

HIGH ACCURACY OF LARGE LANGUAGE MODELS IN DELIVERING GUIDELINE-CONCORDANT BARRETT'S ESOPHAGUS USING REAL-WORLD MULTILINGUAL DATA

Amir Klein

UEG Week Berlin 2025

Poster

Oesophagus

October 5, 2025

Large language models (OpenAI o3-mini-high and Google Gemini 2.0) demonstrated high accuracy (89-98%) in providing guideline-based recommendations for Barrett's esophagus management across three clinical decision tasks when compared to expert gastroenterologists.

KEY POINTS

- In a study of 107 Barrett's esophagus cases, two large language models were trained on 23 cases using ESGE guidelines and tested on 84 cases across three management tasks: pathology revision requests, next action recommendations (surveillance, resection, or eradication), and surveillance interval determination.
- Both models achieved high to near-perfect accuracy: o3-mini-high showed 95% accuracy for surveillance intervals, 97% for action recommendations, and 97% for pathology revision; Gemini 2.0 showed 89%, 98%, and 97% respectively, with no significant differences between models.
- Agreement between the two large language models was higher than agreement between expert gastroenterologists: models agreed on action 99% of the time versus 80% expert agreement, and on intervals 87% versus 69% expert agreement.
- Input data included unaltered esophagogastrosocopy and pathology reports in Hebrew and English; visible lesions were present in 15% of patients and dysplasia in 29%.
- The authors suggest large language models may support clinical decision-making and promote guideline adherence in Barrett's esophagus management, though this was a proof-of-concept study.

VIEW THIS CONTENT ON GUTFLIX

<https://gutflix.eu/search/d/8493d446-a814-11f0-b660-0242ac140006>



AI-generated summary

This summary was generated by an AI large language model based on the content transcript. It is for informational purposes only and should not be considered a substitute for clinical judgment. Always rely on your professional expertise and the full clinical context when making clinical decisions.