MedDecXtract-XAI: Explainable Medical Decision Extraction with Grounded LLM Rationales

Mohamed Elgaar, BSc[1],Hadi Amiri, PhD[1,2], Leo A. Celi, MD[2]
[1] UMass Lowell, Lowell, MA, USA [2] MIT, Cambridge, MA, USA

Introduction: Clinical narratives contain rich information about key decisions (diagnosis, treatment, and care planning) and their corresponding rationales–explanations that justify the medical necessity of the decisions. However, Prior clinical NLP systems largely focus on extracting entities (for example, medications and follow ups) rather than higher-level *decisions* and their *supporting evidence*. This is mainly because the unstructured nature of text obscures the underlying clinical reasoning, making it difficult to computationally analyze medical decisions.[1] MedDecXtract-XAI is an interactive system that extracts medical decision spans from clinical narratives and provides sentence-grounded rationales for each extraction. The system implements a token classifier trained on the MedDec dataset[1] with a local LLM that generates short explanations linked to supporting sentences in the note. The interface enables rapid exploration of decisions, their medical types, and their corresponding rationales.

System Description: MedDecXtract-XAI implements a dual-model architecture: a specialized span extraction model and a rationale generation framework. **First**, the system uses a RoBERTa model to detect and classify ten categories of medical decisions (for example, treatment goals and defining problems) based on the DICTUM taxonomy.[2] This extractor was trained on the MedDec dataset,[1] which provides expert-annotated clinical narratives. **Second**, it incorporates a localized, efficient large language model (Qwen3.5 2B) to provide rationales for each clinical decision. This LLM generates a rationale grounded in specific support sentences from the clinical note when a user hovers over an extracted decision. The UI highlights the cited evidence span or spans used to generate the rationale. For example, the system can extract drug-related decisions (e.g., "*vanco discontinued*" and "*ceftriaxone replaced with zosyn*") and highlight the supporting rationale sentences from the same note (e.g., "*GNR identified in her blood*" and "*concern for possible pseudomonas*") (Figure 1). The interface supports inspection of extracted decision boundaries and their evidence-grounded rationales, enabling efficient review and refinement for research and potential clinical workflow studies.

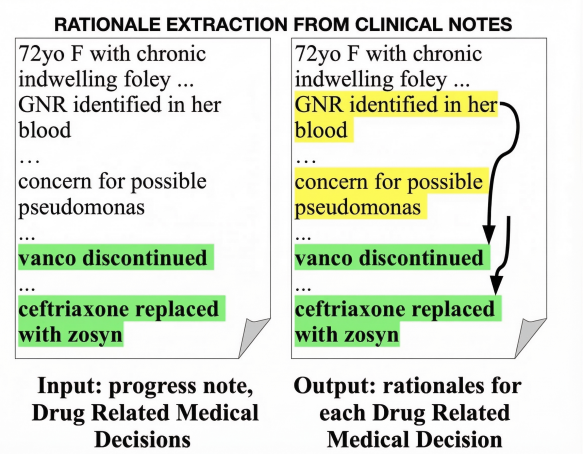

Figure 1: Example of grounded rationale extraction, showing drug-related decisions (green) and supporting rationale sentences (yellow) in a progress note.

We will demonstrate the following components: **Decision extraction:** A fine-tuned RoBERTa token-classification model that identifies and classifies decision spans into ten DICTUM categories as established in.[1] **Grounded rationales:** a local, lightweight LLM (configurable; default: Qwen 2B-class) that produces a brief rationale (*retrieved support sentences*) from the same note. **Human-in-the-loop utility:** inspection of extracted decisions and their rationales for efficient review and refinement for research and potential clinical workflow studies.

Deployment Status: The system is available as a web application: https://mohdelgaar-meddecxtract-2-0.hf.space. It is open source and designed to be lightweight and accessible for real-time extraction of clinical decisions and rationales. We welcome partnerships to evaluate usability and faithfulness of rationales in real-world workflows.

References

1. Elgaar M, Cheng J, Vakil N, Amiri H, Celi LA. MedDec: A dataset for extracting medical decisions from discharge summaries. In: Findings of the Association for Computational Linguistics ACL 2024; 2024. p. 16442-55.
2. Ofstad EH, Frich JC, Schei E, Frankel RM, Gulbrandsen P. What is a medical decision? a taxonomy based on physician statements in hospital encounters: a qualitative study. BMJ open. 2016;6(2):e010098.