Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Performance of a deep learning system for automatic diagnosis of protruding lesions in colon capsule endoscopy: a multicentric study

Miguel José Mascarenhas Saraiva (✉ miguelmascarenhassaraiva@gmail.com)
Centro Hospitalar Universitário de São João, Porto

João Ferreira
Department of Mechanical Engineering, Faculty of Engineering of the University of Porto

Helder Cardoso
Centro Hospitalar Universitário de São João, Porto

João Afonso
Centro Hospitalar Universitário de São João, Porto

Tiago Ribeiro
Centro Hospitalar Universitário de São João, Porto

Patricia Andrade
Centro Hospitalar Universitário de São João, Porto

Marco Parente
Department of Mechanical Engineering, Faculty of Engineering of the University of Porto

Renato Natal
INEGI - Institute of Science and Innovation in Mechanical and Industrial Engineering, Porto

Miguel Mascarenhas Saraiva
ManopH Gastroenterology Clinic, Porto

Guilherme Macedo
WGO Gastroenterology and Hepatology Training Center, Porto

Research Article

Keywords: Colon capsule endoscopy, Artificial intelligence, Convolutional neural network, Colorectal Neoplasia

Posted Date: March 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-284396/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Colon capsule endoscopy (CCE) represents a landmark in minimally invasive exploration of the colonic mucosa for patients with contraindications for a conventional colonoscopy, or for whom the latter exam is unwanted or unfeasible. Colorectal neoplasia is the most common lesion found in CCE. The widespread acceptance of CCE as a non-invasive diagnostic method is particularly important in the setting of colorectal cancer screening. However, reviewing these exams is a time-consuming process as they generate a large number of frames, with the risk of overlooking important lesions. We aimed to develop an artificial intelligence (AI) algorithm using a convolutional neural network (CNN) architecture for the automatic detection of colonic protruding lesions.

A CNN was constructed using an anonymized database of CCE images collected from a total of 124 patients. This database included images of patients with colonic protruding lesions or patients with normal colonic mucosa or with other pathologic findings. A total of 5715 images (2410 protruding lesions, 3305 normal mucosa or other findings) were extracted for CNN development. Two image datasets were created and used for training and validation of the CNN. The performance of the CNN was measured by calculating the area under the receiving operating characteristic curve (AUROC), sensitivity, specificity, positive and negative predictive values (PPV and NPV, respectively).

The AUROC for detection of protruding lesions was 0.99. The sensitivity, specificity, PPV and NPV were 90.0%, 99.1%, 98.6% and 93.2%, respectively. The overall accuracy of the network was 95.3%.

The developed *deep learning* algorithm accurately detected protruding lesions in CCE images. The introduction of AI technology to CCE may increase its diagnostic accuracy and acceptance for the screening of colorectal neoplasia.

Introduction

Capsule endoscopy (CE) is a primary diagnostic tool for the investigation of patients with suspected small bowel disease. Colon capsule endoscopy has been recently introduced as a minimally invasive alternative to conventional colonoscopy for evaluation of the colonic mucosa[1,2]. This system allows to overcome some of the drawbacks associated with conventional colonoscopy, including the potential for pain, use of sedation, and the risk of adverse events, including bleeding and perforation[3]. The clinical application of this tool has been most extensively studied in the setting of colorectal cancer screening, particularly for patients with previous incomplete colonoscopy, or for whom the latter exam is contraindicated, unfeasible or unwanted[4]. Although CCE marks an advance in non-invasive colon evaluation, its sensitivity, particularly for polyp detection, remains suboptimal[5]. Moreover, a single full-length CCE video may produce over 50,000 images, and reviewing these images is a monotonous and time-consuming task, requiring approximately 50 minutes for completion[2]. Furthermore, any given frame may capture only a fragment of a mucosal abnormality and lesions may be depicted in a very small number of frames. Therefore, the risk of overlooking important lesions is not insignificant[2].

The combination of enhanced computational power with large clinical datasets has potentiated the research and development of AI tools for clinical implementation. The application of automated algorithms to diverse medical fields has provided promising results regarding disease identification and classification[6–8]. Convolutional neural networks (CNN) are a type of multi-layered deep learning algorithm tailored for image analysis. The application of these technological solutions to small bowel CE has provided promising results in the detection of several types of lesions[9–12]. The introduction of AI tools for real-time detection of colorectal neoplasia in conventional colonoscopy has suggested a high diagnostic yield for CNN-based algorithms[13]. The impact of AI algorithms for detection of colorectal neoplasia in CCE images has been scarcely evaluated. Enhanced reading of CCE images through the application of these tools may improve the diagnostic accuracy of CCE for colorectal neoplasia, which is currently unsatisfactory[2]. Importantly, the implementation of automated algorithms may help to reduce the time required for reading a single CCE exam. The aim of this study was to develop and validate a CNN-based algorithm for the automatic detection of colonic protruding lesions using CCE images.

Methods

Study design

A multicentric study was performed for development and validation of a CNN for automatic detection of colonic protruding lesions. CCE images were retrospectively collected from the two different institutions: São João University Hospital (Porto, Portugal) and ManopH Gastroenterology Clinic (Porto, Portugal). One hundred and twenty-four CCE exams (124 patients), performed between 2010 and 2020, were included.

The full-length video of all participants was reviewed, and a total of 5715 frames of the colonic mucosa were ultimately extracted.
Inclusion and classification of frames were performed by three gastroenterologists with experience in CCE (MJMS, HC and MMS). A final decision on frame labelling required the agreement of at least two of the three researchers.

This study was approved by the ethics committee of São João University Hospital (No. CE 407/2020). The study protocol was conducted respecting the original and subsequent revisions of the declaration of Helsinki. This study is retrospective and of non-interventional nature. Thus, the output provided by the CNN had no influence on the clinical management of each included patient. Any information susceptible to identify included patients was omitted, and each patient was assigned a random number in order to guarantee effective data anonymization for researchers involved in CNN development. A team with Data Protection Officer (DPO) certification (Maastricht University) confirmed the non-traceability of data and conformity with the general data protection regulation (GDPR).

Colon Capsule Endoscopy Procedure

For all patients, CCE procedures were conducted using the *PillCam™* COLON 2 system (Medtronic, Minneapolis, Minnesota, USA). This system comprises three major components: the endoscopic capsule, an array of sensors connected to a data recorder, and a software for frame revision. The capsule measures 32.3 mm in length and 11.6 mm in width. It has 2 high-resolution cameras, each with a 172º angle of view. The system frame rate varied automatically between 4 and 35 frames per second, depending on bowel motility. Each frame had a resolution of 512 x 512 pixels. The battery of the endoscopic capsule has an estimated life of ≥ 10 hours[2]. This system was launched in 2009 and was not submitted to hardware updates since then. Thus, no significant changes in image quality were evident during this period. The images were reviewed using *PillCam™* software version 9.0 (Medtronic, Minneapolis, Minnesota, USA). Each frame was processed in order to remove information allowing patient identification (name, operating number, date of procedure).

Each patient received bowel preparation according to previously published guidelines[14]. Summarily, patients initiated a clear liquid diet in the day preceding capsule ingestion, with fasting in the night before examination. A solution consisting of polyethylene glycol was used in split-dosage (2 liters in the evening and 2 liters in the morning of capsule ingestion). Prokinetic therapy (10 mg domperidone) was used if the capsule remained in the stomach 1 hour after ingestion, upon real-time image review on the recorder. Two boosters consisting of a sodium phosphate solution were applied after the capsule has entered the small bowel and with a 3-hour interval.

Development of the Convolutional Neural Network

A deep learning CNN was developed for automatic detection of colonic protruding lesions. Protruding lesions included all polyps, epithelial tumors, nodes and subepithelial tumors. From the collected pool of images (*n* = 5715), 2410 showed protruding lesions and 3305 displayed normal mucosa or other mucosal lesions (ulcers, erosions, red spots, angiectasia, varices and lymphangiectasia). This pool of images was split for constitution of training and validation image datasets. The training dataset was composed by 80% of the consecutively extracted images (*n* = 4572). The remaining 20% were used as the validation dataset (*n* = 1143). The validation dataset was used for assessing the performance of the CNN (Fig. 1).

To create the CNN, we used the *Xception* model with its weights trained on *ImageNet* (a large-scale image dataset aimed for use in development of object recognition software). To transfer this learning to our data, we kept the convolutional layers of the model. We removed the last fully connected layers and attached fully connected layers based on the number of classes we used to classify our endoscopic images.

We used two blocks, each having a fully connected layer followed by a dropout layer of 0.3 drop rate. Following these two blocks, we add a dense layer with a size defined as the number of categories to classify. We applied gradient-weighted class activation mapping on the last convolutional layer[15], in order to highlight important features for predicting protruding lesions. The learning rate of 0.0001, batch size of 16, and the number of epochs of 100 was set by trial and error. We used Tensorflow 2.3 and Keras libraries to prepare the data and run the model. The analyses were performed with a computer equipped

with a 2.1 GHz Intel® Xeon® Gold 6130 processor (Intel, Santa Clara, CA, USA) and a double NVIDIA Quadro® RTX™ 4000 graphic processing unit (NVIDIA Corporate, Santa Clara, CA, USA).

Model performance and statistical analysis

The primary outcome measures included sensitivity, specificity, positive and negative predictive values, and accuracy. Moreover, we used receiver operating characteristic (ROC) curve analysis and area under the ROC curve (AUROC) to measure the performance of our model in the distinction between the categories. For each image, the trained CNN calculated the probability for each of the categories (protruding lesions vs. normal colonic mucosa or other findings). A higher probability value translated in a greater confidence in the CNN prediction. The software generated heatmaps that localized features that predicted a class probability (Fig. 2A). The category with the highest probability score was outputted as the CNN's predicted classification (Fig. 2B). The output provided by the network was compared to the specialists' classification *(gold standard)*. Additionally, the image processing performance of the network was determined by calculating the time required for the CNN to provide output for all images in the validation image dataset. Sensitivities, specificities, positive and negative predictive values were obtained using one iteration and are presented as percentages. Statistical analysis was performed using Sci-Kit learn v0.22.2[16].

Results

Construction of the network

One hundred and twenty-four patients were submitted to CCE and enrolled in this study. A total of 5715 frames were extracted, 2410 showing protruding lesions and 3305 showing normal colonic mucosa or other findings. The training dataset was constituted by 80% of the total image pool. The remaining 20% of frames (n = 1143) were used for testing the model. This validation dataset was composed by 482 (42.2%) images with evidence of protruding lesions and 661 (57.8%) images with normal colonic mucosa/other findings. The CNN evaluated each image and predicted a classification (protruding lesions vs. normal mucosa/other lesions) which was compared with the classification provided by gastroenterologists. Repeated inputs of data to the CNN resulted in the improvement of its accuracy (Fig. 3).

Overall performance of the network

The confusion matrix between the trained CNN and expert classifications is shown in Table 1. Overall, the developed model had a sensitivity and specificity for the detection of protruding lesions of 90.0% and 99.1%, respectively. The positive and negative predictive values were, respectively, 98.6% and 93.2%. The overall accuracy of the network was 95.3% (Table 1). The AUROC for detection of protruding lesions was 0.99 (Fig. 4).

Table 1
Confusion matrix and performance marks

		Expert classification	
		Protruding lesion	Normal mucosa
CNN classification	Protruding lesion	434	6
	Normal mucosa	48	655
	Sensitivity	90.0%	
	Specificity	99.1%	
	PPV	98.6%	
	NPV	93.2%	
	Accuracy	95.3%	

Abbreviations: CNN – convolutional neural network; PPV – positive predictive value; NPV – negative predictive value

Computational performance of the CNN

The CNN completed the reading of the testing dataset in 17.5 seconds (approximately 15.4 ms/frame). This translates into an approximated reading rate of 65 frames per second. At this rate, the CNN would complete the revision of a full-length CCE video containing an estimate of 50,000 frames in approximately 13 minutes.

Discussion

The exploration of AI algorithms for application to conventional endoscopic techniques for automatic detection of colorectal neoplasia in conventional endoscopic techniques has been producing promising results over the last decade. The development and implementation of these systems has been recently endorsed (although with reservations) by the European Society of Gastrointestinal Endoscopy[17]. Furthermore, a recent meta-analysis has suggested that the application of AI models for adenoma and polyp's identification may substantially increase the adenoma detection rate and the number of adenomas detected per colonoscopy[18]. These improvements in commonly used performance metrics have shown not to be affected by factors known to influence detection by the human eye, including the size and morphology of the lesions[18].

In our study, we have developed a deep learning tool based on a CNN architecture for automatic detection of protruding lesions in the colonic lumen using CCE images. This study has several highlights. First, our model demonstrated high levels of performance, with a sensitivity of 90.0%, a specificity of 99.1, an accuracy of 95.3% and an AUROC of 0.99. Obtaining fairly high levels of sensitivity and negative

predictive value is paramount for CNN-assisted reading systems, which are designed to lessen the probability of missing lesions, while maintaining a high specificity. Third, our network had a remarkable image processing performance, being capable of reading 65 images per second.

The precise role of CCE in everyday clinical practice is yet to be defined. So far, most studies have focused on its application to colorectal cancer screening. Although colonoscopy remains the undisputed *gold standard*, studies have suggested that CCE could be viewed as a non-invasive complement, rather than substitutive of conventional colonoscopy, particularly in the setting of a previous incomplete colonoscopy[19]. Current guidelines on colorectal cancer screening list CCE as a valid alternative to colonoscopy for the screening of an average-risk population[14]. Studies comparing the diagnostic yield of CCE with another non-invasive screening test, CT colonography, have shown the superiority of CCE[20]. Moreover, when following a first positive fecal-immunological test, CCE may reduce the need for more invasive conventional colonoscopy[21]. Moreover, CCE was shown to have a higher uptake rate (an essential parameter in any population-based screening program) comparing to conventional colonoscopy in a population-based study[22]. However, the use of CCE is hampered by its purely diagnostic character, the need for a rigorous bowel cleansing protocol, as well as the time required for reading each CCE exam.

The development of AI tools for detection of colorectal neoplasia in CCE images has been poorly explored. Automatic detection of these lesions is limited by the poor resolution of CCE images combined with their variable morphology, size and color. To our knowledge, only two other studies have assessed the potential of the application of CNN models to CCE images. Yamada *et al.* was the first to explore the implementation of AI algorithms for the identification of colorectal neoplasia in frames extracted from CCE exams.

Their network was developed using a relatively large pool of CCE images (17,783 frames from 178 patients).
Overall, their algorithm achieved a good performance (AUROC of 0.90)[23]. However, the sensitivity of their model was modest (79%) compared to that of our network. Blanes-Vidal *et al.* adapted a preexisting CNN (*AlexNet*) and trained it for the detection of colorectal polyps. The sensitivity, specificity and accuracy of their work were 97%, 93% and 96%, respectively. In our perspective, the development of these technologies should aim to support a clinical decision rather than substitute the role of the clinician. Therefore, these systems must remain highly sensitive in order to minimize the risk of missing lesions.

Our network demonstrated a high image processing performance (65 frames/second). To date, no value for comparison exists regarding CCE. Nevertheless, these performance marks exceed those published for CNNs applied to other CE systems[10,24]. The development of highly efficient networks may, in the near future, translate into shorter reading times, thus overcoming one of the main drawbacks of CCE. Further well-designed studies are required to assess if a high image processing capacity in experimental settings can be reproduced as an enhanced time efficiency regarding reading times of CCE exams comparing to conventional reading. The combination of enhanced diagnostic accuracy and time efficiency may have a

pivotal role in widening the indications for CCE and its acceptance as a valid screening and diagnostic tool.

This study has several limitations. First, it is a retrospective study. Therefore, further prospective multicentric studies in a real-life setting are desirable to confirm the clinical value of our results. Second, although we included a large number of patients from two distinct medical centers, the number of extracted images is small. We are currently expanding our image pool in order to increase the robustness of our model. The multicentric nature of our work reinforces the validity of our results. Nevertheless, multicentric studies including larger populations are required to ensure the clinical significance of our findings.

In conclusion, we developed a highly sensitive and specific CNN-based model for detection of protruding lesions in CCE images. We believe that the implementation of AI tools to clinical practice will be a crucial step for wider acceptance of CCE for non-invasive screening and diagnosis of colorectal neoplasia.

Declarations

Acknowledgements

The authors acknowledge *Fundação para a Ciência e Tecnologia* (FCT) for supporting the computational costs related to this study through CPCA/A0/7363/2020 grant. This entity had no role in study design, data collection, data analysis, preparation of the manuscript and publishing decision.

Author contributions

Study design (M.J.M.S., J.P.S.F., H.C., J.A., T.R., P.A., M.L.P.P, R.N.J., G.M.). Revision of CCE videos, image extraction and labelling (M.J.M.S., H.C., M.M.S.). Construction and development of the CNN (M.J.M.S., J.P.S.F., J.A., T.R.). Statistical analysis (J.P.S.F.). Data analysis (M.J.M.S., J.P.S.F., T.R.). Drafting of the manuscript (M.J.M.S, T.R.). All authors reviewed the manuscript and agree with its final version.

Competing interests

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Data availability statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1 Eliakim, R. *et al.* Evaluation of the PillCam Colon capsule in the detection of colonic pathology: results of the first multicenter, prospective, comparative study. *Endoscopy* **38**, 963-970, doi:10.1055/s-2006-

944832 (2006).

2 Eliakim, R. *et al.* Prospective multicenter performance evaluation of the second-generation colon capsule compared with colonoscopy. *Endoscopy***41**, 1026-1031, doi:10.1055/s-0029-1215360 (2009).

3 Niikura, R. *et al.* Factors predicting adverse events associated with therapeutic colonoscopy for colorectal neoplasia: a retrospective nationwide study in Japan. *Gastrointest Endosc***84**, 971-982.e976, doi:10.1016/j.gie.2016.05.013 (2016).

4 Spada, C. *et al.* Imaging alternatives to colonoscopy: CT colonography and colon capsule. European Society of Gastrointestinal Endoscopy (ESGE) and European Society of Gastrointestinal and Abdominal Radiology (ESGAR) Guideline - Update 2020. *Endoscopy***52**, 1127-1141, doi:10.1055/a-1258-4819 (2020).

5 Spada, C. *et al.* Accuracy of First- and Second-Generation Colon Capsules in Endoscopic Detection of Colorectal Polyps: A Systematic Review and Meta-analysis. *Clin Gastroenterol Hepatol***14**, 1533-1543.e1538, doi:10.1016/j.cgh.2016.04.038 (2016).

6 Yasaka, K., Akai, H., Abe, O. & Kiryu, S. Deep Learning with Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-enhanced CT: A Preliminary Study. *Radiology***286**, 887-896, doi:10.1148/radiol.2017170706 (2018).

7 Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature***542**, 115-118, doi:10.1038/nature21056 (2017).

8 Gargeya, R. & Leng, T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology***124**, 962-969, doi:10.1016/j.ophtha.2017.02.008 (2017).

9 Aoki, T. *et al.* Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc***89**, 357-363.e352, doi:10.1016/j.gie.2018.10.027 (2019).

10 Aoki, T. *et al.* Automatic detection of blood content in capsule endoscopy images based on a deep convolutional neural network. *J Gastroenterol Hepatol***35**, 1196-1200, doi:10.1111/jgh.14941 (2020).

11 Ding, Z. *et al.* Gastroenterologist-Level Identification of Small-Bowel Diseases and Normal Variants by Capsule Endoscopy Using a Deep-Learning Model. *Gastroenterology***157**, 1044-1054.e1045, doi:10.1053/j.gastro.2019.06.025 (2019).

12 Tsuboi, A. *et al.* Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. *Dig Endosc***32**, 382-390, doi:10.1111/den.13507 (2020).

13 Repici, A. *et al.* Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology***159**, 512-520.e517, doi:10.1053/j.gastro.2020.04.062 (2020).

14 Spada, C. *et al.* Colon capsule endoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy***44**, 527-536, doi:10.1055/s-0031-1291717 (2012).

15 Selvaraju, R. R. *et al.* in *Proceedings of the IEEE international conference on computer vision.* 618-626.

16 Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research***12**, 2825-2830 (2011).

17 Bisschops, R. *et al.* Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019. *Endoscopy***51**, 1155-1179, doi:10.1055/a-1031-7657 (2019).

18 Hassan, C. *et al.* Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. *Gastrointest Endosc***93**, 77-85.e76, doi:10.1016/j.gie.2020.06.059 (2021).

19 Spada, C. *et al.* Colon capsule endoscopy: What we know and what we would like to know. *World J Gastroenterol***20**, 16948-16955, doi:10.3748/wjg.v20.i45.16948 (2014).

20 Cash, B. D. *et al.* Multicentre, prospective, randomised study comparing the diagnostic yield of colon capsule endoscopy versus CT colonography in a screening population (the TOPAZ study). *Gut*, gutjnl-2020-322578, doi:10.1136/gutjnl-2020-322578 (2020).

21 Holleran, G., Leen, R., O'Morain, C. & McNamara, D. Colon capsule endoscopy as possible filter test for colonoscopy selection in a screening population with positive fecal immunology. *Endoscopy***46**, 473-478, doi:10.1055/s-0034-1365402 (2014).

22 Groth, S. *et al.* Capsule colonoscopy increases uptake of colorectal cancer screening. *BMC Gastroenterol***12**, 80, doi:10.1186/1471-230x-12-80 (2012).

23 Yamada, A., Niikura, R., Otani, K., Aoki, T. & Koike, K. Automatic detection of colorectal neoplasia in wireless colon capsule endoscopic images using a deep convolutional neural network. *Endoscopy*, doi:10.1055/a-1266-1066 (2020).

24 Saito, H. *et al.* Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointestinal Endoscopy***92**, 144-151.e141, doi:10.1016/j.gie.2020.01.054 (2020).

Figures

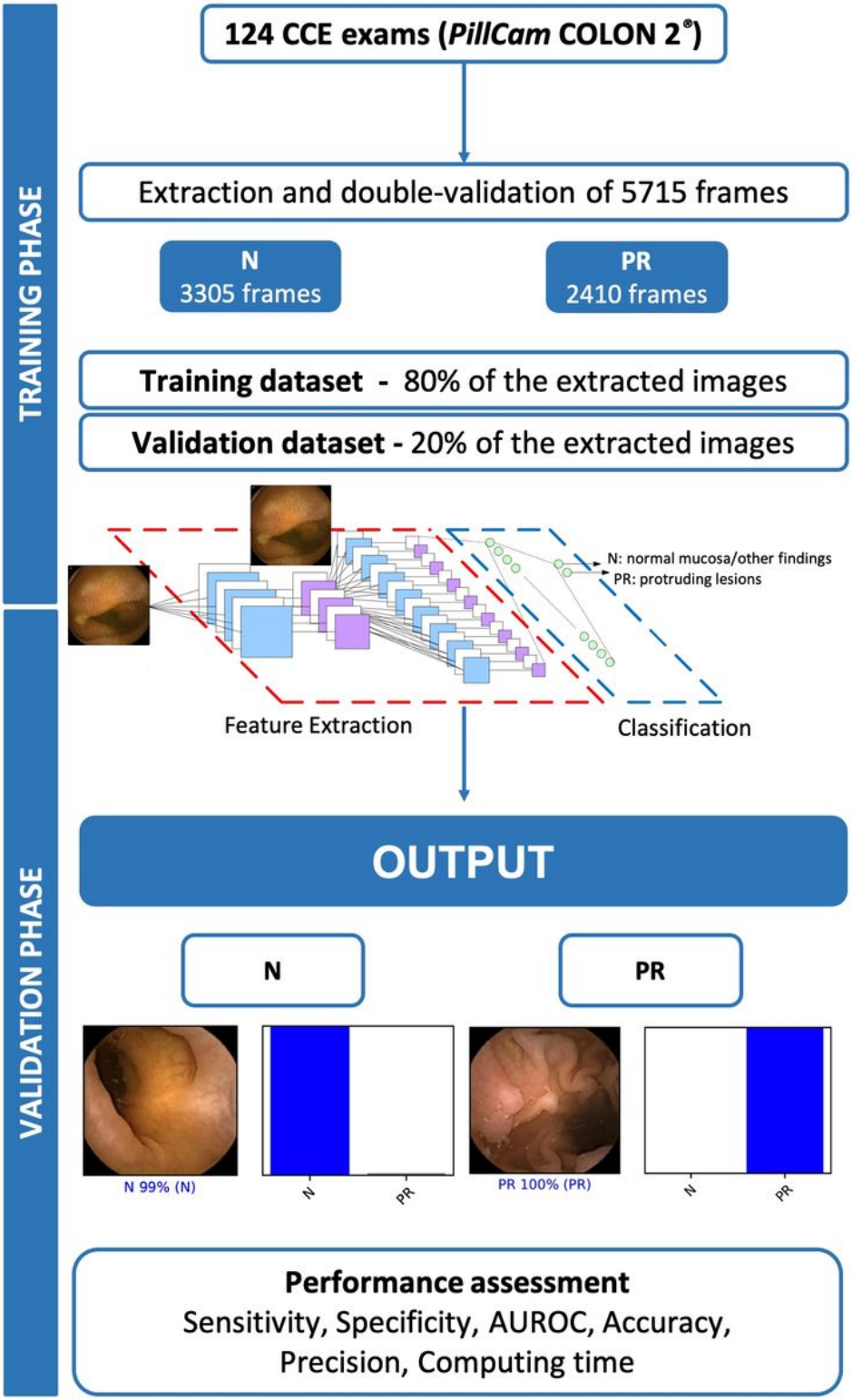

Figure 1

Summary of study design for the training and validation phases. PR – protruding lesion; N – normal mucosa or other findings.

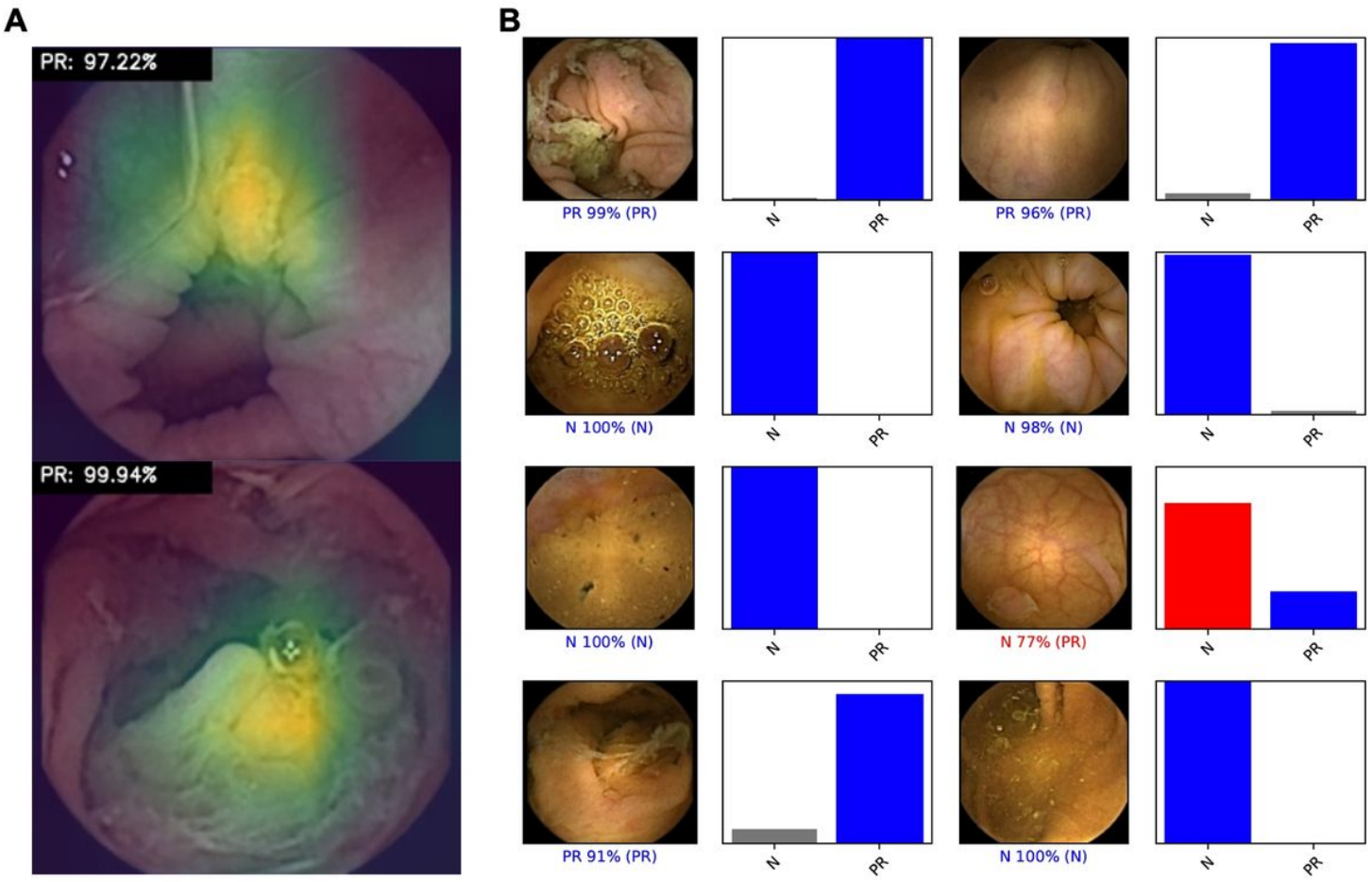

Figure 2

Heatmaps (A) and output (B) obtained from the application of the convolutional neural network. (A) Examples of heatmaps showing CCE features of protruding lesions as identified by the CNN. (B) The bars represent the probability estimated by the network. The finding with the highest probability was outputted as the predicted classification. A blue bar represents a correct prediction. Red bars represent an incorrect prediction. The gold standard classification (specialists' consensus) is reported between brackets. PR – protruding lesion; N – normal mucosa or other findings.

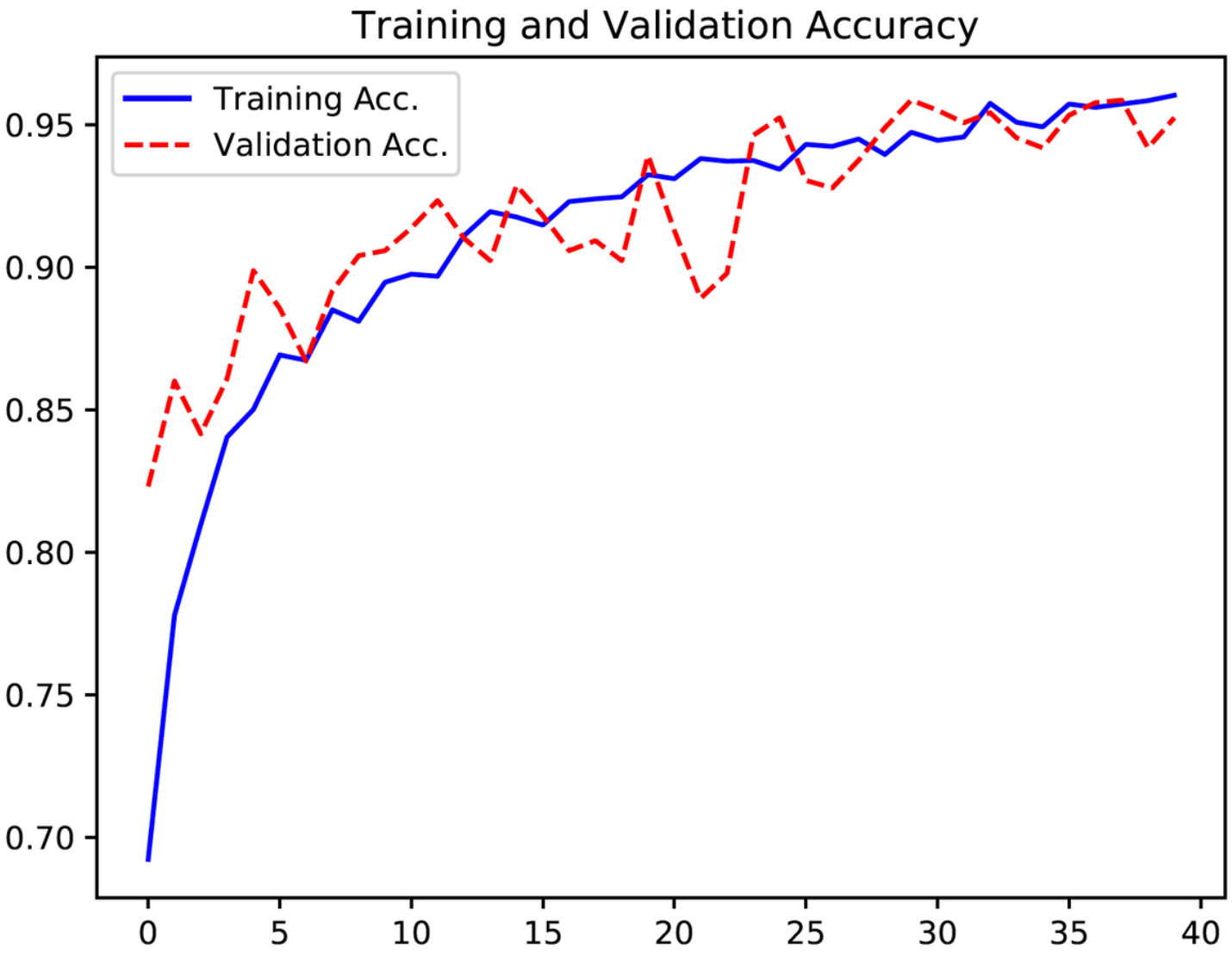

Figure 3

Evolution of the accuracy of the convolutional neural network during training and validation phases, as the training and validation datasets were repeatedly inputted in the neural network.

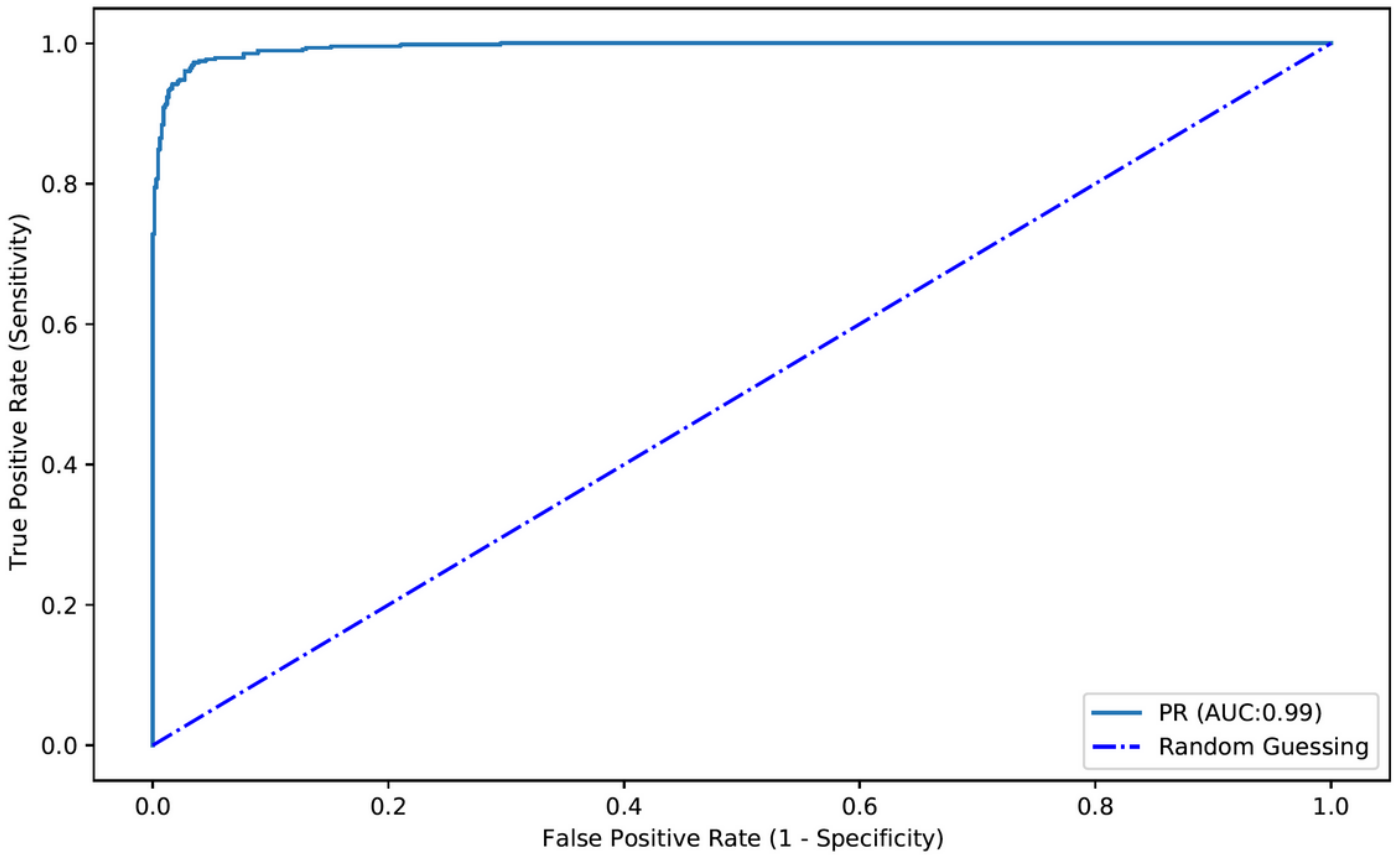

Figure 4

ROC analyses of the network's performance in the detection of protruding lesions vs normal colonic mucosa/other findings. ROC – receiver operator characteristics. PR – protruding lesion.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1SC.docx