



Annual

**Advances and Innovations in Endoscopic Oncology
and Multidisciplinary Gastrointestinal Cancer Care**

How AI is Revolutionizing Endoscopic Cancer Detection Today, and in the Future

Michael B. Wallace, MD, MPH

Fred C. Andersen Professor of Medicine

Division of Gastroenterology and Hepatology

Mayo Clinic Comprehensive Cancer Center



Disclosures

- Grant/Research Support from Boston Scientific, Cosmo Intelligent Medical Devices, FujiFilm, Medtronic, and Olympus.
- Consultant for Boston Scientific, CDX Diagnostics, ColoMAIA, Inc., Cosmo Pharmaceuticals,, Endiatix, FujiFilm, Medtronic, and Microtech Endoscopy.

This presentation and/or comments will be free of any bias toward or promotion of the above referenced companies or their product(s) and/or other business interests.

This presentation and/or comments will provide a balanced, non-promotional, and evidence-based approach to all diagnostic, therapeutic and/or research related content.

This presentation has been peer-reviewed and no conflicts were noted.

Tools to Build Intelligent Systems:

Artificial Intelligence: computer systems that are able to perform tasks that normally require human intelligence, i.e visual perception, speech recognition, decision-making...

Machine learning: an application of AI that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

Deep learning: a subset of machine learning that relies on networks capable of learning unsupervised from data that is unstructured or unlabeled,

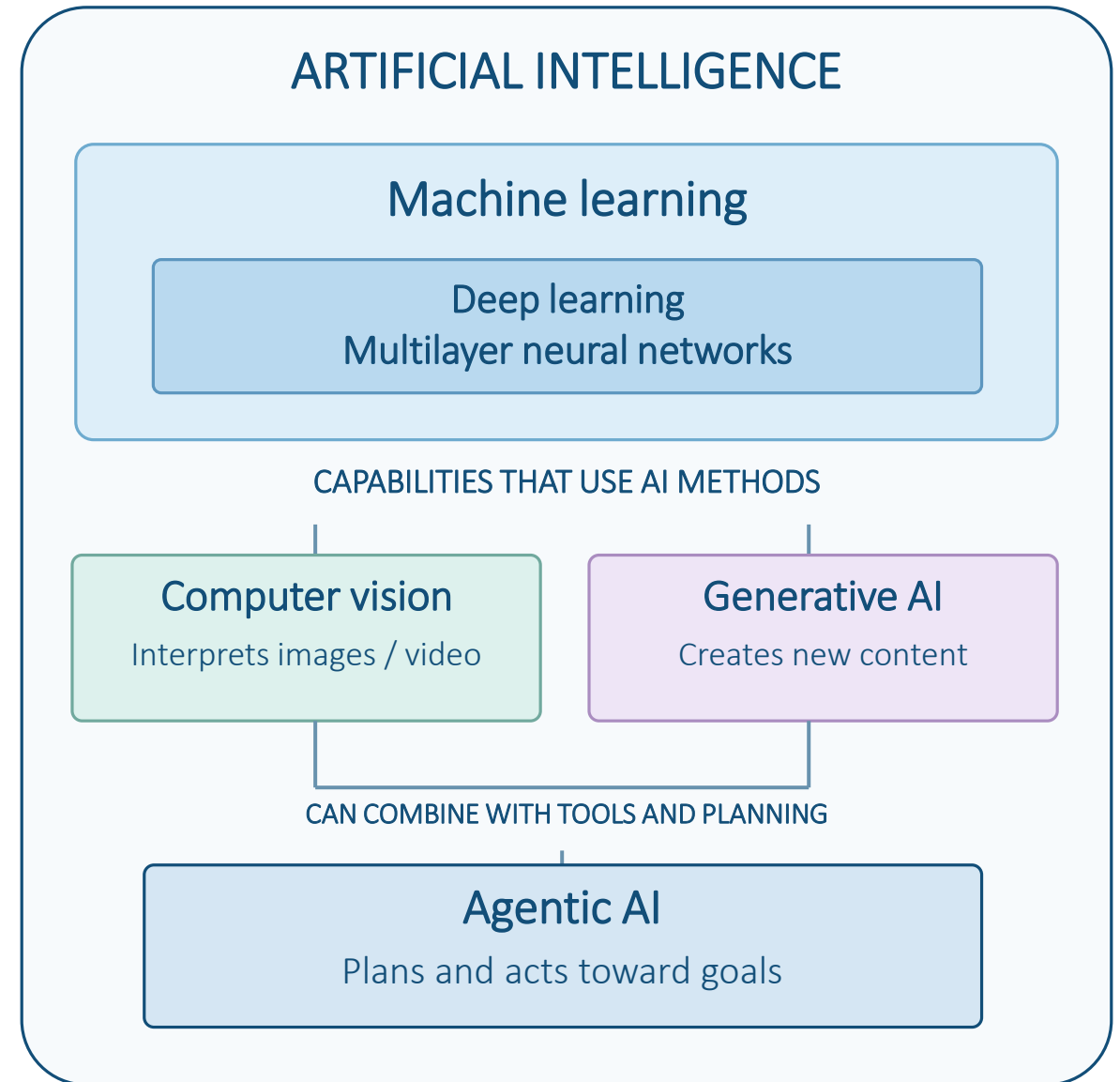
Applications of AI in Practice:

Computer vision: technology that allows computers to “see” and interpret visual content (photos, video..)

Generative AI: Algorithms that use existing data (text, images, video) to generate new text, images, video based on prompts

Agentic AI: Systems that possess the ability to autonomously make decisions, take actions, and adapt in real time based on changing data and environments—without requiring constant human oversight

Terminology in AI



Conceptual map: capabilities overlap; links show use, not subsets.

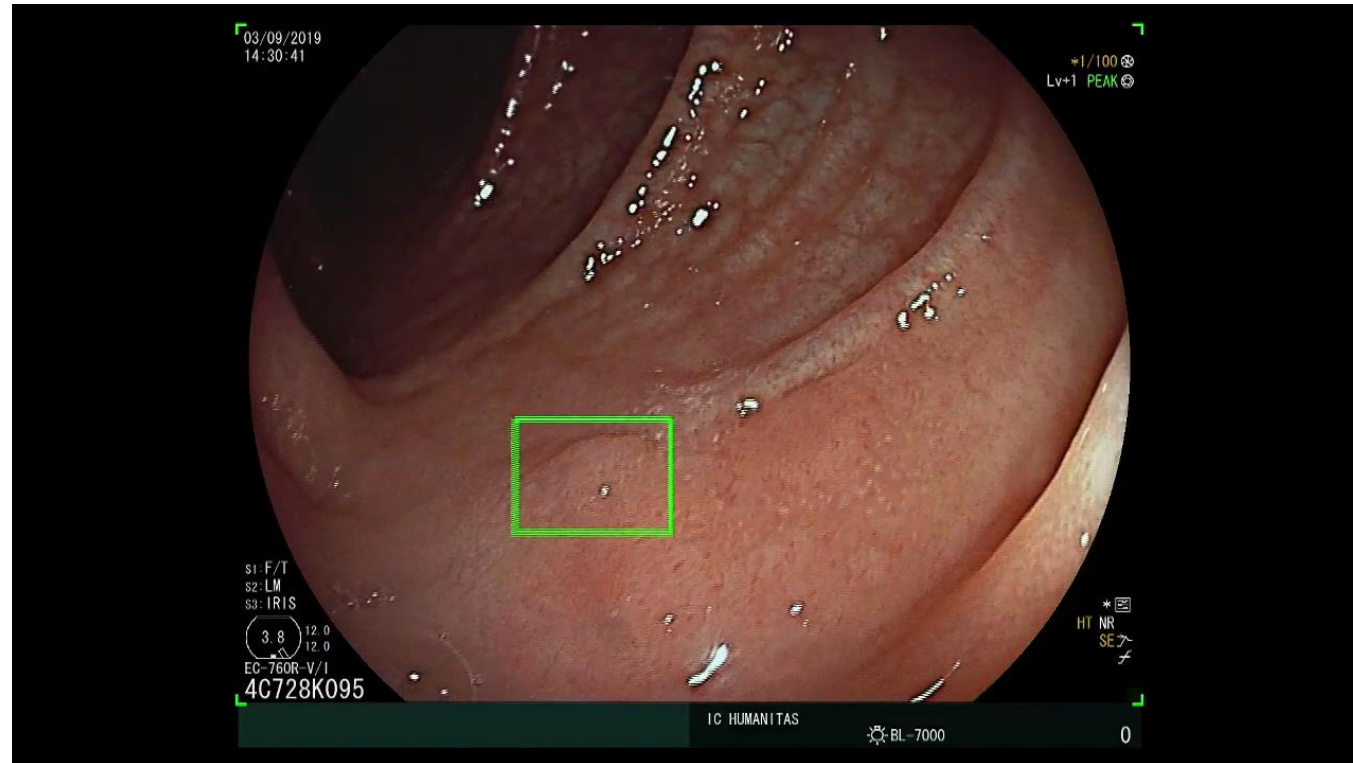
Types of AI and GI Applications

- Computer Vision/Computer Aided Detection (CAD)
 - Endoscopic Lesion Detection (CAdE)/Classification (CAdx)
 - Colon polyps, Barrett's dysplasia, IBD classification
 - Radiographical Image Detection/Classification
 - Quality metrics (bowel prep scores), Withdrawal time
- Generative AI
 - Create common procedural documents (e.g. bowel prep instructions, standard operating procedures) for practice management
 - AI-Scribes, Chatbot responses to common in-box questions
 - Auto-report (endoscopy) generation: Visual Language Models
- Ambient Clinical Intelligence
 - Systems that passively monitor the environment (e.g. doctor-patient conversation, procedure video, wearables) and create documents/reports/analysis

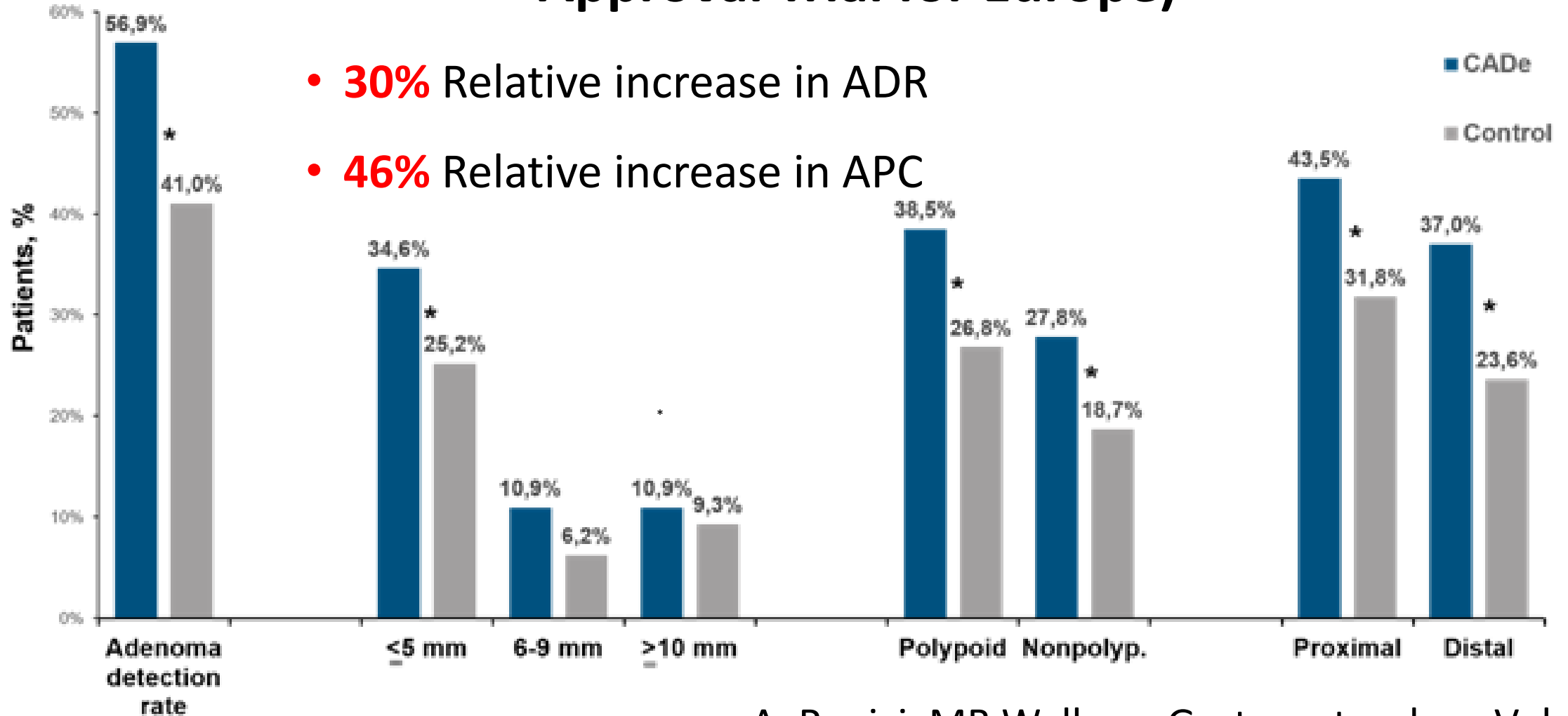
Improving Colon Cancer Prevention Through Better Colonoscopy

- Colonoscopy, with detection or pre-invasive polyps is the most effective method of colorectal cancer (CRC) prevention
 - Limited by “missed” polyps; flat, subtle, behind folds/flexures
 - Can we train AI computer vision to improve detection and reduce CRC?

AI Detection of Flat/Subtle Adenomas



First in Human Randomized Controlled Trial (CE-Mark Approval Trial for Europe)



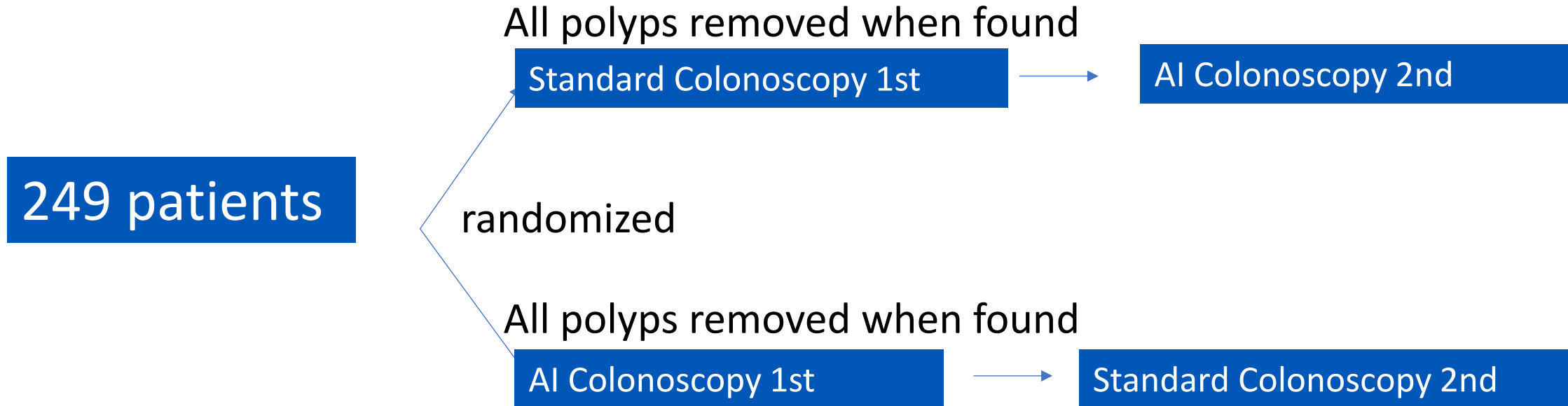
A. Repici, MB Wallace; Gastroenterology Volume 159, 2020, Pages 512-520

ARTIFICIAL INTELLIGENCE

Impact of Artificial Intelligence on Miss Rate of Colorectal Neoplasia



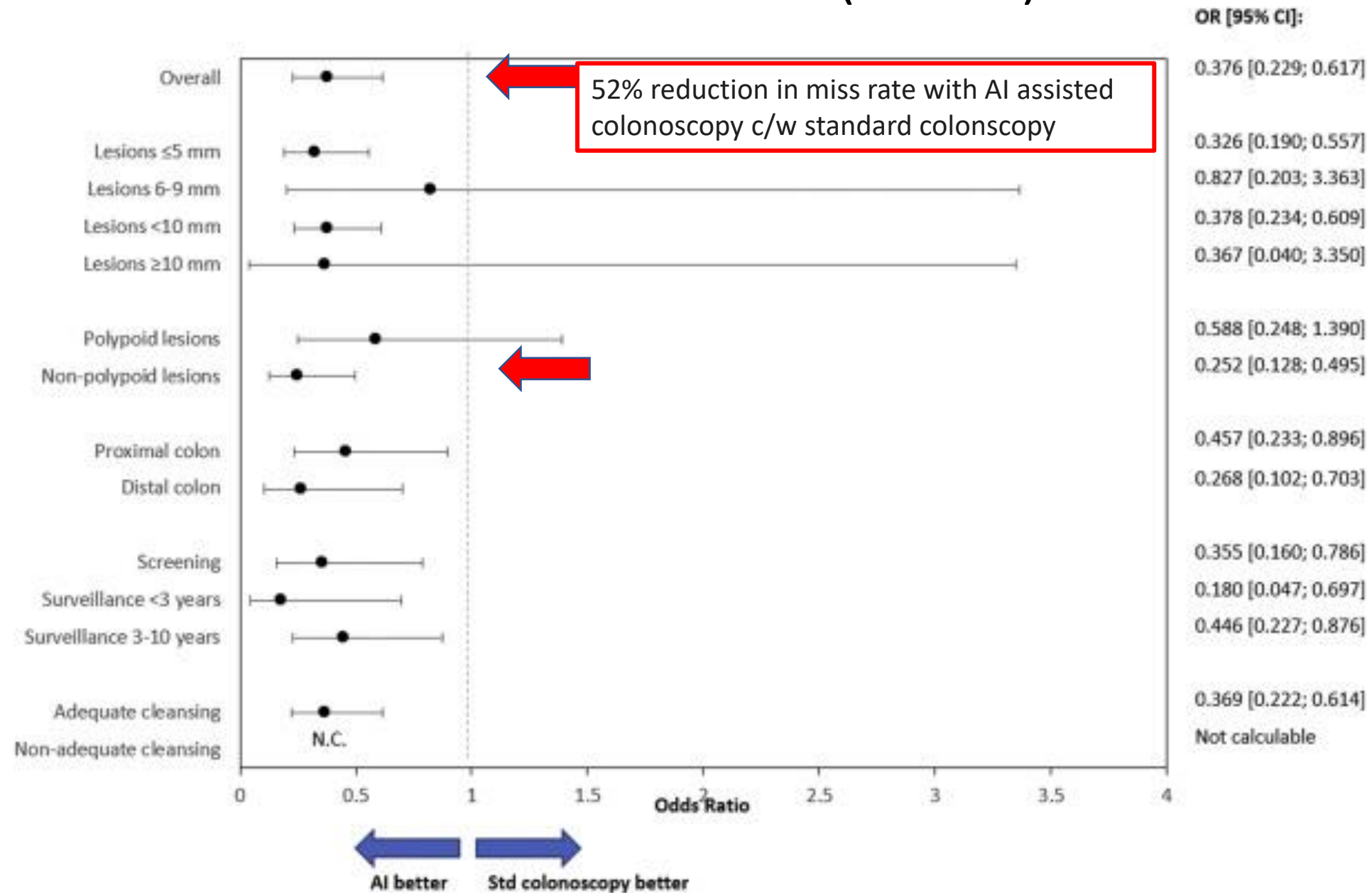
Michael B. Wallace,^{1,2} Prateek Sharma,³ Pradeep Bhandari,⁴ James East,⁵ Giulio Antonelli,^{6,7,8} Roberto Lorenzetti,⁶ Micheal Vieth,⁹ Ilaria Speranza,¹⁰ Marco Spadaccini,⁶ Madhav Desai,⁴ Frank J. Lukens,¹ Genci Babameto,¹¹ Daisy Batista,¹¹ Davinder Singh,¹¹ William Palmer,¹ Francisco Ramirez,¹² Rebecca Palmer,⁵ Tisha Lunsford,¹² Kevin Ruff,¹² Elizabeth Bird-Liebermann,⁵ Victor Ciofoaia,¹¹ Sophie Arndtz,⁴ David Cangemi,¹ Kirsty Puddick,⁴ Gregory Derfus,¹³ Amitpal S. Johal,¹⁴ Mohammed Barawi,¹⁵ Luigi Longo,¹⁶ Luigi Moro,¹⁶ Alessandro Repici,^{17,18} and Cesare Hassan^{17,18}



Primary Outcome:

Miss rate = % adenomas detected only on 2nd exam

Adenoma Miss Rate (AMR)



Meta-Analysis of AI CAdE Trials for ADR (now > 45 RCTs !)

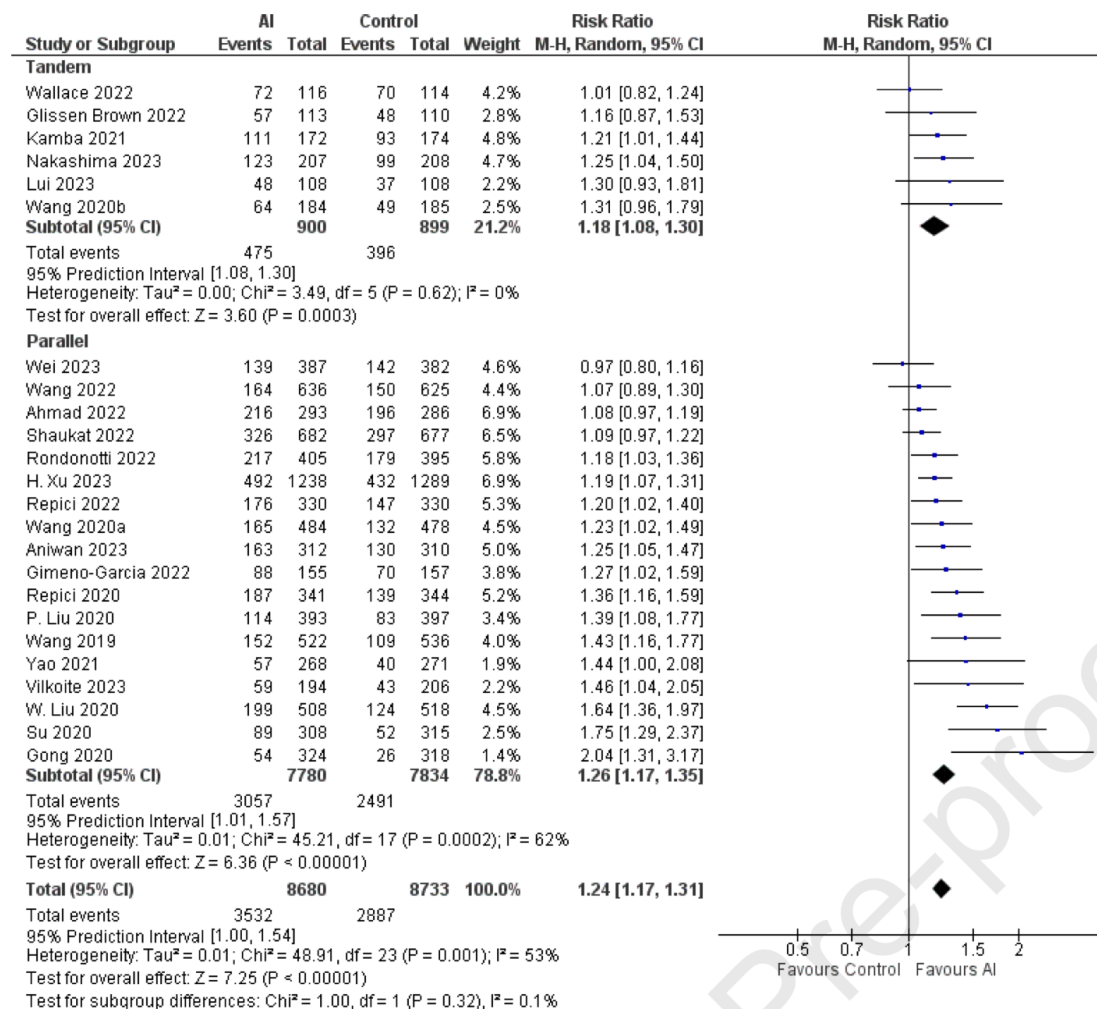
Impact of study design on adenoma detection in the evaluation of artificial intelligence-aided colonoscopy: a systematic review and meta-analysis

Michelle CM. Lee, MD, MEng, Colleen H. Parker, MD, MSc, FRCPC, Louis WC. Liu, MD, PhD, FRCPC, Armin Farahvash, MD, MEng, Thurarshen Jeyalingam, MD, MSc, FRCPC

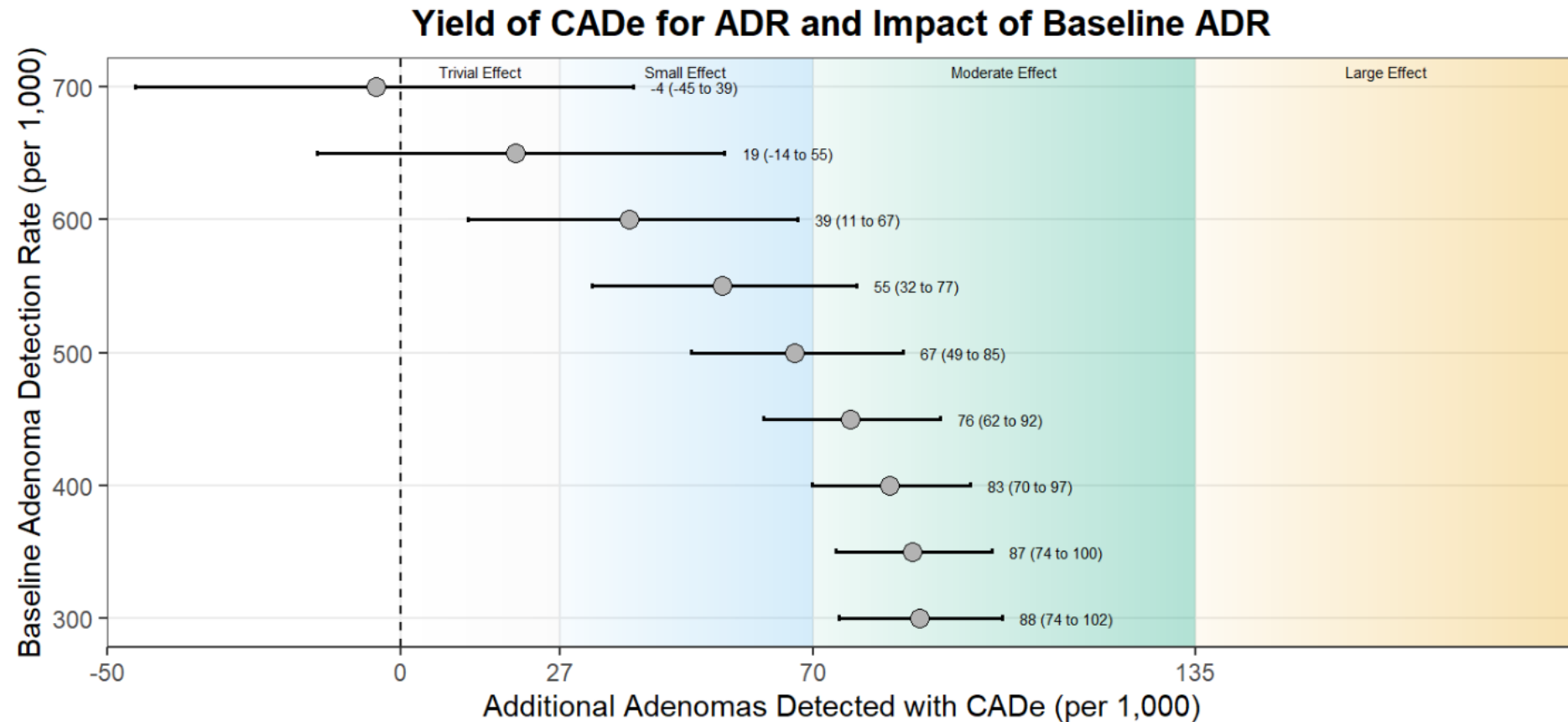
PII: S0016-5107(24)00040-3

DOI: <https://doi.org/10.1016/j.gie.2024.01.021>

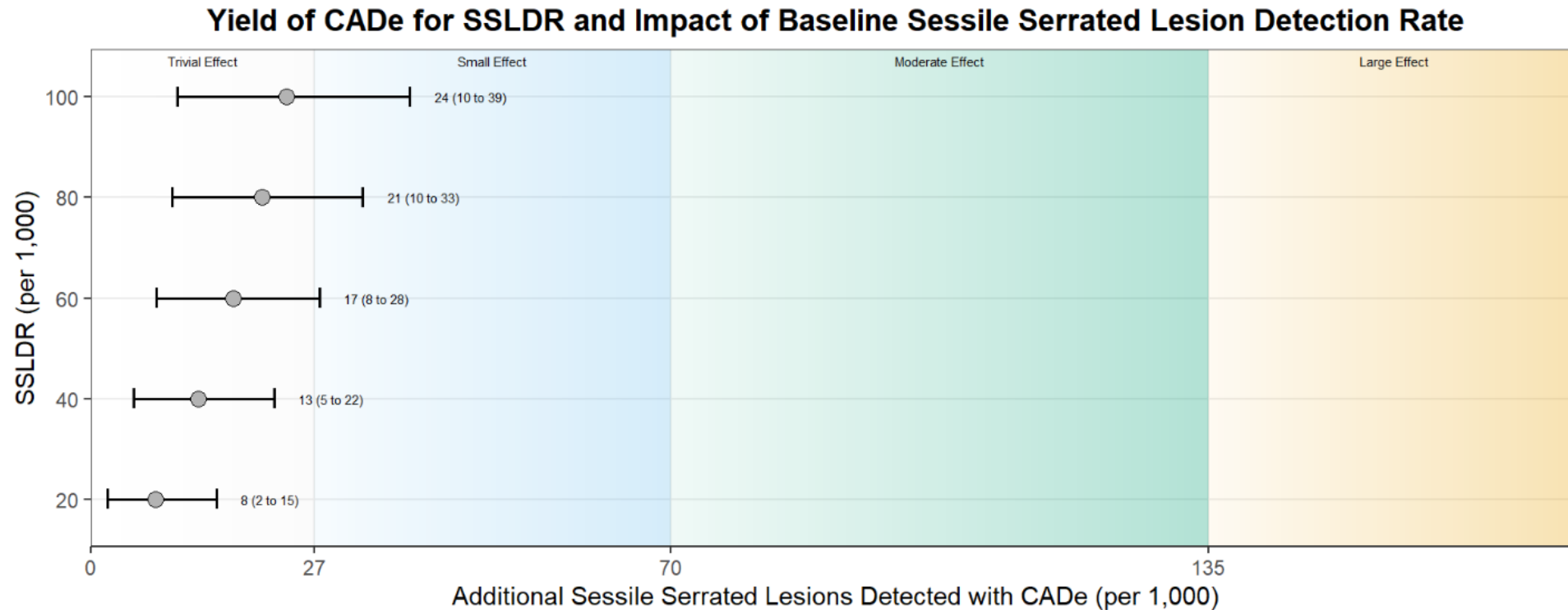
Reference: YMGE 14123



Benefit of AI is Greatest in Low Adenoma Detectors (where it is most needed)



Similar Benefit for Sessile Serrated Lesions



Post FDA Approval Issue

- False alerts/Alert Fatigue
- Detection of clinical low risk lesions
- Deskillling (of experienced endoscopy)
- Never-skilling (of new endoscopists)
- Does it work in the “real world”

Deskilling Impact of CAdE Colonoscopy

Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study

Krzysztof Budzyń, Marcin Romańczyk, Diana Kitala, Paweł Kołodziej, Marek Bugajski, Hans O Adami, Johannes Blom, Marek Buszkiewicz, Natalie Halvorsen, Cesare Hassan, Tomasz Romańczyk, Øyvind Holme, Krzysztof Jarus, Shona Fielding, Melina Kunar, Maria Pellise, Nastazja Pilonis, Michał Filip Kamiński, Mette Kalager, Michael Bretthauer, Yuichi Mori

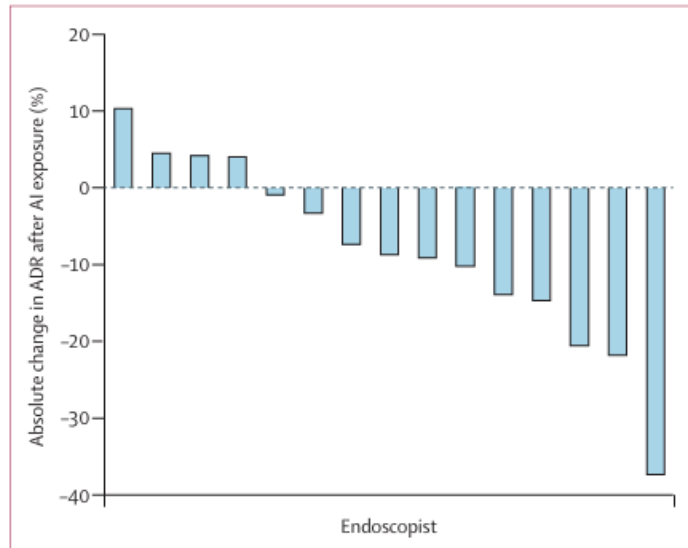
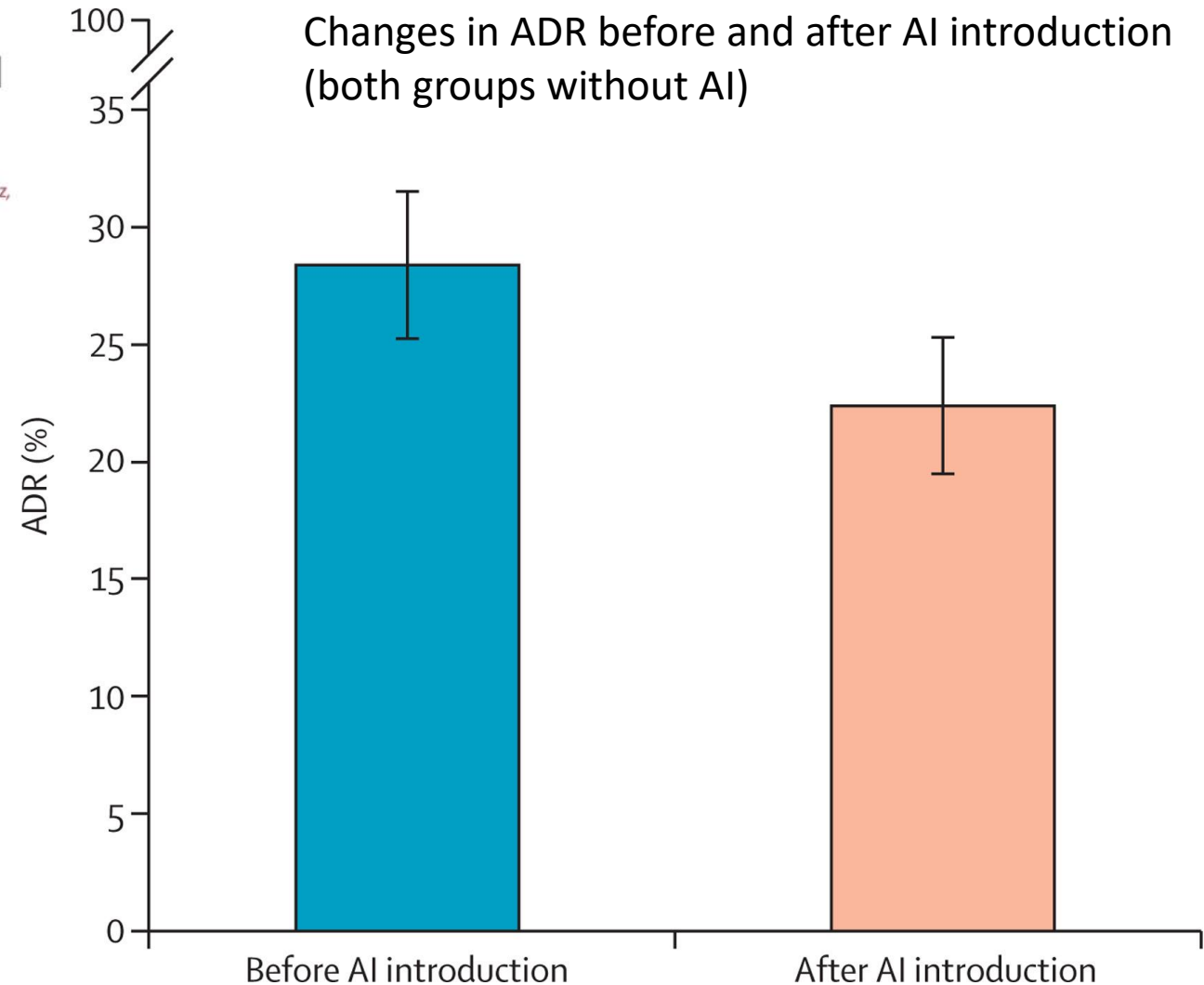


Figure 2: Endoscopist-level change in ADR with standard, non-AI assisted colonoscopy after introduction of AI in colonoscopy at each centre
Four endoscopists who did fewer than ten colonoscopies either before or after AI introduction are not shown. ADR=adenoma detection rate. AI=artificial intelligence.



Evaluating the Impact of FDA Approved CAdE on Adenoma Detection and Endoscopist Desking: A 5-Year Cohort Study

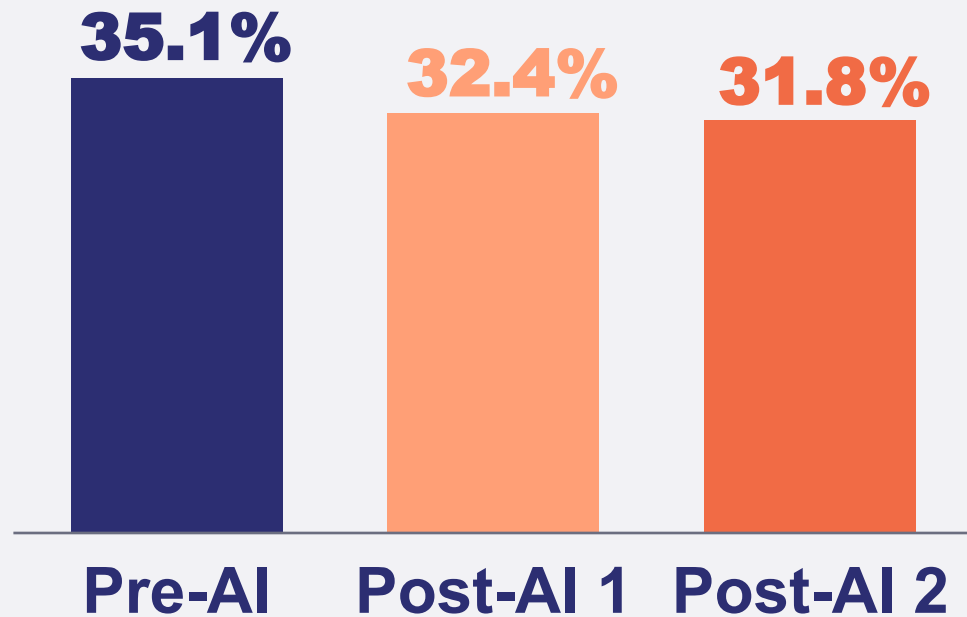


Mario El Hayek, MD; Kapil D. Nayar, MD; Felix A. Lopez Dominici, MD; Albaraa Yazbeck, MD; Michael B. Wallace, MD DDW 2026 (manuscript under review)

Department of Gastroenterology & Hepatology
Mayo Clinic, Jacksonville, FL

No Evidence of Deskillling

Unadjusted ADR by period



Non-AI procedures across 3 time periods (N = 8,250)

Mixed-Effects Analysis

Pre-AI Vs Post-Washout 1

Adjusted OR = 0.94

95% CI: 0.84–1.06 | $p = 0.32$

Pre-AI vs Post-Washout 2

Adjusted OR = 0.95

95% CI: 0.84–1.07 | $p = 0.41$

Do we need CADx to help identify high risk vs low risk lesion?

Interobserver and Intraobserver Agreement for the Laterally Spreading Tumor Classification in Large (>2 cm) Colorectal Laterally Spreading Tumors

Roupen Djibachian, MD^{1,*}, Jérémie Jacques, MD^{2,*}, Victoire Michal, PhD¹, Ludovico Alfaroni, MD³, Robert Bechara, MD⁴, Nicholas G. Burgess, MD^{5,6}, Mariana Figueiredo, MD⁷, Yusuke Fujiyoshi, MD⁸, Lucile Heroin, MD⁹, Michal F. Kaminski, MD^{10,11,12}, Eric Lam, MD¹³, Philippe Leclercq, MD¹⁴, Isabelle Lienhart-Chambon, MD¹⁵, Alexandru Lupu, MD¹⁶, Charles Menard, MD¹⁷, Jeffrey Mosko, MD¹⁸, Nima Mottacki, MD¹⁹, Heiko Pohl, MD^{20,21}, Douglas K. Rex, MD, MACP, MACG²², Enrique Rodriguez De Santiago, MD²³, Yutaka Saito, MD, PhD, FACP²⁴, Amrita Sethi, MD²⁵, Sandro Sferazza, MD²⁶, Neal Shahidi, MD¹³, Daniel von Renteln, MD¹, Dennis Yang, MD, FACP²⁷ and Mathieu Pioche, MD¹⁶

- Low risk → standard polypectomy/piecemeal EMR
- High risk (t1m/T1sm1) → ESD
- Deeply invasive → Surgery
- Unfortunately we humans are not very good at this

Interobserver and Intraobserver Agreement of the LST Classification

Are experts consistent in classifying large (≥20 mm) colorectal laterally spreading tumors?

METHODS



24 expert endoscopists from **4 regions** (N. America, Europe, Asia, Oceania)

Independently classified videos of **46 LSTs (≥20 mm)** twice in randomized order

1,104 total classifications, evaluated with Light κ

RESULTS: INTEROBSERVER AGREEMENT

5-category classification
(LST-GH / LST-G-M / LST-NG-FE / LST-NG-PD / unclassifiable)

K = 0.37
POOR AGREEMENT

3-category classification
(granular / nongranular / unclassifiable)

K = 0.48
MODERATE AGREEMENT

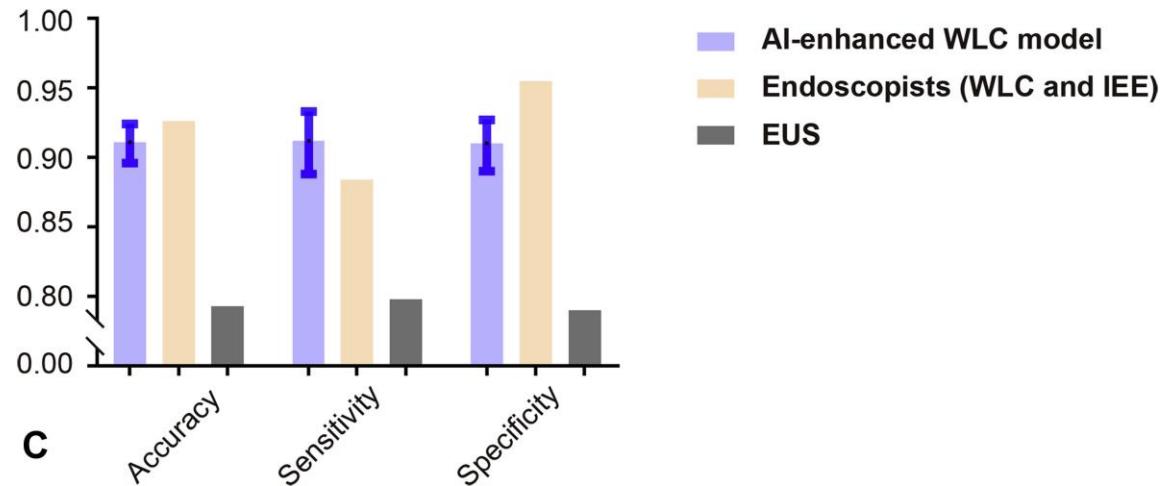
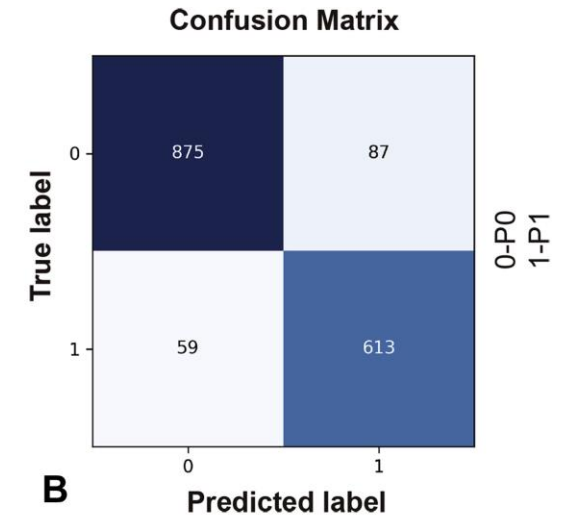
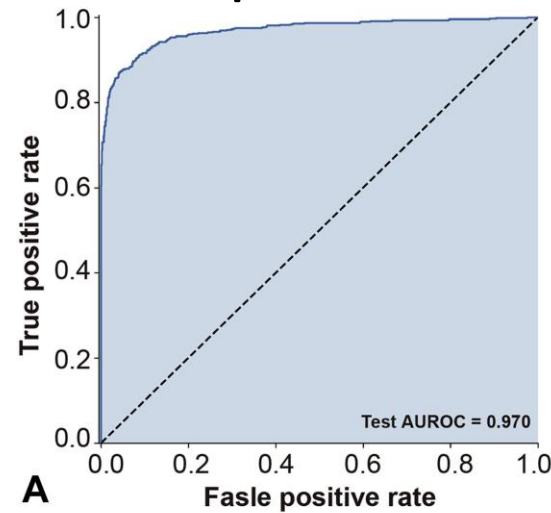
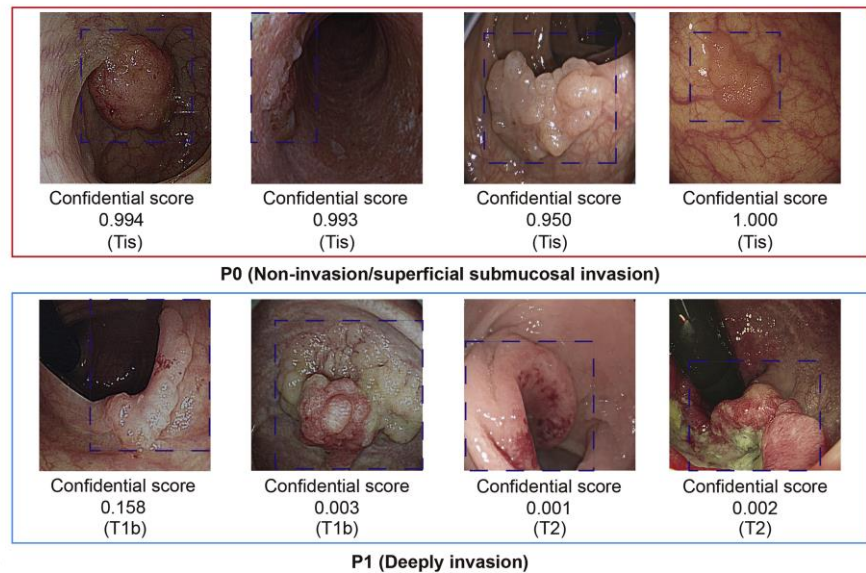
Intraobserver agreement varied widely between endoscopists ($\kappa = 0.28-0.94$); highest reproducibility in Europe and Oceania.

CONCLUSION: The LST classification is unreliable across expert endoscopists. Standardization is needed before relying on subtypes for clinical decisions.

Djibachian et al. *Am J Gastroenterol*. 2026. doi:10.14309/ajg.0000000000004103
© 2024 by The American College of Gastroenterology

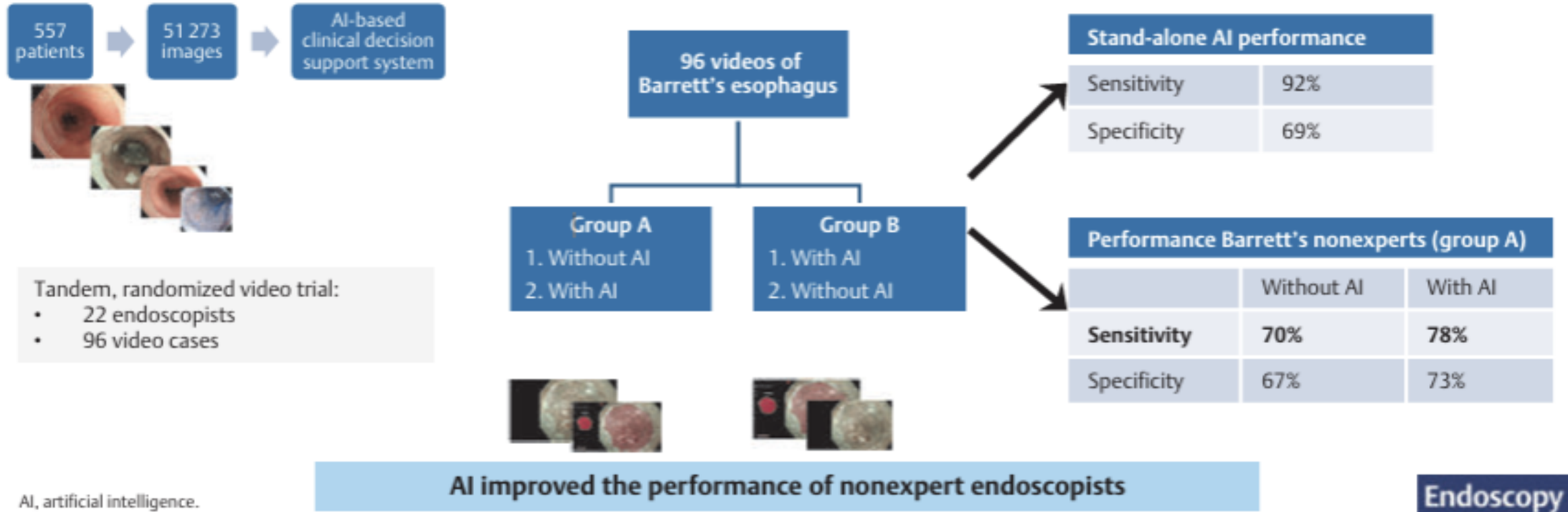
AJG The American Journal of
GASTROENTEROLOGY

Can AI Improve Lesion Classification: vis-à-vis Invasion Depth?



Barrett's Esophagus

AI for evaluation of Barrett's neoplasia – a randomized study



AI in GI-Radiology

Unassisted AI Equivalent to Best-In-Class Radiological Classification of IPMN



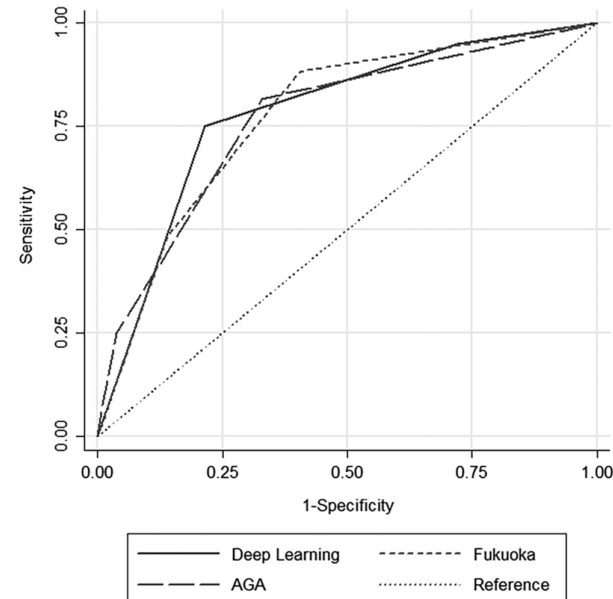
ORIGINAL ARTICLE

Deep Learning to Classify Intraductal Papillary Mucinous Neoplasms Using Magnetic Resonance Imaging

Juan E. Corral, MD,* Sarfaraz Hussein, PhD,† Pujan Kandel, MD,* Candice W. Bolan, MD,‡
Ulas Bagci, PhD, MSc,† and Michael B. Wallace, MD, MPH*

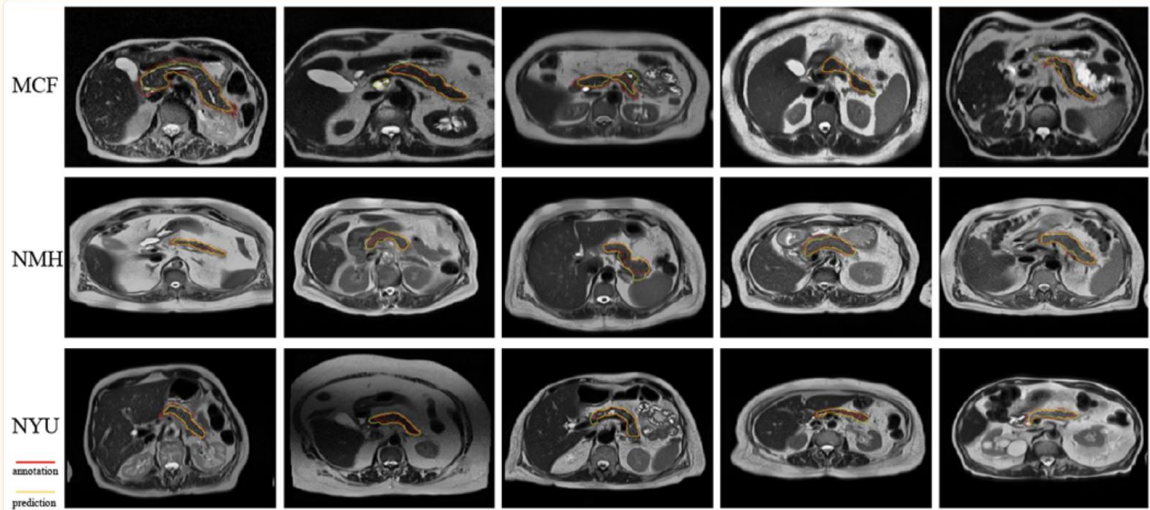
2019. *Pancreas*, 48(6), pp.805-810.

Accuracy for
No/Low Grade vs
High Grade/Cancer



Radiomics Boosts Deep Learning Model for IPMN Classification

[Lanhong Yao](#),¹ [Zheyuan Zhang](#),¹ [Ugur Demir](#),¹ [Elif Keles](#),¹ [Camila Vendrami](#),¹ [Emil Agarunov](#),² [Candice Bolan](#),³
[Ivo Schoots](#),⁴ [Marc Bruno](#),⁴ [Rajesh Keswani](#),¹ [Frank Miller](#),¹ [Tamas Gonda](#),² [Cemal Yazici](#),⁵ [Temel Tirkes](#),⁶
[Michael Wallace](#),⁷ [Concetto Spampinato](#),⁸ and [Ulas Bagci](#)¹



[Fig. 2.](#)
Qualitative pancreas segmentation visualization on multi-center data. Predicted segmentation maps (yellow line) are highly

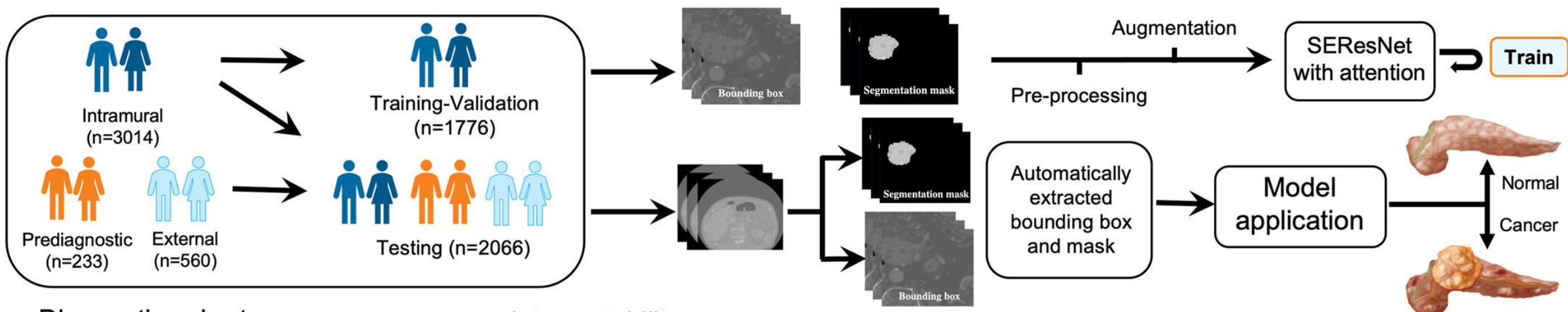
Quantitative comparison (%) for the influence of combining radiomics with deep learning. We can observe that regardless of network structure, combining radiomics with deep learning can impressively leverage the IPMN classification performance. Similarly, combining the deep learning extracted features can also leverage the performance of radiomics for the IPMN classification.

Network	w/o Radiomics				w/ Radiomics			
	ACC	AUC	PR	RC	ACC	AUC	PR	RC
Metrics								
w/o DL	-	-	-	-	71.6	89.7	74.7	75.2
DenseNet	57.4	68.4	55.9	56.2	75.8	87.4	74.7	76.0
ResNet18	48.9	66.6	46.3	48.0	73.7	89.3	76.1	76.7
AlexNet	57.0	64.8	60.3	54.7	77.7	89.8	76.5	75.2
MobileNet	57.0	64.4	57.7	54.6	71.6	89.3	74.7	75.2
ViT	61.3	71.9	56.2	56.6	81.9	89.3	82.4	82.7

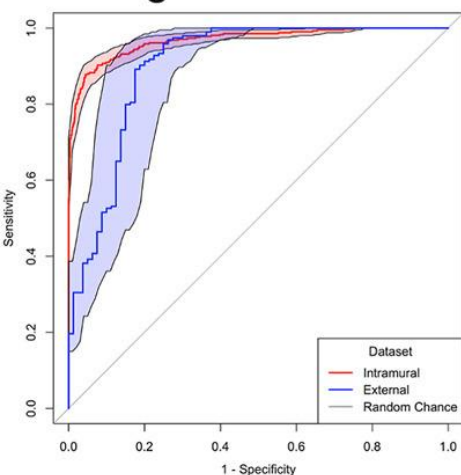
Can We Train AI Algorithm to Detection the Unseen Pancreas Tumor?

- Step 1: Segment the pancreas from other organs
- Step 2: Train computer vision algo on CTs done 1-3 years prior to clinical diagnosis of PDAC
- Step 3: Validate in Case:Control Trial

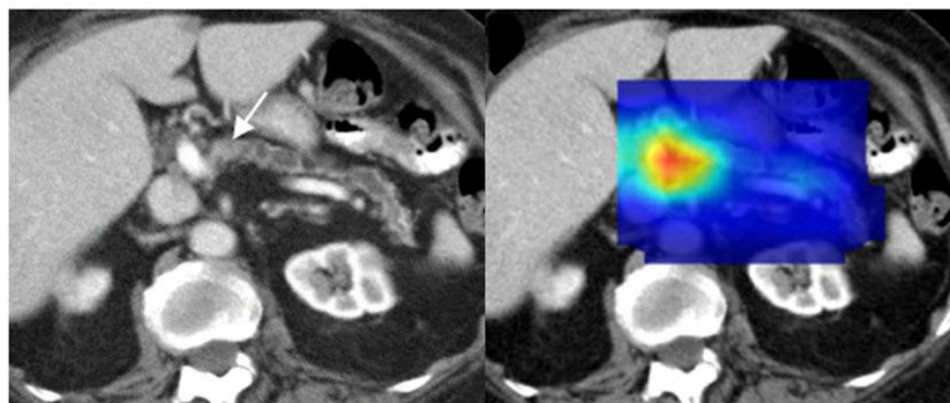
Radiomics Detection of Missed/Occult



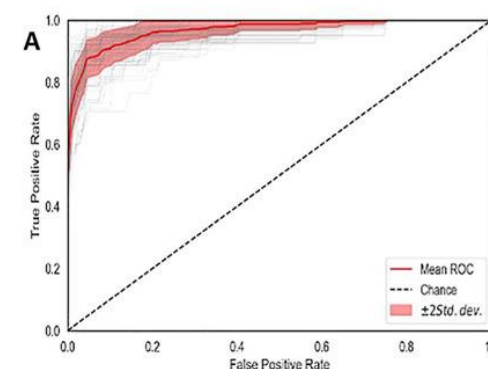
Diagnostic cohorts



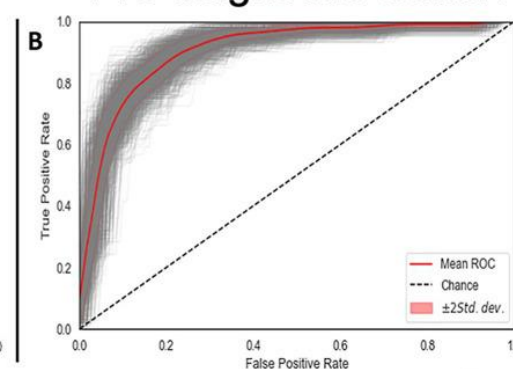
Interpretability



Simulated Cohort



Pre-diagnostic Cohort



Gastroenterology

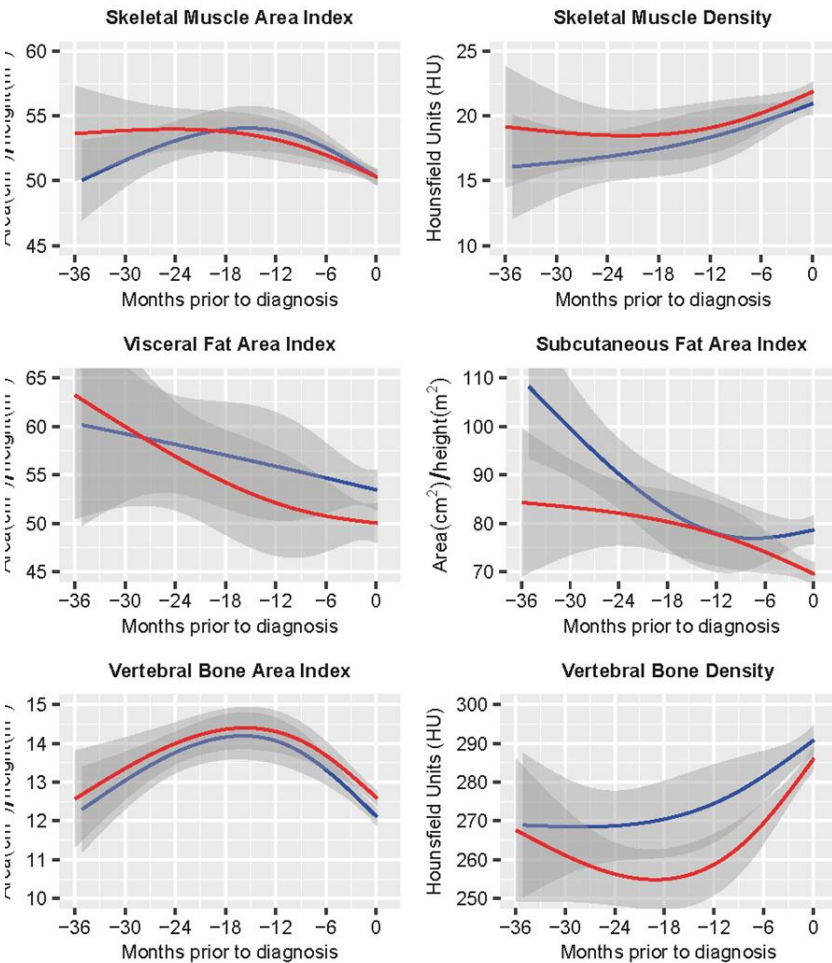
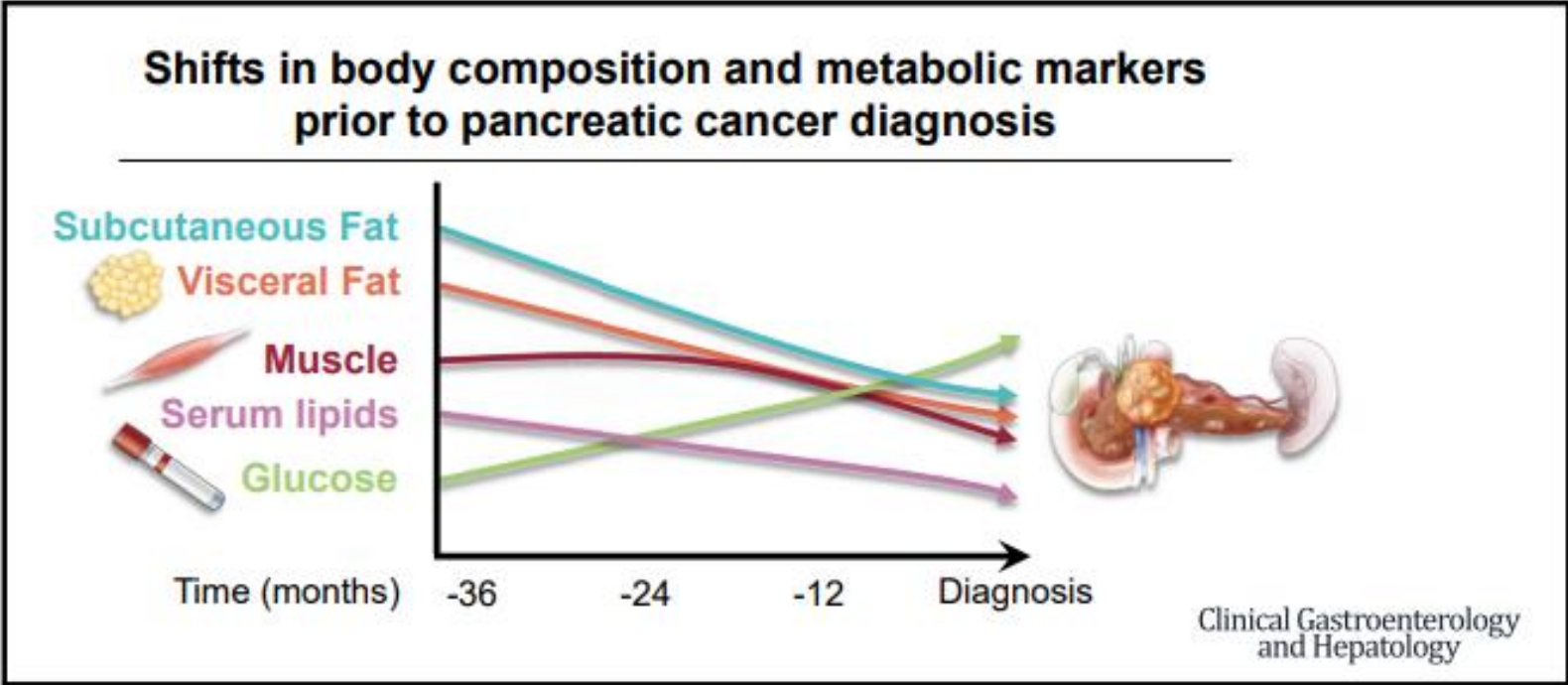
Korfiatis, Panagiotis, et al. "Automated artificial intelligence model trained on a large data set can detect pancreas cancer on diagnostic computed tomography scans as well as visually occult preinvasive cancer on prediagnostic computed tomography scans." *Gastroenterology* 165.6 (2023): 1533-1546.

Temporal Trends in Body Composition and Metabolic Markers Prior to Diagnosis of Pancreatic Ductal Adenocarcinoma

Clinical Gastroenterology and Hepatology 2024; in press

Derk C. F. Klatte,^{1,2} Alexander Weston,³ Yaohua Ma,³ Hanna Sledge,³ Aman Bali,¹ Candice Bolan,⁴ Megan Engels,^{1,2} Jeanin E. van Hooft,² Monique E. van Leerdam,^{2,5} Ahmed Ouni,¹ Michael B. Wallace,¹ and Yan Bi¹

¹Department of Gastroenterology and Hepatology, Mayo Clinic, Jacksonville, Florida; ²Department of Gastroenterology and Hepatology, Leiden University Medical Center, Leiden, The Netherlands; ³Department of Quantitative Health Sciences, Mayo Clinic, Jacksonville, Florida; ⁴Department of Radiology, Mayo Clinic, Jacksonville, Florida; and ⁵Department of Gastrointestinal Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands



Tumor Stage: — Stages I & II — Stages III & IV

Ambient Clinical Intelligence and Electronic Health Records

- Our (Epic) EHR has >20mil people.
- How can I find everyone with a major risk factor for pancreatic cancer and get them screened?
 - Strong family hx (buried deeply in unstructured notes)
 - Genetic mutation (buried deeply in pdf files of WES reports)
 - Rising glucose/falling weight/age > 50 (ENDPAC score)
 - Changing body composition on CTs for other reasons
- Not humanly possible without AI



Identification of pancreatic cancer risk factors from clinical notes using natural language processing



Dhruv Sarwal^a, Liwei Wang^b, Sonal Gandhi^a, Elham Sagheb Hossein Pour^b, Laurens P. Janssens^a, Adriana M. Delgado^a, Karen A. Doering^a, Anup Kumar Mishra^a, Jason D. Greenwood^{c,d}, Hongfang Liu^b, Shounak Majumder^{a,*}

^aDepartment of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN 55905

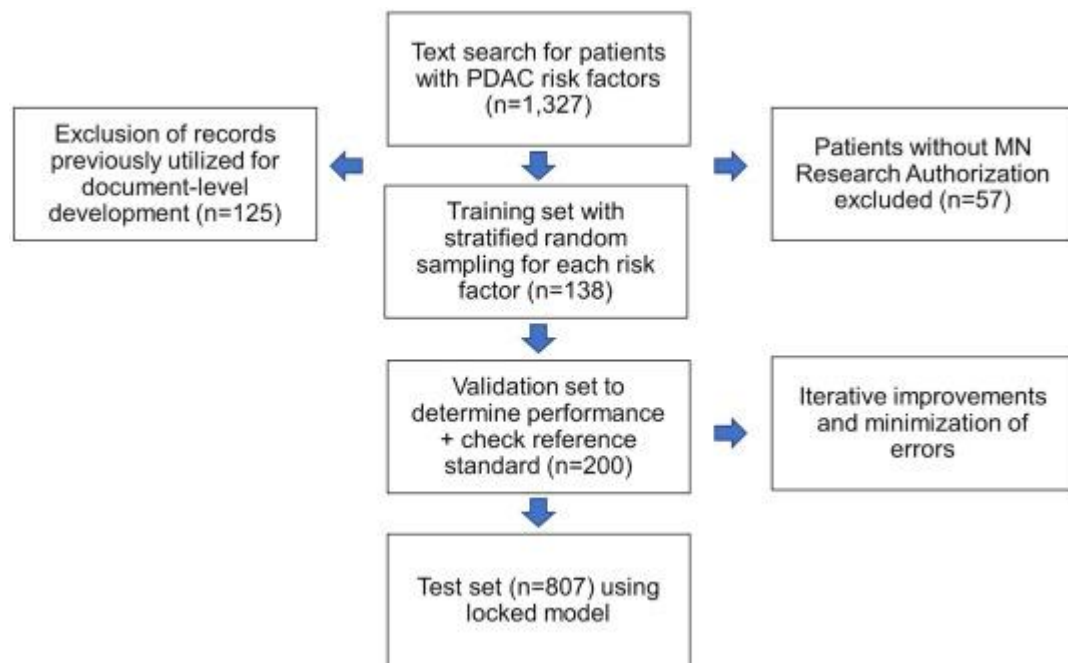


Table 5. Performance of NLP algorithms at patient-level.

		Training	Validation 1	Validation 2	Test
Family History of PDAC	Recall	0.936	0.857	0.864	0.933
	Precision	0.772	0.857	0.950	0.790
	F1-score	0.846	0.857	0.905	0.856
Germline genetic mutation in PDAC susceptibility gene	Recall	0.907	0.667	0.800	0.851
	Precision	0.857	0.320	0.444	0.350
	F1-score	0.881	0.432	0.571	0.496

What's Next

- First we had detection/classification (CAdE/x)
 - Polyps (now in use), Barrett's, Early Gastric Cancer, PDAC, Cystic neoplasia....
 - FDA rules for implementation are the only limiting factor (but getting easier)
- Then we had ambient large language models/generative AI
 - Work flow optimization: note generation
 - EHR integration: identification of risk (ex: PDAC)
- Now we have active “agents”
 - Chat bots to assist with inbox, bowel prep
 - Optimized suggestions for therapy
 - Likely will allow patients direct (without physician) access to advanced knowledge
- Will we have Large “Action” Models (LAMs)
 - Robotic surgery systems trained on the world's best surgeons

Summary

- AI is rapidly transforming medicine
 - GI leading with FDA/CE approved applications
- Computer vision for polyp detection/classification is already in practice
 - Improved efficiency for capsule endoscopy
 - Preliminary work on pancreas cancer detection promising
- Need process to allow “platform” system for multi-application integration (e.g. an “app” store)
- Ambient and Agentic AI will transform how we gather, process and document medical interactions, hopefully for the better!!



Thanks to our team



Universiteit Leiden



Thank you!



Wallace.Michael@mayo.edu