ISCTE IUL
REPOSITÓRIO
INSTITUTO UNIVERSITÁRIO DE LISBOA

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2026-10-27

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:
Mendes, F., Saraiva, M., Cordeiro, J., Mota, J., Martins, M., Cardoso, P....Santander, C. (2025). Artificial intelligence for detection and analysis of motility patterns in high-resolution anorectal manometry: A multicentric multidevice transatlantic study. Gastroenterology (pp. S?118?S?119). Elsevier. https://doi.org/10.1016/S0016-5085(25)01216-8

Further information on publisher's website:
10.1016/S0016-5085(25)01216-8

Publisher's copyright statement:
This is the peer reviewed version of the following article: Mendes, F., Saraiva, M., Cordeiro, J., Mota, J., Martins, M., Cardoso, P....Santander, C. (2025). Artificial intelligence for detection and analysis of motility patterns in high-resolution anorectal manometry: A multicentric multidevice transatlantic study. Gastroenterology (pp. S?118?S?119). Elsevier. https://doi.org/10.1016/S0016-5085(25)01216-8, which has been published in final form at https://dx.doi.org/10.1016/S0016-5085(25)01216-8. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

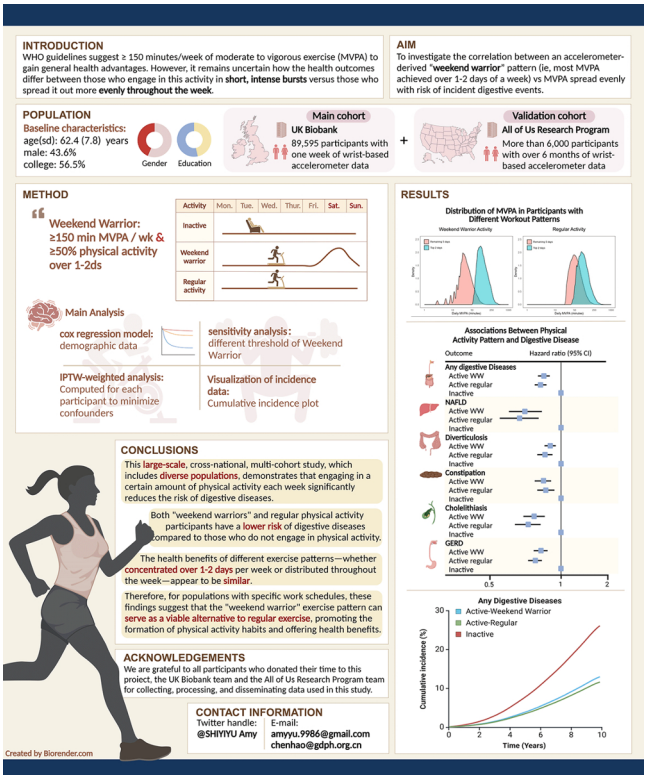
Use policy

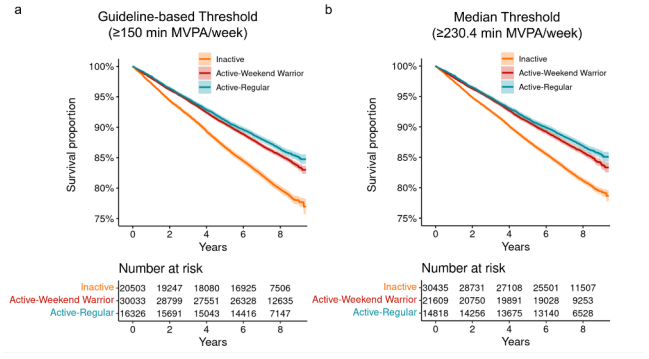
Creative Commons CC BY 4.0
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Serviços de Informação e Documentação, Instituto Universitário de Lisboa (ISCTE-IUL)
Av. das Forças Armadas, Edifício II, 1649-026 Lisboa Portugal
Phone: +(351) 217 903 024 | e-mail: administrador.repositorio@iscte-iul.pt
https://repositorio.iscte-iul.pt





KM PLOT

476

USE OF A LARGE LANGUAGE MODEL TO AID WITH CLINICAL TRIAL RECRUITMENT FOR PATIENTS WITH NON-ADVANCED COLORECTAL POLYPS

Jamie O. Yang, Alicia Maehara, Cleo K. Maehara, Alexandria Uy, Odet Kachikian, Gordon Guyant, Melissa Saul, Jeffrey Dueker, William Hsu, Folasade (Fola) P. May, Robert E. Schoen

Introduction The optimal surveillance colonoscopy guideline for patients with 1 or 2 non-advanced colorectal adenomas is uncertain. FORTE is a national, multicenter clinical trial randomizing participants with 1 or 2 non-advanced adenomas to surveillance colonoscopy at 5 and 10 years versus 10 years. However, clinical trial recruitment is often challenging, as manually identifying eligible participants requires review of large volumes of colonoscopy, pathology, and clinical reports—a process that is resource-intensive and time-consuming, which can hinder trial efficiency. To address this challenge, we developed and evaluated a customized large language model (LLM) that uses artificial intelligence to streamline the identification of potentially eligible trial participants. **Methods** We randomly selected 800 patients who underwent colonoscopy at University of Pittsburgh Medical Center, an academic center in the FORTE clinical trial. Colonoscopy and pathology reports were independently annotated by 3 reviewers and verified by 2 physicians to establish the gold-standard eligibility criteria for FORTE. Four patients were excluded due to missing pathology reports, resulting in 796 patients total (796 colonoscopy and 467 pathology reports). A custom LLM prompt (Llama 3.1 8b) was developed on a subset of 100 patients (Figure). The LLM was tasked with extracting key features from the reports such as polyp count, size, and histology, which was then processed by a customized algorithm to determine trial eligibility. The LLM was tested on the remaining 696 patients to characterize its performance, measured as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy for identifying FORTE-eligible patients. **Results** The LLM demonstrated a sensitivity of 90.4%, specificity of 94.2%, PPV of 64.7%, NPV of 98.8%, and overall accuracy of 93.8% in identifying FORTE-eligible patients (Table). Previously, a regular expression search for "adenomas" in pathology reports was tried as a filter for clinical trial recruitment, with a PPV of 30% for trial eligibility. Transition to the LLM approach improved the PPV to 63.7% (a 116% improvement), which would significantly reduce the number of manual chart reviews needed for clinical trial recruitment by the trial coordinator. **Conclusion** We successfully developed a customized LLM that demonstrated high sensitivity, specificity, and improved PPV in identifying patients eligible for the FORTE clinical trial. Pre-trained LLMs can be rapidly adapted and deployed across varied datasets, offering greater flexibility and scalability. Future work will validate this approach at additional FORTE study sites. By automating identification of patient eligibility, LLMs have the potential to transform clinical trial recruitment, saving time and resources while accelerating trial timelines.

Figure: Large Language Model Training and Validation Data Flow

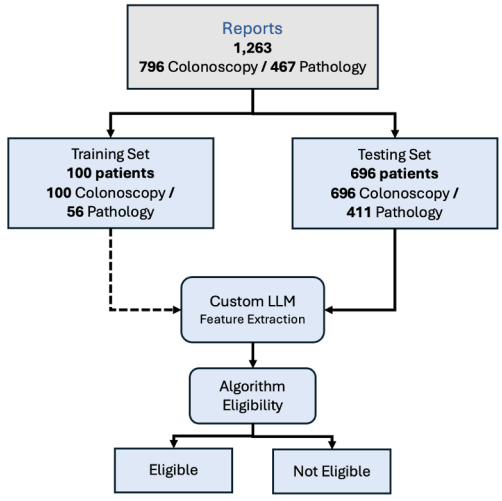


Table: Confusion Matrix Comparing Large Language Model to Manual Chart Review in Identifying Eligible FORTE Clinical Trial Patients

		Manual Chart Review	
		FORTE Eligible Patients	FORTE Non-eligible Patients
Large Language Model	FORTE Eligible Patients	66	36
	FORTE Non-eligible Patients	7	587
Test Parameters			
Sensitivity		90.4%	
Specificity		94.2%	
Positive Predictive Value		64.7%	
Negative Predictive Value		98.8%	
Accuracy		93.8%	

477

ARTIFICIAL INTELLIGENCE FOR DETECTION AND ANALYSIS OF MOTILITY PATTERNS IN HIGH-RESOLUTION ANORECTAL MANOMETRY: A MULTICENTRIC MULTIDEVICE TRANSATLANTIC STUDY

Francisco Mendes, Miguel M. Saraiva, João Rala Cordeiro, Joana Mota, Miguel Martins, Pedro Cardoso, Antonio Miguel M. Pinto da Costa, Ismael e. Martinez, Virginia Matallana Royo, Benjamin Niland, Jack Di Palma, Guilherme Macedo, Cecilio Santander

Introduction Anorectal functional disorders are common in clinical practice, with high-resolution anorectal manometry (HR-ARM) being the gold standard for their evaluation. However, HR-ARM is limited by its restricted accessibility and the complexity of data analysis. The London Protocol and Classification were developed to standardize the classification of anorectal motility patterns. This proof-of-concept study aims to develop and validate an artificial intelligence (AI) model for the identification and differentiation of disorders of tone and contractility in HR-ARM. **Methods** A dataset of 960 HR-ARM exams from two device manufacturers and three tertiary centers in Spain and the United States of America was utilized for the development and testing of multiple machine learning (ML) algorithms. The classification of anorectal motility disorders followed the London Classification. The exams were stratified into training and testing datasets using an 80%/20% patient split design based on label frequency. The models' prediction was compared to a consensus-based classification by two HR-ARM experts (constituting the gold standard for HR-ARM analysis). The testing dataset was used to evaluate model performance through metrics as accuracy and the area under the receiver-operating characteristic curve (AUC-ROC). Model metrics represent the average of 10 independent evaluations. **Results** After fine-tuning, the GradientBoostingClassifier demonstrated the best overall performance, identifying disorders of anal tone and contractility with an accuracy of 85.6% and an AUC-ROC of 0.910. The LGBMClassifier detected anal normotension with hypocontractility with 90.5% accuracy and an AUC-ROC of 0.91. AdaBoostClassifier identified anal hypotension with hypocontractility with 94.9% accuracy and an AUC-ROC of 0.944. Anal hypotension with normal contractility was detected by the LGBMClassifier with 94.1% accuracy and an AUC-ROC of 0.969. Finally, anal hypertension was identified by the GradientBoostClassifier with 96.9% accuracy and an AUC-ROC of 0.909. **Discussion** Various ML models accurately identified and differentiated disorders of anal tone and contractility in HR-ARM studies. The strong performance metrics observed across data from two device manufacturers address the challenge of interoperability, while the inclusion of data from two continents highlights the importance of mitigating

demographic bias. AI models in HR-ARM have the potential to optimize exam accessibility and accuracy, significantly enhancing the management of patients with functional anorectal disorders.

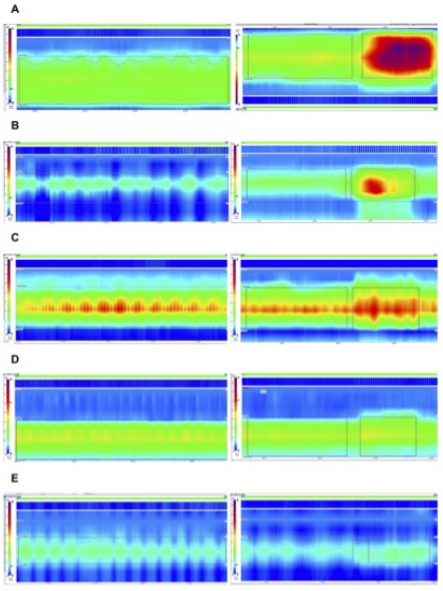


Figure 1 – Evaluation of the resting pressure and squeeze maneuver, with examples of a normal exam (A), anal hypotension with normal contractility (B), anal hypertension (C), anal normotension with hypocontractility (D) and combined anal hypotension and hypocontractility (E).

478

MACHINE-LEARNING BASED MODEL ACCURATELY IDENTIFIES THE PROBABILITY OF CHOLEDOCHOLITHIASIS AND PROVIDES A PERSONALIZED MANAGEMENT PLAN

Arpita N. Jajoo, Babak Moghadas, Jianing Li, Mohammed Rifat Shaik, Nirav Agrawal, Meher Gujral, Melody Tu, Navdeep Singh, Viswanadh Polana, Aditya Chandrashekar, Abhinav Jha, Pranay Marlecha, Sweta Sahu, Sukhmanpreet Sandhu, Grace Rozycki, Alistair Kent, Ivy Mannoh, Mouen A. Khashab, Brian Caffo, Venkata S. Akshintala

Background: The suboptimal accuracy of guidelines for predicting choledocholithiasis (CDL) has led to uncertainty regarding appropriate diagnostic or therapeutic interventions and overutilization of endoscopic retrograde cholangiopancreatography (ERCP). Improved predictive models may help triage patients to modalities associated with lower procedural risk and decreased healthcare expenditure. **Methods:** A prediction model was created utilizing a gradient boosting machine algorithm to quantify the probability of CDL based on clinical presenting features. The model was trained using a cohort of patients presenting to a tertiary care hospital and affiliated community centers with symptoms suggestive of choledocholithiasis. The data split for testing-training was 20-80. Model performance was evaluated by testing accuracy, 10-fold cross-validation, and area under the receiver operating characteristic curve (AUC). A nearest-neighbor imputation approach was implemented to address missing values in the dataset, with an added indicator column to denote whether the original value was missing. A second decision-making tool utilized the probability of CDL to suggest next steps in clinical management, with low risk patients triaged to cholecystectomy with intraoperative cholangiogram (CCY + IOC), intermediate-low risk patients triaged to magnetic resonance cholangiopancreatography (MRCP), intermediate-high risk patients triaged to endoscopic ultrasound (EUS), and high risk patients triaged to ERCP. **Results:** 2342 patients (mean age 56.7 years; 61% female) were included in the model and 46% had CDL. Guided by prior literature and clinical practice, variables most predictive of CDL included the degree of total bilirubin and other liver enzyme elevations, imaging findings, and patient demographics. The model demonstrated 86% accuracy and 85% precision in distinguishing CDL risk, with an AUC of 95%. The decision-making tree suggested MRCP for 14% of patients, ERCP for 36% of patients, EUS for 12% and CCY + IOC for 37% patients. This approach missed only 0.5% of common bile duct (CBD) stones (1/175 test set patients) by suggesting CCY + IOC. It assigned only 3% (12/469) patients without CBD stones to ERCP, compared to 13.8% using ASGE 2019 guidelines (Fig 2). **Conclusions:**These machine-learning models offer an individualized risk assessment for choledocholithiasis and suggested next steps in management based on clinical parameters available at initial presentation. Risk stratification enables low-risk patients to undergo cholecystectomy without further imaging, while reserving advanced imaging (e.g., MRCP) for intermediate-risk patients and prioritizing high-risk patients for ERCP. By improving upon the accuracy of existing guidelines, this model has the potential to reduce procedural morbidity and promote cost-effective, high-value care for patients with biliary disease.

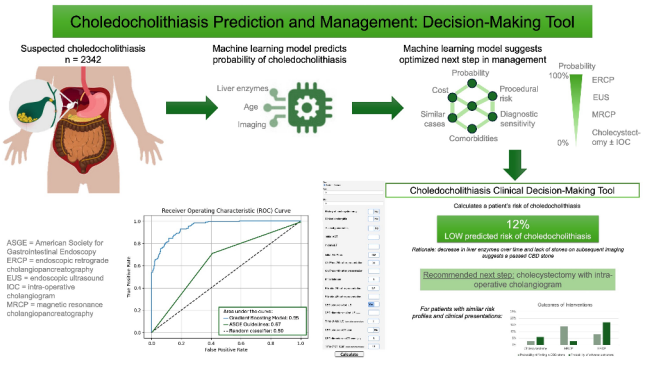


Figure 1. Graphical abstract of a decision making tool that uses machine learning algorithms to predict probability of choledocholithiasis and provide management recommendations

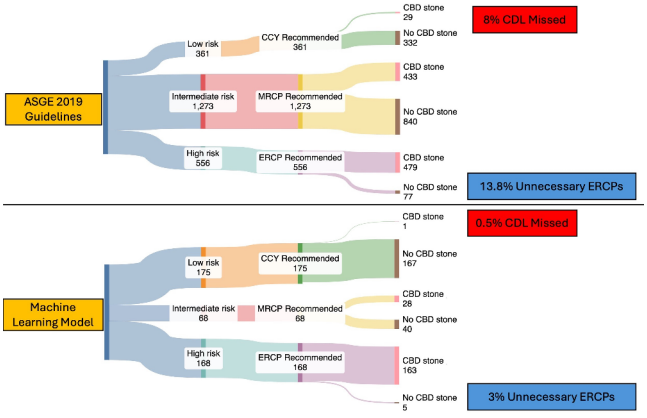


Figure 2. Management recommendations and outcomes of suggested interventions in patients with suspected choledocholithiasis using the machine learning model and ASGE 2019 guidelines. Machine learning model performance was evaluated for a test set of patients. ASGE 2019 guidelines were evaluated for all cohort patients with sufficient data. Inappropriate interventions displayed here include patients with choledocholithiasis who were not recommended for biliary evaluation (CCY instead of MRCP or ERCP) and patients without choledocholithiasis who were recommended for biliary tree evaluation (MRCP or ERCP instead of CCY). ASGE = American society of gastrointestinal endoscopy, CBD = common bile duct, CCY + IOC = cholecystectomy and intraoperative cholangiogram, CDL = choledocholithiasis, ERCP = endoscopic retrograde cholangiopancreatography, MRCP = magnetic resonance .

479

THE CEILING EFFECT: AN INVERSE CORRELATION BETWEEN ARTIFICIAL INTELLIGENCE FOR ADENOMA MISS RATE AND BASELINE PROVIDERS' ADENOMA DETECTION RATE

Tessa Herman, Lisa A. Smith, Amy Gravely, Brian J. Hanson, Francine M. Hawkinson, Michael B. Wallace, Jason A. Dominitz, Daniela Guerrero Vinsard

Background: Artificial intelligence (AI)-assisted colonoscopy with computer-aided detection (CADe) technology aims to improve colonoscopy quality metrics such as adenoma detection rate (ADR) and adenoma miss rate (AMR). The impact of AI on AMR depending on the baseline ADR of the endoscopist is less known. The aim of this study is to evaluate how endoscopist baseline characteristics impact shifts of performance when using AI as measured by right colon (RC) AMR. **Methods**: Prospective, single-center study evaluating the impact of AI on the RC-AMR. Patients undergoing colonoscopies for screening, surveillance, or positive fecal immunochemical test indications (April-August 2024) underwent two inspections of the RC. Patients were randomly assigned to 1st RC inspection with AI versus 2nd RC inspection with AI. The primary endpoint was RC-AMR relative to endoscopist baseline characteristics: baseline ADR, years of experience (junior: <5 years, senior: ≥5 years), and if trainees were involved in the case. Baseline ADR for each endoscopist was obtained from colonoscopies performed between February 2023-2024 for all indications. One endoscopist was excluded from baseline ADR calculation due to insufficient scope volume during that period. Endoscopists were divided into 2 cohorts: "lower baseline" ADR ≤65% and "higher baseline" ADR >65%. Ten endoscopists were included in the study: 7 gastroenterologists and 3 surgeons. **Results**: Average baseline ADR was 66.3% (SD 5.6%). We evaluated a total of 482 colonoscopies: 231 with AI used on the 1st pass and 251 with AI on the 2nd pass. Providers with a "lower baseline" ADR had a significant reduction in RC-AMR of 34.3% when using AI first (15.7% AI in 1st pass vs 50% AI in the 2nd pass respectively; OR 5.38, 95% CI 2.1-13.7, p<0.001). RC-AMR was not statistically different in providers with a "higher baseline" ADR (p=0.58; *Table 1*). RC-AMR in the junior endoscopist cohort was 11.5% lower when AI was used on the 1st pass (18% vs 29.5%, respectively, OR 1.90, 95% CI 1.01-3.59, p=0.04). Otherwise, there were no statistically significant differences seen in RC-AMR with AI used on 1st versus 2nd pass for senior endoscopists or if trainees were involved (*Figure 1*). **Conclusion**: Our study found that AI particularly decreases the RC-AMR for endoscopists with a baseline ADR ≤65% whereas higher-performing endoscopists