Assessment of AI for Colorectal Cancer Imaging Interpretation: A Comparative Analysis Between AI and Radiologist Interpretation

Luis A. Hernández-Flores MD, Eric Acosta-Márquez, Iran Alonso, Daniel Aillaud MD, Karla D. Chávez, Diego Marines-Copado MD, Gabriel Arevalo MD
Houston Methodist Willowbrook

Background/Introduction

Artificial intelligence (AI) is revolutionizing medical diagnostics, with growing application in imaging interpretation. This study evaluates the correlation between AI and an experienced radiologist in interpreting CT scans of patients with colorectal cancer. The aim was to assess the accuracy of AI compared to trained human expertise, focusing on different sections of the colon.

Purpose/Objectives/Hypothesis

The objective is to evaluate the diagnostic accuracy of ChatGPT-4.0 in identifying colon cancer through imaging studies, specifically CT scans, and to compare these findings with diagnoses made by radiologists.

Methods

A 5-months retrospective analysis was conducted on 18 colorectal cancer patients. After excluding three cases due to oral contrast use, 15 cases were included for analysis. Each CT scan was reviewed by an experienced radiologist, and the results were directly compared to those generated by the AI system ChatGPT 4o. A Kappa test was performed to evaluate the correlation between the AI's and the radiologist's interpretations, focusing on key areas of the colon: ascending, transverse, splenic flexure, descending and sigmoid colon..

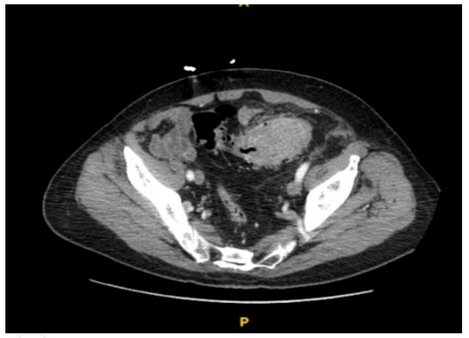
References

1. Liu W, Zhang B, Liu T, Jiang J, Liu Y. Artificial Intelligence in pancreatic image analysis: A Review. Sensors. 2024;24(14):4749. doi:10.3390/s24144749
2. Westwood M, Ramaekers B, Grimm S, et al. Software with artificial intelligence-derived algorithms for analysing CT brain scans in people with a suspected acute stroke: A systematic review and cost-effectiveness analysis. Health Technology Assessment. Published online March 2024:1-204. doi:10.3310/rdpa1487
3. Zhang Q, Huang Z, Jin Y, et al. Total-body PET/CT: A role of Artificial Intelligence? Seminars in Nuclear Medicine. Published online October 2024. doi:10.1053/j.semnuclmed.2024.09.002
4. Quanyang W, Yao H, Sicong W, et al. Artificial Intelligence in lung cancer screening: Detection, classification, prediction, and Prognosis. Cancer Medicine. 2024;13(7). doi:10.1002/cam4.7140
5. Lindgren Belal S, Frantz S, Minarik D, et al. Applications of artificial intelligence in PSMA PET/CT for Prostate Cancer Imaging. Seminars in Nuclear Medicine. 2024;54(1):141-149. doi:10.1053/j.semnuclmed.2023.06.001
6.

Result

Region	#	Kappa Value	*p*-value	Interpretation
All Cases (Total)	15	Varies	0.333	Mixed Agreement
Sigmoid Colon	6	0.857	< 0.001	Strong Agreement
Ascending Colon	4	0.815	0.001	Strong Agreement
Descending Colon	2	Non	N/A	No Correlation
Transverse Colon	2	Non	N/A	No Correlation
Splenic flexure	1	Non	N/A	No Correlation

Table 1: Correlation Human and AI interpretation





Diagnosis from the CT Images:

Image Analysis:

1. **Lesion Identification**:
   - The images display a well-defined mass in the region of the rectum or distal sigmoid colon, marked by the cross.
2. **Location**:
   - The lesion is located in the distal part of the colon, near the rectosigmoid junction.
3. **Potential Diagnosis**:
   - The mass is suggestive of a malignant lesion, likely adenocarcinoma of the rectum or

Figure 1. Example of AI interpretation of CT scan

Results/Implications

The level of agreement between AI and the radiologist varied depending on the location of the colorectal cancer. For the 6 cases of sigmoid colon cancer, the Kappa value was 0.857 (p = 0.001), indicating strong agreement. Similarly, for the 4 cases of ascending colon cancer, the Kappa value was 0.815 (p = 0.001), also showing strong agreement. However, for the 2 cases of descending colon cancer and 1 case of transverse colon cancer, there was no correlation between the AI's and the radiologist's interpretations. Additionally, the single case of splenic flexure cancer demonstrated no correlation as well. The combined analysis of all 15 cases yielded a Kappa of -0.132 (p = 0.333), indicating no statistically significant agreement across all cases between AI and radiologist interpretations.

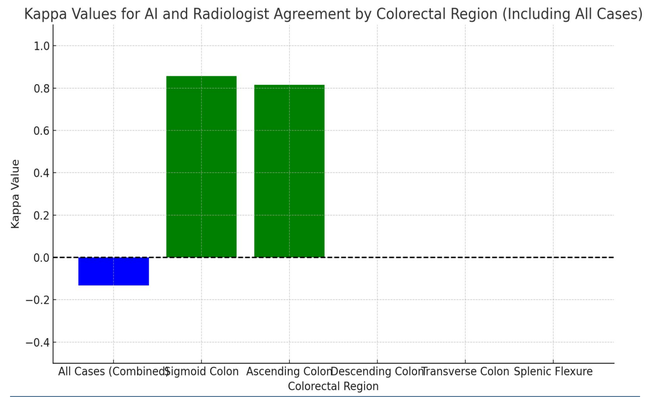


Figure 2: Visual representation of agreement

Conclusions

The AI showed perfect agreement with the radiologist for sigmoid and ascending colon cancer but failed to show consistent agreement in other regions of the colon. These findings suggest that while AI holds potential in imaging interpretation for specific cases, further refinement and research are necessary to ensure broader and reliable applicability across different regions of the colon.