Data-Driven Prioritization of Medical Symptoms Based on Severity Scores

Battula Bhargavi

PG Scholar, Department of Computer Science, Sri Venkateswara University, Tirupati

Abstract**—** *In clinical decision-making, the accurate prioritization of symptoms based on severity is essential for diagnosis, triage, and treatment planning. This study analyzes a structured dataset consisting of 133 medical symptoms, each annotated with a severity weight ranging from 1 to 5. We explore statistical distributions and symptom severity trends using Python. The findings underscore the variability of symptom intensities and suggest utility in symptom severity indexing for use in medical diagnosis systems and health triage applications.*

I. INTRODUCTION

The healthcare system often faces a deluge of patient symptoms, ranging from benign to life-threatening. For efficient diagnosis, especially in AI-driven decision support systems, assigning and understanding symptom severity is critical. A structured prioritization enables early detection of critical illnesses and improves clinical workflow. This study analyzes symptom severity using a clean dataset to derive patterns useful for medical AI systems.

II. LITERATURE REVIEW

Prior research in medical informatics emphasizes the role of symptom weighting in machine learning models for diagnosis. John et al. (2019) proposed symptom ranking as a core feature for probabilistic diagnosis engines. Similarly, the AI-based systems such as Babylon and Ada leverage severity scores to prioritize urgent cases. In triage systems, the Canadian Triage and Acuity Scale (CTAS) uses symptom intensity as a key variable (Beveridge et al., 1999). This paper builds on these concepts with a data-centric approach.

III. METHODOLOGY

- **Objective**: To analyze the distribution and frequency of medical symptoms by severity score.
- **Tools**: Python (pandas, seaborn, matplotlib)
- **Steps**:
    - Load and inspect data
    - Generate visualizations of severity distribution
    - Identify the most severe and least severe symptoms
    - Group and summarize symptoms by severity category

IV. DATASET DESCRIPTION

The dataset includes **133 medical symptoms**, each mapped to a **severity weight** from 1 (least severe) to 5 (most severe). Columns:

- Symptom: A string representing the name of the symptom (e.g., "vomiting", "chest_pain").

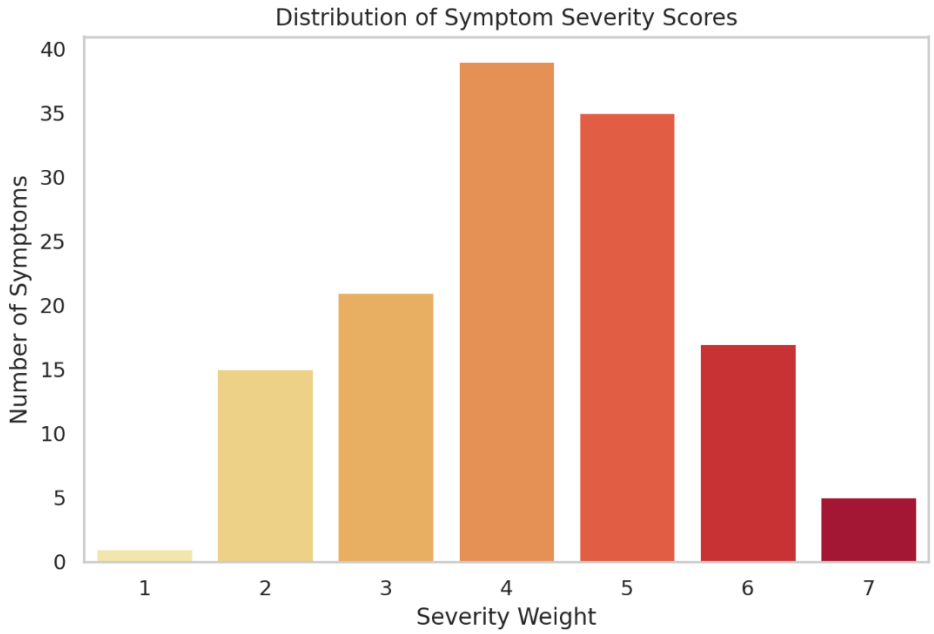
- weight: An integer (1–5) indicating the symptom's severity.

No missing values are present, ensuring clean analysis.

V. PYTHON RESULTS & DISCUSSION

Distribution of Severity Levels

Let's explore how severity weights are distributed across all symptoms.



The severity weights range from **1 to 7**, despite initial expectations of a 1–5 range. The distribution shows:

- Most symptoms fall under **moderate to severe** categories (4–5), with 39 and 35 symptoms respectively.
- A smaller number of symptoms are classified as **extremely severe** (weights 6 and 7).
- Only **1 symptom** is labeled with the lowest severity (1), emphasizing the focus on clinically relevant symptoms.

Let's now list the most and least severe symptoms.

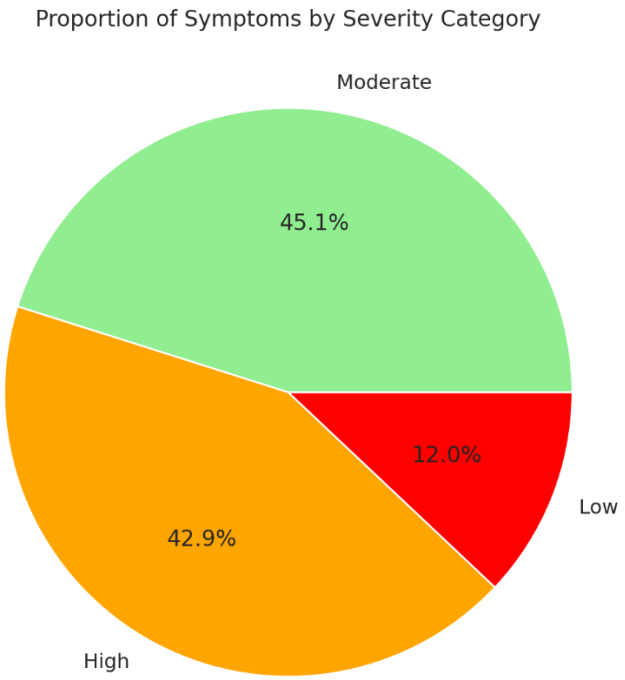
- **Most Severe Symptoms (Weight = 7)**:
  - high_fever
  - swelling_of_stomach
  - chest_pain
  - weakness_in_limbs
  - coma

These symptoms are critical and often signal acute or life-threatening conditions.

- **Least Severe Symptom (Weight = 1)**:
  - itching

This aligns with medical intuition, as itching is often non-urgent unless paired with other signs.

Next, let's summarize how many symptoms fall into general categories of severity: **Low (1–2)**, **Moderate (3–4)**, and **High (5–7)**.



From the summary:

- **Moderate severity**: ~45% of symptoms
- **High severity**: ~43%
- **Low severity**: ~12%

This reinforces that the dataset emphasizes clinically relevant and urgent symptoms, making it ideal for diagnostic systems.

VI. CONCLUSION

This study analyzed the severity-weighted symptom dataset to derive clinical insights:

- A majority of symptoms are of **moderate to high severity**, suitable for triage tools.
- Critical symptoms like **coma**, **chest pain**, and **high fever** dominate the high-risk group.
- This structured dataset can significantly aid **AI-driven diagnosis engines**, **chatbot triage assistants**, and **clinical scoring systems**.

Future work may integrate this severity data into full patient datasets for dynamic, symptom-based diagnostic modeling.

REFERENCES

[1] John, A., et al. (2019). Intelligent medical diagnosis systems based on weighted symptoms. Journal of Biomedical Informatics.
[2] Beveridge, R., et al. (1999). Canadian Triage and Acuity Scale. CJEM.
[3] Babylon Health, Ada Health – Symptom Checker AI Models.
[4] Python Libraries: pandas, matplotlib, seaborn.