

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

Mohamed Elgaar, BSc,¹ Hadi Amiri, PhD,^{1,2} Leo A. Celi, MD²
¹UMass Lowell, Lowell, MA, USA ²MIT, Cambridge, MA, USA

Introduction: Clinical narratives contain rich information about key decisions (diagnosis, treatment, and care planning) and corresponding rationales that justify their medical necessity. 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], [2]} MedDecXtract-XAI extracts medical decision spans and sentence-grounded rationales from clinical narratives. It combines a RoBERTa token classifier trained on MedDec,^[1] a dataset of 451 expert-annotated, deidentified MIMIC-III discharge summaries, with a local LLM for rationale extraction and explanation. The interface lets users inspect decisions and supporting sentences.

System Description: MedDecXtract-XAI has two components. **First**, the system uses a RoBERTa model to detect and classify ten DICTUM categories of medical decisions: contact, gathering information, defining problem, treatment goal, drug, therapeutic procedure, evaluating test result, deferment, advice/precaution, and legal/insurance.^[3] This extractor was trained on the MedDec dataset,^[1] and prior evaluation reports 79.9% token-level accuracy and 34.8 exact-match span F1, the strongest result among evaluated baselines. **Second**, a local LLM (Qwen3.5 2B) returns two linked outputs for each decision: rationale sentences from the note and a brief hover explanation of why they support the decision. For example, the system extracts drug-related decisions (e.g., “*vanco discontinued*” and “*ceftriaxone replaced with zosyn*”) and highlights supporting sentences from the same note (e.g., “*GNR identified in her blood*” and “*concern for possible pseudomonas*”) (Figure 1).

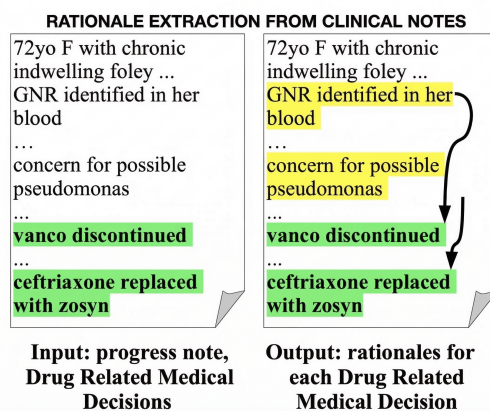


Figure 1: A drug-related decisions (green) and supporting rationale sentences (yellow) in a progress-note excerpt.

Decision extraction: a fine-tuned token-classification model that identifies decision spans into ten DICTUM categories.^[1] **Grounded rationales:** extraction of supporting sentences, with a local LLM providing a brief explanation of the sentence-decision link on hover. **Human-in-the-loop utility:** inspection of outputs for clinical workflow studies.

Deployment Status: The open source system is at <https://mohdelgaar-meddecxtract-2-0.hf.space> and applicable to real-time extraction of clinical decision rationale sentences. The public demo is intended for de-identified or synthetic text. We welcome partnerships to evaluate usability, and accuracy rates in real-world workflows.

Acknowledgment: This work was supported in part by a UMass Lowell seed grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding institution.

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] Elgaar M, Cheng J, Vakil N, Sadrolashrafi M, Mohtarami M, Wong A, et al. Overview of the Medical Decision Extraction, Analysis, and Classification Task (MedExACT) of BioNLP 2026. In: Proceedings of the 25th Workshop on Biomedical Language Processing (Shared Tasks). Association for Computational Linguistics; 2026.
- [3] 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.