MediCouncil: Multi-Agent LLM Council for Symptom Triage

Done by:

Bhavatharini C (71762233004)

Gayathri SK   (71762233014)

DATASET EXPLANATION

The **Disease and Symptoms Dataset** is a structured medical dataset used to model the relationship between patient-reported symptoms and diagnosed diseases. Each record represents a patient case, where symptoms are encoded as binary values indicating their presence or absence, and mapped to a single disease outcome. The dataset supports medical diagnosis, symptom-based disease classification, and clinical decision-support systems.

- **Total features:** 378
- **Target feature:** 1 (diseases)
- **Symptom features:** 377
- **Encoding:** Binary (1 = symptom present, 0 = symptom absent)

Symptom Coverage Description

The symptom features collectively cover a wide range of medical domains, including:

- **Psychological & Behavioral Symptoms**
  (e.g., anxiety, depression, panic disorder, hallucinations, low self-esteem, obsessions)
- **Respiratory Symptoms**
  (e.g., shortness of breath, wheezing, cough, abnormal breathing sounds, apnea)
- **Cardiovascular Symptoms**
  (e.g., palpitations, irregular heartbeat, chest pain, decreased or increased heart rate)
- **Gastrointestinal Symptoms**
  (e.g., nausea, vomiting, diarrhea, abdominal pain, blood in stool, constipation)
- **Neurological Symptoms**
  (e.g., dizziness, seizures, fainting, weakness, memory disturbance, slurred speech)
- **Musculoskeletal Symptoms**
  (e.g., joint pain, stiffness, swelling, cramps, limb weakness)
- **Genitourinary & Reproductive Symptoms**
  (e.g., painful urination, vaginal discharge, menstrual irregularities, prostate symptoms)
- **Dermatological Symptoms**
  (e.g., skin rash, lesions, swelling, itching, dryness, abnormal moles)
- **ENT & Ophthalmological Symptoms**
  (e.g., sore throat, nasal congestion, ear pain, vision problems, eye redness)
- **Systemic & General Symptoms**
  (e.g., fever, fatigue, chills, weight changes, flu-like syndrome)

EXPLORATORY DATA ANALYSIS

- **Dataset Shape & Column Overview**

```
Dataset Shape: (246945, 378)

Column Names:
 Index(['diseases', 'anxiety and nervousness', 'depression',
       'shortness of breath', 'depressive or psychotic symptoms',
       'sharp chest pain', 'dizziness', 'insomnia',
       'abnormal involuntary movements', 'chest tightness',
       ...
       'stuttering or stammering', 'problems with orgasm', 'nose deformity',
       'lump over jaw', 'sore in nose', 'hip weakness', 'back swelling',
       'ankle stiffness or tightness', 'ankle weakness', 'neck weakness'],
      dtype='object', length=378)
```

Inference:

- Original dataset size: **246,945 records × 378 columns**
- After removing duplicates: **189,647 records**
- Total symptoms captured: **377**
- One target column: **diseases**
- All symptom columns are binary/int64, indicating **presence (1) or absence (0)** of symptoms.

Conclusion:

- This is a **large-scale, high-dimensional medical dataset**.
- The high number of symptom features enables **fine-grained symptom pattern recognition**.
- Suitable for both **classical ML models** and **AI-driven symptom triage systems**.

- **Data Types Distribution**

```
Data Types:
 diseases                          object
anxiety and nervousness           int64
depression                        int64
shortness of breath               int64
depressive or psychotic symptoms  int64
                                  ...
hip weakness                      int64
back swelling                     int64
ankle stiffness or tightness      int64
ankle weakness                    int64
neck weakness                     int64
```

Inference:

- The diseases column is categorical (object).
- All symptom features are numerical (int64).
- Binary encoding simplifies:
    - Feature engineering
    - Model training
    - Interpretability

Conclusion:
The dataset is **clean and ML-ready**, requiring minimal preprocessing and enabling efficient model benchmarking.

- **Duplicate Removal & Target Summary**

```
Length: 378, dtype: object
Duplicate Rows: 57298
Shape After Removing Duplicates: (189647, 378)
Target Column: diseases
Total Symptoms: 377
```
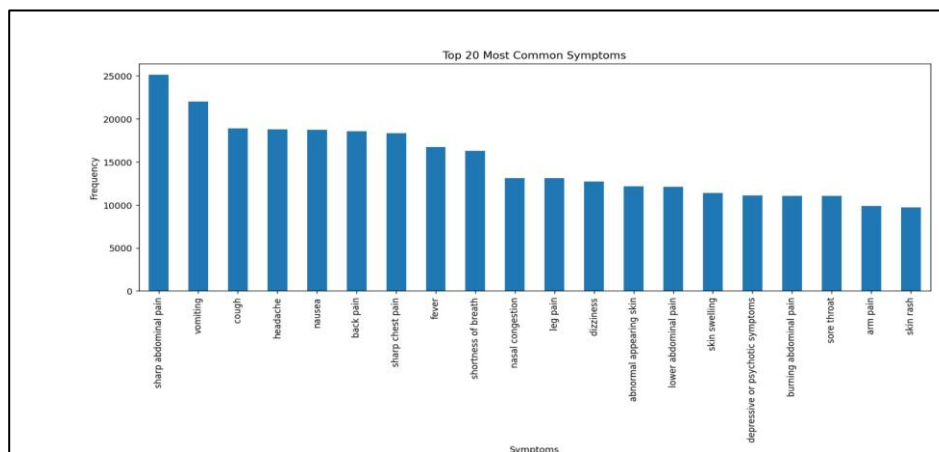
Inference:

- Duplicate rows detected: **57,298**
- Duplicate removal significantly reduced dataset size without losing feature richness.
- Ensures **data quality and unbiased learning**.
- Target variable (diseases) clearly defined for supervised learning.

Conclusion:
Removing duplicates improves **model generalization**, prevents overfitting, and ensures reliable evaluation.

- **Top 20 Most Common Symptoms**

Inference:

- The dataset is **symptom-heavy**, with certain symptoms occurring far more frequently than others.
- **Sharp abdominal pain**, **vomiting**, **cough**, **headache**, and **nausea** are the most dominant symptoms.
- Respiratory-related symptoms (**shortness of breath**, **nasal congestion**, **fever**) appear frequently, indicating strong representation of **infectious and respiratory conditions**.
- Pain-related symptoms (abdominal pain, back pain, chest pain, leg pain) dominate, showing that **pain severity is a critical indicator** for triage decisions.
- Mental health indicators (**depressive or psychotic symptoms**, **dizziness**) are also present, highlighting **multi-domain symptom coverage** (physical + psychological).

Conclusion:
This distribution supports building a **robust triage model**, as it includes both **common, low-risk symptoms** and **high-risk red-flag symptoms**, essential for urgency classification.

- **Top 20 Most Frequent Diseases**
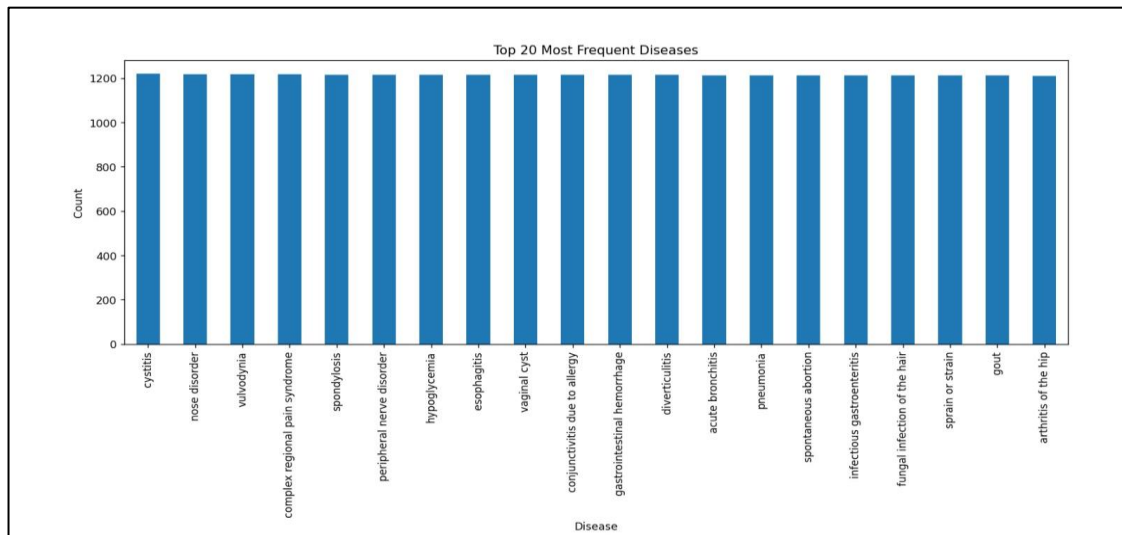
Inference:

- Disease distribution is **fairly balanced** across the top 20 diseases.
- No extreme class dominance → reduced risk of **class imbalance bias**.
- Includes diverse disease types:
  - Infectious (gastroenteritis, pneumonia)
  - Musculoskeletal (sprain, arthritis, spondylosis)
  - Neurological (peripheral nerve disorder)
  - ENT & dermatological conditions

Conclusion:
Balanced disease frequency supports **multi-class classification** and **fair triage evaluation**, especially for benchmarking ML and AI models.

REQUIREMENTS ANALYSIS

Functional Requirements

- **FR1 – Symptom Intake:**
  The system shall allow users to enter symptoms along with optional context such as age, sex, duration, and severity through a web interface.
- **FR2 – Input Validation:**
  The system shall validate mandatory fields, handle invalid values, and normalize symptom text into standardized medical terms.
- **FR3 – Feature Generation:**
  The system shall transform normalized symptoms and context into a structured, machine-readable format (symptom vector / JSON).
- **FR4 – Baseline ML Inference:**
  The system shall generate baseline predictions using classical machine learning models such as Naive Bayes, Logistic Regression, and Random Forest.
- **FR5 – LLM Council Inference:**
  The system shall invoke multiple specialized LLM agents (triage, primary-care severity, red-flag detection) to produce independent assessments.
- **FR6 – Safety & Consensus Decision:**
  The system shall combine ML and LLM outputs using a consensus mechanism and apply a safety override to escalate cases when red-flag symptoms are detected.

- **FR7 – Triage Output:**
  The system shall return a triage result including risk level, urgency category, and recommended next action (self-care, clinic visit, or emergency care).
- **FR8 – Explainability:**
  The system shall provide a user-readable explanation highlighting key symptoms and reasons, along with a confidence indicator and medical disclaimer.
- **FR9 – Logging:**
  The system shall log inputs, individual model outputs, final decisions, and timestamps for auditing and evaluation purposes.
- **FR10 – Admin & Reporting:**
  The system shall allow authorized administrators to access aggregated logs and generate evaluation reports such as accuracy and safety metrics.
- **FR11 – Error Handling:**
  The system shall handle invalid inputs, model failures, or service downtime gracefully and return clear error messages without exposing sensitive information.
- **FR12 – API Endpoint:**
  The system shall expose a /triage API endpoint that accepts symptom data and returns triage results in JSON format for UI and testing integration.

Non-Functional Requirements

- **NFR1 Performance:**
  The system shall return a triage response within an acceptable response time under normal operating conditions.
- **NFR2 Reliability:**
  The system shall provide consistent and correct triage results and continue to operate safely even if one model or agent fails.
- **NFR3 Safety:**
  The system shall prioritize conservative triage decisions and escalate cases when red-flag symptoms or low confidence are detected.
- **NFR4 Security and Privacy:**
  The system shall protect user data through secure communication and restrict access to authorized users only.
- **NFR5 Explainability and Usability:**
  The system shall present triage outcomes with clear explanations and confidence indicators that are easy for users to understand.