

Deep learning and capsule endoscopy: Automatic multi-brand and multi-device panendoscopic detection of vascular lesions

OPEN
ACCESS

Authors

Miguel Mascarenhas¹, Miguel Martins², João Afonso¹, Tiago Ribeiro¹, Pedro Cardoso¹, Francisco Mendes¹, Patrícia Andrade¹, Helder Cardoso¹, Miguel Mascarenhas-Saraiva³, João Ferreira⁴, Guilherme Macedo¹

Institutions

- 1 Gastroenterology, Centro Hospitalar Universitário de São João, Porto, Portugal
- 2 Gastroenterology, Hospital São João, Porto, Portugal
- 3 Gastroenterology, ManopH Gastroenterology Clinic, Porto, Portugal
- 4 Department of Mechanical Engineering., University of Porto Faculty of Engineering, Porto, Portugal

Key words

Small bowel endoscopy, Small intestinal bleeding, Endoscopy Upper GI Tract, Non-variceal bleeding, Endoscopy Lower GI Tract, Lower GI bleeding

received 31.5.2023

accepted after revision 21.12.2023

accepted manuscript online 2.1.2024

Bibliography

Endosc Int Open 2024; 12: E570–E578

DOI 10.1055/a-2236-7849

ISSN 2364-3722

© 2024. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Corresponding author

Dr. Miguel Mascarenhas, Centro Hospitalar Universitário de São João, Gastroenterology, Rua Oliveira Martins 104, 4200-427 Porto, Portugal
miguelmascarenhasaraiva@gmail.com

Additional material is available at
<https://doi.org/10.1055/a-2236-7849>

ABSTRACT

Background and study aims Capsule endoscopy (CE) is commonly used as the initial exam for suspected mid-gastrointestinal bleeding after normal upper and lower endoscopy. Although the assessment of the small bowel is the primary focus of CE, detecting upstream or downstream vascular lesions may also be clinically significant. This study aimed to develop and test a convolutional neural network (CNN)-based model for panendoscopic automatic detection of vascular lesions during CE.

Patients and methods A multicentric AI model development study was based on 1022 CE exams. Our group used 34655 frames from seven types of CE devices, of which 11091 were considered to have vascular lesions (angiectasia or varices) after triple validation. We divided data into a training and a validation set, and the latter was used to evaluate the model's performance. At the time of division, all frames from a given patient were assigned to the same dataset. Our primary outcome measures were sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), and an area under the precision-recall curve (AUC-PR).

Results Sensitivity and specificity were 86.4% and 98.3%, respectively. PPV was 95.2%, while the NPV was 95.0%. Overall accuracy was 95.0%. The AUC-PR value was 0.96. The CNN processed 115 frames per second.

Conclusions This is the first proof-of-concept artificial intelligence deep learning model developed for pan-endoscopic automatic detection of vascular lesions during CE. The diagnostic performance of this CNN in multi-brand devices addresses an essential issue of technological interoperability, allowing it to be replicated in multiple technological settings.

Introduction

The development of capsule endoscopy (CE) enhanced examination of the small bowel [1]. Today, it is commonly used as

the initial exam in situations of suspected mid-gastrointestinal bleeding, after normal upper and lower endoscopy [2]. It is minimally invasive, has a higher diagnostic yield than other noninvasive methods, and it has proven to be cost-effective in

these clinical scenarios [3,4,5]. Nonetheless, it is time-consuming and error-prone, because crucial frames might be overlooked, especially if there are just a few of them [6].

Vascular lesions are the most common cause of gastrointestinal bleeding, not only in the small bowel, but also in other locations [7]. Although the assessment of the small bowel is the primary focus of CE, detection of upstream or downstream lesions in other areas of the gastrointestinal system may also be clinically significant. In fact, because the majority of gastrointestinal bleeding occurs beyond the duodenal ampulla or distal to the ileocecal valve, it might be considered a second examination of the upper or lower digestive tract, especially if the initial examination did not yield conclusive results [7].

The introduction of dual-camera capsule has prompted to discuss a CE-based panendoscopic evaluation of the digestive tract, especially for colorectal cancer screening and for Crohn's assessment [8,9]. It would be beneficial if CE allowed for complete assessment of the whole digestive tract, excluding all potential vascular lesions while avoiding repeat exams. Nonetheless, it is associated with an even greater reading burden which, along with its significant cost and lack of experience in the majority of centers, may limit its use in clinical practice, mainly in low-volume ones [9].

Convolutional neural networks (CNNs) have revolutionized image pattern recognition. This type of deep learning technology was inspired by the neural architecture of the human cortex and it emulates the neurobiological process of accomplishing complex tasks combining multiple layers of interconnected neurons [10]. Many articles have been published using this type of artificial intelligence (AI) system in different image-based procedures, including CE. Currently, there is published research in the field of automatic detection of vascular lesions during CE, in the small bowel, which has high overall accuracy [11, 12, 13, 14, 15]. These algorithms can not only identify different types of vascular lesions (red spots, angiectasia and/or varices), but also predict their likelihood of bleeding, according to Saurin classification [14,16]. Nonetheless, no studies on panendoscopic assessment of vascular lesions have been published.

This study aimed to develop and test a CNN-based algorithm for panendoscopic automatic detection of vascular lesions during CE.

Patients and methods

Study design

A multicenter retrospective cohort study was conducted in two different centers (Centro Hospitalar Universitário de São João and ManopH Gastroenterology Clinic, both in Porto, Portugal), which included 1188 CE and colon CE (CCEs) which were performed between June 2011 and August 2023.

The project was developed without direct intervention on patients; therefore, their clinical management was not affected. To protect patient identity, identifying information was omitted and random numbers were allocated to each one. A legal team with Data Protection Officer (DPO) certification (Maastricht University) ensured data protection, regarding to

its non-traceability as well as compliance with General Data Protection Regulation.

Capsule endoscopy protocol

Seven different CE devices were used for CE procedures: PillCam COLON (Medtronic Corp., Minneapolis, Minnesota, United States), PillCam Crohn's Capsule (Medtronic Corp., Minneapolis, Minnesota, United States), PillCam SB1 (Medtronic Corp., Minneapolis, Minnesota, United States), PillCam SB3 (Medtronic Corp., Minneapolis, Minnesota, United States), OMOM HD Capsule (JINSHAN Co., Yubei, Chongqing, China), Olympus Endocapsule 10 (Olympus Corp., Tokyo, Japan), and MiroCam (Intromedic Corp., Seoul, South Korea). PillCam COLON 1, PillCam Crohn's, PillCam SB1 and PillCam SB3 images were examined with PillCam Software version 9 (Medtronic, Minneapolis, Minnesota, United States), while OMOM HD images were reviewed with Vue Smart Software (Jinshan Science & Technology Co, Chongqing, Yubei, China), Olympus with EC-10 System (Olympus) and MiroCam with MiroView Software. To protect patient identification, image processing was used to erase personal information (name, operation number, and procedure date). Each frame was then labeled with a sequential number.

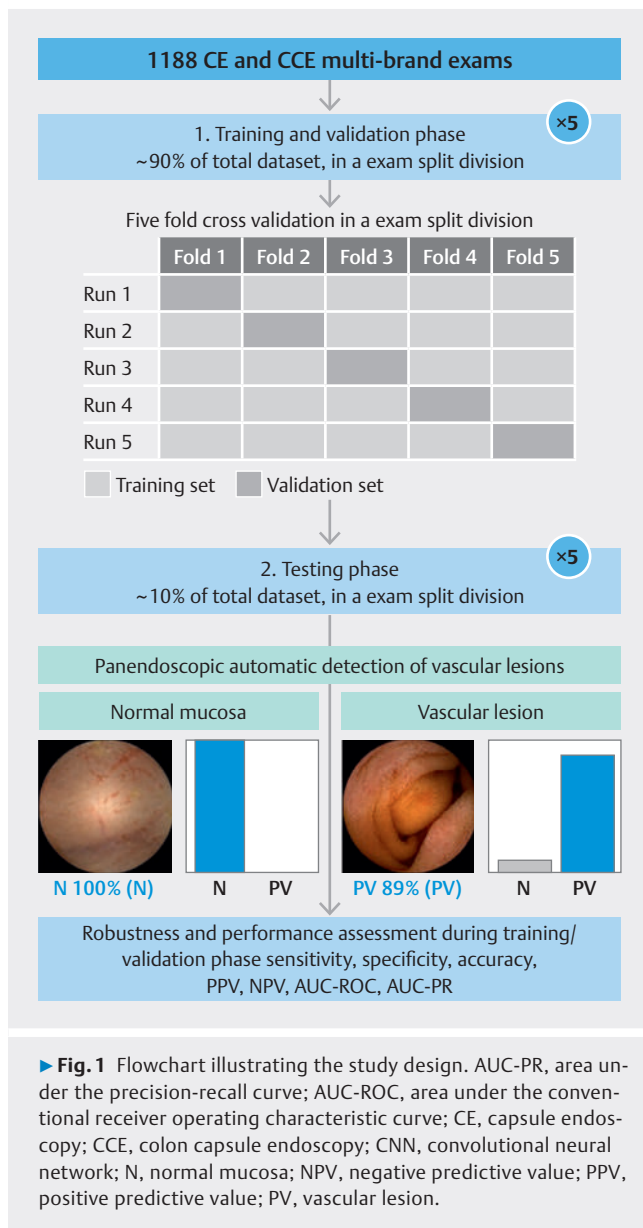
The European Society of Gastrointestinal Endoscopy recommendations were followed for bowel preparation [7]. Patients were advised to have a clear liquid diet the day before taking the capsule, and to fast the night before the examination. Prior to ingestion, patients underwent bowel preparation, which involved taking 2 L of polyethylene glycol (PEG) solution. For the PillCam Crohn's capsule, patients were given 2 L of PEG solution the night before the procedure and another 2 L on the morning of the procedure. An anti-foaming agent, namely simethicone, was used, and if the capsule remained in the stomach for more than 1 hour after ingestion (which implied image review on the patient's data recorder), domperidone 10 mg was given.

Categorization of lesions

The existence of vascular lesions, defined as angiectasias (tortuous and clustered capillary dilatations, resulting in well-defined brilliant red lesions) or varices (elevated venous dilatations with serpiginous appearance), was subsequently assessed in each frame. The images were separated into two groups: those with normal mucosa and those with vascular lesions. A consensus among three experienced gastroenterologists in CE was required for the final inclusion of each frame. A total of 152,312 frames, from seven types of CE devices, were used to develop the CNN, of which 14,942 contained pleomorphic vascular lesions.

Development of the CNN and performance analysis

We constructed a deep learning CNN to automatically detect vascular lesions, allowing for panendoscopic assessment of the presence of this lesion throughout the gastrointestinal system. This was accomplished by a two-step process. First, we used 90% of the dataset to perform a 5-fold cross-validation, during training and validation, to ascertain the robustness and assess the global performance of the CNN. Second, the remaining 10% of the dataset was used for testing with the average model



resulting from the five training sessions of the cross-validation. During this phase, the test set was used to screen eventual discrepancies in the algorithm. The whole process was iterated five times in total, using different combinations. ▶ **Fig. 1** shows a graphical flowchart of the research design.

The CNN was built using the RegNetY model [17]. Weights between units were trained using ImageNet, a large-scale image dataset created for object software recognition. We kept its convolutional layers in order to transfer its learning to our model. We deleted the last fully connected layers from our own classifier of dense and dropout layers. Each of the two blocks we utilized had completely connected layers first, followed by dropout layers with a 0.2 drop rate. After that, we added a dense layer the size of which was defined the number of classification groups (two: normal or vascular lesions). Trial and error were used to determine the learning rate (ranging between 0.0000625 and 0.0005), batch size (128) and the num-

ber of epochs (20). PyTorch and scikit libraries were used to prepare the model. During training standard data augmentation techniques, such as picture rotations and mirroring were used. A 2.1 GHz Intel Xeon Gold 6130 processor (Intel, Santa Clara, California, United States) and double NVIDIA Quadro RTX 80000 graphic processing unit (NVIDIA Corp, Santa Clara, California, United States) powered the computer.

The algorithm calculated the probability of each frame being considered normal and the probability of having a vascular lesion. Each frame was assigned to one of the aforementioned categories, and the one with the highest probability was selected (**Supplementary Fig. 1**). We generated heatmaps to identify the features that contributed the most to the CNN prediction (▶ **Fig. 2**). The algorithm's final classification was compared with the equivalent evaluation supplied by the three expert gastroenterologists, with the latter considered the gold standard.

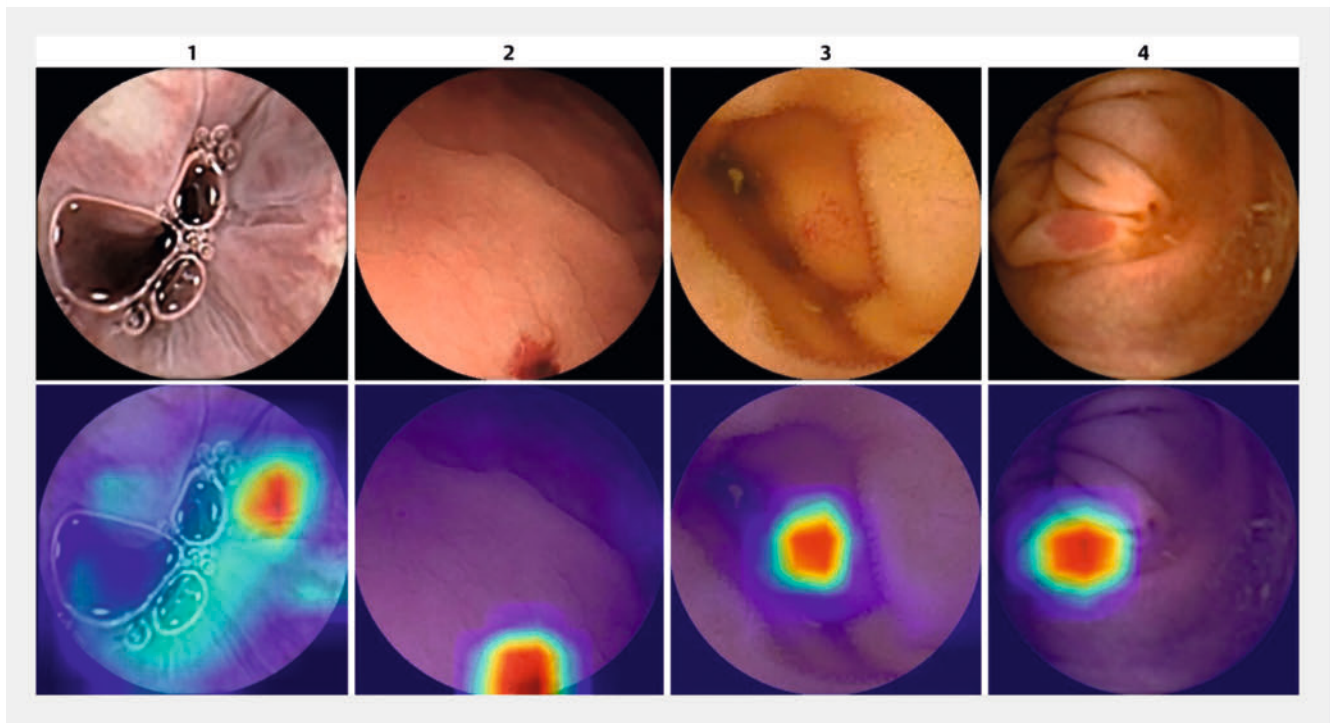
Training and validation dataset

First, we performed a 5-fold cross-validation, to assess robustness and assess the global performance of the CNN. From the total dataset, 90% of data ($n = 1069$ exams) was divided 5-fold with equivalent dimensions. For this division, we utilized the StratifiedGroupKFold method, ensuring that images from the same procedure were grouped together within each fold, while also ensuring that lesions were diversely represented. We conducted a total of five separate runs. In each of these runs, four folds were designated to train the model, while the remaining one was used to validate it. The folds used to train and validate the CNN changed within each run. This process was iterated a total of five times. ▶ **Table 1** lists the number of frames, patients, devices, regions (esophagus, stomach, small bowel and colon), and pleomorphic vascular lesions contained in each fold. ▶ **Table 2** lists the number of exams and corresponding number of frames for each device during the 5-fold cross-validation experiment and the test set.

After completing five iterations of this 5-fold cross-validation, we calculated the mean sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV). We also computed the mean area under the conventional receiver operating characteristic (AUC-ROC) curve and area under the precision-recall curve (AUC-PR) for each one. We chose to calculate booths (precision-recall calculation, in addition to conventional ROC curve), due the higher proportion of normal mucosa frames (true negatives) over frames containing vascular lesions (true positives), which could lead to a misinterpretation of the ROC curve [18].

Test dataset

In a second step, the testing phase involved the remaining 10% of the dataset ($n = 119$ exams), employing the average model resulting from the five training sessions in the cross-validation. Frames from a single exam were assigned to either the training/validation or testing set, ruling out the possibility of their inclusion in both. We repeated this process in five iterations, with different random combinations. In this phase, the algorithm



► **Fig. 2** Examples of generated heatmaps showing how CNN distinguishes a vascular lesion. 1-esophagus, 2-stomach, 3-small bowel, 4-colon.

was scrutinized for potential discrepancies through the examination of the independent test set.

During the testing phase, we calculated the mean sensitivity, specificity, accuracy, NPV and PPV. In addition, the mean AUC-ROC and AUC-PR were also calculated. Furthermore, we determined the CNN computational performance through the determination of processing time of all frames within the test set. We performed statistical analysis using scikit-learn v0.22.2 [19]. All the outcomes derived from this five-iteration process are presented as means along with their respective 95% confidence intervals.

Results

A total of 1188 exams were included for the development and testing of the CNN, 1097 corresponding to small bowel CE (PillCam SB3, n=941 OMOM HD Capsule, n=152; Olympus Endocapsule, n=1; PillCam SB1, n=2; MiroCam, n=1), and 210 to devices allowing CCE (PillCam Crohn's Capsule, n=192; PillCam COLON, n=18). From these exams, 152312 images were ultimately validated and incorporated in the dataset, from which 14942 showed vascular lesions (angiectasias or varices).

Training and validation dataset

► **Table 3** shows the results obtained from the five iterations of this five-fold cross-validation experiment. The mean sensitivity was 87.5% (IC95% 81.5–93.6%) and median specificity was 99.5% (IC95% 99.3–99.7%). The mean PPV and NPV were 94.9% (IC95% 93.1–96.8%) and 98.6% (IC95% 98.0–99.3%), respectively. Mean global accuracy was 98.4% (IC95% 97.7–

99.1 %). The mean AUC-ROC was 0.987 (IC95% 0.980–0.995) while the mean AUC-PR was 0.998 (IC95% 0.997–1.000) (► **Fig. 3**).

Test set

The testing dataset comprised an independent group of images (10% of the full dataset). **Supplementary Table 1** displays the metrics of the test set for each of the five iterations performed.

The model's mean sensitivity and specificity were 72.8% (IC95% 55.8–89.6%) and 99.0% (IC95% 98.5–99.5%), respectively. The PPV was 83.3% (IC95% 76.1–90.4%), while the NPV was 97.8% (IC95% 96.0–99.8%). The algorithm's overall accuracy was 97.0 (IC95% 94.8–99.2%). The AUC-ROC value was 0.984 (IC95% 0.9772–0.991), while AUC-PR value was 1.000.

There were two main reasons why the model was typically incorrect: existence of large air bubbles and inadequate cleansing during CE (**Supplementary Fig. 2**).

The CNN algorithm processed each frame in 26±3 milliseconds.

Discussion

This was the first study to evaluate the application of AI deep learning models in panendoscopic automatic detection of vascular lesions, not only in the small intestine, but also in other gastrointestinal topographies. This model not only performed well in all of the evaluated outcomes, but the results also suggest that it could possibly be used effectively with various types of CE devices. We believe that these results are promising and might contribute to implementation of AI-assisted panendo-

► **Table 1** Number of frames, patients, devices, regions (esophagus, stomach, small bowel and colon) and pleomorphic vascular (PV) lesions presented within each fold, during the 5-fold cross-validation experiment (five iterations total) and the test set (five iterations total).

		Frames (n)	Patients (n)	Devices (n)	Regions (n)	PV lesions (n)
Iteration 1	Fold 1	30889	226	4	4	2634
	Fold 2	23848	222	4	4	5662
	Fold 3	33063	228	5	4	1834
	Fold 4	23324	237	5	4	2155
	Fold 5	28045	156	6	4	1516
	Test set	13143	119	4	4	1141
Iteration 2	Fold 1	23951	231	4	4	3898
	Fold 2	24024	225	4	4	2016
	Fold 3	31039	214	5	4	3459
	Fold 4	40274	245	3	4	3243
	Fold 5	21617	154	6	4	1665
	Test set	11407	119	5	4	661
Iteration 3	Fold 1	29035	231	5	4	1684
	Fold 2	24753	222	4	4	2658
	Fold 3	37927	233	4	4	4454
	Fold 4	17612	155	4	4	1428
	Fold 5	27551	228	5	4	3775
	Test set	15434	119	5	4	943
Iteration 4	Fold 1	40446	213	5	4	4490
	Fold 2	25285	219	4	4	2927
	Fold 3	26678	201	4	4	1109
	Fold 4	21997	207	4	4	3054
	Fold 5	22657	229	5	4	1883
	Test set	15249	119	6	4	1479
Iteration 5	Fold 1	27056	226	4	4	2323
	Fold 2	32663	226	6	4	2506
	Fold 3	28752	229	5	4	4955
	Fold 4	28836	234	5	4	3004
	Fold 5	18127	154	4	4	1511
	Test set	16878	119	4	4	643

PV, pleomorphic vascular lesions.

scopic CE in routine clinical practice, independent of the device brand.

There are a few methodologic details concerning this study that should be highlighted. Because each exam's frames were assigned to a single fold in the cross-validation experiment and to a single dataset (training or testing) in the subsequent phase of assessing CNN global performance, the risk of overfitting was reduced. When frames from the same patient are given to both groups, the probability of having similar images and producing

too accurate prediction measures increases. Exam split design improves the external validity of the results, as well as inclusion of CE exams from two distinct high-volume centers. In addition, the CNN was developed using frames from different types of CE devices, including not only one but even two camera capsules, which may improve its effectiveness in real-world clinical practice. Furthermore, in the 5-fold cross-validation experiment involving different patients and device distributions, the model demonstrated excellent median diagnostic performance me-

► **Table 2** Number of exams and corresponding number of frames for each device during the 5-fold cross-validation experiment (five iterations total) and the test set (five iterations total).

		Fold 1 Exams (frames)	Fold 2 Exams (frames)	Fold 3 Exams (frames)	Fold 4 Exams (frames)	Fold 5 Exams (frames)	Test set Ex- ams (frames)
Iteration 1	PillCam COLON 1	1 (28)	5 (1636)	4 (206)	3 (653)	1 (34)	2 (128)
	PillCam Crohn's	43 (14818)	24 (5557)	31 (6577)	33 (9154)	27 (11508)	17 (8056)
	PillCam SB1	0 (0)	0 (0)	0 (0)	1 (10)	1 (10)	0 (0)
	PillCam SB3	158 (12825)	169 (13721)	161 (25080)	166 (12576)	111 (15075)	88 (4834)
	OMON	24 (3218)	24 (2934)	31 (1197)	34 (931)	15 (1391)	12 (125)
	Olympus	0 (0)	0 (0)	0 (0)	0 (0)	1 (27)	0 (0)
	Mirocam	0 (0)	0 (0)	1 (3)	0 (0)	0 (0)	0 (0)
Iteration 2	PillCam COLON 1	3 (319)	4 (563)	4 (331)	0 (0)	3 (883)	2 (589)
	PillCam Crohn's	30 (9480)	41 (10735)	30 (8468)	41 (15776)	27 (10236)	6 (975)
	PillCam SB1	0 (0)	0 (0)	1 (10)	0 (0)	1 (10)	0 (0)
	PillCam SB3	161 (9618)	163 (11404)	153 (20322)	179 (24006)	101 (10242)	96 (8519)
	OMON	37 (4534)	17 (1322)	26 (1908)	25 (492)	21 (219)	14 (1321)
	Olympus	0 (0)	0 (0)	0 (0)	0 (0)	1 (27)	0 (0)
	Mirocam	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (3)
Iteration 3	PillCam COLON 1	5 (1146)	2 (178)	1 (150)	3 (56)	5 (1155)	0 (0)
	PillCam Crohn's	37 (6590)	33 (10704)	44 (14021)	20 (6703)	28 (11020)	13 (6632)
	PillCam SB1	1 (10)	0 (0)	0 (0)	0 (0)	1 (10)	0 (0)
	PillCam SB3	160 (19723)	160 (12180)	162 (21353)	107 (9918)	170 (12323)	94 (8614)
	OMON	28 (1566)	27 (1691)	26 (2403)	25 (935)	24 (3043)	10 (158)
	Olympus	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (27)
	Mirocam	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (3)

trics. This implies that the CNN performance remains robust, regardless of the type of CE device employed. The development of a proficient deep learning model with this many (seven) different brands of CE devices marks a noteworthy achievement

which, to the best of our knowledge, has not previously been documented. Addressing this important interoperability barrier may increase the technology readiness level (TLR), allowing

► **Table 2** (Continuation)

		Fold 1 Exams (frames)	Fold 2 Exams (frames)	Fold 3 Exams (frames)	Fold 4 Exams (frames)	Fold 5 Exams (frames)	Test set Ex- ams (frames)
Iteration 4	PillCam COLON1	1 (588)	3 (453)	1 (150)	6 (1333)	4 (152)	1 (9)
	PillCam Crohn's	35 (15992)	33 (10215)	27 (9392)	26 (5951)	35 (8600)	19 (5520)
	PillCam SB1	1 (10)	0 (0)	0 (0)	0 (0)	1 (10)	0 (0)
	PillCam SB3	154 (19985)	153 (13481)	149 (14536)	155 (14254)	158 (13158)	84 (8697)
	OMON	22 (3871)	30 (1136)	24 (2600)	20 (459)	31 (737)	13 (993)
	Olympus	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (27)
	Mirocam	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (3)
Iteration 5	PillCam COLON1	2 (408)	4 (1193)	3 (257)	4 (696)	2 (130)	1 (1)
	PillCam Crohn's	44 (14649)	30 (10093)	38 (13128)	30 (7186)	21 (7960)	12 (2654)
	PillCam SB1	0 (0)	0 (0)	1 (10)	1 (10)	0 (0)	0 (0)
	PillCam SB3	163 (11546)	161 (20614)	158 (10535)	169 (17910)	110 (9579)	92 (13927)
	OMON	17 (453)	29 (733)	29 (4822)	30 (3034)	21 (458)	14 (296)
	Olympus	0 (0)	1 (27)	0 (0)	0 (0)	0 (0)	0 (0)
	Mirocam	0 (0)	1 (3)	0 (0)	0 (0)	0 (0)	0 (0)

► **Table 3** Five-fold cross-validation with exam split (repeated a total of five iterations).

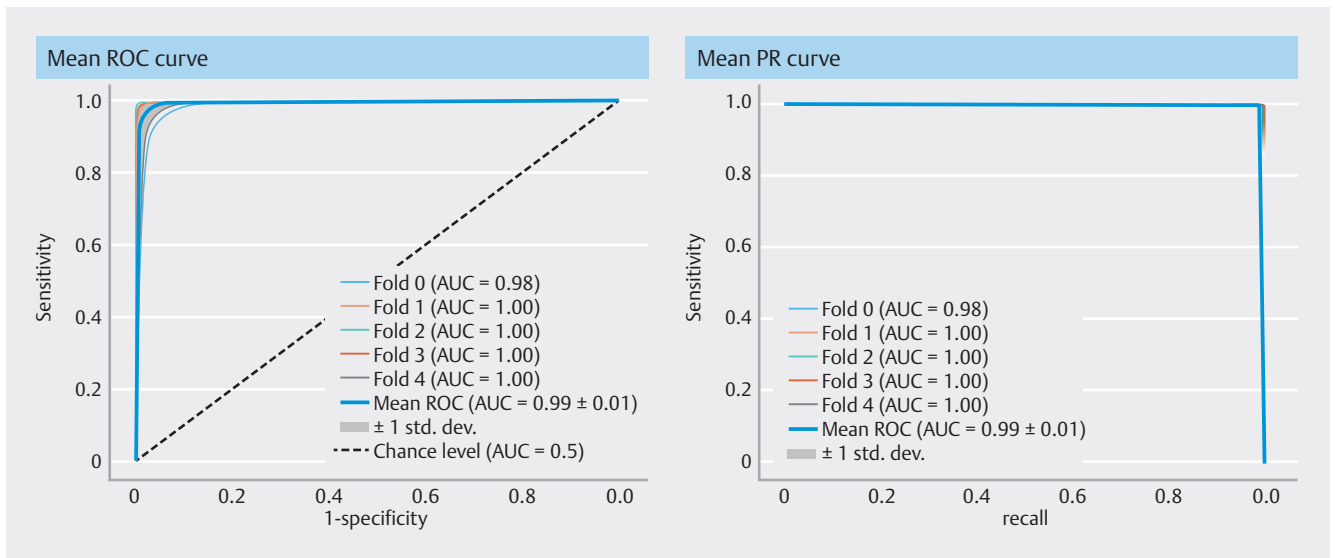
	Sensitivity %	Specificity %	PPV %	NPV %	Accuracy %	AUC-ROC	AUC-PR
Iteration 1	87.9	99.4	93.4	98.3	98.1	0.984	0.998
Iteration 2	91.1	99.5	95.8	98.9	98.5	0.990	0.998
Iteration 3	93.3	99.7	96.1	99.4	99.2	0.996	1.000
Iteration 4	83.2	99.4	92.8	98.1	97.7	0.986	0.998
Iteration 5	82.0	99.7	96.2	98.4	98.2	0.980	0.998

AUC-PR, area under the precision-recall curve; AUC-ROC, area under the conventional receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value.

for earlier implementation of AI-assisted gastroenterology procedures into routine clinical practice.

The study has some limitations. First, it was conducted retrospectively, which may introduce selection bias because the studied sample may not be as representative as it should be. Second, the study included a relatively small number of frames, mainly in the test dataset, which could also compromise the ex-

ternal validity of our findings. To corroborate these results, prospective and multicenter studies are required before introducing these deep learning models in clinical practice. Third, achieving excellent performance outcomes with still frames may not guarantee comparable performance with video segments or full-length videos. Nonetheless, we hypothesize that the algorithm computational performance, with a reading rate



► **Fig. 3** Representative example of an area under the conventional receiver operating characteristic curve (AUC-ROC) (1) and an area under the precision-recall curve (AUC-PR) (2) of CNN performance in detection of vascular lesions in iteration 3 of the training/validation phase. Precision (on the y axis), also known as positive predictive value, is related to the proportion of cases in which the CNN algorithm was correct. Recall (on the x axis), also known as sensitivity, is related to the proportion of frames containing vascular lesion that were retrieved by the CNN model. A higher precision indicates a lower false-positive rate, whereas a higher recall means a lower false-negative rate. The higher the precision and recall, the bigger the AUC-PR.

of approximately 38 frames per second, gives it the capacity to adjust to real-life settings. Although our results look promising, more studies are needed to determine whether the use of AI models is cost-effective. Fourth, by comparing the performance metrics obtained from cross-validation during training/validation with those derived from the testing set, we observed a slight decrease in sensitivity and PNV in the latter. This discrepancy could be attributed to various factors. On the one hand, despite our efforts to mitigate overfitting during training, it cannot be entirely ruled out. On the other hand, differences in representation between the validation and test sets may also contribute to this variation.

Research on AI and CE is increasing exponentially. However, most studies focus on the development of deep learning models in automatic identification of a specific type of lesion in either the small bowel or the colon. In the small bowel, there are very accurate deep learning models capable of detecting different types of vascular lesions, as well as predicting their bleeding risk accuracy [14]. In the colon, although the vast majority of retrospective studies focus on the detection of protruding lesions, there are already published AI algorithms not only for automatic detection of blood or hematic residues [20]. In addition, there is also a published trinary network aiming to detect and differentiate blood from normal colonic mucosa and from mucosa lesions (including ulcers and erosions, vascular lesions and protruding lesions) with high sensitivity, specificity, and accuracy [21].

Panendoscopic evaluation of the entire gastrointestinal tract is still in the developmental stage, even though it has a wide range of potential and exponential growth is anticipated in it. To our knowledge, there are no published papers that reporting

on development of a deep learning algorithm to detect vascular lesions, not only in the small bowel and colon, but also in the esophagus and stomach, allowing a true panendoscopic evaluation of the entire digestive tract mucosa. This may be important in clinical practice in patients who present with overt gastrointestinal bleeding. Our results demonstrated not only exceptional CNN robustness, but also high global performance levels with 98% overall accuracy, supporting AI use in a live healthcare practice environment.

Conclusions

In conclusion, this was the first proof-of concept AI deep learning model, worldwide, that was developed and validated for panendoscopic automatic detection of vascular lesions during CE. The high diagnostic performance of this CNN in multibrand devices addresses an important issue of technological interoperability, allowing it to be replicated in multiple technological settings. The enhancement in diagnostic efficiency of CE provided by AI, combined with increased interest in minimally invasive techniques, may contribute to increased access to this diagnostic method, thus promoting its performance when a purely diagnostic endoscopic exploration is expected.

Conflict of Interest

João Ferreira is a paid employee of DigestAID. The remaining authors have no conflict of interest to declare.

References

- [1] Triester SL, Leighton JA, Leontiadis GI et al. A meta-analysis of the yield of capsule endoscopy compared to other diagnostic modalities in patients with obscure gastrointestinal bleeding. *Am J Gastroenterol* 2005; 100: 2407–2418
- [2] Liao Z, Gao R, Xu C et al. Indications and detection, completion, and retention rates of small-bowel capsule endoscopy: a systematic review. *Gastrointest Endosc* 2010; 71: 280–286 doi:10.1016/j.gie.2009.09.031
- [3] Teshima CW, Kuipers EJ, van Zanten SV et al. Double balloon enteroscopy and capsule endoscopy for obscure gastrointestinal bleeding: an updated meta-analysis. *J Gastroenterol Hepatol* 2011; 26: 796–801 doi:10.1111/j.1440-1746.2010.06530.x
- [4] Marmo R, Rotondano G, Rondonotti E et al. Capsule enteroscopy vs. other diagnostic procedures in diagnosing obscure gastrointestinal bleeding: a cost-effectiveness study. *Eur J Gastroenterol Hepatol* 2007; 19: 535–542 doi:10.1097/MEG.0b013e32812144dd
- [5] Otani K, Watanabe T, Shimada S et al. Clinical utility of capsule endoscopy and double-balloon enteroscopy in the management of obscure gastrointestinal bleeding. *Digestion* 2018; 97: 52–58 doi:10.1159/000484218
- [6] Wang A, Banerjee S, Barth BA et al. Wireless capsule endoscopy. *Gastrointest Endosc* 2013; 78: 805–815 doi:10.1016/j.gie.2013.06.026
- [7] Rondonotti E, Spada C, Adler S et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Technical Review. *Endoscopy* 2018; 50: 423–446 doi:10.1055/a-0576-0566
- [8] Cortegoso Valdivia P, Elosua A, Houdeville C et al. Clinical feasibility of panintestinal (or panenteric) capsule endoscopy: a systematic review. *Eur J Gastroenterol Hepatol* 2021; 33: 949–955 doi:10.1097/MEG.0000000000002200
- [9] Ribeiro T, Fernández-Urien I, Cardoso H. Colon capsule endoscopy and artificial intelligence: a perfect match por panendoscopy. In: Mascarenhas MCH, Macedo G. *Artificial Intelligence in Capsule Endoscopy*. Academic Press; 2023: 255–269 doi:10.1016/B978-0-323-99647-1.00007-1
- [10] AmishaMalik P, Pathania M et al. Overview of artificial intelligence in medicine. *J Family Med Prim Care* 2019; 8: 2328–2331 doi:10.4103/jfmpc.jfmpc_440_19
- [11] Leenhardt R, Vasseur P, Li C et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc* 2019; 89: 189–194
- [12] Noya F, Alvarez-Gonzalez MA, Benitez R. Automated angiodysplasia detection from wireless capsule endoscopy. *Annu Int Conf IEEE Eng Med Biol Soc* 2017; 2017: 3158–3161 doi:10.1109/EMBC.2017.8037527
- [13] Otani K, Nakada A, Kurose Y et al. Automatic detection of different types of small-bowel lesions on capsule endoscopy images using a newly developed deep convolutional neural network. *Endoscopy* 2020; 52: 786–791 doi:10.1055/a-1167-8157
- [14] Ribeiro T, Saraiva MM, Ferreira JPS et al. Artificial intelligence and capsule endoscopy: automatic detection of vascular lesions using a convolutional neural network. *Ann Gastroenterol* 2021; 34: 820–828 doi:10.20524/aog.2021.0653
- [15] Tsuboi A, Oka S, Aoyama K et al. Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. *Dig Endosc* 2020; 32: 382–390 doi:10.1111/den.13507
- [16] Saurin JC, Delvaux M, Gaudin JL et al. Diagnostic value of endoscopic capsule in patients with obscure digestive bleeding: blinded comparison with video push-enteroscopy. *Endoscopy* 2003; 35: 576–584
- [17] Radosavovic I, Kosaraju RP, Girshick RB et al. Designing Network Design Spaces. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 10425–10433
- [18] Fu GH, Yi LZ, Pan J. Tuning model parameters in class-imbalanced learning with precision-recall curve. *Biom J* 2019; 61: 652–664 doi:10.1002/bimj.201800148
- [19] Pedregosa F, V. Michel, Thirion B et al. Scikit-learn: Machine Learning in Python. *J Mach Learn* 2011; 12: 2825–2830
- [20] Mascarenhas Saraiva M, Ferreira JPS, Cardoso H et al. Artificial intelligence and colon capsule endoscopy: automatic detection of blood in colon capsule endoscopy using a convolutional neural network. *Endosc Int Open* 2021; 9: E1264–E1268
- [21] Mascarenhas M, Ribeiro T, Afonso J et al. Deep learning and colon capsule endoscopy: automatic detection of blood and colonic mucosal lesions using a convolutional neural network. *Endosc Int Open* 2022; 10: E171–E177 doi:10.1055/a-1675-1941