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

Speaking the Same Language: Utilizing Large Language Models to Automate Rad-Path Results Concordance Classification Across a Multi-System Healthcare Network

Othria Ahmed, MD, Northwell Health; Nicholas Mynarski, MD; Ritesh D. Patel, MD

Introduction/Background

With the rise of diagnostic imaging within a high-volume healthcare network, demand for streamlined integration of diagnostic data is crucial to clinical workflow and education in Radiology. Rad-Path Results is a fully automated system designed to aggregate pathology reports across multiple information systems, identify relevant imaging across our RIS system and email radiologists throughout training to encourage review via an easily accessible interface. Within the current program framework, radiologists can categorize provided matches as concordant or discordant. This feedback is utilized to train the Python-based algorithm to improve accuracy over time. With over 244,000 matches made since its deployment in 2023 and only 19,309 cases reviewed, manual categorization has become increasingly time-intensive which discourages its routine use.

Methods/Intervention

Further integrating Rad-Path Results in the clinical process required streamlining the current program to reduce radiologist workload. AI, specifically large language model (LLM), was explored to decrease the labor-intensive aspect of manual categorization which would simultaneously refine Rad-Path's ability to provide matches. A structured Gemini 2.5 prompt was created to analyze matched radiology and pathology reports and assign them into 3 categories: concordant, discordant, and indeterminant. While the original web interface utilized "irrelevant" to describe inappropriate matches, "indeterminant" allowed for manual interpretation of possibly ambiguous findings to decrease the rate of inaccurate categorization.

Results/Outcome

A pilot trial was conducted using 65 mixed concordant and discordant matched CT, US, and MR cases. Reports were fed into the prompted LLM and results were compared against manually assigned categories by radiologists. The LLM accurately matched 55 cases, 5 were misclassified, and 5 were flagged indeterminant (Figure 3). After manual review, cases were predominantly flagged due to inability to recognize nuanced differences in verbiage and unstructured reports.

Conclusion

These preliminary findings demonstrate the feasibility of utilizing LLM's to automate classification of Rad-Path matched reports with a high-degree of accuracy. While some limitations remain, this approach demonstrated promising potential. Continued refinement of the LLM could significantly reduce manual input, improve efficiency, and provide quality control of Rad-Path.

Statement of Impact

Integrating LLMs into Radiology-Pathology correlation offers a scalable solution to reduce manual review burden and increase accuracy in high-volume healthcare systems.

Figures 1-3

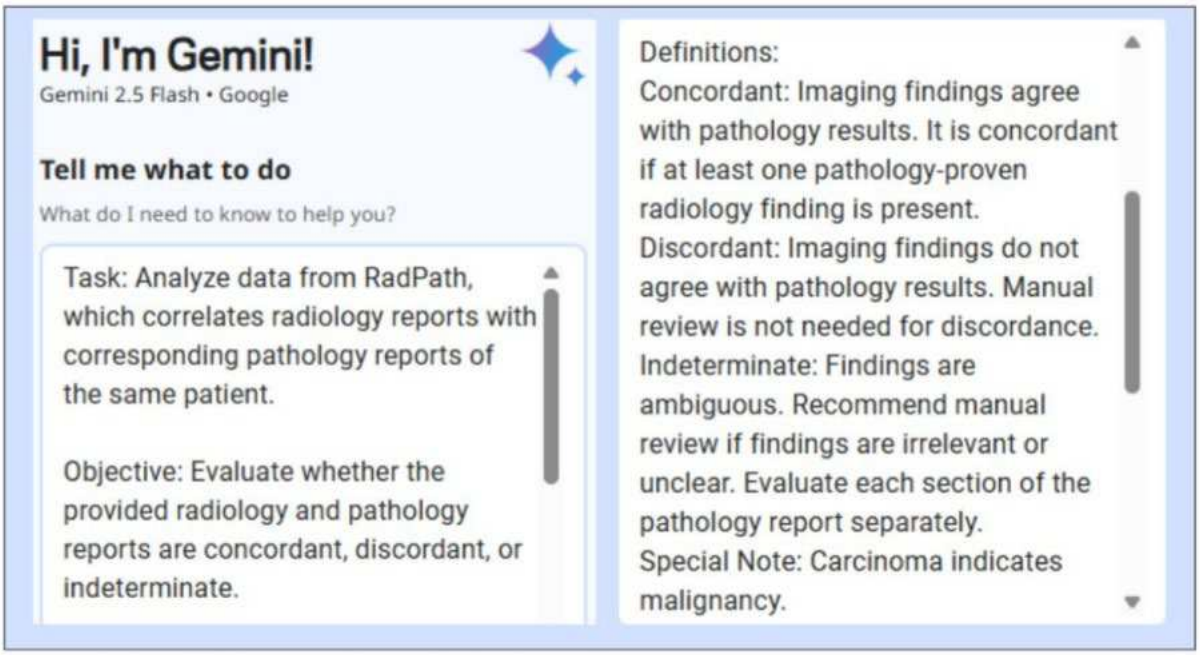

Figure 1. Rad-Path Concordance LLM Prompt
Using Gemini 2.5, a prompt was developed to take single radiology reports and their associated pathology report to evaluate concordance based on predetermined definitions. Output would be presented in a standardized summary to ensure readability and efficiency of responses.

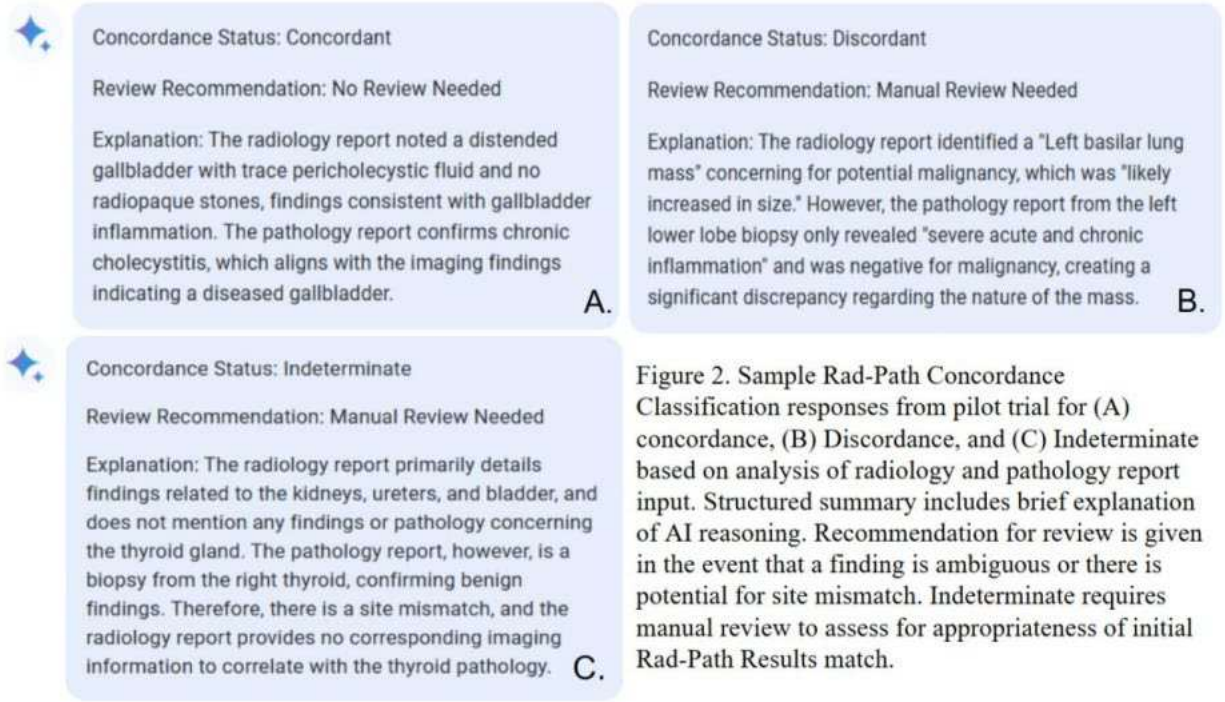

Figure 2. Sample Rad-Path Concordance Classification responses from pilot trial for (A) concordance, (B) Discordance, and (C) Indeterminate based on analysis of radiology and pathology report input. Structured summary includes brief explanation of AI reasoning. Recommendation for review is given in the event that a finding is ambiguous or there is potential for site mismatch. Indeterminate requires manual review to assess for appropriateness of initial Rad-Path Results match.

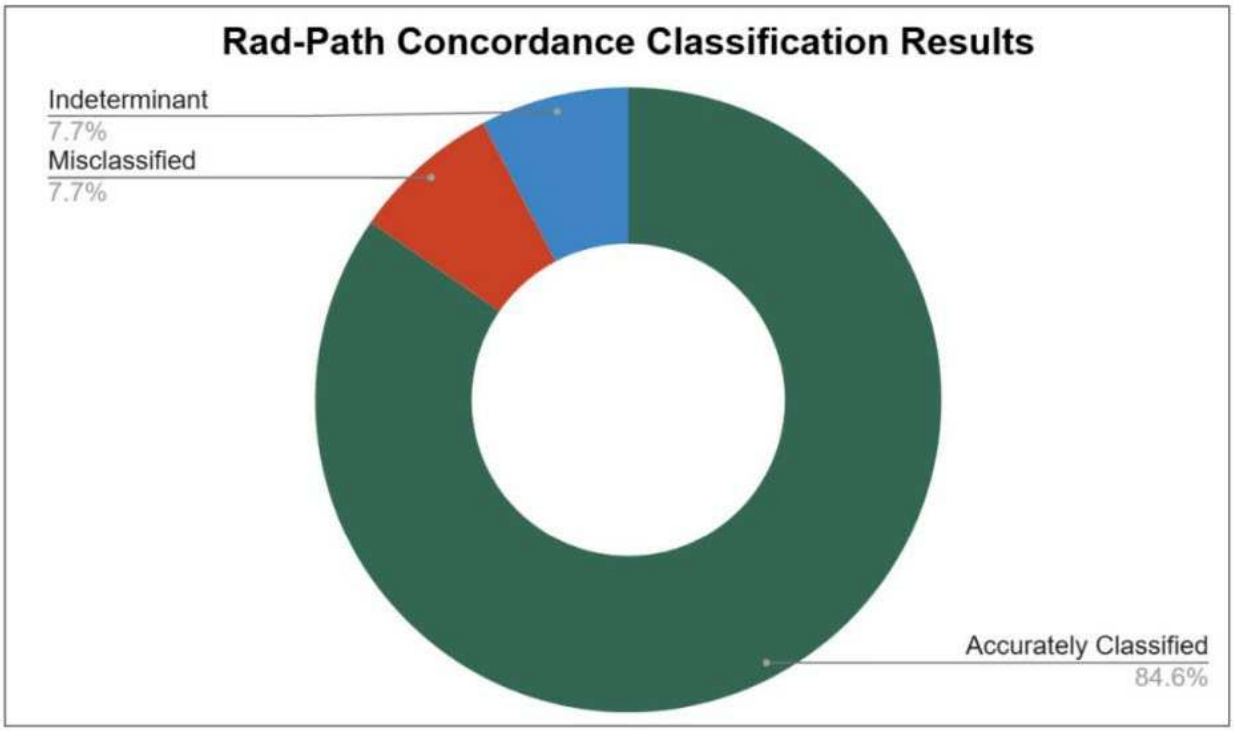

Figure 3. Rad-Path Classification Results
65 previous Rad-Path matched CT,US, and MR imaging cases were input into the prompted Gemini 2.5 model. Cases were mixed concordant and discordant as manually categorized by radiologists and then compared again LLM responses. Indeterminate cases were reviewed by radiologists for reasoning. Results showed high accuracy classification of cases across modalities.

Keywords
Diagnostic Automation; AI Workflow Integration; Large Language Model; Radiology; Pathology; Concordance Classification