Analysis of Published Case Reports on Lymphoma Using Large Language Models

Experimental hematology/oncology

Fabio Dennstädt, Alin Handra, Nikola Cihoric

Center for Artificial Intelligence in Radiation Oncology (CAIRO), Inselspital, Bern University Hospital, University of Bern

INTRODUCTION

Lymphoma is a complex group of cancers with a wide range of clinical presentations.
While randomized clinical trials are the gold standard for evidence-based medicine, they often do not include the diverse and unique patient scenarios found in case reports. A vast repository of clinical information from these reports remains untapped. This study aimed to develop and validate an automated pipeline using large language models (LLMs) to perform a large-scale analysis of all lymphoma case reports published on PubMed.

METHODS

We conducted a comprehensive search of the PubMed database to gather all relevant case reports. A structured questionnaire with 51 questions was designed to extract key data points, including patient demographics, lymphoma subtype, diagnostic methods, disease location, treatments administered, and patient outcomes.
The LLM Rombos-LLM-V2.6-Qwen-14b, deployed with the data-element-extractor Python package, was used to systematically analyze the title and abstract of each publication and answer the predefined questions. To evaluate the LLM's performance a ground truth dataset of 298 manually labeled reports was used before proceeding with the large-scale quantitative analysis of all available case reports. A methodologicaloverview is shown in Figure 1.

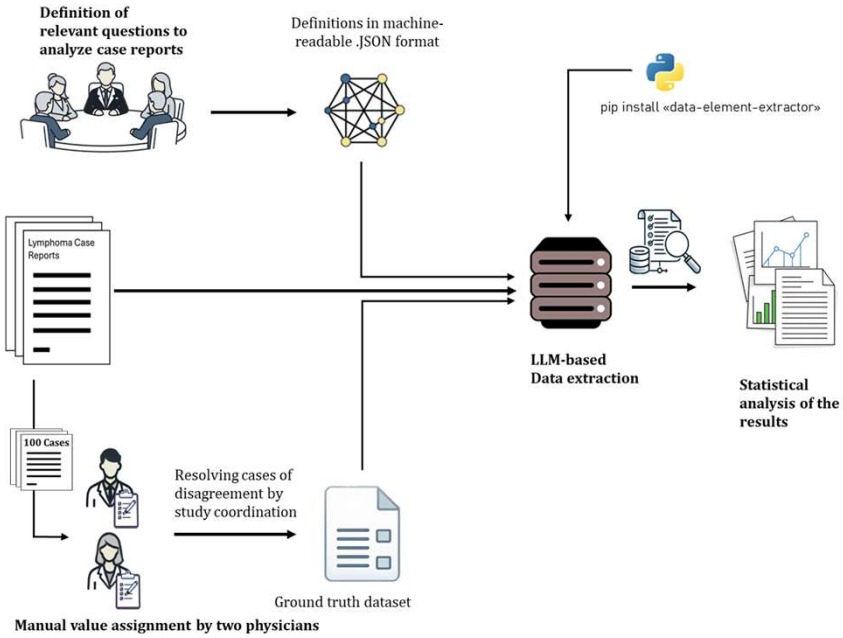

Figure 1: Schematic illustration of the study design.

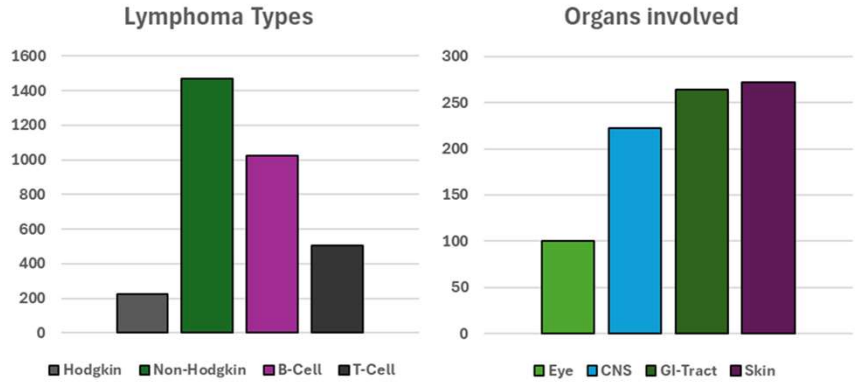

Figure 2: Number of case reports identified in the domains of Lymphoma Types and Organs involved (selected key categories; the total analysis included 51 questions.).

RESULTS

The search yielded 10,681 publications. Validation of the AI-based data extraction on a subset of 298 reports demonstrated a high overall accuracy of 96.1% and an F1-score of 80.1%. The AI pipeline identified 3,347 publications as individual patient case reports. Key findings from these reports include:
Demographics: 40.8% of patients were male, 30.4% female. The median age was 52 years, with 17.2% of patients being under 18.
Diagnosis: 11.1% of cases were Hodgkin lymphoma. PET-CT scans were reported in 3.5%, and bone marrow biopsies were done in 6.7%. Most frequently involved organs by lymphoma were the skin (8.1%), gastrointestinal tract (7.9%), and central nervous system (6.6%).
Therapy and Outcome: Chemotherapy was given in 20.0% of cases, radiotherapy in 8.9%, and surgery in 3.6%. The patient's death was reported in 16.2% of the analyzed case reports.

CONCLUSION

This is the first systematic analysis of all lymphoma case reports on PubMed using generative AI. Our results confirm that LLMs can accurately and efficiently extract a wide range of clinical data from case reports on a massive scale. This innovative method provides a powerful tool for creating dynamic, "living" case report registries. Such registries can offer detailed clinical evidence to deepen our understanding of the complex and varied landscape of lymphoma.

Presented at SOHC 2025 from 19 – 21 November 2025