

MediCouncil: Multi-Agent LLM Council for Symptom Triage

DONE BY:

BHAVATHARINI C- 71762233004

GAYATHRI SK - 71762233014

Statistical Tools and Results

1. Dataset Summary

The dataset used for this study consists of symptom-based medical records represented in a structured tabular format. Each row corresponds to one patient case (observation), where the input features are a set of symptom indicators and the output label is the diagnosed disease category.

- Number of records (N): <code>df.shape[0]</code>
- Number of symptom features (M): <code>df.shape[1]</code>
- Target variable: <code>TARGET_COL</code>
- Feature type: Binary (0/1) symptom indicators
- Task type: Multi-class classification (disease prediction) and decision-support triage mapping.

This dataset format is suitable for both statistical analysis (feature relevance, symptom prevalence) and machine learning baselines (Logistic Regression, Naive Bayes, Random Forest).

2. Data Quality and Preprocessing Checks

2.1 Missing Value Analysis

A missing value analysis was performed to ensure dataset reliability and to identify any preprocessing requirements. The missing-value counts were computed for each column and exported for reference.

```
3.1 Missing values summary

missing = df.isna().sum().sort_values(ascending=False)
missing[missing > 0].to_csv(f"{RESULTS_DIR}/missing_values.csv")
missing.head(20)

✓ 0.2s

diseases
stiffness all over
infertility
painful menstruation
unpredictable menstruation
heavy menstrual flow
long menstrual periods
nosebleed
back mass or lump
low back cramps or spasms
muscle cramps, contractures, or spasms
back cramps or spasms
sweating
burning chest pain
uterine contractions
recent pregnancy
fears and phobias
painful sinuses
sinus congestion
flu-like syndrome
dtype: int64
```

Observation:

No missing values were found.

2.2 Binary Feature Validation

Since most symptom columns are expected to be binary, a validation was conducted to ensure each symptom feature only contains values from the set {0,1}. Non-binary columns (if any) were identified for cleaning or conversion.

3. Descriptive Statistical Tools and Results

3.1 Class Distribution (Target Variable)

To understand dataset balance and disease representation, the frequency distribution of target classes was calculated. This helps identify class imbalance, which can affect model training and evaluation.

```
3.2 Class distribution

class_counts = df[TARGET_COL].value_counts()
class_counts.to_csv(f"{RESULTS_DIR}/class_distribution.csv")

print(class_counts.head(20))

plt.figure(figsize=(12,6))
sns.countplot(y=df[TARGET_COL], order=class_counts.index)
plt.title("Class Distribution (Disease Counts)")
plt.tight_layout()
plt.savefig(f"{RESULTS_DIR}/class_distribution.png", dpi=200)
plt.show()

✓ 12.7s

diseases
cystitis                1219
vulvodynbia              1218
nose disorder             1218
complex regional pain syndrome 1217
spondylosis               1216
hypoglycemia               1215
peripheral nerve disorder 1215
esophagitis                1215
vaginal cyst                1215
conjunctivitis due to allergy 1215
diverticulitis              1214
gastrointestinal hemorrhage 1214
acute bronchitis             1213
pneumonia                  1212
fungal infection of the hair 1212
infectious gastroenteritis   1212
spontaneous abortion          1212
sprain or strain             1212
gout                        1211
arthritis of the hip           1210
Name: count, dtype: int64
```

Key Findings:

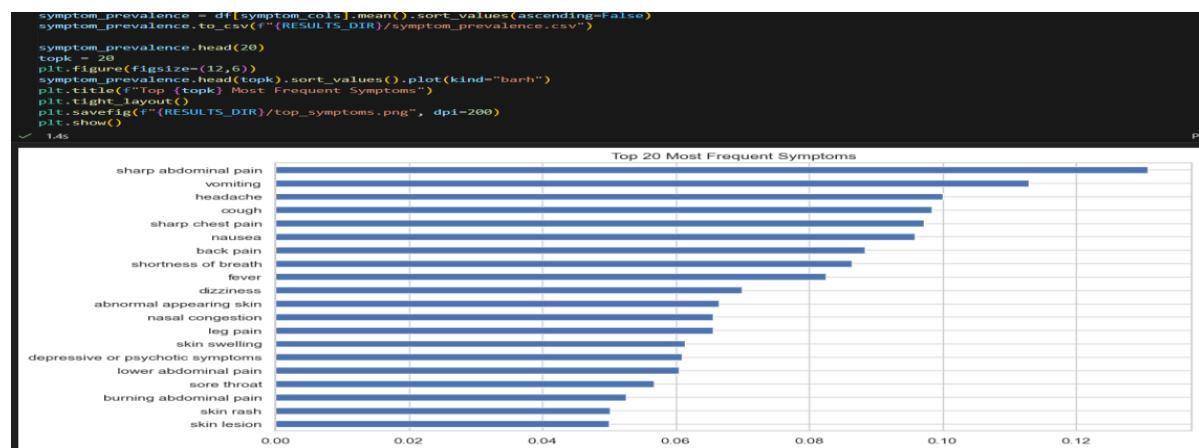
- Target variable: **diseases** (multi-class disease label).
- Balanced class distribution (important): The most frequent diseases in the top-20 list have counts tightly clustered between 1210 and 1219, indicating the dataset is highly balanced across these classes (no major dominance by a single disease).

- Most frequent class in the shown output: *cystitis* with 1219 samples.
- Next most frequent classes (shown): *vulvodynia* (1218), *nose disorder* (1218), *complex regional pain syndrome* (1217), *spondylosis* (1216).
- Least frequent class within the shown top-20 list: *arthritis of the hip* with 1210 samples.
- Practical implication: Because the distribution is balanced, evaluation metrics like accuracy and macro-F1 will be meaningful, and the model is less likely to be biased toward a small set of dominant classes.

Overall, the class distribution appears nearly uniform (\approx 1210–1219 samples per class in the top-20), suggesting minimal class imbalance and supporting stable multi-class training and fair evaluation across diseases.

3.2 Symptom Prevalence Analysis

Symptom prevalence statistics were computed to identify which symptoms occur most frequently across the entire dataset. This helps understand the dataset's dominant symptom patterns and can guide feature selection and explainability.



Key findings (Overall symptom prevalence)

From the “Top 20 Most Frequent Symptoms” bar plot, the dataset is dominated by a small group of high-frequency symptoms, mostly related to pain, gastrointestinal issues, respiratory complaints, and common infection-like symptoms.

Most frequent symptoms (Top group):

- sharp abdominal pain (highest prevalence in the dataset).

- vomiting, headache, cough, sharp chest pain, nausea appear next and form the second-highest prevalence group.
- back pain, shortness of breath, fever, and dizziness are also highly prevalent and occur across many records.

Other frequently present symptoms (still in top-20):

- abnormal appearing skin
- nasal congestion
- leg pain
- skin swelling
- depressive or psychotic symptoms
- lower abdominal pain
- sore throat
- burning abdominal pain
- skin rash
- skin lesion

Interpretation :

- These results show the dataset includes general symptoms that commonly overlap across multiple diseases (e.g., headache, fever, cough), which means single symptoms are not enough to identify a disease—models must learn symptom combinations and patterns.
- The frequent presence of both respiratory (cough, shortness of breath, fever, nasal congestion, sore throat) and gastrointestinal symptoms suggests the dataset covers diverse conditions rather than one medical domain.

3.3 Symptom Prevalence by Disease (Group Statistics)

To improve interpretability, symptom prevalence was also computed class-wise (per disease). This reveals which symptoms are most representative of each disease class.

```

5.2 Per-disease symptom prevalence (top symptoms for each disease)

# Mean of 0/1 columns grouped by disease
grouped = df.groupby(TARGET_COL)[symptom_cols].mean()

# Save full table (can be large)
grouped.to_csv(f"{RESULTS_DIR}/symptom_prevalence_by_disease.csv")

# Show top 10 symptoms for one example disease
example_disease = grouped.index[0]
grouped.loc[example_disease].sort_values(ascending=False).head(15)

✓ 1.4s

palpitations          0.797101
burning abdominal pain 0.789855
shortness of breath    0.760870
arm swelling           0.760870
sharp abdominal pain   0.565217
back pain              0.471014
anxiety and nervousness 0.000000
nosebleed              0.000000
mass on eyelid          0.000000
sweating                 0.000000
frequent menstruation    0.000000
infertility              0.000000
painful menstruation     0.000000
unpredictable menstruation 0.000000
heavy menstrual flow      0.000000
Name: abdominal aortic aneurysm, dtype: float64

```

Key findings (Per-disease prevalence)

In your disease-wise symptom prevalence example (for abdominal aortic aneurysm), the top symptoms shown are:

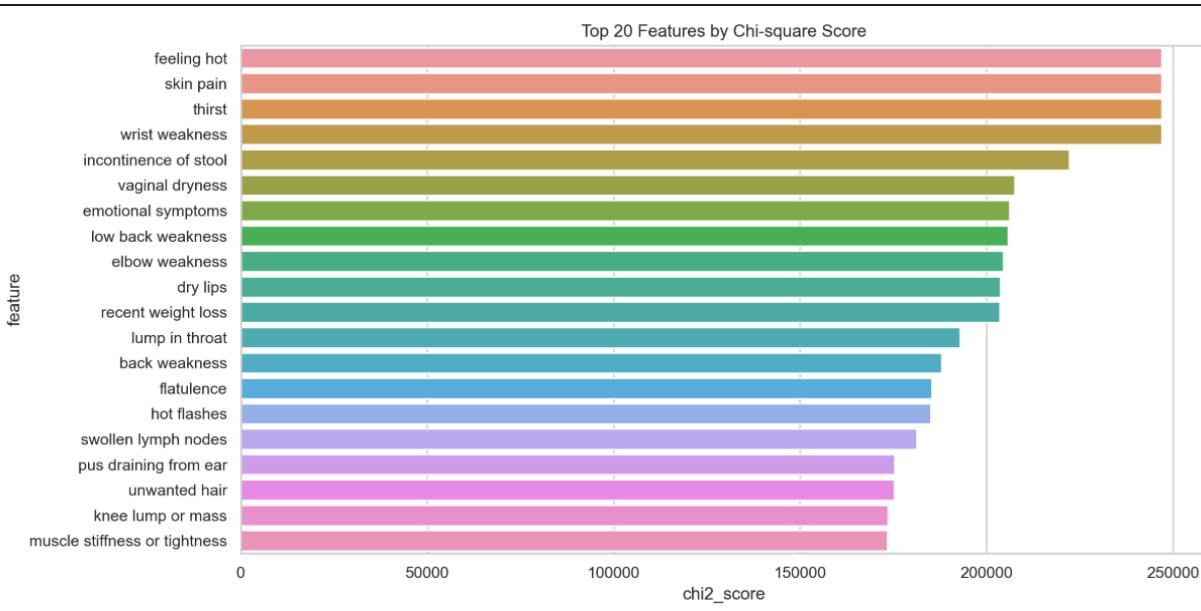
- palpitations (~0.797)
- burning abdominal pain (~0.790)
- shortness of breath (~0.761)
- arm swelling (~0.761)
- sharp abdominal pain (~0.565)
- back pain (~0.471)

Many other symptoms in the printed list show 0.0 prevalence for that disease in the displayed sample, indicating that for a given disease, only a subset of symptoms are strongly associated.

4. Feature Relevance Statistical Tools and Results

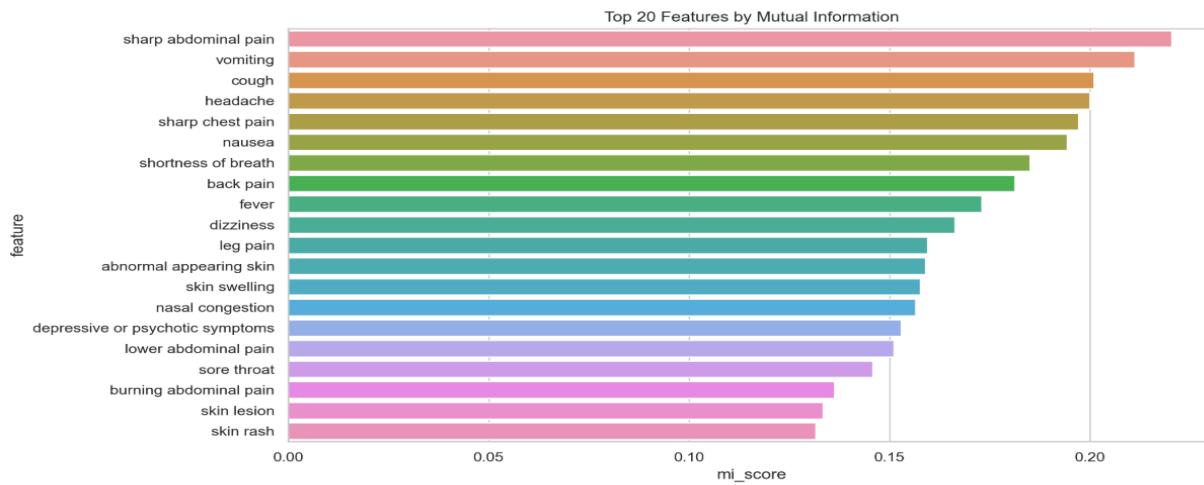
4.1 Chi-Square Test (Feature vs Class)

Since symptom features are binary (0/1), the Chi-square test of independence was used to measure statistical association between each symptom feature and the disease label. Features with higher Chi-square scores are more informative for distinguishing classes.



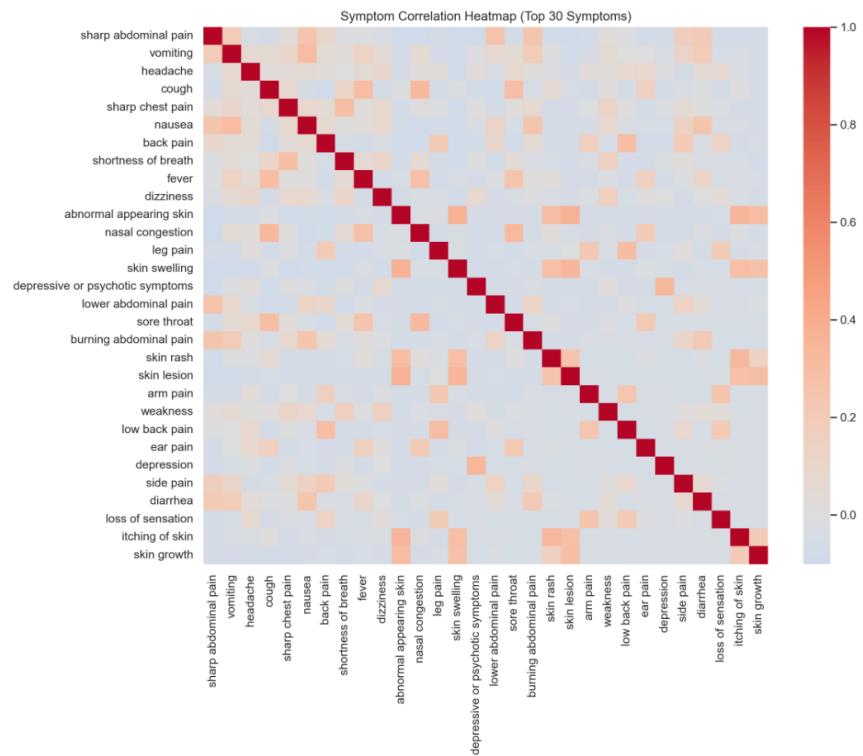
4.2 Mutual Information (MI)

Mutual Information (MI) was computed as an additional feature relevance measure. Unlike Chi-square, MI can better capture general dependency patterns between features and the target label.



5. Correlation Analysis (Symptom Co-occurrence)

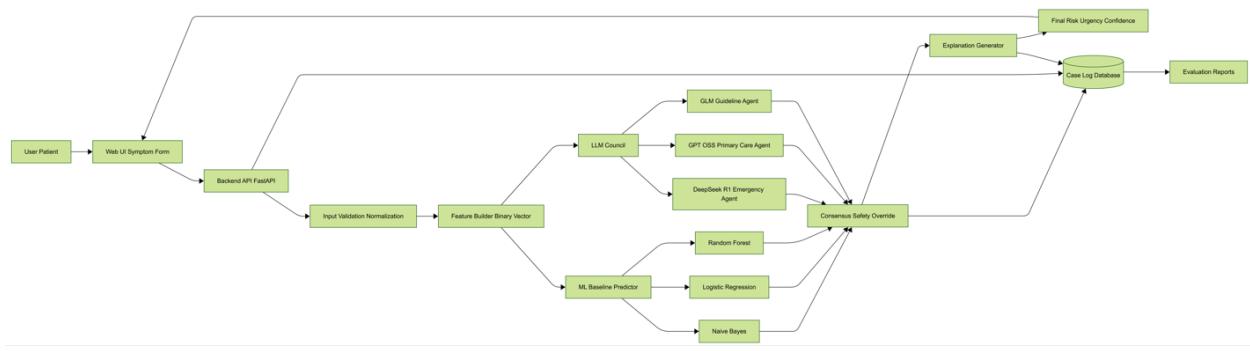
To understand symptom co-occurrence patterns, correlation analysis was performed on the top N most frequent symptoms. This helps detect symptom clusters (e.g., chest pain + shortness of breath) and supports clinical grouping in the explanation layer.



Algorithms/Design Diagrams

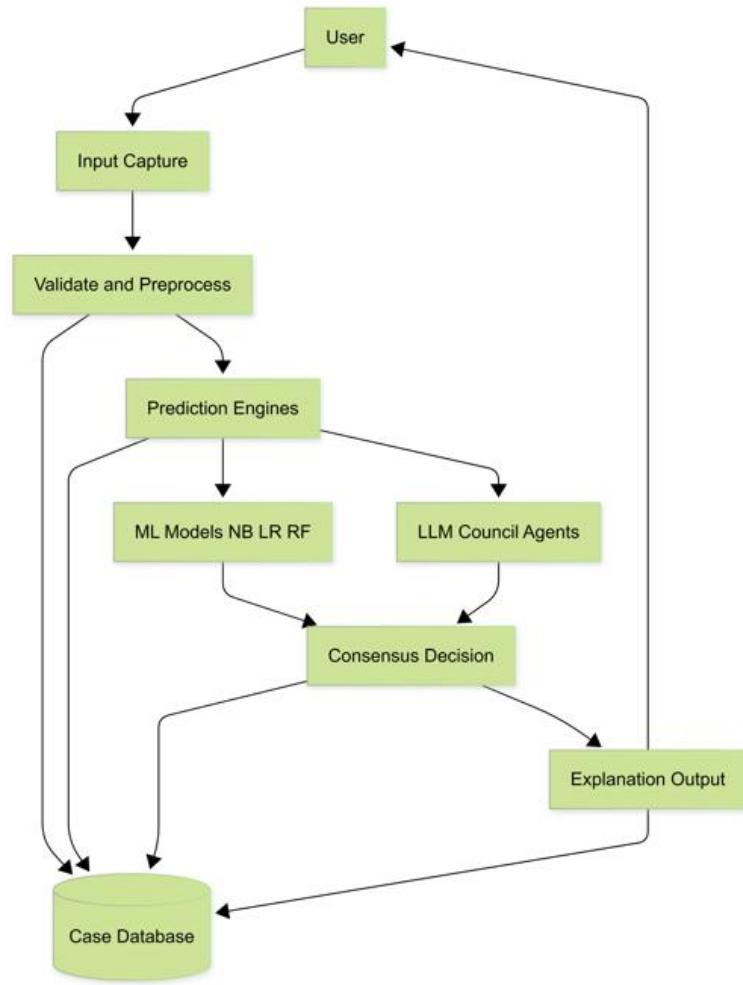
System Architecture

- The user enters symptoms through a web interface, which sends data to the backend API.
- The backend validates inputs and converts symptoms into a binary feature vector.
- Machine Learning models and LLM agents process the data in parallel.
- A consensus and safety module generates the final risk, urgency, explanation, and stores logs.



Data Flow Diagram (Level-1)

- User symptoms are captured and preprocessed by the system.
- The processed data is sent to ML models and LLM agents for prediction.
- Outputs are combined using a consensus decision process.
- Final results and logs are stored and returned to the user.



Algorithm Flowchart – Explanation

- User input is validated and formatted into a structured case.
- ML models and LLM agents generate risk scores simultaneously.
- Emergency conditions are checked first, followed by weighted scoring.
- The final risk level, confidence, and explanation are generated and displayed.

