

Implementation of a Prospective, FAIR-Compliant Multimodal Video Databank for AI Development and Clinical Outcomes Research in Gastrointestinal Endoscopy



Megan Oleksiw¹, Roupen Djinbachian^{1,2}, Claire Gefflot¹, Victoire Michal¹, Nikoo Mahsayekhi¹, Mira Kandikar-Bloch¹, Robert Battat^{1,2}, Eric Cristea¹, Samira Hanin¹, Alan Barkun³, Mickael Bouin^{1,2}, Taleen Aïda Ashikian¹, Melina Vassiliou⁴, Daniel von Renteln^{1,2}

¹Montreal University Hospital Research Center, Montreal, Quebec, Canada; ³Division of Gastroenterology, Montreal University Hospital Center, Montreal, Quebec, Canada; ³Division of Gastroenterology, McGill University and McGill University Health Center, Montreal, Quebec, Canada; ⁴Division of Surgery, McGill University and McGill University Health Center, Montreal, Quebec, Canada

Background

Existing endoscopic databases for AI development rely on retrospective sampling, isolated images, or lack clinical context, limiting their utility for artificial intelligence (AI) validation and clinical research. This study establishes a prospective endoscopy video databank with standardized annotations enabling AI development and clinical outcomes research.

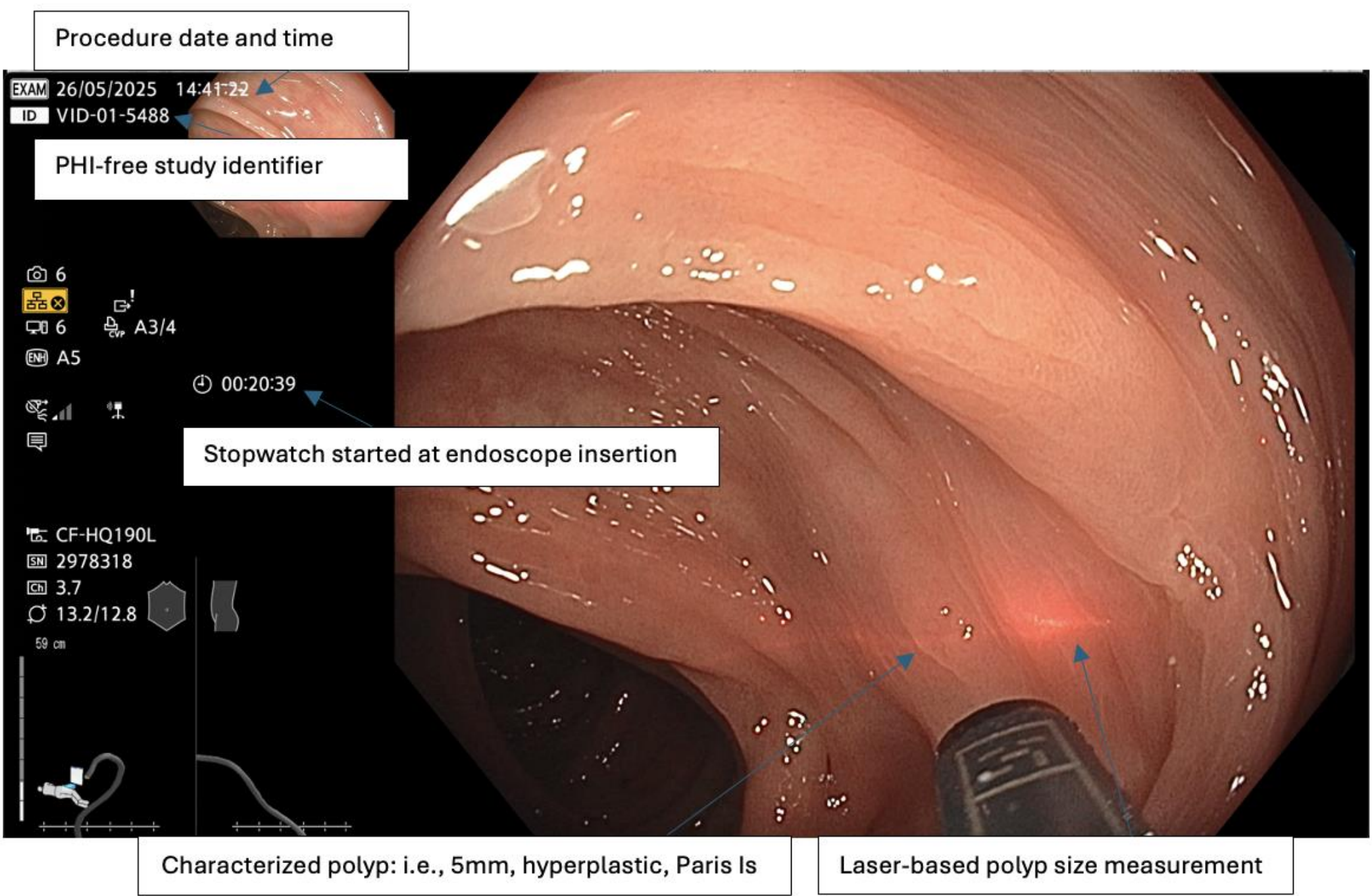
Methods

Prospective video databank established at the *Montreal University Hospital Center* (NCT06822616):

- Enrollment of patients (June 2022-ongoing) undergoing colonoscopy and EGD, providing informed consent for data collection and subsequent use in AI system development
- **Video recordings de-identified** from acquisition, without protected health information, all technical specifications optimized for AI training (full-length, full-frame videos in 1920x1080-pixel resolution and 60fps, and 10-bit color depth).
- Videos linked with **structured, real-time clinical documentation**:
 - **450-740 variables per procedure** including anatomical landmarks, lesion characteristics, intervention timing, disease-specific scoring (i.e., IBD activity scores, Prague classification for Barrett's), optical diagnoses, AI-assisted detection outputs, and quality metrics.
 - **Timestamps link video frames** directly to clinical events, enabling frame-level navigation
 - **Pathology results** subsequently integrated, establishing ground-truth correlation for each visualized lesion

Automated algorithms calculate procedure quality metrics: adenoma detection rates, withdrawal times, etc.

Figure 1. Still-frame from recorded, de-identified and time-stamped full-length video recording



Crohn's Disease		Ulcerative Colitis	
Patient characteristics (N=524)		Patient characteristics (N=455)	
Age, mean (SD)	49.1 (15.8)	Age, mean (SD)	50.2 (16.3)
Female, n (%)	287 (54.8)	Female, n (%)	225 (49.5)
First degree family history of IBD, n (%)	119 (22.7)	First degree family history of IBD, n (%)	66 (14.5)
Ethnicity*, n (%)		Ethnicity*, n (%)	
North American	438 (83.6)	North American	352 (77.4)
European	45 (8.6)	European	39 (8.6)
Caribbean	9 (1.7)	Caribbean	8 (1.8)
Latin, Central, and South American	6 (1.1)	Latin, Central, and South American	10 (2.2)
African	17 (3.2)	African	23 (5.1)
Asian	16 (3.1)	Asian	26 (5.7)
Indigenous	-	Indigenous	1 (0.2)
Jewish	1 (0.2)	Jewish	2 (0.4)
Unknown	1 (0.2)	Unknown	5 (1.1)
Disease characteristics (N=605 procedures)		Disease characteristics (N=515 procedures)	
Disease location, when active, n (%)		Extent of disease, when active, n (%)	
Ileal	119 (19.7)	<15 cm (E1: proctitis)	217 (42.1)
Ileo-colonic	25 (4.1)	15-60 cm (E2: left-sided colitis)	80 (15.5)
Colonic	90 (14.9)	>60 cm (E3: pancolitis)	11 (2.1)
Not applicable (Inactive)	371 (61.3)	Not applicable (Inactive)	207 (40.2)
Endoscopic inflammation scores, n (%)		Endoscopic inflammation scores, n (%)	
Inactive: SES-CD <3	36 (6.0)	Inactive: MES 0	270 (52.4)
Mild: SES-CD 3-6	124 (20.5)	With histological inflammation	14 (5.2)
Moderate: SES-CD 7-15	64 (10.6)	Mild: MES 1	109 (21.2)
Severe: SES-CD ≥15	14 (2.3)	Moderate: MES 2	85 (16.5)
Missing	367 (60.7)	Severe: MES 3	34 (6.6)
Patients with prior surgery, n (%)	214 (35.4)	Missing	17 (3.3)
Rutgeerts score, n (%)		Patients with prior surgery, n (%)	21 (4.1)
i0	103 (17.0)		
i1	16 (2.6)		
i2	8 (1.3)		
i2a	28 (4.6)		
i2b	21 (3.5)		
i3	2 (0.3)		
i4	18 (3.0)		
Missing	409 (67.6)		
Stenosis present, n (%)			
Single, passable	37 (6.1)		
Multiple, passable	11 (1.8)		
Not passable	28 (4.6)		
Missing	529 (87.4)		

SD, standard deviation; IBD, inflammatory bowel disease; SES-CD, Simple Endoscopic Score for Crohn's Disease; MES, Mayo Endoscopic Subscore; *Patients may provide multiple ethnicities.

Figure 2. Overview of the databank data collection, management, and quality-control pipeline

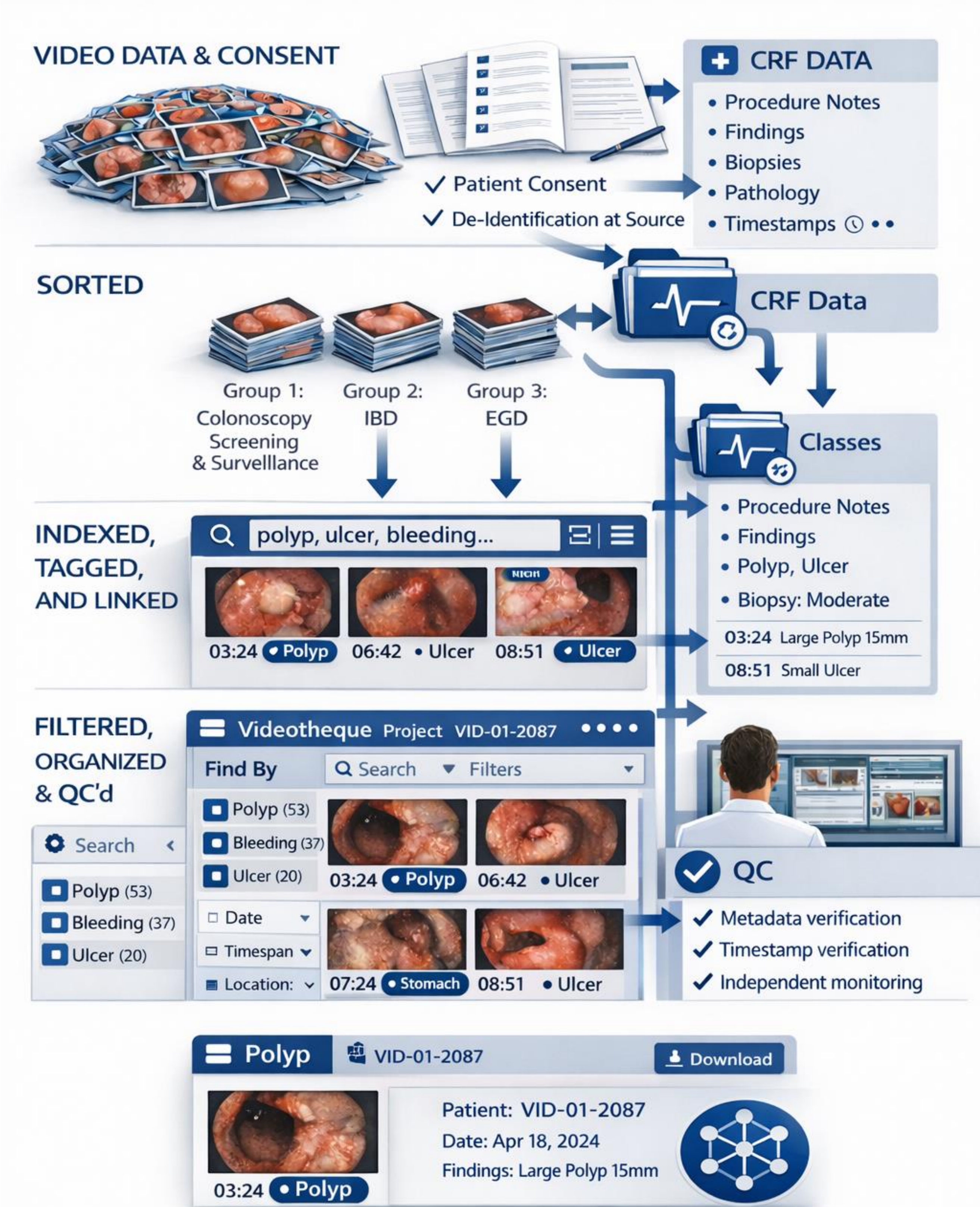


Table 1. Patient and procedure characteristics (IBD sub-cohort)

Results

As of December 2025, the databank includes **8739 endoscopic procedures in 7139 patients** (mean age 62.3, 51.7% female):

- **6433 colonoscopies** and **2306 EGDs**
 - 1291 colonoscopies in IBD patients (Table 1)
- Performed by **49 endoscopists**
- **8301 polyps** were detected, timestamped, individually characterized, and correlated to pathology results
- **8631 biopsies** timestamped, annotated, correlated to pathology
 - Polyps and biopsies annotated with **up to 49 and 45 variables**

- The databank has **supported several AI validation studies**:
 - For AI-based polyp detection^{1,2} and diagnosis^{3,4}
- Participating endoscopists demonstrated 12% relative ADR improvement over 24 months through automated ADR monitoring and performance feedback

Conclusion

This prospective, FAIR-compliant databank of >8500 endoscopic procedures provides a **high-quality, scalable foundation for the development and validation of AI systems in endoscopy and clinical outcomes research.**

Expansion to four international centers targeting 20,000 procedures by 2027 is underway. Future enhancements include digital pathology integration, multi-omics and longitudinal data for IBD patients and registry linkage for long-term cancer incidence tracking. This video and clinical data capture platform enables rigorous AI systems validation, human-AI interaction studies and high-quality labelled data for AI systems development.

Contact: Daniel von Renteln, email: danielrenteln@gmail.com

1. Huang F, et al. AJG 2026; 2. Djinbachian R, et al. Gut 2025; 3. Oleksiw M, et al. Endoscopy 2026; 4. Oleksiw M, et al. CGH 2026.