







Automatic textual description of colorectal polyp features: explainable artificial intelligence



Authors

Ayla Thijssen^{1,2} , Ramon-Michel Schreuder³ , Roger Fonollà⁴, Quirine van der Zander^{1,2} , Thom Scheeve⁴ , Bjorn Winkens^{5,6}, Sharmila Subramaniam⁷ , Pradeep Bhandari⁷, Peter de With⁴, Ad Masclee¹, Fons van der Sommen⁴ , Erik Schoon^{2,3}

Institutions

- 1 Maastricht University Medical Center, Division of Gastroenterology and Hepatology, Maastricht, Netherlands
- 2 Maastricht University, GROW School for Oncology and Reproduction, Maastricht, Netherlands
- 3 Catharina Hospital, Division of Gastroenterology and Hepatology, Eindhoven, Netherlands
- 4 Eindhoven University of Technology, Department of Electrical Engineering, Eindhoven, Netherlands
- 5 Maastricht University, Department of Methodology and Statistics, Maastricht, Netherlands
- 6 Maastricht University, CAPHRI, Care and Public Health Research Institute
- 7 Portsmouth Hospitals University NHS Trust, Division of Gastroenterology and Hepatology, Portsmouth, United Kingdom

submitted 1.2.2023

accepted after revision 6.4.2023

accepted manuscript online 11.4.2023

Bibliography

Endosc Int Open 2023; 11: E513–E518

DOI 10.1055/a-2071-6652

ISSN 2364-3722

© 2023. 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

Ayla Thijssen, Maastricht University, GROW School of Oncology and Reproduction, PO Box 5800, 6202 AZ Maastricht, Netherlands
Phone: +31433881844
a.thijssen@maastrichtuniversity.nl

ABSTRACT

Computer-aided diagnosis systems (CADx) can improve colorectal polyp (CRP) optical diagnosis. For integration into clinical practice, better understanding of artificial intelligence (AI) by endoscopists is needed. We aimed to develop an explainable AI CADx capable of automatically generating textual descriptions of CRPs. For training and testing of this CADx, textual descriptions of CRP size and features according to the Blue Light Imaging (BLI) Adenoma Serrated International Classification (BASIC) were used, describing CRP surface, pit pattern, and vessels. CADx was tested using BLI images of 55 CRPs. Reference descriptions with agreement by at least five out of six expert endoscopists were used as gold standard. CADx performance was analyzed by calculating agreement between the CADx generated descriptions and reference descriptions. CADx development for automatic textual description of CRP features succeeded. Gwet's AC1 values comparing the reference and generated descriptions per CRP feature were: size 0.496, surface-mucus 0.930, surface-regularity 0.926, surface-depression 0.940, pits-features 0.921, pits-type 0.957, pits-distribution 0.167, and vessels 0.778. CADx performance differed per CRP feature and was particularly high for surface descriptors while size and pits-distribution description need improvement. Explainable AI can help comprehend reasoning behind CADx diagnoses and therefore facilitate integration into clinical practice and increase trust in AI.

Introduction

The optical diagnosis of colorectal polyps (CRPs), the *in vivo* assessment of histopathology by endoscopists, is currently inadequate. Image enhancement techniques such as blue light imaging (BLI) (Fujifilm, Tokyo, Japan) have been developed to improve visualization of the mucosal surface and microvasculature. Furthermore, optical diagnosis can be improved by using clinical characterization models to structurally assess CRP characteristics. One such characterization model is the BLI Adenoma Serrated International Classification (BASIC), which can be used to differentiate among hyperplastic polyps, sessile serrated lesions (SSLs), adenomas, and colorectal carcinomas (CRCs) [1].

The performance of optical diagnosis by endoscopists using characterization models has been studied extensively, showing promising results [2]. However, the variety of image enhancement techniques and corresponding characterization models involve considerable learning curves [3]. Optical diagnosis by endoscopists could be improved without this disadvantage using computer-aided diagnosis systems (CADx) based on artificial intelligence (AI).

Several CADxs have been developed for characterizing CRPs. However, trust of endoscopists in these systems is lacking and their “black box” nature makes it difficult for endoscopists to understand CADx outcomes, which is necessary for integration into clinical practice [4]. A solution can be to automatically generate textual descriptions of an image based on deep learning, as a way of explainable AI. Users are allowed to understand and trust CADx output by explainable AI, which is a set of processes and methods depicting the AI model. Explainable AI creates outcome parameters described as text, facilitating more transparent, fair, and accurate AI-guided decision making. In medical imaging for radiology, explainable AI has already been applied successfully. Furthermore, by implementing the “diagnose-and-leave” strategy for diminutive (≤ 5 mm) hyperplastic CRPs in the rectosigmoid and “resect-and-discard” strategy for diminutive adenomas [5], accurately reporting CRP features

becomes vital due to the lack of a histopathological confirmation of the endoscopists’ optical diagnosis.

A CADx differentiating between benign (hyperplastic) and (pre)malignant (SSLs, adenomas, and CRCs) CRPs was previously developed successfully [6]. Using the fundamentals of that CADx, this study aimed to develop CADTexD (Computer-Aided Diagnosis with Textual Descriptions), a CADx capable of automatically generating textual descriptions of CRP features based on BASIC [7].

Methods

This multicenter study was conducted prospectively at the Maastricht University Medical Center+ (MUMC+), Catharina Hospital Eindhoven, both in the Netherlands, and Queen Alexandra Hospital, Portsmouth, United Kingdom. The Department of Electrical Engineering at Eindhoven University of Technology developed CADTexD. The study was conducted according to the Declaration of Helsinki and the General Data Protection Regulation and approved by the Medical Research Ethics Committee of MUMC+ (METC2019-1231).

Image and reference description database

For the description of CRP features, no gold standard is available, which makes it difficult to objectively rate the descriptions generated by CADTexD. In the current study, textual descriptions of CRP features according to BASIC were used for training and testing of CADTexD. In addition, CADTexD was trained with CRP histopathology results. BASIC consists of three descriptors: surface, pits, and blood vessels (► **Table 1**). The surface descriptor contains a description of surface mucus, regularity, and depression. The pits descriptor contains a description of pits features, type, and distribution. For the vessel descriptor, vessels were described as either present or not. CRP size has been described in the categories of diminutive, small, or large. For CRP size, solely histopathology was used as gold standard.

For the CADTexD training dataset, 95 of 468 included CRPs were selected based on the availability of CRP textual descrip-

► **Table 1** BLI Adenoma Serrated International Classification of colorectal polyp features as used in the reference and generated descriptions [1].

BASIC descriptor		Hyperplastic polyp	Adenoma	SSL	Cancer
Surface	Mucus	Without mucus	Without mucus	With mucus	Without mucus
	Regularity	Regular	Regular/irregular	Regular/irregular	Irregular
	Depression	No depression, no pseudo-depression	No depression, pseudo-depression	No depression, no pseudo-depression	True depression, no pseudo-depression
Pits	Features	Without features	With features	With features	With features
	Type	Round	Not round	Round, dark/not dark	Round/not round
	Distribution	Homogenous	Homogenous/heterogenous	Homogenous/heterogenous	Heterogenous
Vessels	Present	With/without vessels	With vessels	With/without vessels	With vessels

BLI, blue light imaging; BASIC, BLI Adenoma Serrated International Classification; SSL, sessile serrated lesion.

tions to be used as input. For these 95 CRPs (35 hyperplastic polyps, 12 SSLs, and 48 adenomas), 6525 textual descriptions and 507 images in BLI were available.

The CADTexD test dataset consisted of 60 CRPs with one image in BLI for each CRP. Five CRPs were removed due to data inconsistency, meaning inability to use a textual description as gold standard because this did not match histopathology.

The remaining 55 CRPs (15 hyperplastic polyps, 3 SSLs, 36 adenomas, and 1 CRC) were optically diagnosed by 19 endoscopists, resulting in 1045 reference descriptions containing the BASIC features to be used as gold standard for comparison with the generated descriptions by CADTexD. Of these 19 endoscopists, six were considered experts. These experts from the international BLI expert group were familiar with using BLI and BASIC, and each had conducted over 2000 lifetime colonoscopies. The 13 novice endoscopists were less familiar with using BLI and BASIC, and each had conducted fewer than 400 lifetime colonoscopies. All endoscopists were blinded for histopathology.

Both datasets consisted of images obtained from bowel cancer screening colonoscopies or from surveillance colonoscopies. Expert pathologists reported the CRP histopathology according to the revised Vienna classification.

CADTexD details

The model employed in this study is composed of two sub-models. The first model consists of the CRP image module that translates the CRP image into understandable information for the network. The second model, which employs the Bidirectional Encoder Representations from Transformers (BERT) module, allows for CRP text description to be learned. In the training set of the BERT module, only BLI images were used because this image enhancement mode is suitable for the application of BASIC. A more extensive explanation of CADTexD development and performance analyzed with technical metrics can be found in Fonollà et al. [7].

CADTexD was tested by letting the algorithm generate textual descriptions as output for the same images of 55 CRPs in BLI that were optically diagnosed by the 19 endoscopists.

Study outcomes

In this study, the descriptions generated by CADTexD were compared to reference descriptions by endoscopists. To determine whether the reference descriptions were of sufficient quality, interobserver agreement between the reference descriptions was calculated. This was performed for each CRP feature separately.

Statistical analysis

The interobserver agreement in the reference descriptions by all endoscopists was calculated using Fleiss' kappa with 95% confidence intervals (CIs). Values of Fleiss' kappa were interpreted according to the strength of agreement classification from Landis and Koch.

Reference descriptions with agreement on a CRP feature description by five or six experts were compared to the descriptions generated by CADTexD by calculating the raw proportion

of agreement with corresponding Gwet's chance-corrected agreement coefficient (AC1) values with 95% CI. Gwet's AC1 was used because of a lack of variation in the data (high proportion of one category of the CRP feature) and the problem this poses for inter-rater reliability estimation for Cohen's Kappa [8].

Statistical analyses were performed with IBM SPSS Statistics for Windows version 27 (IBM Corp., Armonk, New York, United States), the online Vassarstats calculator (<https://vassarstats.net/kappa.html>), and AgreeStat 360.

Results

Interobserver agreement reference descriptions

The interobserver agreement for optical diagnosis of the CRPs by the 19 endoscopists was slight to moderate for all separate descriptors, with Fleiss' kappa ranging from 0.089 to 0.411 (► **Table 2**). After exclusion of reference descriptions by novices, interobserver agreement increased slightly but remained moderate at most for all descriptors, with Fleiss' kappa ranging from 0.032 to 0.538.

Due to this relatively low interobserver agreement even between experts, solely CRP feature descriptions with agreement by at least five of six experts involved were included to evaluate CADTexD performance. For CRP size, histopathology results were used as reference.

CADTexD performance

CADTexD was able to automatically generate a textual description of CRP features based on a CRP image in BLI (► **Fig. 1**).

Results for CADTexD performance are displayed in ► **Table 3**. Agreement between the generated and reference descriptions was almost perfect for all surface descriptors with a proportion of agreement and Gwet's AC1 of 93.5% and 0.930 for the mucus description, 93.3% and 0.926 for the regularity description, and 94.6% and 0.940 for the depression description, respectively. Similarly, almost perfect agreement was observed for description of pits features (proportion of agreement 92.7% and Gwet's AC1 0.921) and description of pits type (proportion of agreement 96.0% and Gwet's AC1 0.957).

Substantial agreement was seen in the comparison of the generated and reference description of the CRP vessels, with a proportion of agreement of 84.1% and Gwet's AC1 of 0.778.

Comparison of the generated and reference descriptions of CRP size and pits distribution showed less agreement, with a proportion of agreement of 61.8% and Gwet's AC1 of 0.496 and proportion of agreement of 58.3% and Gwet's AC1 of 0.167, respectively.

Discussion

This study demonstrated that development of a CADx capable of automatically generating a textual description of a CRP image based on BASIC is possible. CADTexD performed particularly well in the description of CRP surface features when comparing the generated descriptions by CADTexD with reference descriptions agreed upon by expert endoscopists.

► **Table 2** Interobserver agreement for colorectal polyp feature descriptions by both 19 novice and expert endoscopists and six expert endoscopists.

Colorectal polyp feature	No. colorectal polyps	Fleiss' kappa novice and expert reference descriptions (95% CI)	No. colorectal polyps	Fleiss' kappa expert reference descriptions (95% CI)
Surface – mucus	43	0.243 (0.220–0.266)	53	0.474 (0.404–0.543)
Surface – regularity	54	0.411 (0.391–0.431)	55	0.529 (0.461–0.597)
Surface – depression	55	0.333 (0.315–0.350)	55	0.502 (0.437–0.567)
Pits – features	41	0.101 (0.078–0.125)	44	0.032 (-0.044–0.109)
Pits – type	14	0.315 (0.282–0.347)	37	0.436 (0.374–0.498)
Pits – distribution	13	0.089 (0.048–0.131)	39	0.233 (0.152–0.314)
Vessels	43	0.259 (0.236–0.282)	50	0.538 (0.467–0.610)

CI, confidence interval.



Automatically generated textual description
 a **small** polyp, surface without mucus regular, pits with features non-rounded **heterogenous**, with vessels

Reference description
 a diminutive polyp, surface without mucus regular no depression, pits with features non-rounded homogenous, with vessels

► **Fig. 1** Automatically generated textual description by the computer-aided diagnosis system CADTexD for an adenoma in blue light imaging, with the corresponding reference description agreed upon by five of six expert endoscopists. Words displayed in red do not correspond with the reference description.

Real-time optical diagnosis of CRPs by endoscopists remains challenging. To the best of our knowledge, all CADxs developed to assist endoscopists in the optical diagnosis of CRPs so far create an output of the CRP characterization in categories such as benign or premalignant, non-neoplastic or neoplastic, and hyperplastic polyp or adenoma. CADxs with explainable AI could not only improve the optical diagnosis of endoscopists, they also have the potential to teach endoscopists more about specific CRP features leading to the corresponding diagnosis.

With increasing research on CADxs for CRP characterization, implementation of the “diagnose-and-leave” and “resect-and-discard” treatment strategies in clinical practice comes within reach. However, implementation of these strategies is not possible without acceptance of and trust in AI-based systems by endoscopists. For this reason, a demand exists for opening the AI “black box” [4]. The black box refers to the information that leads to the outcome of the AI-based system and is unknown for deep learning systems. CADTexD is innovative regarding the use of explainable AI, providing additional information to the AI outcome and, thus, illuminating part of that black box. The automatically generated description of colorectal polyp features could help endoscopists to understand why a CADx provides the diagnosis. It becomes possible for the endoscopist to check whether they agree with the description of all colorectal polyp features and thus the reasoning of CADTexD. Trust in AI can thus be improved.

With successful implementation of the proposed treatment strategies in the near future, histopathological examination of diminutive CRPs (≤ 5 mm) will no longer take place. This change in clinical practice again highlights the importance of adequate optical diagnostic reporting by endoscopists. A strength of CADTexD is its ability to aid the endoscopist in making a precise description of CRP features. Beaulieu et al. (2012) showed that endoscopy reports in a quality audit were lacking important CRP information such as CRP size in 34.2% and CRP resection method in 20.7% of colonoscopies [9]. Automatically generated textual descriptions merely need to be checked by endoscopists and, therefore, contribute to saving administration time. Moreover, automatically generated textual descriptions of CRPs could be a first step toward generating fully automated endoscopy reports and thus improving efficiency of colonoscopy practice. This could facilitate the acceptance of AI-based systems by endoscopists, particularly because it counteracts the fear that use of AI-based systems in clinical practice prolongs the procedure duration [10].

This study has some limitations. First, a CADx can only be as strong as the gold standard used as input for the algorithm. For the description of a CRP, no gold standard exists other than that the description needs to match the outcome of the histopatho-

► Table 3 Interobserver agreement of colorectal polyp feature descriptions between CADTexD-generated descriptions and reference descriptions by at least five of six expert endoscopists.

Colorectal polyp feature	No. colorectal polyps	Proportion of agreement, % (95% CI)	Gwet's AC1 (95% CI)
Size ¹	55	61.8 (47.7–74.3)	0.496 (0.299–0.692)
Surface – mucus	46	93.5 (81.1–98.3)	0.930 (0.847–1.000)
Surface – regularity	45	93.3 (80.7–98.3)	0.926 (0.836–1.000)
Surface – depression	55	94.6 (83.9–98.6)	0.940 (0.869–1.000)
Pits – features	41	92.7 (79.0–98.1)	0.921 (0.826–1.000)
Pits – type	25	96.0 (77.7–99.9)	0.957 (0.866–1.000)
Pits – distribution	24	58.3 (36.9–77.2)	0.167 (-0.250–0.583)
Vessels	44	84.1 (69.3–92.8)	0.778 (0.602–0.955)

CI, confidence interval; AC, agreement coefficient; CADTexD, computer-aided diagnosis with textual descriptions
¹ For CRP size, histopathology results were used as reference.

logical assessment. Our study showed relatively low interobserver agreement on CRP descriptions by endoscopists based on BASIC, even when solely including expert endoscopists. This complicates the use of these reference descriptions as a reliable gold standard and led to a smaller number of CRPs used with CADTexD testing in this study. The description of pits features and pits distribution showed lowest Fleiss' kappa values, even among experts. This indicates that especially these features are difficult to describe and might require simplification in the classification model BASIC.

As a result of agreement by fewer than five of six experts in the description of the excluded CRPs, a relatively low number of CRPs were included in the analysis especially for pits type and pits distribution, indicating the difficulty of recognition and accurate description of these CRP features.

Clinical validation of BASIC showed high diagnostic performance [2]. Within the same group of 19 endoscopists as in this study, however, a previous study showed no significant increase in diagnostic performance when using BASIC after following an additional previously validated BLI and BASIC training module [6], even though both accuracy and sensitivity based on intuition were much lower than in the BASIC validation study. This comparison indicates that use of BASIC might not be preferred. Other clinical characterization models such as JNET should be considered for future studies, aiming at increasing interobserver agreement between endoscopists.

CADTexD performed poorly for CRP size description. Accurate CRP size estimation is important when adopting the treatment strategies for diminutive CRPs. Correct endoscopic CRP size estimation has been shown to be challenging. AI tools such as virtual scales can also be a solution for objectively sizing CRPs. Another limitation was the small CADTexD training dataset, although containing a reasonable distribution of CRP types.

In addition, CADTexD provides a textual description of CRP features without generating the explicit histopathology prediction (hyperplastic polyp, SSL, or adenoma). With the promising

results of CADTexD focused on CRP features, the next step is to incorporate the resulting diagnosis.

Conclusions

In conclusion, this study presents a CADx automatically generating textual descriptions of CRPs in BLI with acceptable performance, but needs further improvement. The descriptive output could increase trust of endoscopists in CADx, easing implementation into clinical practice. Future studies should focus on optimization of the gold standard using characterization models other than BASIC.

Competing interests

F. van der Sommen, P.H.N. de With, A.A.M. Masclee, and E.J. Schoon report a joined research grant from the Dutch Cancer Society for the submitted work. Besides this grant, the authors have no conflicts of interest related to the content of this manuscript. With regard to any other disclosures: Q.E.W. van der Zander was supported by Fujifilm to attend scientific meetings. S. Subramaniam has received speaker fees from Fujifilm. P. Bhandari has received research grants and speaker fees from Olympus, Fujifilm, Pentax, 3-D matrix, and Boston Scientific. A.A.M. Masclee was supported by a health care efficiency grant from ZonMw, an unrestricted research grant from Will Pharma S.A., a restricted educational grant from Ferring B.V., a research grant from Pentax Europe, research funding from Allegan and Grünenthal, and gave scientific advice to Bayer, Kyowa Kirin, and Takeda, outside the submitted work. E.J. Schoon received research support and speakers' fees from Fujifilm, outside the submitted work. All other authors declare that they have no conflicts of interest.

Funding

KWF Kankerbestrijding 12639
<http://dx.doi.org/10.13039/501100004622>

References

- [1] Subramaniam S, Hayee B, Aepi P et al. Optical diagnosis of colorectal polyps with Blue Light Imaging using a new international classification. *United European Gastroenterol J* 2019; 7: 316–325
- [2] Rondonotti E, Hassan C, Andrealli A et al. Clinical validation of BASIC classification for the resect and discard strategy for diminutive colorectal polyps. *Clin Gastroenterol Hepatol* 2020; 18: 2357–2365
- [3] Neumann H, Vieth M, Fry LC et al. Learning curve of virtual chromoendoscopy for the prediction of hyperplastic and adenomatous colorectal lesions: a prospective 2-center study. *Gastrointest Endosc* 2013; 78: 115–120
- [4] Poon Alf, Sung JJY. Opening the black box of AI-Medicine. *J Gastroenterol Hepatol* 2021; 36: 581–584
- [5] Houwen B, Hassan C, Coupe VMH et al. Definition of competence standards for optical diagnosis of diminutive colorectal polyps: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2022; 54: 88–99
- [6] van der Zander QEW, Schreuder RM, Fonolla R et al. Optical diagnosis of colorectal polyp images using a newly developed computer-aided diagnosis system (CADx) compared with intuitive optical diagnosis. *Endoscopy* 2021; 53: 1219–1226
- [7] Fonolla R, van der Zander QEW, Schreuder RM et al. Automatic image and text-based description for colorectal polyps using BASIC classification. *Artif Intell Med* 2021; 121: 102178
- [8] Gwet K. Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical methods for inter-rater reliability assessment* 2002; 1: 1–6
- [9] Beaulieu D, Barkun A, Martel M. Quality audit of colonoscopy reports amongst patients screened or surveilled for colorectal neoplasia. *World J Gastroenterol* 2012; 18: 3551
- [10] Messmann H, Ebigbo A, Hassan C et al. How to integrate artificial intelligence in gastrointestinal practice. *Gastroenterology* 2022; 162: 1583–1586