

# A technical review of artificial intelligence as applied to gastrointestinal endoscopy: clarifying the terminology



## Authors

Alanna Ebigbo<sup>1,\*</sup>, Christoph Palm<sup>2,3,\*</sup>, Andreas Probst<sup>1</sup>, Robert Mendel<sup>2,3</sup>, Johannes Manzeneder<sup>1</sup>, Friederike Prinz<sup>1</sup>, Luis A. de Souza<sup>2,4</sup>, João P. Papa<sup>5</sup>, Peter Siersema<sup>6</sup>, Helmut Messmann<sup>1</sup>

## Institutions

- 1 Department of Gastroenterology, Universitätsklinikum Augsburg, Germany
- 2 Regensburg Medical Image Computing (ReMIC), Ostbayerische Technische Hochschule Regensburg (OTH Regensburg) – Germany
- 3 Regensburg Center of Health Sciences and Technology, OTH Regensburg – Germany
- 4 Department of Computing, Federal University of São Carlos – Brazil
- 5 Department of Computing, São Paulo State University – Brazil
- 6 Department of Gastroenterology and Hepatology, Radboud University Medical Center, Nijmegen, The Netherlands

submitted 11.6.2019

accepted after revision 31.7.2019

## Bibliography

DOI <https://doi.org/10.1055/a-1010-5705> |

Endoscopy International Open 2019; 07: E1616–E1623

© Georg Thieme Verlag KG Stuttgart · New York

eISSN 2196-9736

## Corresponding author

Dr. Alanna Ebigbo, Universitätsklinikum Augsburg, Stenglinstr. 2, 86156 Augsburg, Germany

Fax: +00498214002748

[Alanna.ebigbo@gmx.de](mailto:Alanna.ebigbo@gmx.de)

[Alanna.ebigbo@uk-augsburg.de](mailto:Alanna.ebigbo@uk-augsburg.de)

## ABSTRACT

**Background and aim** The growing number of publications on the application of artificial intelligence (AI) in medicine underlines the enormous importance and potential of this emerging field of research.

In gastrointestinal endoscopy, AI has been applied to all segments of the gastrointestinal tract most importantly in the detection and characterization of colorectal polyps. However, AI research has been published also in the stomach and esophagus for both neoplastic and non-neoplastic disorders.

The various technical as well as medical aspects of AI, however, remain confusing especially for non-expert physicians. This physician-engineer co-authored review explains the basic technical aspects of AI and provides a comprehensive overview of recent publications on AI in gastrointestinal endoscopy. Finally, a basic insight is offered into understanding publications on AI in gastrointestinal endoscopy.

## Introduction

In the past few years, artificial intelligence (AI) has gained tremendous momentum in the medical domain [1, 2]. Various AI applications are currently undergoing intensive research with the ultimate goal of improving the quality of diagnosis made in clinical routine.

AI can have a wide range of applications in gastrointestinal endoscopy, especially in detection and classification of dysplastic and neoplastic lesions [3, 4]. The correct interpretation of such lesions or disease entities can be extremely challenging even for experienced physicians. Considering the excellent di-

agnostic performance of AI in well-defined scopes, the demand on computer-aided diagnosis (CAD) support is increasing.

Although AI research in gastrointestinal endoscopy is still mostly preclinical and engineer-driven, recently real-life clinical studies have also been published [5]. However, the technical aspects of AI and the different methods of Machine Learning (ML) and CAD, summed up under the term AI, remain confusing and sometimes incomprehensible for physicians. Because AI will have an enormous impact on medicine in general and gastrointestinal endoscopy in particular, it is important for endoscopists to understand at least the basic technical and clinical implications of AI.

\* Drs. Ebigbo and Palm: These authors contributed equally.

In this physician-engineer co-authored review article, we provide a comprehensive overview of the state-of-the-art of ML and AI in gastrointestinal endoscopy.

## Technical aspects of AI and machine learning

The general task of software development is to code a computer program on the basis of an algorithm, which generates to a specific input a defined output. Machine learning changes this paradigm, because parts of the computer program remain undetermined. After coding, these parts are defined using input data and a training procedure to “learn” from these data, e.g. the class of an object. The main goal is to find a generalizable model which holds true even for new data samples, which were not included in the training data. With such a model, new data samples can be correctly processed as well and, thus, the computer can learn to cope with new situations.

The generic term “artificial intelligence” is now established for all procedures that include such a learning component (► **Fig. 1**). However, all methods used in practice are still not “intelligent” in a human way of reasoning, but rather deal with different sorts of pattern recognition. In general, three types of learning procedures have to be differentiated [6]:

1. Supervised learning: here, the computer learns from known patterns;
2. Unsupervised learning: the computer finds common features in unknown patterns;
3. Reinforcement learning: the computer learns from trial and error.

### ML using hand-crafted features

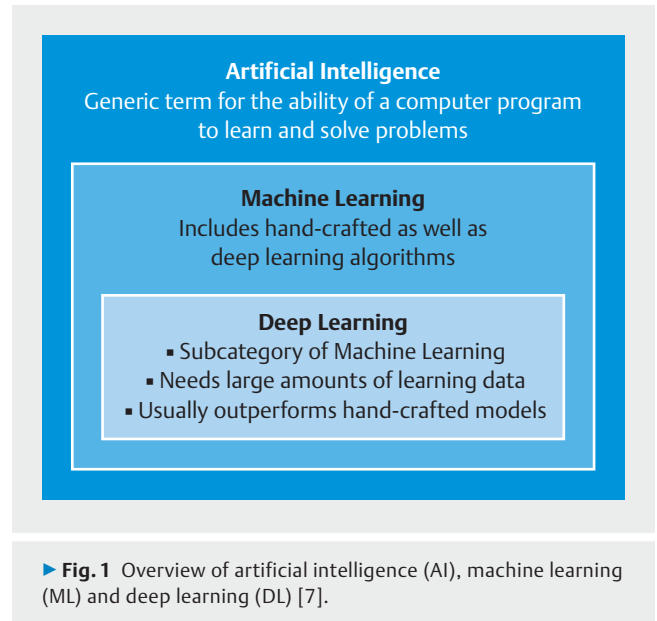
For many years, machine learning from images focused mainly on hand-crafted features, where the computer scientist coded a mathematical description of patterns, e.g. color and texture. During training, a classifier learned to distinguish between the features of different classes and used this knowledge to determine the class of a new image sample.

### ML using deep learning

In recent years, the paradigm of hand-crafted features has changed to “deep learning” (DL) methods where not only the classifier but also the features are learned by an artificial neural network (ANN) [7].

In general, an ANN consists of layers of neurons with all neurons of adjacent layers being connected. Therefore, in a fully connected neural network, the outputs of the neurons of one layer serve as input for the next layer. Each connection is associated with a weight. These weights are the features learned during the training procedure. Mathematically, each neuron realizes the scalar product of weights and input values followed by a non-linear sigmoidal activation function. DL architectures provide a large number of layers and, thus, have to learn a large amount of weights.

In the image-understanding-domain, DL is based on convolutional neural networks (CNN). The raw data from the image are the input values for the first layer. Unlike the fully connected networks, a series of convolutions are computed in each

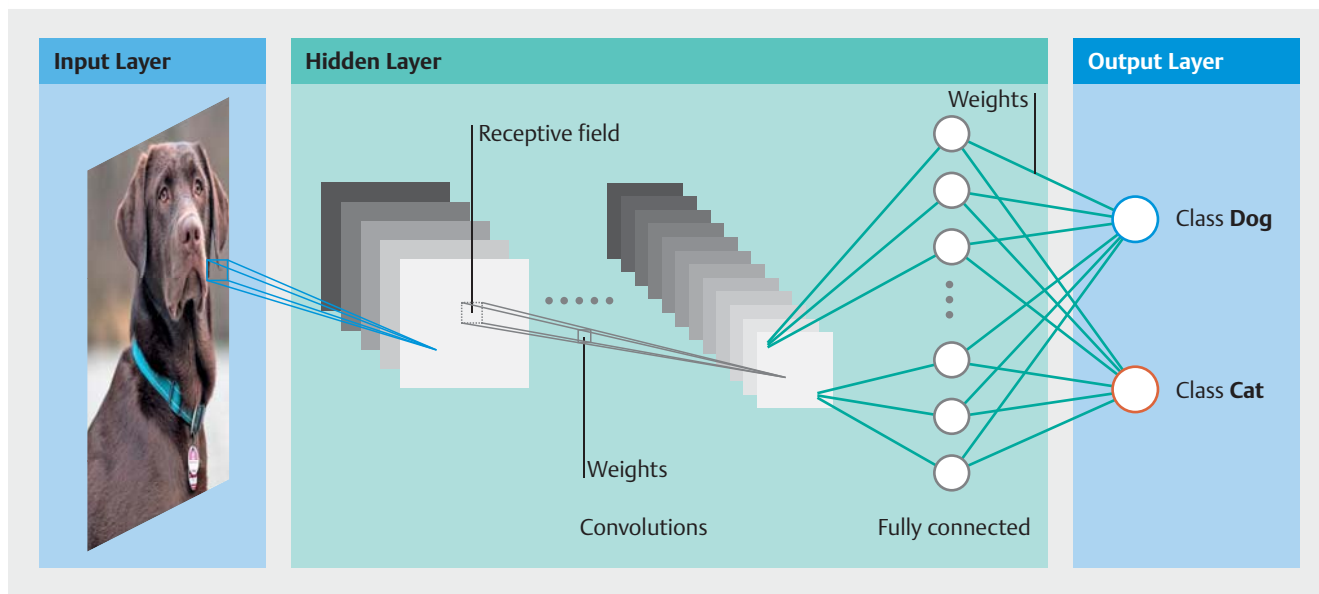


layer (► **Fig. 2**). The learned weights of a CNN are the elements of the convolution kernels. Because the kernels take a small receptive field of an image into account and remain constant for all image positions, the number of weights is reduced significantly compared to fully connected networks. CNN architectures use these basic convolution modules and complement them with different kinds of sigmoidal activation functions, pooling operations and other elements. Over the last years, a large number of CNN architectures for different tasks have been introduced allowing, for example, for very deep networks with 100 layers or even more such as residual nets [8] or effecting an encoder-decoder approach for pixel-wise classification such as U-Net [9].

## General clinical applications of AI in gastrointestinal endoscopy

Although AI applications first were described in non-neoplastic disorders [10], the focus has shifted mainly to malignant or neoplastic gastrointestinal disease. The most common examples include detection and classification of polyps, adenomas or carcinomas in the colon during screening colonoscopy. As mentioned, AI has been shown to have potential indications in benign or non-neoplastic conditions as well. For example, diagnosis of *Helicobacter pylori* infection with AI may have a practical benefit, particularly in high-prevalence regions, and has been demonstrated using still images [10, 11]. A further interesting application is assessment of gastrointestinal ulcers with the aim of predicting risk of rebleeding [12].

AI applications can be subdivided into tasks or assignments based on clinical challenges that physicians face in everyday practice (► **Table 1**). These tasks will be described in further detail below.



► **Fig. 2** Deep learning (DL) based on convolutional neural networks (CNN) showing the input layer with raw data of the image, the hidden layer with a series of convolutions computed for each layer and the classification of the image in the output layer.

► **Table 1** Brief summary of AI applications.

AI tasks	Comments
Frame detection task	Frames are individual pictures in a sequence of images; in this task, AI detects frames with suspicious objects which need closer examination; for example, during colonoscopy, the detection of frames bearing an adenoma or polyp.
Object detection task	AI recognizes and identifies a region of interest (ROI) (such as a dysplastic lesion in BE) during an endoscopic examination.
Classification task	AI categorizes detected lesions into classes such as neoplastic vs. non-neoplastic or adenomatous vs. hyperplastic
Segmentation task	AI delineates the outer margin or border of a detected lesion and correctly differentiates between pathological and normal at the interface between the lesion and the healthy tissue.
Task combinations	AI can ultimately combine these tasks described above in one work-flow, for example the detection and classification of a colorectal polyp followed by the delineation of the outer margin of the lesion.

BE, Barrett's esophagus.

### Frame detection task

Frames are individual pictures in a sequence of images presented at a particular speed called frames per second. A particular number of frames per second is blended by the human eye into moving images. In real time, during an endoscopic examination, or at least in a video of such an examination, frames with suspicious objects that need closer examination have to be detected. The goal of this task is to prevent the endoscopist from missing an object such as a polyp [5].

### Object detection task

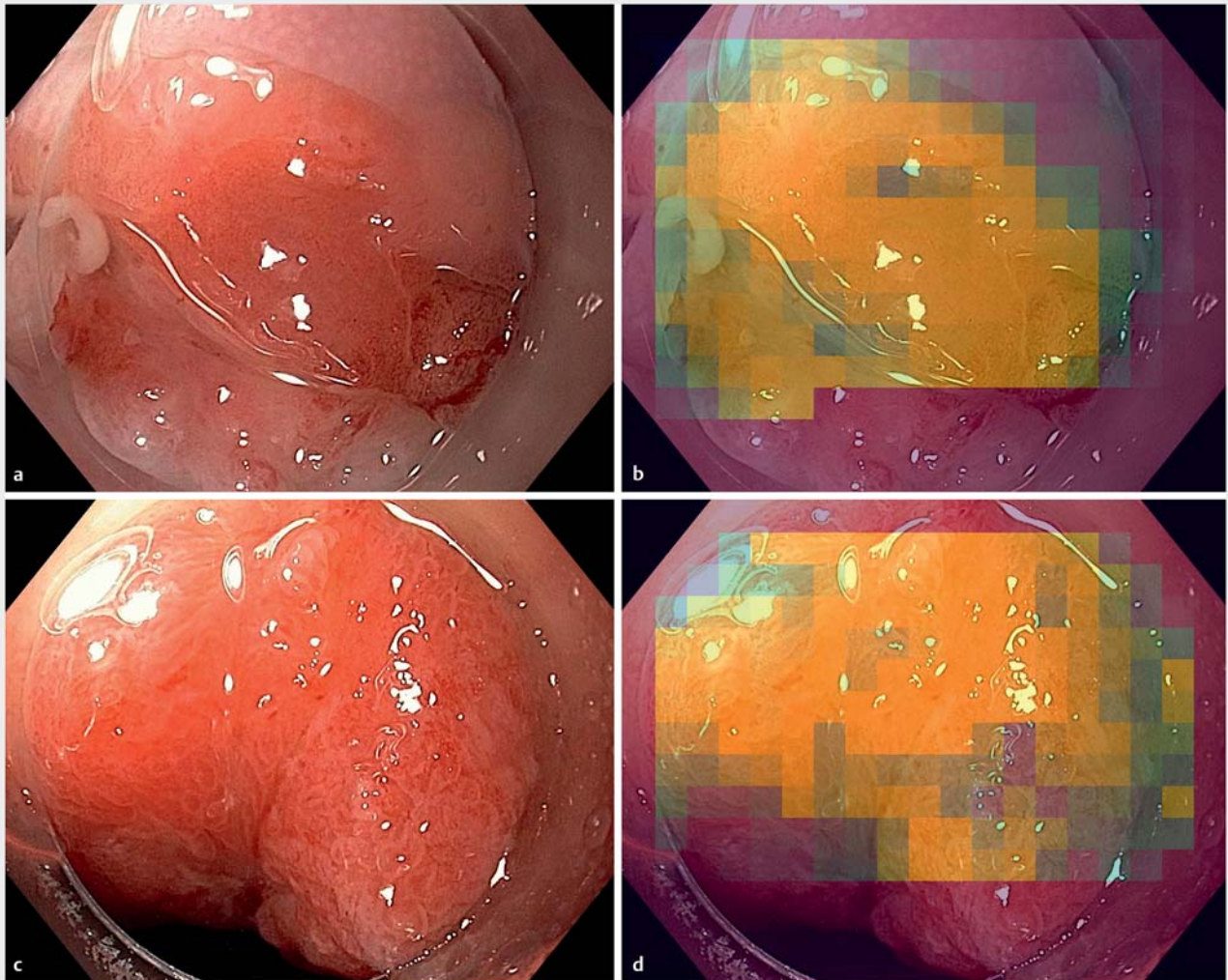
A still image with a suspicious region may be detected automatically during an examination or recognized by the examiner. AI can be trained to recognize and identify a region of interest (ROI) during an endoscopic examination. A ROI could be a polyp – as in detection of adenomas during screening colonoscopy [13] – or a dysplastic lesion, as in detection of focal lesions during assessment of Barrett's esophagus (BE) [14].

### Classification task

Having detected a lesion, AI can be assigned the task of categorizing the lesion into different classes (► **Fig. 3**). For example, in BE, AI is able to classify a detected ROI into two categories, neoplastic vs. non-neoplastic [15–17], with the potential of assisting the physician in deciding which therapy to implement.

Another application of the classification task in AI can be found in the colon, whereby a detected polyp is further subclassified into adenomatous vs. hyperplastic [18]. This could have an important clinical implication for “optical diagnosis” in the resect-and-discard or diagnose-and-leave strategy for diminutive polyps. In the context of AI, the authors prefer the term “computer vision diagnosis” to refer to diagnosis of lesions based on image analysis.

The classification task could also involve other aspects of a lesion's morphology such as its invasion depth. The invasion depth of a malignant gastrointestinal lesion could have a significant impact on the therapeutic process. AI with deep neural



► **Fig. 3** Automatic tumor classification and segmentation on two endoscopic images (a, c) are shown by colored contours (c, d) overlaid on the original images as so-called heat maps.

networks has been shown to predict invasion depth of stomach cancer with excellent performance scores, especially when compared with non-expert physicians [19].

### Segmentation task

Segmentation or delineation of outer margins or borders of a gastrointestinal lesion is usually done by experts with the help of image-enhanced endoscopy and/or virtual or conventional chromoendoscopy [20]. Non-experts or less experienced endoscopists may find this task more difficult and could benefit from AI-assisted delineation. The segmentation or delineation task has been successfully demonstrated in still images of early esophageal and gastric cancer [17,21] and provides a tissue class determined for each pixel. In the colon, the segmentation task is less important than the detection and classification tasks.

### Task combinations

Regarding machine learning methods, some of the tasks described above are solved at the same time. For example, ROI determination in a still image (object detection task) combined with determination of the ROI class (classification task) using object detection procedures like single-shot multibox detectors [14]. Other approaches solve the segmentation task as the classification of small image patches [15–17].

### Clinical studies and data on AI/ML

The AI tasks of detection, classification and segmentation, described above, have been implemented in CAD research. ► **Table 2** provides a short overview of some clinical studies in which AI has been applied in various regions of the gastrointestinal tract. In the interpretation of clinical studies on AI, it should be noted that most studies have used endoscopic still images rather than more complex video sequences. Also, a distinction

► **Table 2** Selected studies of use of AI in the gastrointestinal tract.

Reference/year	Organ/disease	AI application task	ML- modality	Outcome
Ebigbo A, et al; 2018 [17]	Barrett's esophagus	Classification: cancer vs. non-cancer	DL/CNN	Sensitivity 97 % and Specificity 88 %; outperformed human endoscopists
Horie Y, et al; 2018 [25]	Esophageal SCC	Detection of cancer and classification into superficial and advanced cancer	DL/CNN	Sensitivity of 98 % in the detection of cancer and a diagnostic accuracy of 98 % in the differentiation between superficial and advanced cancer
Kanesaka, et al; 2018 [28]	Gastric cancer	Identification of cancer on NBI images; delineation task	CNN	Accuracy of 96 % and 73,8 % respectively in the identification and delineation tasks.
Zhu Y, et al; 2019 [19]	Gastric cancer	Evaluation of the invasion depth of gastric cancer	CNN	Overall accuracy of 89.16 % which was significantly higher than that of human endoscopists
Nakashima, et al; 2018 [11]	<i>H. pylori</i> gastritis	Optical diagnosis of <i>H. pylori</i> gastritis	CNN	Sensitivity/specificity >96 %
Wang P, et al; 2019 [29]	Colonic polyps	Real-time automatic polyp detection	CNN	Significant increase in detection of diminutive adenomas and hyperplastic polyps (29.1 % vs 20.3 %, $P < 0.001$ )
Mori Y, et al; 2018 [18]	Colonic polyps	Detection task; Real-Time identification of diminutive polyps	CNN	Pathologic prediction rate of 98,1 %

DL, deep learning; CNN, convolutional neural network; SCC, squamous cell carcinoma.

needs to be made between hand-crafted models and DL algorithms because although DL needs far more learning data, it has the capacity to outperform more conventional hand-crafted algorithms.

## Esophagus

### Barrett's esophagus

BE is particularly challenging because of the difficulty endoscopists, especially non-experts, encounter during its assessment [22]. Detection of focal lesions as well as differentiation between non-dysplastic lesions, low-grade dysplasia, high-grade dysplasia, and adenocarcinoma can be extremely difficult [23].

Mendel et al. published a deep learning approach for analysis of BE [24]. Fifty endoscopic white light (WL) images of Barrett's cancer as well as 50 non-cancer images from an open access data base (Endoscopic Vision Challenge MICCAI 2015) were analyzed with CNNs. The system achieved a sensitivity and specificity of 94% and 88% respectively.

The same study group went further to publish a clinical paper on the classification and segmentation task in early Barrett's adenocarcinoma using deep learning [17]. Ebigbo et al. prospectively collected and analyzed 71 high-definition WL and NBI images of early (T1a) Barrett's cancer and non-dysplastic Barrett's. A sensitivity and specificity of 97% and 88% respectively was achieved in the classification of images into cancer or non-cancer. Results for the open access data base of 100 images were enhanced to sensitivity and specificity of 92% and 100%, respectively. Furthermore, the CAD model achieved a high correlation with expert annotations of cancer margins in the segmentation task with a Dice-coefficient of 0.72. Interest-

ingly, the CAD model was significantly more accurate than non-expert endoscopists who evaluated the same images.

The same open-access data set of 100 images was used by Ghatwary et al. using a deep learning-based object detection method, resulting in sensitivity and specificity of 96% and 92%, respectively [14].

In the ARGOS project by de Groof et al., a CAD model was developed using supervised learning of hand-crafted features based on color and texture [16]. Using 40 prospectively collected WL images of Barrett's cancer and 20 images of non-dysplastic BE, the CAD system had sensitivity and specificity of 95% and 85% in identification of an image as neoplastic or non-neoplastic, respectively. Furthermore, the system showed a high level of overlap with delineation of tumor margins provided by expert endoscopists.

### Squamous cell carcinoma

Horie et al. demonstrated the diagnostic evaluation of esophageal cancer by using CNN which was trained on 8428 high-resolution images and finally tested on 49 esophageal cancers (41 SCC and 8 adenocarcinomas) and 50 non-esophageal cancers [25]. The CNN system correctly detected cancer with a sensitivity of 98% and distinguished superficial from advanced cancer with a diagnostic accuracy of 98%.

Zhao et al. developed a CAD model to classify intrapapillary capillary loops (IPCL) for detection and classification of squamous cell carcinoma. A total of 1383 lesions were assessed with high-resolution endoscopes using magnification NBI [26]. The CAD system was based on a double-labelling fully convolutional network (FCN). Mean diagnostic accuracy of the model was 89.2% and 93% at the lesion and pixel levels, respectively, and performed significantly better than endoscopists.

► **Table 3** Understanding AI research: characteristics of publications.

Characteristics	Comments
Origin of images Self-acquired vs. open access database	Images generated by clinicians specifically for an AI study rather than images taken from an open-access data base may provide more accurate answers to the study hypothesis. However, an open-access database could have the advantage of improved comparability when other AI methods or studies are used on images from the same open access data base.
Quantity of images for training	Generally, the more images used in an AI study, the more accurate the results may be. However, it is not possible to make a blanket statement about the number of images needed for a high-quality research paper. To increase the quantity of training data AI researchers sometimes make use of many small subsegments of the original image. Additionally, the number of training images may be increased due to augmentation. For this, small variations of the original images are computed to simulate variations of the real-world. Standard augmentation procedures are rotation, translation and mirroring along the horizontal and vertical axis. Also, changes in contrast, brightness, hue and saturation may be applied in a randomized fashion, while the original images remain the same.
Validation and cross-validation	The true performance of an AI system has to be proven on data of the daily routine in a clinic over a long period without data selection. Since these long-term evaluations are not available yet, a fixed number of image data have to be used for training and validation. But images used for validation should never be used for training. However, testing the AI system on one validation data set only might lead to an over- or underestimation of the true performance, depending on the data separation. Therefore, only cross validation yields a statistically robust quality measure. In a cross-validation setting, the performance is assessed several times for different partitionings of the data strictly separating training and validation data, respectively. Then, the overall performance is given by the mean of all sub-experiments. Common choices of the number of sub-experiments are five or ten. But also $N$ sub-experiments for $N$ patients are used called leave-one-patient-out cross-validation.
Real-time analysis of real-life images	The analysis of real-life images in real-time comes closer to the clinical reality than the analysis of optimally collected images. The latter may lead to an over estimation of the performance ability of an AI system.
Comparison with the human expert	Controlled trials comparing the AI system in real-time with the human expert on the same set of test images may provide useful information on the performance ability of the AI system since the human expert remains the gold standard of the computer vision diagnosis.
Deep learning (DL)	AI research using DL seems to have higher potential than systems which rely on hand-crafted features only. Therefore, most recent AI studies have made use of DL algorithms.

## Stomach

Most clinical AI studies in the stomach focus on detection and characterization of gastric cancer. Hirasawa et al. trained a CNN-based system with more than 13,000 high-resolution WL, NBI and indigo carmine-stained images of gastric cancer [27]. On a second set of 2296 stomach images, a sensitivity of 92.2 % was achieved. However, a positive predictive value of only 30.6% was reached, showing that non-cancerous lesions were incorrectly identified as cancer.

In a further study on detection, Kanesaka et al. used a CNN to identify gastric cancer on magnified NBI-images with an accuracy of 96% [28]. In the delineation task, the performance of area concordance, on a block basis, demonstrated an accuracy of  $73.8\% \pm 10.9\%$ .

In characterization of gastric cancer, Zhu Y et al. applied a CNN to evaluate invasion depth on high-definition WL cancer images. The CNN-CAD system achieved an overall accuracy of 89.16%, which was significantly higher than that of human endoscopists [19].

In non-cancerous disorders, various studies have shown promising results, especially in the stomach. Itoh et al. were able to detect and diagnose *H. pylori* gastritis on WL images with a sensitivity and specificity above 85% [10]. Nakashima et al. optimized these results using blue-light imaging (BLI) and linked color imaging (LCI): sensitivity and specificity improved to

above 96% [11]. Finally, Wong et al. used machine learning to derive a predictive score which was subsequently validated in patients with *H. pylori*-negative ulcers [12].

## Colon

The greatest progress in endoscopic application of AI has been made in the colon, where AI has come close to clinical implementation in real-life settings. In an open, non-blinded trial, Wang et al. randomized 1038 patients to undergo diagnostic colonoscopy with or without assistance of a real-time automatic polyp detection system. The AI system was trained using a deep CNN, which resulted in a significant increase in the adenoma detection rate (29.1 % vs 20.3 %,  $P < 0.001$ ), especially due to a higher number of diminutive adenomas and hyperplastic polyps found [5].

A further study by Sánchez et al. used hand-crafted features (textons) inspired by Kudo's pit pattern classification to distinguish between dysplastic and non-dysplastic polyps [29]. This was the first report on AI diagnosis using HD-WL images. Interestingly, the overall diagnostic performance of the system was comparable to that achieved by endoscopists using the Kudo and NICE classification during colonoscopy as well as an expert endoscopist who evaluated polyp images off-site, after colonoscopy.

Various other AI studies using deep learning CNN models have produced excellent results in real-time identification and localization of polyps during colonoscopy as well as in differentiation between adenomatous and hyperplastic polyps identified during colonoscopy [4, 18, 30].

## The way ahead

Most studies on AI have been of a retrospective design using still images in non-clinical settings. These situations do not mimic real-life sufficiently enough to include the limitations and pitfalls of poor or difficult-to-analyze images often encountered in daily routine. Clinical trials of AI must progress to the next step, which involves real-life situations in daily endoscopic routine. Prospective analysis of video images, which is more similar to real-life situations, may be a good start. A further exciting possibility would be to demonstrate implementation of CAD and AI in all three tasks (detection, classification and delineation) during the same procedure.

Given the complex and interdisciplinary nature of medical AI research, non-AI experts as well as medical journals may have difficulty assessing papers or publications on AI and its applications. There are certain characteristics of AI papers that should be looked out for while reading, assessing or evaluating a paper or publication on endoscopic AI applications. Generally, the more images used in an AI study, the more accurate the results may be. However, by using small segments of the original image as well as implementing the principles of augmentation, the quantity of training data may be increased considerably. In validation of an AI model, cross-validation, whereby the performance is assessed several times for different partitionings of the data strictly separating training and validation data, yields statistically more robust results. Finally, clinical studies demonstrating use of AI in a real-life setting come closer to reality than studies done on high-quality, hand-picked images. These issues are highlighted in ► **Table 3**.

## Conclusion

Endoscopic AI research has shown the incredible potential CAD has in diagnostic medicine as a whole and endoscopy in particular. Concepts such as computer vision biopsies may be made feasible by AI. The assistance of endoscopists in the classical tasks of detection, characterization, and segmentation will probably be the primary application of AI. However, more studies and clinical trials showing implementation of AI in real-life settings are needed.

## Competing interests

None

## References

- [1] Esteva A, Krupel B, Novoa RA et al. Dermatologist-level classification of skin-cancer with deep neural networks. *Nature* 2017; 542: 115–118
- [2] Baker JA, Rosen EL, Lo JY et al. Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting architectural distortion. *AJR Am J Roentgenol* 2003; 81: 1083–1088
- [3] Kominami Y, Yoshida S, Tanaka S et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest Endosc* 2016; 83: 643–649
- [4] Urban G, Tripathi P, Alkayali T et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology* 2018; 155: 1069–1078
- [5] Wang P, Berzin TM, Glissen Brown JR et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019; doi:10.1136/gutjnl-2018-317500 [Published Online First: 27 February 2019]
- [6] Topol E. *Deep Medicine*. Hachette Book Group USA; 2019
- [7] Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press; 2016
- [8] He K, Zhang X, Ren S et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 770–778
- [9] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. 2015: 234–241
- [10] Itoh T, Kawahira H, Nakashima H et al. Deep learning analyzes *Helicobacter pylori* infection by upper gastrointestinal endoscopy images. *Endoscopy International Open* 2018; 06: E139–E144
- [11] Nakashima H, Kawahira H, Kawachi H et al. Artificial intelligence diagnosis of *Helicobacter pylori* infection using blue laser imaging-bright and linked color imaging: a single-center prospective study. *Ann Gastroenterol* 2018; 31: 462–468
- [12] Wong GL, Ma AJ, Deng H et al. Machine learning model to predict recurrent ulcer bleeding in patients with history of idiopathic gastroduodenal ulcer bleeding. *Aliment Pharmacol Ther* 2019; 49: 912–918
- [13] Misawa M, Kudo SE, Mori Y et al. Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology* 2018; 154: 2027–2029
- [14] Ghatwary N, Zolgharni M, Ye X. Early esophageal adenocarcinoma detection using deep learning methods. *Int J Comp Assist Radiol Surg* 2019; 14: 611–621
- [15] Van der Sommen F, Zinger S, Curvers WL et al. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. *Endoscopy* 2016; 48: 617–62
- [16] de Groof J, van der Sommen F, van der Putten J et al. The Argos project: The development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. *United European Gastroenterol J* 2019; 7: 538–547
- [17] Ebigbo A, Mendel R, Probst A et al. Computer-aided diagnosis using deep learning in the evaluation of early oesophageal adenocarcinoma. *Gut* 2019; 68: 1143–1145
- [18] Mori Y, Kudo SE, Misawa M et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. *Ann Intern Med* 2018; 169: 357–366
- [19] Zhu Y, Wang QC, Xu MD et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. *Gastrointest Endosc* 2019; 89: 806–815

- [20] Iizuka T, Kikuchi D, Hoteya S et al. The acetic acid + indigocarmine method in the delineation of gastric cancer. *J Gastroenterol Hepatol* 2008; 23: 1358–1361
- [21] Kanesaka T, Lee T-C, Uedo N et al. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. *Gastrointest Endosc* 2018; 87: 1339–1344
- [22] Scholvinck DW, van der Meulen K, Bergman JJ et al. Detection of lesions in dysplastic Barrett's esophagus by community and expert endoscopists. *Endoscopy* 2017; 49: 113–120
- [23] Sharma P, Bergman JJ, Goda K et al. Development and validation of a classification system to identify high-grade dysplasia and esophageal adenocarcinoma in Barrett's esophagus using narrow-band imaging. *Gastroenterology* 2016; 150: 591–598
- [24] Mendel R, Ebigbo A, Probst A et al. Barrett's esophagus analysis using convolutional neural networks. In: *Bildverarbeitung für die Medizin*. Springer; 2017: 80–85
- [25] Horie Y, Yoshio T, Aoyama K et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc* 2018; 89: 25–32
- [26] Zhao Y-Y, Xue D-X, Wang Y-L et al. Computer-assisted diagnosis of early esophageal squamous cell carcinoma using narrow-band imaging magnifying endoscopy. *Endoscopy* 2019; 51: 333–341
- [27] Hirasawa T, Aoyama K, Tanimoto T et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* 2018; 21: 653–660
- [28] Kanesaka T, Lee T-C, Uedo N et al. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. *Gastrointest Endosc* 2018; 87: 1339–1344
- [29] Sánchez-Montes C, Sánchez FJ, Berra J et al. Computer-aided prediction of polyp histology on white light colonoscopy using surface pattern analysis. *Endoscopy* 2019; 51: 261–265
- [30] Byrne MF, Chapados N, Soudan F et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019; 68: 94–100