SIIM CAIMI25
Conference on Artificial Intelligence in Medical Imaging
OCT 20-21 | UNIVERSITY OF CALIFORNIA, IRVINE

Reasoning-Based Large Language Model Study for Automated Drain Management in Interventional Radiology

Krishnaveni Parvataneni, Massachusetts General Hospital; Satvik Tripathi; Felix Dorfner; Christopher P. Bridge, PhD; Dania Daye, MD, PhD

Introduction/Background

Effective abscess drain management is critical for patient outcomes and care coordination in Interventional Radiology (IR). Current documentation methods rely heavily on manual chart review, which is a time-intensive and error-prone bottleneck. This study evaluates the use of Large Language Models (LLMs) to extract key procedural information from free-text IR procedural notes automatically, and introduces an agentic framework designed to emulate clinical reasoning by deciding whether to extract more data, recommend an action (such as drain removal or exchange) with rationale, or stop once a complete decision has been made.

Methods/Intervention

A large set of abscess drain-related procedural notes were collected from the EMR (n=1990), covering placement, removal, and exchange. The reports were processed using a reasoning-based loop using the Qwen1.5-72B-Chat-AWQ model, an open-source LLM. The model was prompted to simulate a clinical reasoning workflow, making iterative decisions at each step on whether to extract another data field, make a recommendation on drain removal and exchange, or end the evaluation as a result of sufficient context. Sixteen structured fields were targeted, such as drain type, purulence, output volume, and presence of a fistula. Model outputs were compared against expert-annotated ground truth using accuracy, precision, recall, F1 score, and Cohen's Kappa.

Results/Outcome

The model achieved 95.96% overall accuracy across extracted fields. For all non-binary fields, Cohen's Kappa exceeded 0.89, indicating near-perfect agreement with human annotations. Binary classification tasks yielded an average recall of 0.9720 and an average precision of 0.9181. In the decision tasks central to the agentic loop, Qwen predicted "Should Remove Drain" with 83% accuracy (Recall=0.8375, Precision=0.7917, F1=0.814) and "Should Exchange Drain" with 94% accuracy (Recall=1.000, Precision=0.9211, F1=0.9589). These results demonstrate that the model could dynamically adjust its reasoning path to emulate human-like clinical decision patterns.

Conclusion

An LLM reasoning-based approach can accurately extract structured data and recommend drain management actions from IR notes, offering a scalable solution to reduce manual workload and enhance consistency in clinical decision-making.

Statement of Impact

This study demonstrates a practical, privacy-preserving LLM framework that mimics clinical reasoning to streamline abscess drain management, improving efficiency and standardization in interventional radiology workflows.

Keywords
Large Language Models; EMR Data Structuring; Abscess Drains; Clinical Reasoning; Interventional Radiology