Large Language Models for Patient-Centered Care: A Proof of Concept Study for a Drain Management and Clinical Follow-Up Database in Interventional Radiology

Krishnaveni Parvataneni, Research Intern, Massachusetts General Hospital
Felix Dorfner; Meredith Preziosi, FNP-BC; Patrick Sutphin, MD; Christopher Bridge, PhD; Dania Daye, MD, PhD

Introduction

Effective management of abscess drains is essential for patient care and resource utilization in Interventional Radiology.

Hypothesis

Our study aims to create a drain management database using Large Language Models to extract key information from medical notes, reducing reliance on manual chart reviews for abscess drain management in Interventional Radiology.

Methods

In this IRB-approved, HIPAA-compliant study, we identified patients who underwent drain procedures in Interventional Radiology at a quaternary referral hospital. All procedural reports were collected from the EMR. We used the vLLM library to prompt one of three pre-trained publicly-available Large Language Models (LLMs), Qwen1.5-72B-Chat-AWQ, LLAMA-2-13b-chat-hf, and Mistral-7B-Instruct-v0.2, to extract 15 key data fields from the procedural notes using zero-shot prompting (prompting without examples). These fields were Number Drains, Drain Types, Output Volume, Material Quality, Drain Indication, Attending Operator, Drain Location, Is New Drain, Is Abscess Drain, Exchange/Reposition/Upsize, Injection, Removal, Purulent, Fistula, and Persistent Collection. Performance of this new model was compared to traditional NLP algorithms, including spaCy and Regular Expression (RegEx). Model performance was assessed using F1-Score for columns with binary classification, and Accuracy for columns with information extraction tasks.

Results

We identified a total of 1,990 reports that met our inclusion criteria from the period between October 2, 2022, and December 8, 2023. Each model's performance, including processing time and accuracy, was recorded: Qwen took 30 minutes, LLAMA 15 minutes, and Mistral 10 minutes to process 1,990 cases using 4 Nvidia A100 GPUs. Qwen had the highest accuracy (96.89%), outperforming LLAMA (85.10%) and Mistral (94.52%), and was chosen for its superior performance.
Overall, the utilization of LLMs significantly improved the reliability of data extraction from unstructured medical notes compared to traditional NLP methods such as RegEx and spaCy. The Qwen model achieved a notable accuracy of 96.89%, with an F1-Score of 0.9722, compared to 29.33% accuracy and an F1-Score of 0.7246 for the spaCy/RegEx model.

Conclusion

Our study demonstrates that using LLMs for automated data extraction in a clinical setting, particularly in IR abscess drain management, can enhance the efficiency and accuracy of information retrieval from unstructured medical notes. This approach has the potential to improve patient care and optimize resource utilization in Interventional Radiology.

Keywords

Artificial Intelligence/Machine Learning; Clinical Workflow & Productivity; Emerging Technologies; Quality Improvement & Quality Assurance