom presentations across large datasets. The project's RAG-based chatbot ensures dynamic and contextual responses, making healthcare assistance more interactive. As for our primary users, it will consist mostly of those looking to learn and extract valuable insights from large amounts of medical data. While correlation does not imply causation, valuable information can still be taken away from our visualizations and chatbot experience.

However, it is essential to recognize its ethical limitations. The visualization tool and AI-powered chatbot are designed to serve as reference tools rather than substitutes for professional medical advice. This tool only aims at empowering users with insights but should always be used in conjunction with expert medical consultation. **It is simply a tool to use for educational and research purposes.**

Lastly, there are existing healthcare applications like WebMD Symptom Checker that provide symptom assessments, but they often rely on rule-based decision trees. Our project integrates machine learning, RAG, and interactive visualizations to create a more dynamic and accurate healthcare reference tool. Additionally, the visualization tool helps users explore symptom-diagnosis relationships, offering insights that most symptom checkers lack.

Realness

To create a healthcare application relevant to the real world, we are utilizing two real-world datasets that provide medical data. These datasets allow us to integrate data visualization, retrieval-augmented generation (RAG), and vector similarity search for symptom analysis and intelligent chatbot responses.

The Healthcare Documentation Database is a structured dataset available in CSV format with a cardinality of approximately 1,800 and a degree of 6, containing patient symptoms and relevant medical keywords. These visualizations will allow users to explore symptoms that occur together the most and track the evolution of diagnoses over time to identify seasonal trends and emerging health risks. Additionally, clustering algorithms will help group related symptoms, enabling smart filtering and more intuitive exploration of medical conditions. The dataset's structured nature makes it ideal for identifying patterns in healthcare data, making it useful for both patients and professionals seeking insights into common medical conditions.

The Healthcare NLP Dataset is available in CSV format with a cardinality of approximately 15,000 and degree of 4, containing medical question-answer pairs related to symptoms, diagnoses, and treatments. Unlike simple FAQ-based models, we will use vector similarity search through the Elasticsearch API to transform this dataset into a retrieval-augmented chatbot that provides personalized responses. Instead of relying on keyword matching, our

chatbot will use embeddings to retrieve the most relevant medical cases. This dataset bridges the gap between raw medical records and natural language understanding, allowing users to receive dynamic, evidence-based recommendations.

Project Functionality

Our website will contain 3 main web pages, each offering a unique service to the User -

1. Patient Information Search Engine - This page will allow for users to **search** for patient records containing specific keywords, symptoms, ages, genders, etc. This will provide an interface for users to filter for specific attributes they would like to retrieve.
2. Keyword Graph visualization - Various illnesses and diseases contain shared symptoms and attributes between one another. We hope to provide an insightful visualization that highlights correlations and links between symptoms and illnesses. The graph will cluster symptoms that are often found together depicting their connections.
3. RAG Chatbot - Using our dataset of common questions and answers pertaining to different illnesses and symptoms, we hope to build an interactive chatbot to answer questions. Using a vector database to **store** our content and semantic similarity search methods, we will append integral information to our LLM prompts, ensuring accurate and insightful answers are provided to the user.

Each web page will ensure to have an appealing user experience that clearly conveys important data and information to the users.

CRUD Operations

We are storing the following entities in our database.
1. Users information
2. User chat instances
3. Patient diagnosis records (from dataset)
4. Question-answer symptoms (from dataset)
5. Each user query log from our datasets.

In each user chat instance, we will store a list of user messages and LLM responses so that users can view their past conversation history.

Create (C) -
● Creating new users
● Creating new user chats
● Creating new query logs

Read (R) -
- Retrieving an old chat
- Viewing old query logs
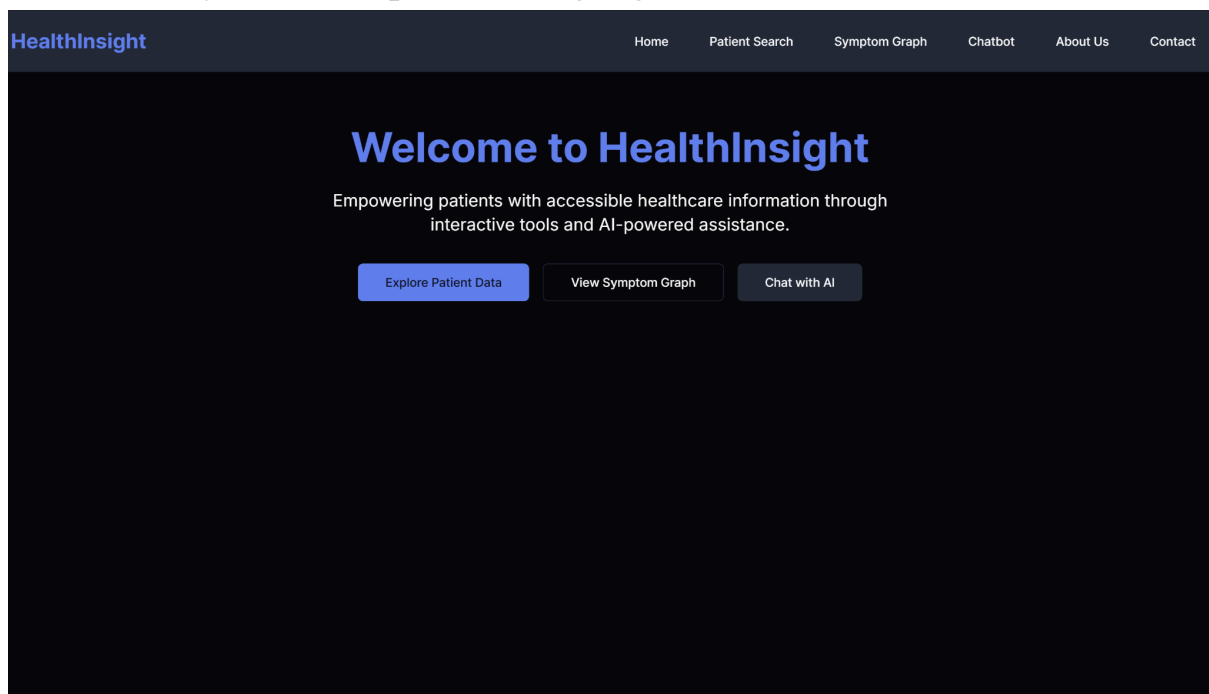- Looking at user information

Update (U) -
- Updating user information
- Queries on a chat will be updated to store the history

Delete (D) -
- Deleting users
- Deleting user chats
- Deleting user query logs

A low-fidelity UI mockup: built using Figma

Patient Information Search

Search for symptoms, conditions, or patient attributes...

Age   Gender   Fever   Cough   Headache   Fatigue

Min Age
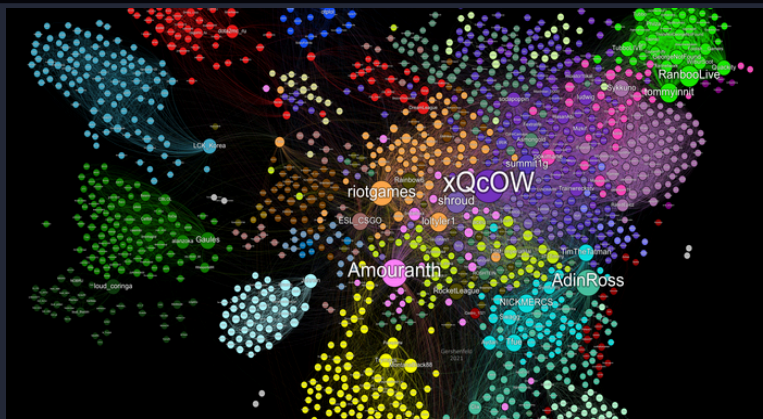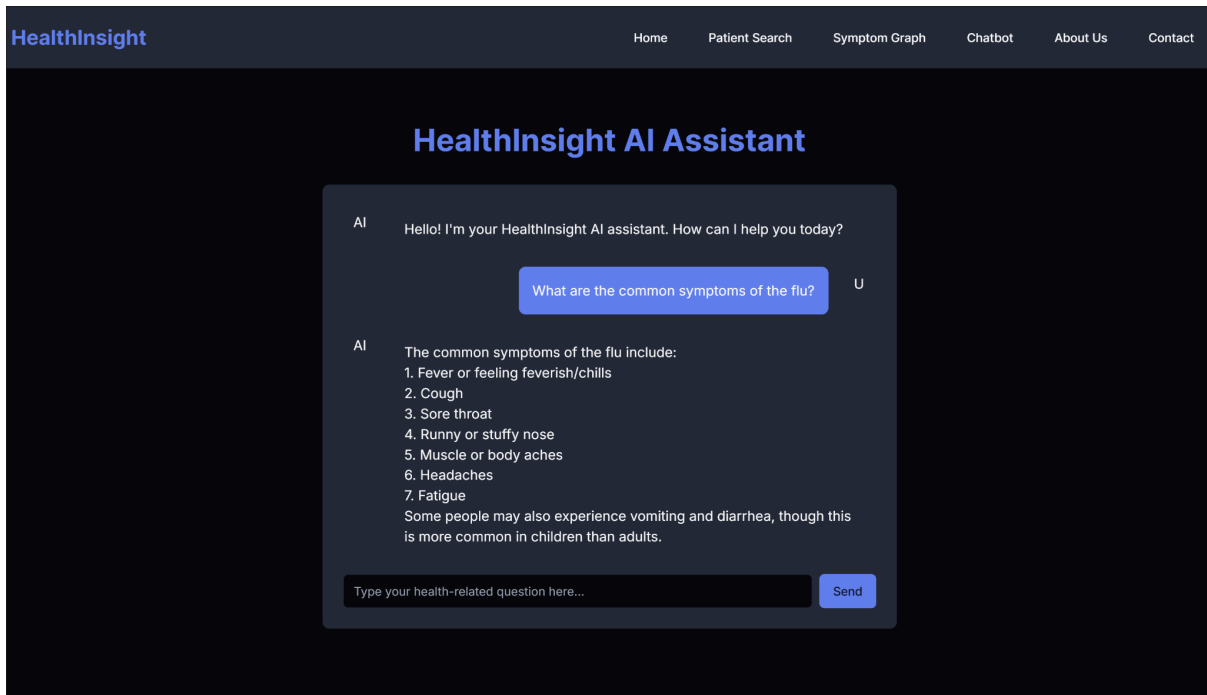
Max Age

Clear Filters

Search

Symptom-Diagnosis Graph Visualization

This interactive graph visualizes the relationships between symptoms and diagnoses. Explore connections, identify patterns, and gain insights into how different health conditions are interrelated.

Project Work Distribution

- **Aditya Saxena** – Develop the search engine using Elasticsearch.
- **Aditya Raju & Nischay Singh** – Build graph visualizations using Gephi, mapping symptom-diagnosis relationships.
- **Aryamaan Sen & Aditya Raju** – Implement the chatbot using Retrieval-Augmented Generation (RAG) for patient Q&A.
- **Aditya Saxena, Nischay Singh** – Develop the front-end for user interaction with the search engine, chatbot, and interactive graphs.