

# Deep learning and minimally invasive inflammatory activity assessment: a proof-of-concept study for development and score correlation of a panendoscopy convolutional network

Pedro Cardoso\*, Miguel Mascarenhas\*<sup>ID</sup>, João Afonso, Tiago Ribeiro, Francisco Mendes, Miguel Martins, Patrícia Andrade<sup>ID</sup>, Hélder Cardoso, Miguel Mascarenhas Saraiva, João P.S. Ferreira and Guilherme Macedo

## Abstract

**Background:** Capsule endoscopy (CE) is a valuable tool for assessing inflammation in patients with Crohn's disease (CD). The current standard for evaluating inflammation are validated scores (and clinical laboratory values) like Lewis score (LS), Capsule Endoscopy Crohn's Disease Activity Index (CECDAI), and ELIAKIM. Recent advances in artificial intelligence (AI) have made it possible to automatically select the most relevant frames in CE.

**Objectives:** In this proof-of-concept study, our objective was to develop an automated scoring system using CE images to objectively grade inflammation.

**Design:** Pan-enteric CE videos (PillCam Crohn's) performed in CD patients between 09/2020 and 01/2023 were retrospectively reviewed and LS, CECDAI, and ELIAKIM scores were calculated.

**Methods:** We developed a convolutional neural network-based automated score consisting of the percentage of positive frames selected by the algorithm (for small bowel and colon separately). We correlated clinical data and the validated scores with the artificial intelligence-generated score (AIS).

**Results:** A total of 61 patients were included. The median LS was 225 (0–6006), CECDAI was 6 (0–33), ELIAKIM was 4 (0–38), and SB\_AIS was 0.5659 (0–29.45). We found a strong correlation between SB\_AIS and LS, CECDAI, and ELIAKIM scores (Spearman's  $r=0.751$ ,  $r=0.707$ ,  $r=0.655$ ,  $p=0.001$ ). We found a strong correlation between LS and ELIAKIM ( $r=0.768$ ,  $p=0.001$ ) and a very strong correlation between CECDAI and LS ( $r=0.854$ ,  $p=0.001$ ) and CECDAI and ELIAKIM scores ( $r=0.827$ ,  $p=0.001$ ).

**Conclusion:** Our study showed that the AI-generated score had a strong correlation with validated scores indicating that it could serve as an objective and efficient method for evaluating inflammation in CD patients. As a preliminary study, our findings provide a promising basis for future refining of a CE score that may accurately correlate with prognostic factors and aid in the management and treatment of CD patients.

## Plain language summary

### Artificial intelligence in Crohn's disease: the development of an automated score for disease activity evaluation

This study introduces an innovative AI-based approach to evaluate Crohn's Disease. The AI system automatically analyzes images from capsule endoscopy, focusing on finding ulcers and erosions to measure disease activity. The research reveals a robust correlation

Ther Adv Gastroenterol

2024, Vol. 17: 1–11

DOI: 10.1177/  
17562848241251569

© The Author(s), 2024.  
Article reuse guidelines:  
sagepub.com/journals-  
permissions

Correspondence to:

**Miguel Mascarenhas Saraiva**

Department of  
Gastroenterology, São  
João University Hospital,  
Alameda Professor  
Hernâni Monteiro, 4200-  
427 Porto, Portugal

WGO Training Center,  
Porto, Portugal

Faculty of Medicine of  
the University of Porto,  
Alameda Professor  
Hernâni Monteiro, 4200-  
427 Porto, Portugal

**miguelmascar  
enhassaraiva@gmail.com**

**Pedro Cardoso  
Miguel Mascarenhas**

**João Afonso**

**Tiago Ribeiro**

**Francisco Mendes**

**Miguel Martins**

**Patrícia Andrade**

**Hélder Cardoso**

**Guilherme Macedo**

Department of  
Gastroenterology, São  
João University Hospital,  
Porto, Portugal

WGO Training Center,  
Porto, Portugal

Faculty of Medicine of the  
University of Porto, Porto,  
Portugal

**João P.S. Ferreira**

Faculty of Engineering,  
University of Porto, Porto,  
Portugal

Institute of Science and  
Innovation in Mechanical  
and Industrial Engineering,  
Porto, Portugal

\*These authors  
contributed equally.

between the AI-generated score assessing inflammation in the small bowel and traditional clinical scores. This suggests that the AI solution could be a quicker and more consistent way to evaluate Crohn's Disease, speeding up the evaluation process and reducing manual scoring variability. While promising, the study acknowledges limitations and emphasizes the need for further validation with larger groups of patients. Overall, it represents a crucial step toward integrating AI into gastroenterology, offering a glimpse into a future of more objective and personalized Crohn's Disease evaluation.

**Keywords:** artificial intelligence, capsule endoscopy, Crohn's disease, inflammatory bowel disease, panendoscopy inflammatory assessment

Received: 24 August 2023; revised manuscript accepted: 14 April 2024.

### Introduction

Inflammatory bowel disease (IBD) is a chronic inflammatory disease of the gastrointestinal tract that affects millions of people worldwide. The assessment and management of IBD can be challenging, requiring a comprehensive evaluation of disease activity and response to treatment.

Capsule endoscopy (CE) is a minimally invasive procedure initially designed to assess the small bowel (SB), demonstrating a high diagnostic yield in detecting SB lesions.<sup>1,2</sup>

CE is typically safe and well tolerated, presenting few contraindications but a clinical evaluation of the risk of capsule retention is advisable. This is especially relevant for patients with established Crohn's disease (ECD), where the likelihood of retention is increased, particularly when known stricturing phenotype or obstructive symptoms are evident. A meta-analysis evaluated the risk of capsule retention in patients with suspected Crohn's disease (SCD) and ECD. The authors found a retention rate in the overall CD cohort of 3.3% (SCD = 2.3% and ECD = 4.6%). However, the concern for retention should not preclude CE utilization, as long as patients have undergone appropriate patency testing.<sup>3</sup>

The role of CE in patients with SCD and ECD has expanded over the years. Its applications encompass the diagnosis of SB CD, evaluation of disease activity and therapy response, objective assessment of mucosal healing, and detection of postoperative recurrence. Indeed, interest in evaluating the entire SB and colon has grown in recent years.<sup>4</sup>

The introduction of pan-enteric capsule endoscopy (PCE), enabling simultaneous enteric and

colonic evaluation, has raised expectations that this modality may allow a comfortable and accurate evaluation of the gastrointestinal involvement in IBD within a single examination.<sup>5,6</sup> Indeed, recent studies have emphasized the potential role of PCE in IBD patients, particularly its utility in monitoring patients with CD.<sup>7</sup>

CD, with its discontinuous nature and varied disease location, benefits from a pan-enteric approach. PCE, for example, PillCam Crohn (PCC)'s<sup>®</sup>, offers a convenient method for simultaneously assessing SB and colonic lesions, allowing for the evaluation of disease severity, extent, and distribution.<sup>8</sup> Numerous studies have evaluated the application of VCE in patients with CD and have consistently demonstrated that PCE achieves a high diagnostic yield for lesions detection throughout the entire gastrointestinal tract.<sup>9–12</sup>

To standardize reporting of CE exams and enhance reproducibility and inter-observer agreement, scoring systems have been developed.<sup>13</sup> Lewis's score (LS) assesses CD inflammation in the SB. It assesses ulcerations, erythema, and mucosal abnormalities, showing excellent inter-observer agreement and correlation with inflammatory markers, providing a standardized and objective evaluation of disease activity.<sup>14</sup> Similar to the Lewis score (LS), the Capsule Endoscopy Crohn's Disease Activity Index (CECD AI) evaluates inflammation in the SB using three main parameters: inflammation, extent of disease, and strictures. The score is separately calculated for proximal and distal segments. In contrast to previously mentioned scores, the ELIAKIM score is a novel PCC's capsule score that evaluates mucosal inflammation not only in the small

intestine but also in the colon. The entire bowel is divided into five segments, with the small intestine divided into three tertiles and the colon into right and left segments. The score takes into account the most common and severe lesions, disease extent, and the presence of strictures, providing a comprehensive assessment of CD activity.<sup>15</sup>

Traditional scoring systems, such as the LS, CECDAI, and ELIAKIM, have been used to evaluate disease activity in IBD. However, these scoring systems have limitations, including subjectivity and variability, and being a time-consuming task.

Over the past few years, significant efforts have been dedicated to developing and implementing artificial intelligence (AI) tools for automated image analysis in the field of gastroenterology.<sup>16</sup> Based on convolutional neural network (CNN) models, AI-powered systems have been developed to decrease reading times and improve lesion detection.<sup>17</sup> To this date, evidence regarding the application of deep learning modules to pan-endoscopy systems is still in its early stages and comes mainly from retrospective studies including a small number of patients and with limited datasets.<sup>18–22</sup>

Published as a proof-of-concept study by this group conducted by Ferreira *et al.*, a deep learning model was developed for automatic detection of both SB and colonic ulcers and erosions using PCC's® capsule images. This model achieved encouraging results for lesion detection with a sensitivity of 90% and a specificity of 96%.<sup>22</sup>

The main objective of this article is to explore the application of AI in assessing disease activity in IBD, including the development of an automated AI score based on CE images selected by CNN and comparing it with laboratory data and existing scoring systems.

## Materials and methods

### Study design

We performed a retrospective analysis of clinical data of patients with CD undergoing PCC's between September 2020 and January 2023 at a single center (São João University Hospital,

Porto, Portugal). The reporting of this study conforms to the Strengthening the Reporting of Observational Studies in Epidemiology statement.<sup>23</sup> In all, 61 were enrolled patients in this study. All procedures were recorded as a video file. Each full-length video was reviewed, and images retrieved from these examinations were used. These images comprised still frames extracted by the decomposition of each video. The segmentation of each video into frames was performed using dedicated video software (VLC media player, Paris, France). Each exam was analyzed by a pre-developed and validated CNN for automatic detection of ulcers and erosions in PCC images.<sup>22</sup>

A total of 1,659,175 frames of enteric and colonic mucosa were ultimately extracted. Each exam video was reviewed by three readers (PC, MM, and FM) and inflammation was scored in all studies using LS, CECDAI, and ELIAKIM. A final decision on each score attribution required the agreement of at least two of the three researchers.

Laboratory parameters of these patients were also collected, such as relative C-reactive protein (CRP) and fecal calprotectin (FCP).

### CE procedure

PCE procedures were conducted using the PCC system (Medtronic, Dublin, Ireland). The images were reviewed using PillCam™ software version 9.0 (Medtronic). Each frame was processed to remove any information allowing patient identification (name, operating number, date of procedure). The bowel preparation protocol followed previously published guidelines.<sup>24</sup>

### Development of the CNN

A developed CNN for automatic detection of ulcers or erosions in the enteric or colonic mucosa published in previous works was used.<sup>22</sup> From the collected pool of images ( $n=1,659,175$ ), 55,317 displayed ulcers and erosions. The remaining ( $n=1,603,858$ ) showed normal SB or colonic mucosa. For each image, CNN estimated the probability for each category: ulcers or erosions *versus* normal mucosa. The category with the highest probability score was outputted as the CNN's predicted classification.

### *Development of the automated AI score*

The videos were subsequently analyzed by software and image processing to detect ulcers and erosions. Based on the total number of images selected that display ulcers and erosions (in SB and colon, distinctively) and the total number of frames per video, a score was calculated. This score was determined by comparing the count of images with findings to the overall number of frames. The score reflects a fraction, giving an idea of the proportion of affected images relative to the total number of images in the video. The score was calculated for the SB (AIS\_Small Bowel), the colon (AIS\_Colon), and for both (AIS\_total). A potential challenge for the algorithm lies in the risk of counting the same ulcer or erosion multiple times, especially when the capsule crosses a segment repeatedly. To address this issue, cross-matching procedures were integrated into both the training and test datasets, effectively eliminating redundant images of identical ulcers. In addition, to reduce overfitting bias, patients were methodically divided between the training and validation datasets.

### *Model performance and statistical analysis*

Statistical analysis was performed with SPSS Version 29.0 statistic software package, developed by IBM Corporation, Armonk, NY, USA. Data were expressed as means. Variables minimum and maximum were calculated. For correlation analysis, Spearman correlation was used to examine the relationship between the variables of interest. A value of  $p < 0.01$  was considered statistically significant.

## Results

### *PCC characteristics*

In all, 61 patients CD patients underwent PCC examination. The capsule reached the cecum in all cases. The examination was considered completed (visualization of rectum or toilet) in 55/61 (90%). The bowel preparation was assessed in all cases and was considered satisfactory in the SB in all cases.

### *CNN characteristics*

The CNN for automatic detection of ulcers or erosions used a total of 1,659,175 images, 617,743 of enteric mucosa, and 986,115 of

colonic mucosa. It identified 20,787 ulcers or erosions of enteric mucosa and 34,530 of colonic mucosa. Figure 1 presents an example of the output of the CNN. Figure 2 presents the heatmaps produced by the neural network which are visual representations that highlight the important features in the input image, in this case, ulcers and erosions. These heatmaps help in interpreting CNN's decision-making process by highlighting these regions.

### *Capsule activity scores*

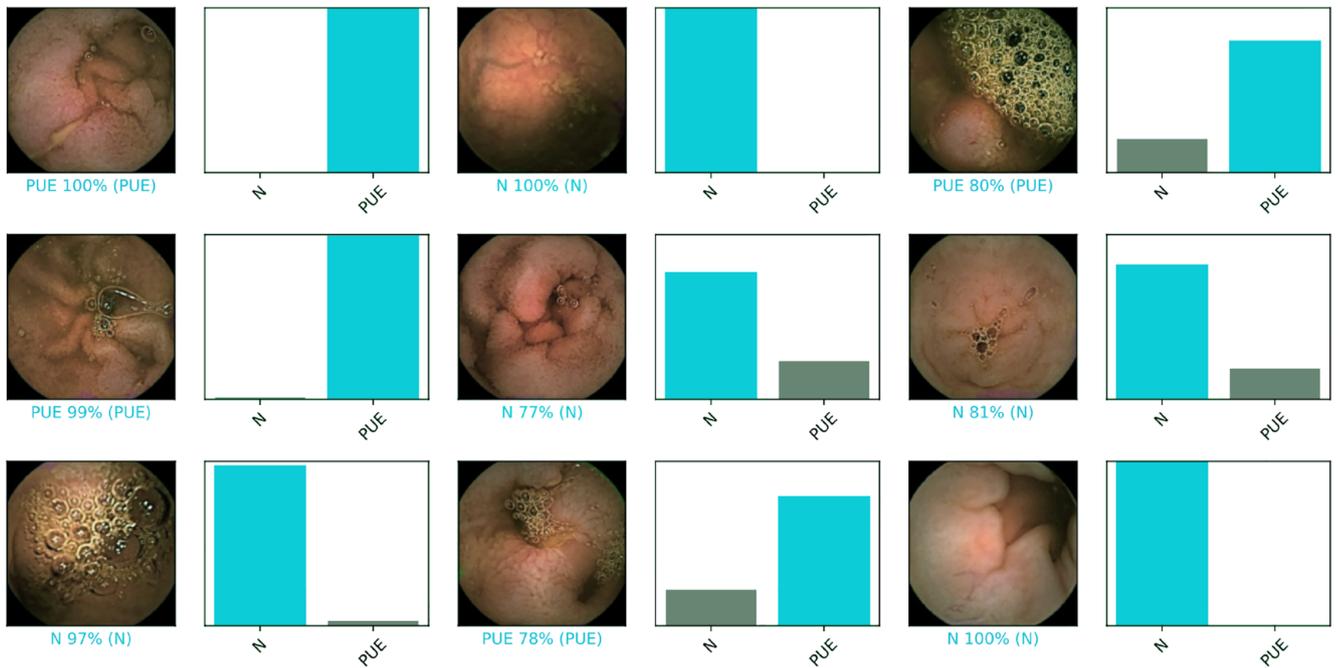
The median LS was 225 (0–6060), CECDAI was 6 (0–33), ELIAKIM was 4 (0–38), and AIS\_Small Bowel was 0.5659 (0–29.45). The median, minimum, and maximum values are shown in Table 1.

To compare the automated score generated by the CNN, we compared it with the laboratory data and with the already validated scores.

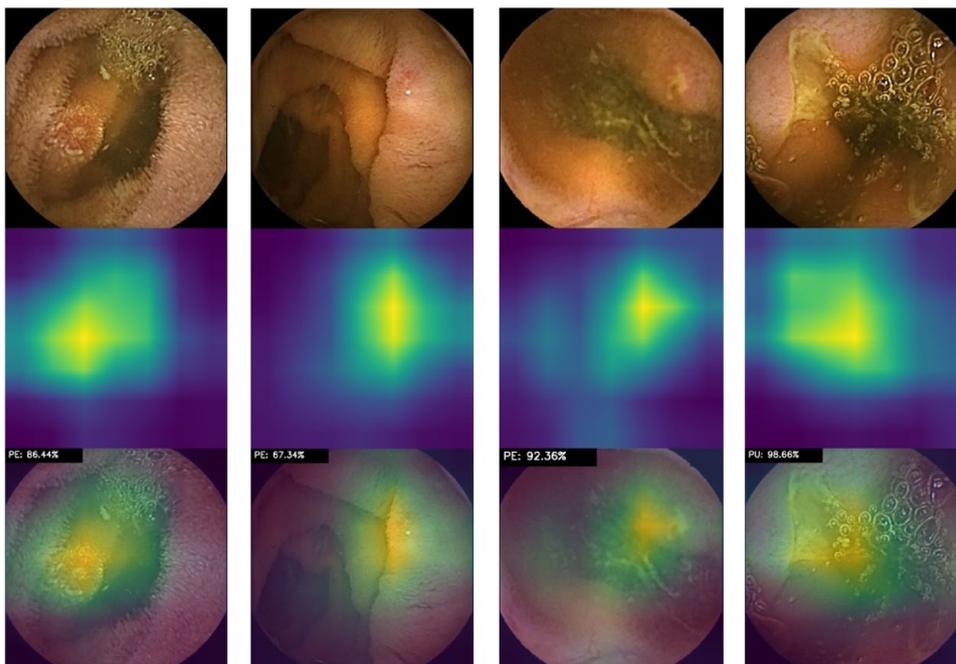
Since the variables were nonparametric in nature, we utilized Spearman correlation for the analysis. Regarding the interpretation, the Spearman coefficient varies between  $-1$  and  $1$  and the closer  $\rho$  is to  $-1$  or  $1$ , the stronger the correlation (values above  $0$  are positive). The authors considered the following range of values:  $\rho$   $0-0.19$  = very weak,  $0.2-0.39$  = weak,  $0.4-0.59$  = moderate,  $0.6-0.79$  = strong, and  $0.8-0.99$  very strong.<sup>25</sup> Table 2 shows the correlations between all the variables.

It is important to mention that due to the retrospective nature of the study, the timing of blood sample collection was not standardized for all patients. In fact, there was a considerable variation in the time interval between sample collection and CE. For this reason, the researchers established a reasonable minimum time frame of 3 months to ensure meaningful comparisons between these values and the findings in the PCC. This explains that the comparison between RCP and FCP was performed for only 36 and 35 patients, respectively.

We started to analyze the laboratory values like CRP and FCP. Regarding CRP, we found virtually no correlations with the other variables. In what concerns FCP, we found Spearman correlations close to  $0.5$  ( $p < 0.05$ ) with the known scores (low to moderated association).



**Figure 1.** Output obtained from the application of the CNN. A blue bar represents a correct prediction. The red bars represent an incorrect prediction. The category with the highest probability was outputted as CNN's prediction. CNN, convolutional neural network; N, normal mucosa; PUE, ulcers and erosions of the enteric and colonic mucosa.



**Figure 2.** Heatmaps produced by the CNN: highlighted regions of ulcers and erosions. CNN, convolutional neural network.

**Table 1.** Descriptive statistics of variables.

	Median	Min	Max
CRP (mg/dL)	7.53	0	29
FCP (µg/g)	364	1	1302
Lewis score	225	0	6060
CECDAI	6	0	33
ELIAKIM	4	0	38
AIS_Small Bowel	0.566	0	29.45
AIS_Colon	0.985	0	34.80
AIS_total	2.248	0	62.59

CECDAI, Capsule Endoscopy Crohn's Disease Activity Index; CRP, C-reactive protein; FCP, fecal calprotectin.

**Table 2.** Spearman rank correlation coefficient between variables.

	PCR	CPF	LS	CECDAI	ELIAKIM	AIS_Small Bowel	AIS_Colon	AIS_total
RCP								
<i>r</i>	1	0.056	0.155	0.178	0.295	0.026	0.254	0.313
<i>p</i>		0.750	0.368	0.298	0.081	0.881	0.135	0.063
FCP								
<i>r</i>	0.056	1	0.508**	0.520**	0.515**	0.347*	0.108	0.296
<i>p</i>	0.750		0.002	0.001	0.002	0.041	0.536	0.084
LS								
<i>r</i>	0.155	0.508**	1	0.854**	0.768**	0.751**	-0.059	0.472**
<i>P</i>	0.368	0.002		<0.001	<0.001	<0.001	0.650	<0.001
CECDAI								
<i>r</i>	0.178	0.520**	0.854**	1	0.827**	0.707**	-0.060	0.398**
<i>p</i>	0.298	0.001	<0.001		<0.001	<0.001	0.645	0.001
ELIAKIM								
<i>r</i>	0.295	0.515**	0.768**	0.827**	1	0.655**	0.103	0.517**
<i>p</i>	0.081	0.002	<0.001	<0.001		<0.001	0.430	<0.001
AIS_Small Bowel								
<i>r</i>	0.026	0.347*	0.751**	0.707**	0.655**	1	0.019	0.593**
<i>p</i>	0.881	0.041	<0.001	<0.001	<0.001		0.884	<0.001

(Continued)

**Table 2.** (Continued)

	PCR	CPF	LS	CECDAI	ELIAKIM	AIS_Small Bowel	AIS_Colon	AIS_total
AIS_Colon	PCR	CPF	LS	CECDAI	ELIAKIM	AIS_Small Bowel	AIS_Colon	
<i>r</i>	0.254	0.108	-0.059	-0.060	0.103	0.019	1	0.694**
<i>p</i>	0.135	0.536	0.650	0.645	0.430	0.884		<0.001
AIS_total	PCR	CPF	LS	CECDAI	ELIAKIM	AIS_Small Bowel	AIS_Colon	AIS_total
<i>r</i>	0.313	0.296	0.472**	0.398**	0.517**	0.593**	0.694**	1
<i>p</i>	0.063	0.084	<0.001	0.001	<0.001	<0.001	<0.001	
*Correlation is significant at the 0.05 level (two-tailed).								
**Correlation is significant at the 0.01 level (two-tailed).								
CECDAI, Capsule Endoscopy Crohn's Disease Activity Index; FCP, fecal calprotectin; LS, Lewis score.								

When analyzing the validated scores (LS, CECDAI, and ELIAKIM), we found that they correlate well. We found a very strong correlation between LS and CECDAI (Spearman's  $r=0.854$ ,  $p<0.001$ ), CECDAI and ELIAKIM scores (Spearman's  $r=0.827$ ,  $p<0.001$ ), and a strong correlation between LS and ELIAKIM (Spearman's  $r=0.768$ ,  $p<0.001$ ).

We found a strong correlation between SB\_AIS and LS, CECDAI, and ELIAKIM scores (Spearman's  $r=0.751$ ,  $r=0.707$ ,  $r=0.655$ ,  $p=0.001$ ). We did not find any correlation between the generated score for the colon segment (AIS\_Colon) and the other scores. For the score corresponding to both the SB segment and colon (AIS\_total), we found statistically significant correlations, although not very strong, in particular, a positive but weak correlation with LS (Spearman's  $r=0.472$ ,  $p<0.001$ ), CECDAI (Spearman's  $r=0.398$ ,  $p=0.001$ ), and a moderated positive correlation with ELIAKIM (Spearman's  $r=0.517$ ,  $p<0.001$ ).

## Discussion

The idea of a comprehensive pan-enteric examination, particularly for evaluating conditions like CD, arose with the advent of colon CE.<sup>26</sup> Given that PCE enables the assessment of the entire gastrointestinal tract, the concept of a single, minimally invasive panendoscopy has become an enticing prospect. PCE is a convenient and

accurate method for assessing both the SB and colon, reducing the need for separate evaluations.

As previously mentioned, there are several validated scores for evaluating disease activity in CE. The development of a new score is questioned when multiple validated scores already exist and is valid to think about the chance of developing an AI tool for swift computation of existing scores. Some are dedicated to assessing the mucosa of the small intestine (LS and CECDAI), while others provide a pan-enteric evaluation (ELIAKIM). However, the application of these scores involves certain challenges. In manual reading and scoring, we need to acknowledge subjectivity and variability among different observers. Furthermore, it is often a time-consuming task, especially the panendoscopy capsules (double headed cameras) in which a single exam can take anywhere from 30 to 90 min. As stated, the traditional scoring systems (LS and CECDAI) are dedicated to assessing only the small intestine mucosa, and this limited scope can make it challenging to obtain a comprehensive assessment of disease activity.

The establishment of scoring systems has indeed introduced standardized reporting, but we believe that there is space for improvement. The use of AI algorithms in assessing disease activity in IBD, in particular CD, has shown promise for accurate and less biased disease evaluation.<sup>27,28</sup> The use of CNN in CE has shown that

the implementation of these technologies may improve the accuracy, efficiency, and reproducibility of IBD evaluations.<sup>18–21</sup>

Based on a pre-developed and validated CNN,<sup>22</sup> we developed an automated score for disease activity assessment and compared it with clinical variables and scores. In this study, we compared that score with clinical laboratory markers and other established scores. While the other scores were based on villous appearance, edema, the presence of ulcers, strictures, stenosis, and the extent and severity of lesions, the score we created is based on the count of images with ulcers and erosions relative to the total number of frames in the video. This automated score provides an objective and standardized assessment of inflammation, reducing subjectivity and variability. In practice, the number of frames with lesions will, indirectly, reflect the extent and severity of the different lesions, as well as the presence of stenosis (more selected frames from delayed passage).

In this research, our primary method for estimating inflammatory burden centered on identifying ulcers and erosions, with a specific emphasis on quantifying the number of frames featuring those lesions. We acknowledge that this approach may lack complete accuracy, as other relevant endoscopic aspects, notably strictures, were not considered. However, the development of CNN for strictures identification presents complexity, given the impracticality of a frame-by-frame assessment. In future works, we will address this limitation by incorporating all relevant endoscopic features that could reflect inflammatory activity. Our goal is to develop an AI tool based on diverse endoscopic aspects across the entire gastrointestinal tract, aiming to automatically provide a comprehensive estimate of pan-enteric inflammatory activity. Although the application of additional landmarks poses technical challenges, this aspect is already undergoing development.

Another important point to emphasize is the reason behind the separate analysis of SB and colon. It is true that one of the objectives of this study is to perform a pan-endoscopic evaluation but the division into the SB and colon makes sense from a methodological perspective. In fact, considering the anatomical differences between the two segments, from the standpoint of neural network development, it is possible to create much more

accurate and robust CNNs by conducting a segmented evaluation as described. However, given the high reading rate of each CNN and the fact that both networks are coupled in the algorithm, this division does not interfere with the overall accuracy. Therefore, this methodological division is part of a data science strategy that guarantees the best performance of the network.

The results of the study showed a strong correlation between the automated AI score in the SB segment and validated scoring systems. This strong correlation may indicate that the AI-based core provides results that are consistent and aligned with the assessments made by well-established and validated scores which can mean that this score might serve as an effective and efficient alternative for evaluating CD activity. In addition, the speed at which the neural network evaluates thousands of frames can significantly speed up the evaluation process, saving time. Furthermore, the use of AI-assisted technologies eliminates inter-observer variability, leading to increased consistency and reproducibility.

This study lays a promising foundation for future research. The researchers are enthusiastic about exploring ways to enhance the score's performance for disease activity assessment, with the ultimate goal of creating a superior tool compared to existing scores. Further validation studies with larger and diverse patient cohorts are necessary to affirm this automated score's reliability and generalizability.

Despite the potential value of these technologies, we must acknowledge some limitations. First, the retrospective nature of this study limits its generalization potential. Second, the nature of this score is mainly quantitative (fraction of images with ulcers and erosions in relation to the total number of frames) which can limit the ability to capture the full spectrum of disease activity. In addition, this is the first study with the use of the score to assess the disease activity and therefore has a limited number of patients. It is essential to validate the AI score in a clinical setting in a prospective manner to determine its correlation with clinical symptoms, response to treatment, and the need for escalation or de-escalation of therapy. This validation would provide valuable insights into the clinical utility and reliability of the score in guiding clinical decision-making and optimizing patient care.<sup>29–31</sup>

Despite the acknowledged limitations, this study suggests that the development of an automated score holds significant potential as an objective and efficient tool for assessing disease activity in patients with CD. In fact, the use of AI in PCE for assessing inflammation in IBD patients is a promising field that has the potential to revolutionize disease management. Ultimately, the goal of the group is to develop clinical scores and interfaces that aggregate all patient information, including clinical status, laboratory values, endoscopic exams (including CE), radiological exams, and others, to optimize disease treatment and improve quality of life.<sup>27</sup> This approach would provide a comprehensive and integrated view of the patient's disease status, allowing for more informed clinical decision-making, toward personalized medicine.

### Conclusion

In conclusion, our study suggests that the application of AI tools to assess IBD, in particular CD, can have potential benefits. Our group explored the development of an automated score using CNN for CE image analysis. Preliminary findings indicate a correlation with established clinical scores, suggesting the possibility of objective disease activity evaluation. However, it is important to acknowledge uncertainties, and further research with consideration of more endoscopic features and larger cohorts is needed to determine the real-world effectiveness of this AI-based approach. These emerging technologies might improve the evaluation and management of CD patients, but their practical implications require careful consideration and validation.

### Declarations

#### *Ethics approval and consent to participate*

This study was approved by the ethics committee of Centro Hospitalar Universitário de São João/Faculdade de Medicina da Universidade do Porto (CE 41/2021) and was conducted respecting the declaration of Helsinki. This study is retrospective and of a non-interventional nature. Any information deemed to potentially identify the subjects was omitted.

#### *Consent for publication*

Not applicable.

#### *Author contributions*

**Pedro Cardoso:** Data curation; Investigation; Writing – original draft; Writing – review & editing.

**Miguel Mascarenhas:** Conceptualization; Data curation; Investigation; Methodology; Project administration; Supervision; Writing – original draft; Writing – review & editing.

**João Afonso:** Conceptualization; Data curation; Investigation.

**Tiago Ribeiro:** Data curation; Investigation.

**Francisco Mendes:** Data curation; Investigation.

**Miguel Martins:** Data curation; Investigation.

**Patrícia Andrade:** Supervision; Validation.

**Hélder Cardoso:** Supervision; Validation.

**Miguel Mascarenhas Saraiva:** Supervision; Validation.

**João P.S. Ferreira:** Formal analysis; Methodology; Software; Validation; Writing – review & editing.

**Guilherme Macedo:** Supervision.

#### *Acknowledgements*

None.

#### *Funding*

The authors received no financial support for the research, authorship, and/or publication of this article.

#### *Competing interests*

J.P.S.F. is currently an employee of DigestAID – Digestive Artificial Intelligence Development.

#### *Availability of data and materials*

Not applicable.

#### **ORCID iDs**

Miguel Mascarenhas  <https://orcid.org/0000-0002-0340-0830>

Patrícia Andrade  <https://orcid.org/0000-0003-1334-0947>

#### **References**

1. Odeyinka O, Alhashimi R, Thoota S, *et al.* The role of capsule endoscopy in Crohn's disease: a review. *Cureus* 2022; 14: e27242.

2. Lahat A and Veisman I. Capsule endoscopy in Crohn's disease – from a relative contraindication to habitual monitoring tool. *Diagnostics (Basel)* 2021; 11: 1737.
3. Pasha SF, Pennazio M, Rondonotti E, *et al.* Capsule retention in Crohn's disease: a meta-analysis. *Inflamm Bowel Dis* 2020; 26: 33–42.
4. Sakurai T, Omori T, Tanaka H, *et al.* Multicenter prospective registration study of efficacy and safety of capsule endoscopy in Crohn's disease in Japan (SPREAD-J study). *J Gastroenterol* 2023; 58: 1003–1014.
5. 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.
6. Na SY and Lim YJ. Capsule endoscopy in inflammatory bowel disease: when? To whom? *Diagnostics (Basel)* 2021; 11: 2240.
7. Skamnelos A, Lazaridis N, Vlachou E, *et al.* The role of small-bowel endoscopy in inflammatory bowel disease: an updated review on the state-of-the-art in 2021. *Ann Gastroenterol* 2021; 34: 599–611.
8. Annese V and Manetti N. Capsule endoscopy in Crohn's disease: is there enough light in the tunnel? *J Crohns Colitis* 2014; 8: 1598–1600.
9. Nam JH, Lee KH and Lim YJ. Examination of entire gastrointestinal tract: a perspective of mouth to anus (M2A) capsule endoscopy. *Diagnostics* 2021; 11: 1367.
10. Yamada K, Nakamura M, Yamamura T, *et al.* Diagnostic yield of colon capsule endoscopy for Crohn's disease lesions in the whole gastrointestinal tract. *BMC Gastroenterol* 2021; 21: 75.
11. Tamilarasan AG, Tran Y, Paramsothy S, *et al.* The diagnostic yield of pan-enteric capsule endoscopy in inflammatory bowel disease: a systematic review and meta-analysis. *J Gastroenterol Hepatol* 2022; 37: 2207–2216.
12. Tai FWD, Ellul P, Elosua A, *et al.* Panenteric capsule endoscopy identifies proximal small bowel disease guiding upstaging and treatment intensification in Crohn's disease: a European multicentre observational cohort study. *United European Gastroenterol J* 2021; 9: 248–255.
13. Limdi JK, Picco M and Farraye FA. A review of endoscopic scoring systems and their importance in a treat-to-target approach in inflammatory bowel disease (with videos). *Gastrointest Endosc* 2020; 91: 733–745.
14. Gralnek IM, Defranchis R, Seidman E, *et al.* Development of a capsule endoscopy scoring index for small bowel mucosal inflammatory change. *Aliment Pharmacol Ther* 2008; 27: 146–154.
15. Eliakim R, Yablecovitch D, Lahat A, *et al.* A novel PillCam Crohn's capsule score (Eliakim score) for quantification of mucosal inflammation in Crohn's disease. *United European Gastroenterol J* 2020; 8: 544–551.
16. Le Berre C, Sandborn WJ, Aridhi S, *et al.* Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology* 2020; 158: 76–94.e2.
17. Parigi TL, Mastrorocco E, Da Rio L, *et al.* Evolution and new horizons of endoscopy in inflammatory bowel diseases. *J Clin Med* 2022; 11: 872.
18. Ding Z, Shi H, Zhang H, *et al.* Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology* 2019; 157: 1044–1054.e5.
19. Klang E, Barash Y, Margalit RY, *et al.* Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020; 91: 606–613.e2.
20. Barash Y, Azaria L, Soffer S, *et al.* Ulcer severity grading in video capsule images of patients with Crohn's disease: an ordinal neural network solution. *Gastrointest Endosc* 2021; 93: 187–192.
21. Klang E, Grinman A, Soffer S, *et al.* Automated detection of Crohn's disease intestinal strictures on capsule endoscopy images using deep neural networks. *J Crohns Colitis* 2021; 15: 749–756.
22. Ferreira JPS, de Mascarenhas Saraiva M, Afonso JPL, *et al.* Identification of ulcers and erosions by the novel Pillcam™ Crohn's capsule using a convolutional neural network: a multicentre pilot study. *J Crohns Colitis* 2022; 16: 169–172.
23. von Elm E, Altman DG, Egger M, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007; 335: 806–808.
24. Spada C, Hassan C, Galmiche JP, *et al.* Colon capsule endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) guideline. *Endoscopy* 2012; 44: 527–536.

25. Schober P, Boer C and Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg* 2018; 126: 1763–1768.
26. Eliakim R and Adler SN. Colon PillCam: why not just take a pill? *Dig Dis Sci* 2015; 60: 660–663.
27. Stidham RW and Takenaka K. Artificial intelligence for disease assessment in inflammatory bowel disease: how will it change our practice? *Gastroenterology* 2022; 162: 1493–1506.
28. de Arce EP, Quera R, Núñez FP, *et al.* Role of capsule endoscopy in inflammatory bowel disease: Anything new? *Artif Intell Gastrointest Endosc* 2021; 2: 136–148.
29. Chang Y, Wang Z, Sun HB, *et al.* Artificial intelligence in inflammatory bowel disease endoscopy: advanced development and new horizons. *Gastroenterol Res Pract* 2023; 2023: 3228832.
30. Gubatan J, Levitte S, Patel A, *et al.* Artificial intelligence applications in inflammatory bowel disease: emerging technologies and future directions. *World J Gastroenterol* 2021; 27: 1920–1935.
31. Takenaka K, Kawamoto A, Okamoto R, *et al.* Artificial intelligence for endoscopy in inflammatory bowel disease. *Intest Res* 2022; 20: 165–170.

Visit Sage journals online  
[journals.sagepub.com/  
home/tag](https://journals.sagepub.com/home/tag)

 Sage journals