Context-Aware Symptom Inference in Cancer Care Using LLMs

Joonwoo Park, MS; Alaa Albashayreh, PhD, MSHI, RN
University of Iowa

IOWA

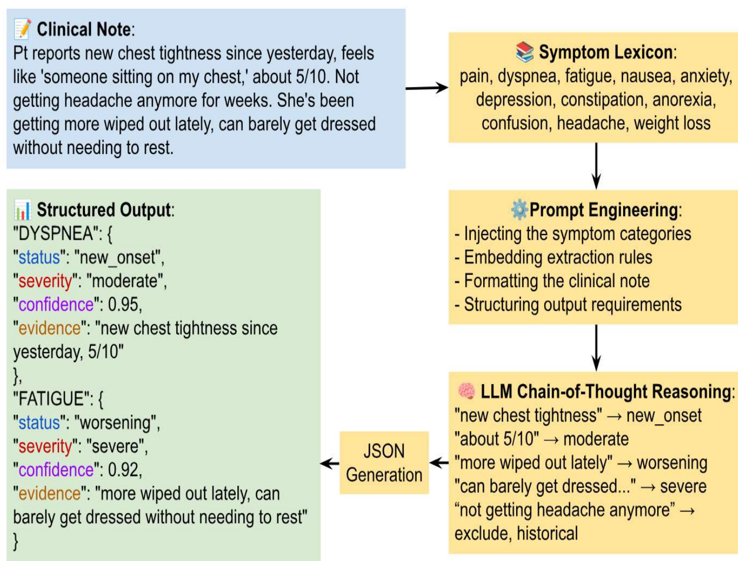
Introduction

- Accurate assessment of symptom burden is essential for cancer care, yet implementation remains limited due to workflow barriers and incomplete documentation.
- Existing natural language processing (NLP) systems can identify symptom mentions but fail to differentiate active vs. historical symptoms, current vs. resolved conditions, and educational content vs. actual patient experiences.
- Traditional rule-based approaches lack the contextual understanding needed for severity assessment, temporal progression tracking, and negation detection.
- Comprehensive symptom surveillance requires systems that can extract presence, severity, progression status, and temporal context from unstructured clinical narratives.
- **Objective**: We developed SympInferAI, a large language model (LLM)-based system using chain-of-thought prompting to extract structured symptom information from clinical notes, enabling scalable, context-aware symptom surveillance for cancer patients.

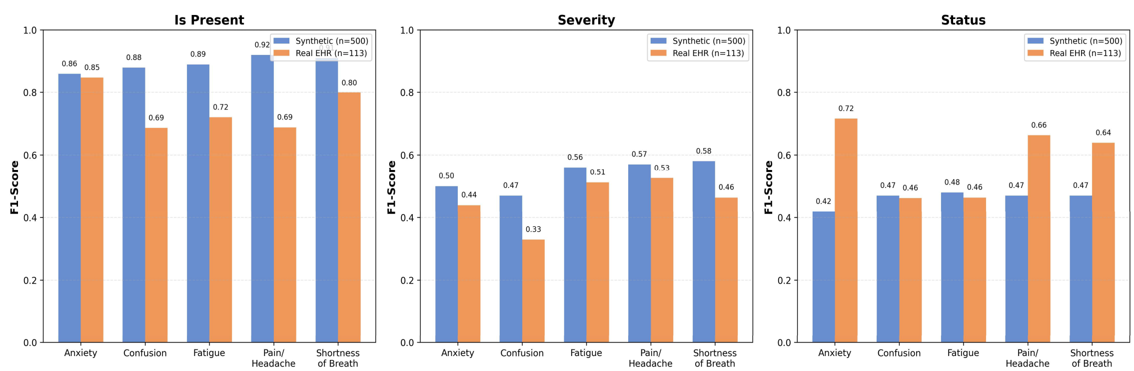
Methods and Materials

- **Model**: Llama-3.1-8B-Instruct with zero-shot chain-of-thought prompting. Hyperparameters optimized using Optuna. Structured JSON outputs with symptom presence, progression status, severity, and supporting evidence.
- **Target symptoms**: SympInferAI targets five high-priority symptoms: pain/headache, fatigue, shortness of breath, anxiety, and confusion.
- **Validation**: Synthetic notes (*n*=500) systematically generated using Claude 3.7 Sonnet, and real EHR notes (*n*=113) from cancer patients.
- **Annotation**: All notes independently annotated by two trained healthcare researchers with consensus resolution (inter-annotator agreement *κ*=0.86).

SympInferAI – Clinical Symptom Inference Pipeline



SympInferAI performance varied by attribute and dataset, with progression status benefiting most from real clinical context.



Results

- SympInferAI achieved high F1-scores (0.86-0.92) on synthetic notes and maintained strong performance (0.69-0.85) on real EHR notes, with anxiety demonstrating the strongest overall performance.
- Progression status improved on EHR notes (F1=0.60) compared to synthetic notes (F1=0.47), with anxiety showing the greatest improvement (0.72 vs 0.42).
- Moderate severity classification (F1: 0.47-0.58) across both datasets highlights the complexity of capturing linguistic cues and functional impact descriptions in clinical documentation.

Conclusions

- SympInferAI demonstrates that LLM-based extraction with chain-of-thought prompting can reliably detect symptom presence (F1: 0.69-0.92) and capture progression patterns from clinical narratives.
- The superior performance on real EHR notes for temporal reasoning highlights the value of rich clinical context. While severity classification remains challenging, the system's evidence-based, explainable outputs address critical requirements for clinical AI adoption.
- **Clinical impact**: This framework enables scalable, automated symptom surveillance to support evidence-based, person-centered care in oncology.
- **Future work**: expand annotated dataset and fine-tune models on domain-specific clinical narratives.

This work was funded by the Iowa Initiative for AI.

Let's Connect!

Alaa-Albashayreh@uiowa.edu