

# A multisystem-compatible deep learning-based algorithm for detection and characterization of angiectasias in small-bowel capsule endoscopy. A proof-of-concept study.

Charles Houdeville, Marc Souchaud, Romain Leenhardt, Hanneke Beaumont, Robert Benamouzig, Mark Mcalindon, Sylvie Grimbert, Dominique Lamarque, Richard Makins, Jean-Christophe Saurin, et al.

## ▶ To cite this version:

Charles Houdeville, Marc Souchaud, Romain Leenhardt, Hanneke Beaumont, Robert Benamouzig, et al.. A multisystem-compatible deep learning-based algorithm for detection and characterization of angiectasias in small-bowel capsule endoscopy. A proof-of-concept study.. Digestive and Liver Disease, 2021, 10.1016/j.dld.2021.08.026 . hal-03352234

# HAL Id: hal-03352234 https://hal.science/hal-03352234

Submitted on 5 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

#### A multisystem-compatible deep learning-based algorithm for detection and characterization of angiectasias in small-bowel capsule endoscopy. A proof-of-concept study.

Charles Houdeville<sup>1</sup>, Marc Souchaud<sup>2</sup>, Romain Leenhardt<sup>1,2</sup>, Hanneke Beaumont<sup>6</sup>, Robert Benamouzig<sup>8</sup>, Mark McAlindon<sup>4</sup>, Sylvie Grimbert<sup>9</sup>, Dominique Lamarque<sup>3</sup>, Richard Makins<sup>5</sup>, Jean-Christophe Saurin<sup>7</sup>, Aymeric Histace<sup>2</sup>, Xavier Dray<sup>1,2</sup>

- 1. Sorbonne Université, Centre d'Endoscopie Digestive, Hôpital Saint-Antoine, APHP, Paris, France
- 2. ETIS UMR 8051 (CY Paris Cergy University, ENSEA, CNRS), Cergy, France
- 3. Université de Versailles St-Quentin en Yvelines, Service d'Hépato-Gastroentérologie, Hôpital Ambroise Paré
- 4. Royal Hallamshire Hospital, Sheffield Teaching Hospitals, Sheffield, United Kingdom
- 5. Cheltenham General Hospital, Cheltenham, Gloucestershire, UK
- 6. Amsterdam Universitair Medische Centra, Amsterdam, The Netherlands
- 7. Department of Hepato-gastroenterology, E. Herriot Hospital, Hospices civils de Lyon, Lyon, France
- 8. Department of Hepato-gastroenterology, Hôpital Avicenne, APHP, Bobigny, France
- 9. Department of Hepato-gastroenterology, Groupe hospitalier Diaconesses Croix Saint Simon, Paris, France

#### **Corresponding author**

Xavier Dray, Centre d'Endoscopie Digestive Hôpital Saint-Antoine 184 rue du Faubourg Saint-Antoine 75012 Paris Mail: xavier.dray@aphp.fr Tel: (0033) 1 49 28 21 70

#### **Disclosure statement**

Xavier Dray, Aymeric Histace and Romain Leenhardt are co-founders and shareholders of Augmented Endoscopy.

Xavier Dray has received lecture fees from Bouchara Recordati, Fujifilm, Medtronic, MSD and Pfizer, and has acted as a consultant for Alfasigma, Boston Scientific, Norgine, and Pentax.

Richard Makins has received travel and accommodation expenses from Intromedic to teach capsule endoscopy internationally and in the United Kingdom.

Mark McAlindon has received research materials from Given Imaging, Jinshan Science, Intromedic, Ankon and has acted as a consultant for Medtronic.

Robert Benamouzig is working for Alfasigma, Medtronics, and Bayer (board)

The others authors do not have disclosures

#### Keywords

capsule endoscopy; small bowel; artificial intelligence; deep learning; neural networks; algorithms

#### Abbreviations

AI: Artificial Intelligence CD: crohn's disease CE: capsule endoscopy DL: deep learning GI: gastrointestinal GIA: gastrointestinal angiectasia (or angiodysplasia) ms: millisecond NPV: Negative Predictive Value OGIB: obscure gastrointestinal bleeding PPV: Positive Predictive Value **ROI:** Region of Interest SB: small bowel SBCE: small bowel capsule endoscopy Se: sensitivity Sp: specificity **TP:** True Positive

#### Abstract

**Background and aims.** Current artificial intelligence (AI)-based solutions for capsule endoscopy (CE) interpretation are proprietary. We aimed to evaluate an AI solution trained on a specific CE system (Pillcam, Medtronic) for the detection of angiectasias on images captured by a different proprietary system (MiroCam, Intromedic).

**Material and Methods**. An advanced AI solution (Convolutional neural network), previously trained on on Pillcam<sup>®</sup> small bowell images, was evaluated on independent datasets with more than 1200 Pillcam<sup>-</sup> and MiroCam<sup>-</sup> still frames (equally distributed, with or without angiectasias). Images were reviewed by experts before and after AI interpretation.

**Results**. Sensitivity for the diagnosis of angiectasia was 97.4% with Pillcam images and 96.1% with Mirocam images, with specificity of 98.8% and 97.8%, respectively. Performances regarding the delineation of regions of interest and the characterization of angiectasias were similar in both groups (all above 95%). Processing time was significantly shorter with Mirocam (20.7 ms) than with Pillcam images (24.6 ms, p<0.0001), possibly related to technical differences between systems.

**Conclusion**. This proof-of-concept study on still images paves the way for the development of resource-sparing, "universal" CE databases and AI solutions for CE interpretation.

#### Introduction

Capsule endoscopy (CE) is recommended for investigating the small bowel (SB) in patients with gastrointestinal (GI) bleeding, iron deficiency anemia, or suspected Crohn's disease, after normal upper GI endoscopy and colonoscopy. (1) Gastroenterologists spend about 30 to 120 min to review and interpret full-length small-bowel capsule endoscopy (SBCE) recordings that encompass a mean number of 50,000 frames. The interpretation of one CE leads to an accuracy decrease of the practician, preventing the reading of several capsules. (2,3)

In attempt to relieve physicians from the tedious, time-consuming task of SBCE reading, our team has designed in 2019 a proof-of-concept study where a deep learning (DL)-based artificial intelligence (AI) solution was shown highly sensitive and highly specific to detect gastrointestinal angiectasias (GIA, also known as angiodysplasias), a most common SB vascular lesion. (4) This development was based on datasets of still images captured from one specific third-generation SBCE system. Since then, DL-based solutions have flourished from many academic and industrial research teams. (5,6) These solutions allow a reading time of a full-length SBCE recording in about 2 to 5 minutes. Some solutions are now commercially available (7), but most are proprietary, dedicated to the automated SBCE reading of one specific brand/company. Nevertheless, we believe that, visually speaking, SB findings (and specifically GIA) are very similar, whatever the proprietary CE system used for capture.

The aim of this study was to evaluate a DL-based solution trained on images captured by a specific CE system (namely, Pillcam SB3, Medtronic, Minneapolis, MN, USA), for the detection of GIA on still images captured not only by the same system, but also by a different one (namely MiroCam, Intromedic, Seoul, South Korea).

#### **Materials and Methods**

Institutional Review Board approval was obtained (Sorbonne University, n°20210209162452, February 9-2021) for this retrospective, non-interventional study on de-identified images.

#### Initial training and testing, on image datasets built from one specific CE proprietary system

The initial database and the DL-based prototype AI solution have been described previously. Briefly, de-identified annotated SBCE still frames were extracted from a multicenter (12 French Endoscopy unit) database called CAD-CAP, in which 25,000 still frames were collected from 4166 de-identified, third-generation, SB-CE videos (PillCam<sup>®</sup> SB3 system, Medtronic, Minneapolis, MN, USA. (8)

GIAs were defined as "*clearly demarcated, bright-red, flat lesions, consisting of tortuous and clustered capillary dilatations, within the mucosal layer*" (2) and then reviewed and validated by three senior CE experts with more than 200 readings each (9).Convolutional neural network (CNN)–based approaches were used for DL The GPU used was a Quadro RTX® 3000 (NVIDIA, Santa Clara, CA, USA) with a 6GB RAM.

Dataset A1 with 300 GIA images was used for data augmentation and DL, and dataset A2 with 300 different GIA images was used for testing. Similarly, control datasets were created with 300 normal images each (N1 for DL, N2 for testing). After training on datasets A1 and N1, the DL-based approach demonstrated a 100% sensitivity and a 96% specificity when

tested on datasets A2 and N2, and also precisely located regions of interest (ROI) within images containing GIA. (10)

Since then, this DL-based solution has been refined. It is now able to detect (discriminating still frames encompassing lesions from normal frames), to delineate (box around ROI), to characterize findings (labelling GIA and ulcerations, and displaying a percentage of confidence in the proposed diagnosis), and to score SB video recordings in terms of cleanliness.(11)

#### Additional testing on image datasets built from two different CE proprietary systems

In the current study, new frames were extracted, selected, reviewed and validated based on the same methodology as for the initial training/testing datasets. Four additional image datasets (A3, N3, A4, N4) were thus created from different SBCE video recording, independent from the datasets (A1, N1, A2, N2) of the initial study (8). Datasets A3 (with 305 still frames with GIA), and N3 (with 321 normal, control still frames) were created from Pillcam SB3 recordings. Datasets A4 (with 308 still frames with GIA), and N4 (with 321 normal, control still frames) were created from MiroCam1200 recordings.

# Diagnostic performances of the AI solution for frame detection, ROI delineation, and characterization.

All AI outputs were reviewed by three CE experts (1), according to the definitions of true positive (TP) findings in terms of GIA frame detection, delineation of ROI and characterization (*table 1, figure 1*)

The primary endpoint of the study was the sensitivity (Se) of the DL-based algorithm in terms of GIA frame detection, on both A3/N3 (Pillcam) and A4/N4 (MiroCam) datasets. Specificity (Sp), positive (PPV) and negative (NPV) predictive values were calculated as well. Diagnostic performances (TP rate, Se, Sp, PPV, NPV) for ROI delineation and characterization, level (%) of confidence in diagnostic of GIA, and time for processing (calculation plus display on screen), were secondary endpoints

#### **Statistics**

Results were expressed as percentages with 95% confidence intervals ("C.I.) for categorical variables and as means  $\pm$  standard deviations (SD) for continuous variables. Fisher's exact test and Student test were used to compare categorical and continuous data between groups, respectively.

#### Results

TP finding rates were above 95% (**table 2**) for all tasks: GIA frame detection (Pillcam 97.4% vs Mirocam 96.1%, p=0.50), delineation of ROI (99.7% for both systems, p>0.99) and characterization of GIA (99.7% for both systems, p>0.99). The AI solutions also performed well when considering all three tasks combined (Pillcam 96.7% vs Mirocam 95.5%, p=0.53).

Similarly, the AI solutions demonstrated high diagnostic performances whatever the dataset used for evaluation (**table 3**): Se of 97.4% with Pillcam images and 96.1% with Mirocamimages, Sp of 98.8% and 97.8%, PPV of 98.8% and 97.7%, and NPV of 97.6% and 96.3%, respectively.

The mean percentage of confidence in diagnosing GIA was similar in both groups (82.5  $\pm$  22.9% with Pillcam images vs 81.6  $\pm$  24.8% in Mirocam images, p 0.51).

The mean processing time was significantly shorter with Mirocam images  $(20.7 \pm 6.5 \text{ ms})$  compared to what obtained with Pillcam images  $(24.6 \pm 7.0 \text{ ms}, \text{ p} < 0.0001)$ .

#### Discussion

To the best of our knowledge, this is the first study where a DL-based AI solution trained on CE specific proprietary image datasets (namely Pillcam SB3) demonstrated excellent diagnostic performances for frame detection, delineation of ROI and characterization of ROI, not only on an independent dataset of GIA images from the same proprietary system (overall, Se of 97.4%, Sp of 98.8%), but also (overall, Se of 96.1%, Sp of 97.8%) on an independent dataset of GIA images captured from a different CE system (namely MiroCam). The system was fast (mean processing time of 20 to 25 ms per image) and displayed a high level of confidence in the diagnostic of GIA (above 80% in most cases).

Our findings with the Pillcam- system are similar to that found for AI-based GIA detection by Noya et al. in 2017 (Se of 89%, Se of 97%) (12) and by our group in 2019 (Se of 100%, Sp of 96%) (8), thus providing some external validity. Current Se for GIA detection of our DLbased solution is slightly lower – but still over 95% – compared to the initial study (100%). This is possibly because the solution has evolved and now addresses more varied and complex tasks (detection, delineation, and characterization of GIA and inflammatory lesions, plus cleanliness evaluation). Diagnostic performances (for all endpoints) were slightly lower when tested on the A4/N4 datasets (MiroCam images) than when tested on the A3/N3 datasets (Pillcam images). We have three main hypotheses for explaining these subtle variations. First, training and testing on images from the same trademark likely makes a difference. Second, technical characteristics between devices such as different fields of view, resolutions (320x320 for MiroCam®, 340x340 for Pillcam®), contrasts, brightness, blurriness or opacity may also have an impact on performances (table 4)(13). Third, using a convolution process may also explain why the processing time was significantly shorter with MiroCam ® images than with Pillcam® images (in relation with heavier calculations in frames with higher resolution in the latter system). Still, overall, these differences were not statistically significant, and diagnostic performances were all above 95%.

Our study has several strengths. Image datasets were built from several centers across Europe. All images were reviewed (before and after testing) and categorized (for all clinical outcomes) by experts (14) according to well-defined criteria (for GIA definition and for outcome measurements). Still, we must acknowledge several limitations as well. First, the evaluation was retrospective, with potential bias. Among those, GIA were selected to be typical (4) and possibly in a visually clean environment. GIA detection in a real-life setting, in a prospective manner, may be more challenging. Second, the evaluation is based on still frames and not on video recordings which is a source of bias. However, the likelihood for a DL-based solution to capture a GIA is certainly higher when lesions are present on several consecutive frames on a video sequence than once on a single, selected, still frame. Still, another study at the video level would be of interest. Third, only GIA were considered because it is the more common and emblematic lesion seen in SB CE at the stage of a proof-of-concept study. Demonstration of similar results with more varied types of findings is needed.

Overall, these preliminary findings on still images pave the way for a "killing two birds with one stone" approach in the development of AI solutions for SBCE interpretation. A similar strategy has been used successfully for DL-based solutions for polyp detection in colonoscopy, developed on a proprietary video database, claimed to work well with all major endoscopy brands (15) and now commercially available (GI Genius-, Medtronic). Among advantages, this approach allows building resource-sparing, "universal", CE databases and AI solutions. Such solutions could be either plug-and-play boxes (as for colonoscopy) or on-line platforms (Software as a Service, SaaS, as proposed with our solution) where CE video recordings from any brand/company could be uploaded and semi-automatically interpreted.

#### References

1. Rondonotti E, Spada C, Adler S, May A, Despott EJ, Koulaouzidis A, et al. Smallbowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Technical Review. Endoscopy. avr 2018;50(4):423-46.

2. Koulaouzidis A, Dabos K, Philipper M, Toth E, Keuchel M. How should we do colon capsule endoscopy reading: a practical guide. Clin Med Insights Gastroenterol. janv 2021;14:263177452110019.

3. Beg S, Card T, Sidhu R, Wronska E, Ragunath K, Ching H-L, et al. The impact of reader fatigue on the accuracy of capsule endoscopy interpretation. Digestive and Liver Disease. août 2021;53(8):1028-33.

4. Leenhardt R, Li C, Koulaouzidis A, Cavallaro F, Cholet F, Eliakim R, et al. Nomenclature and semantic description of vascular lesions in small bowel capsule endoscopy: an international Delphi consensus statement. Endosc Int Open. mars 2019;07(03):E372-9.

5. Iakovidis DK, Koulaouzidis A. Software for enhanced video capsule endoscopy: challenges for essential progress. Nat Rev Gastroenterol Hepatol. mars 2015;12(3):172-86.

6. Eliakim R. Where do I see minimally invasive endoscopy in 2020: clock is ticking. Ann Transl Med. mai 2017;5(9):202.

7. Dray X, Iakovidis D, Houdeville C, Jover R, Diamantis D, Histace A, et al. Artificial intelligence in small bowel capsule endoscopy - current status, challenges and future promise. J Gastroenterol Hepatol. janv 2021;36(1):12-9.

8. Leenhardt R, Li C, Le Mouel J-P, Rahmi G, Saurin JC, Cholet F, et al. CAD-CAP: a 25,000-image database serving the development of artificial intelligence for capsule endoscopy. Endosc Int Open. mars 2020;8(3):E415-20.

9. Sidhu R, Chetcuti Zammit S, Baltes P, Carretero C, Despott EJ, Murino A, et al. Curriculum for small-bowel capsule endoscopy and device-assisted enteroscopy training in Europe: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy. août 2020;52(08):669-86.

10. Leenhardt R, Vasseur P, Li C, Saurin JC, Rahmi G, Cholet F, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. Gastrointestinal Endoscopy. janv 2019;89(1):189-94.

11. Leenhardt R, Souchaud M, Houist G, Le Mouel J-P, Saurin J-C, Cholet F, et al. A Neural Network-based Algorithm for Assessing the Cleanliness of Small Bowel during Capsule Endoscopy. Endoscopy. 2 nov 2020;

12. Noya F, Alvarez-Gonzalez MA, Benitez R. Automated angiodysplasia detection from wireless capsule endoscopy. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [Internet]. Seogwipo: IEEE; 2017 [cité 6 janv 2021]. p. 3158-61. Disponible sur: https://ieeexplore.ieee.org/document/8037527/

13. Yung DE, Plevris JN, Leenhardt R, Dray X, Koulaouzidis A. Poor Quality of Small Bowel Capsule Endoscopy Images Has a Significant Negative Effect in the Diagnosis of Small Bowel Malignancy. CEG. oct 2020;Volume 13:475-84.

14. Rondonotti E, Pennazio M, Toth E, Koulaouzidis A. How to read small bowel capsule endoscopy: a practical guide for everyday use. Endosc Int Open. oct 2020;08(10):E1220-4.

15. Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, et al. Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. Gastroenterology. août 2020;159(2):512-520.e7.

## Table 1: definitions of true positive findings for various study endpoints

Outcome	Definition of true positive
GIA frame detection	Any output (box) of the AI solution within a frame containing a GIA, wherever the box is placed, whatever label is displayed (GIA, or not) (see figures 1A, 1C, 1D and 1F for examples).
Delineation of ROI	Delineating box covering at least 50% of the GIA, being no larger than twice and no smaller than half the area of the GIA, whatever label is specified (GIA, or not) ( <i>see figures 1A, 1C, and 1D for examples</i> ).
Characterization	"Angiectasia" label is displayed within any frame where a GIA is present (wherever the box is placed) (see figures 1A and 1D for examples).

GIA: gastrointestinal angiectasia

<b>Tuble 21</b> The positive minings in mages containing anglectastas				
	Pillcam <sup>®</sup> SB3	Mirocam1200®	р	
	A3 dataset (n=305)	A4 dataset (n=308)		
Frame detection	297/305	296/308		
	97.4%	96.1%	0.50	
	<sub>95%</sub> C.I.[95.6%; 99.2%]	<sub>95%</sub> C.I.[93.9%; 98.3%]		
<b>Delination of ROI*</b>	296/297	295/296		
	99.7%	99.7%	>0.99	
	<sub>95%</sub> C.I.[99.1%; 100.0%]	<sub>95%</sub> C.I.[99.1%; 100.0%]		
Characterization*	296/297	295/296		
	99.7%	99.7%	>0.99	
	<sub>95%</sub> C.I.[99.1%; 100.0%]	<sub>95%</sub> C.I.[99.1%; 100.0%]		
All three combined	295/305	294/308		
	96.7%	95.5%	0.53	
	<sub>95%</sub> C.I.[94.7%; 98.7%]	<sub>95%</sub> C.I.[93.2%; 97.8%]		

### **Table 2:** True positive findings in images containing angiectasias

\* among true positive frame detections only

	Pillcam <sup>®</sup> SB3	Mirocam1200®	
	A3/N3 datasets (n=626)	A4/N4 datasets (n=629)	
Sensitivity	97.4%	96.1%	
	C.I. <sub>95%</sub> [96.2%; 98.7%]	C.I. <sub>95%</sub> [94.6%;97.6%]	
Specificity	98.8%	97.8%	
	C.I. <sub>95%</sub> [98.0%;99.7%]	C.I. <sub>95%</sub> [96.7%;99.0%]	
PPV	98.7%	97.7%	
	$C.I{_{95\%}}[97.8\%; 99.6\%]$	C.I. <sub>95%</sub> [96.5%; 98.9%]	
NPV	97.5%	96,3%	
	C.I. <sub>95%</sub> [96.3%; 98.7%]	C.I. <sub>95%</sub> [94.8% ;97.8%]	
Level of confidence	82 504 + 22 004	81 604 + 24 804	
in angiectasia diagnosis*	$02.3\% \pm 22.9\%$	01.0% ± 24.8%	
Calculation time (ms)	$24.6 \pm 7.0$	$20.7 \pm 6.5$	

**Table 3:** Diagnostic performances in terms of detection of frames with angiectasia

PPV: positive predictive value

NPV: negative predictive value

\* among images with true positive characterisation only

Table 4: Main technical differences between	Pillcam <sup>®</sup> SB3 (Medtronic, MN, USA) and
MiroCam <sup>®</sup> (Seoul, South Korean). (6)	

	Pillcam <sup>®</sup> SB3	MiroCam®
Sensor	CMOS	CMOS
Light emitting diodes	4	6
Resolution	340 x 340	320 x 320
Field of view	156°	170°
Capture rate	2 to 6 frames per second (adaptative)	3 frames per second
Transmission	Radiofrequency communication	Human body electrical communication

CMOS: Complementary Metal Oxide Semiconductor

Figure 1. Examples of artificial intelligence-based outputs.

A. Pillcam<sup>®</sup> (Medtronic) image with angiectasia; True positive frame detection, true positive delineation of region of interest, true positive frame characterization.

B. Pillcam<sup>®</sup> (Medtronic) normal image, false positive frame detection.

C. Pillcam<sup>®</sup> (Medtronic) image with angiectasia; True positive frame detection, true positive delineation of region of interest, false negative characterization.

D. Pillcam<sup>®</sup> (Medtronic), false positive frame characterization. A red sport were described by the expert while a angiodysplsia were detected.

E. Pillcam<sup>®</sup> (Medtronic), false detection and characterization. The angiodysplasia was correctly detected, with good delineation and characterization. However an inflammatory lesion was also detected in the same place.

F. Mirocam<sup>®</sup>(Intromedic) image with angiectasia; True positive frame detection, true positive delineation of region of interest, true positive frame characterization.

G. Mirocam<sup>®</sup>(Intromedic) normal image, false positive frame detection.

H. Mirocam<sup>®</sup>(Intromedic)) image with angiectasia; True positive frame detection, false negative delineation of region of interest (the box is more than twice larger the angiectasia), false negative characterization.

I. Mirocam<sup>®</sup>(Intromedic), two angiodysplasias were described by the expert while only one were detected

J. Mirocam<sup>®</sup>(Intromedic) normal image, false positive frame detection. The expert described a bubble, while an angiodysplasia were detected.

