IN-CONTEXT LEARNING ENABLES ACCURATE, LABEL-EFFICIENT CLASSIFICATION OF B-CELL LYMPHOMAS USING VISION-LANGUAGE MODELS

MOBINA SHRESTHA, MD Sentara Albemarle Medical Center, USACS

M Shrestha[1], B Mandal[2] and V Mandal[3]

1. Sentara Albemarle Medical Center, Elizabeth City, NC United States
2. Kansas State University, Manhattan, KS United States
3. University of Missouri-Columbia, United States

US Acute Care Solutions

Introduction

Accurate classification of B-cell lymphomas such as Chronic Lymphocytic Leukemia (CLL), Follicular Lymphoma (FL), and Mantle Cell Lymphoma (MCL) is vital for prognosis and treatment but difficult due to overlapping morphology. Even with ancillary tests, diagnosis outside specialized centers is challenging. Artificial Intelligence (AI) offers promise but requires large amounts of labeled datasets, in-context learning (ICL), a newer paradigm within AI however offers an alternative by enabling predictions from just a few labeled examples without retraining.

Aim

This study aims to evaluate the application of vision-language models (VLMs) with ICL for B-cell lymphoma classification in digital histopathology. We assess whether ICL can accurately distinguish CLL, FL, and MCL using only a few labeled examples at inference, without model retraining. The goal is to evaluate ICL as a fast hypothesis-testing approach that helps physicians validate diagnostic impressions in lymphoma cases, especially where expert pathology review is limited.

Objectives

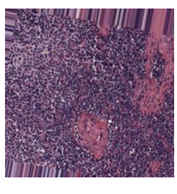
To evaluate the performance of four VLMs (GPT-4o, Paligemma, CLIP, and ALIGN) on H&E-stained images of three B-cell lymphoma subtypes: CLL, FL, and MCL. A dataset of 150 images (50 per subtype) was tested under 0-shot, 3-shot, 5-shot, and 10-shot ICL settings, with performance measured using weighted F1 score and 95% CI.

Schematic Workflow

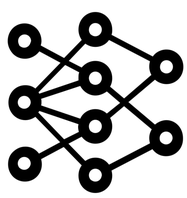
Prompt

You are a Hematopathology assistant and are evaluating an H&E-stained biopsy. Start by analyzing the overall architecture and cellular composition, then evaluate the overall pattern. Note any preserved follicles, mantle zones, or proliferation centers. Then assess the cytology, focusing on lymphocyte size, nuclear shape, and chromatin texture. Based on these findings, select the most appropriate diagnosis from the options below:

(a) Chronic Lymphocytic Leukemia (CLL)
(b) Follicular Lymphoma (FL)
(c) Mantle Cell Lymphoma (MCL)



Zero/
Few-shot Prompting



GPT-4o
Paligemma
CLIP
ALIGN

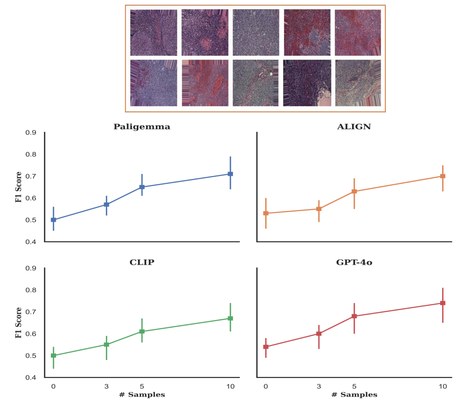
Sample GPT-4o Model Response

GT: Follicular Lymphoma (FL)

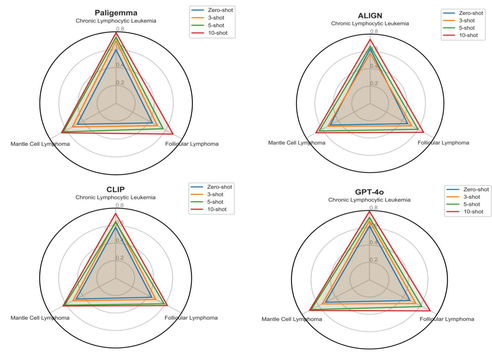
Predicted: Follicular Lymphoma (FL)

CORRECT

Results



F1 score trends for VLMs across increasing shot settings



Radar plots with F1 score distribution across lymphoma subtypes, CLL, FL, and MCL for VLMs: Paligemma, ALIGN, CLIP and GPT-4o.

Summary

ICL improved B-cell lymphoma classification across all vision-language models, with the largest performance gains between zero-shot and 5-shot settings. GPT-4o achieved the highest accuracy overall, while CLL was consistently well-classified and FL benefited most from few-shot prompting.

Discussion

- Overall trend: Performance improved with few-shot prompting; biggest jump from 0 → 5 shots, modest gains after.

- GPT-4o: F1 0.54 → 0.74 (95% CI: 0.65–0.81) from 0 to 10-shot (best overall).
- Paligemma: F1 0.50 → 0.71 (0.64–0.79), narrowing gap with GPT-4o at 10-shot.
- CLIP & ALIGN: Plateaued earlier, F1 ~0.67–0.70 at 10-shot.
- Subtype performance:
  - CLL: Strong across all models, F1 up to 0.79.
  - FL: Most improved with shots, e.g., GPT-4o 0.48 → 0.72.
  - MCL: Intermediate results, GPT-4o 0.71, ALIGN 0.68.

ICL boosts accuracy without retraining and can support hypothesis testing for physicians.

References

1. Patel, H., Shah, H., Patel, G., & Patel, A. (2024). Hematologic cancer diagnosis and classification using machine and deep learning: State-of-the-art techniques and emerging research directives. *Artificial Intelligence in Medicine*, 102883.
2. Steiner, D. F., MacDonald, R., Liu, Y., Truszkowski, P., Hipp, J. D., Gammage, C., ... & Stumpe, M. C. (2018). Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *The American journal of surgical pathology*, *42*(12), 1636-1646.
3. Mandal, B., Khanal, S., & Caragea, D. (2024, May). Contrastive learning for multimodal classification of crisis related tweets. In Proceedings of the ACM Web Conference 2024 (pp. 4555-4564).
4. Shrestha, M., Mandal, B., Mandal, V., & Shrestha, A. (2025). In-Context Learning for Label-Efficient Cancer Image Classification in Oncology. *Informatics in Medicine Unlocked*.

Read full paper here

SCAN ME!

SOHO2025