MINIREVIEWS

Artificial intelligence for the early detection of colorectal cancer: A comprehensive review of its advantages and misconceptions

Michelle Viscaino, Javier Torres Bustos, Pablo Muñoz, Cecilia Auat Cheein, Fernando Auat Cheein

ORCID number: Michelle Viscaino 0000-0001-6693-7865; Javier Torres Bustos 0000-0003-1333-3894; Pablo Muñoz 0000-0001-5007-7864; Cecilia Auat Cheein 0000-0003-2042-6108; Fernando Auat Cheein 0000-0002-6347-7696.

Conflict-of-interest statement: The authors deny any conflict of interest.

Michelle Viscaino, Javier Torres Bustos, Department of Electronic Engineering, Universidad Tecnica Federico Santa Maria, Valpaiso 2340000, Chile

Pablo Muñoz, Hospital Clinico, University of Chile, Santiago 8380456, Chile

Cecilia Auat Cheein, Facultad de Medicina, Universidad Nacional de Santiago del Estero, Santiago del Estero 4200, Argentina

Fernando Auat Cheein, Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaiso 2340000, Chile

Corresponding author: Fernando Auat Cheein, PhD, Associate Professor, Department of Electronic Engineering, Universidad Técnica Federico Santa María, Av. España 1680, Valparaiso 2340000, Chile. fernando.auat@usm.cl

Abstract

Colorectal cancer (CRC) was the second-ranked worldwide type of cancer during 2020 due to the crude mortality rate of 12.0 per 100000 inhabitants. It can be prevented if glandular tissue (adenomatous polyps) is detected early. Colonoscopy has been strongly recommended as a screening test for both early cancer and adenomatous polyps. However, it has some limitations that include the high polyp miss rate for smaller (< 10 mm) or flat polyps, which are easily missed during visual inspection. Due to the rapid advancement of technology, artificial intelligence (AI) has been a thriving area in different fields, including medicine. Particularly, in gastroenterology AI software has been included in computer-aided systems for diagnosis and to improve the assertiveness of automatic polyp detection and its classification as a preventive method for CRC. This article provides an overview of recent research focusing on AI tools and their applications in the early detection of CRC and adenomatous polyps, as well as an insightful analysis of the main advantages and misconceptions in the field.

Key Words: Artificial intelligence; Machine learning; Deep learning; Medical images; Colorectal cancer; Colorectal polyps

Core Tip: Artificial intelligence-based (AI) methods have demonstrated high performance in classification, object detection, and segmentation tasks. Through multidisciplinary and collaborative work between clinicians and technicians, the advantages of AI have been successfully applied in automatic polyp detection and classification. The new AI-based systems present a better polyp detection rate and contribute to better clinical decision-making for preventing colorectal cancer (CRC). This article provides an overview of recent research focusing on AI and its applications in the early detection of CRC and adenomatous polyps.

INTRODUCTION

Colorectal cancer (CRC) is a common malignancy. Worldwide, in 2020, it ranked third among neoplasms, with incidences of 1931590 cases, representing 10% of neoplasms. However, in terms of mortality, it ranked second for the same year after lung cancer due to the crude mortality rate of 12.0 per 100000 inhabitants, prevailing in the male population[1].

Although CRC remains among the ten most frequent cancers, in a retrospective description, it was observed that at a global level, between 2000 and 2019 its ranking was stable in high-income countries, in which it maintained second place as a cause of death from neoplasia. However, in the remaining countries, it gradually increased; hence, in 2019, CRC was the 3rd leading cause of cancer death in upper-middle-income countries, the 4th leading cause in lower-middle-income countries, and the 5th leading cause in countries where income was low[2]. It is expected that by 2035, in those countries where it remains stable, the CRC mortality rate will decrease due to the application of early detection programs that are being implemented, the active participation of the population, and the prioritization of education on this matter. Nevertheless, it is expected that by 2035, in countries with low incomes, the mortality rate will continue to increase due mainly to late diagnosis and limited access to treatment if these indicators are not strategically addressed in time[3].

The risk of CRC in the general population is not uniform, and it is associated with factors such as a family history of CRC, lifestyles and eating habits and above all, the presence of polyps, either isolated or associated with genetic polyposis syndromes[4]. CRC can then be prevented by modifying diet and lifestyle, as well as early detection and timely treatment. Various studies have shown that screening tests facilitate the detection of precursor lesions in early stages. This, added to their subsequent elimination, promotes a reduction in CRC incidence and mortality[5-7].

Colonoscopy is the gold standard procedure for the diagnosis of large intestine (colon) and rectal diseases. The World Gastroenterology Organization establishes that both the sensitivity and the specificity of colonoscopy for the detection of polyps and colon cancer is 95%[4]. The United States Preventive Services Task Force determined that colonoscopy has a sensitivity between 89% and 98% for detecting adenomas of 10 mm and larger. For adenomas of 6 mm or more, the sensitivity ranges from 75% to 93%, while the specificity found was 89%[8], in which case a screening test for CRC is recommended. Additionally, in joint work between the American Cancer Society, the United States Multisocial Working Group and the American College of Radiology in 2008, colonoscopy was strongly recommended as a screening test designed to detect both early cancer and adenomatous polyps if resources were available and patients were willing to undergo an invasive test[5]. Similarly, the National Comprehensive Cancer Network recommends and promotes the application of colonoscopy for the detection of adenomatous polyps and early stages of CRC[9]. It should be noted that colonoscopy is a procedure that depends fundamentally on physician observation. In recent decades, technology has been incorporated into the inspection procedure, known as computer-aided systems.

Computer-aided detection/diagnosis systems (CADe/CADx) have been proposed, developed, and clinically used since 1966, especially in thoracic and breast imaging as well as in the cancer risk assessment[10]. The progress of computational resources and medical imaging devices has enabled CADe/CADx systems to support tasks in other areas, such as endoscopic examination[11]. CADe aims to find or localize abnormal or suspicious regions, increasing the detection rate while reducing the false negative rate (FNR). Additionally, CADx provides a second objective opinion regarding the assessment of a disease from image-based information. In the early stages of both systems, their algorithms were predominantly based on feature extraction methods engineered by domain experts[12]. However, the widespread progress of diseases and variability of cases have rendered these methods obsolete and have opened the research to new and improved methods. In particular, artificial intelligence (AI) has provided tools and algorithms capable of achieving high performance in terms of accuracy, sensitivity, and specificity to face tasks related to feature extraction, classification, detection, and region segmentation.

This work focuses on the main contributions of AI in gastroenterology, in particular, to the early detection of CRC through polyp detection and classification. We focus on those works that enhance the performance of endoscopic tests, which allow for direct visualization of existing lesions in the mucosa of the colon and rectum. With the numerous applications of AI and the growing interest in AI-related topics, some misconceptions have been discovered that are worth analysing.

WORLD OF AI

Since the term AI was first used in 1956, it has been a thriving field with relevant applications in several areas, including medicine[10]. The term AI refers to technology that allows computer systems to perform tasks that normally require human skills. The field of AI is broad and includes different fields, such as robotics, computer vision, natural language processing and machine learning, as shown in Figure 1. Often, such areas overlap to deliver more advanced features and capabilities. In medicine, robotic devices are increasingly being used in minimally invasive surgical procedures, such as robotic-assisted surgery for patients with CRC[13]. Natural language processing is another crucial AI area used to make the machine read, understand, and interpret human language. In the treatment of CRC, natural language processing has been useful for extracting relevant clinical information from scanned colonoscopy and pathology reports that would otherwise have to be extracted manually[14]. Computer vision and image processing have also been helpful in colonoscopy exploration, enhancing the visualization of lesion tissues[15]. However, from all AI fields, machine learning is the most widely used in three areas of medicine: Early detection and diagnosis, treatment, and outcome prediction and prognosis evaluation[16]. Gastrointestinal endoscopy has advanced in all three areas, but there is a clear trend in the detection and classification of polyps (see Wang *et al*[11], Nogueira-Rodríguez *et al* [17] and the references therein).

The following subsections focus on analysing the most prominent AI-based works on endoscopic tests without ignoring a brief review of the most commonly used machine learning algorithms (including deep neural networks) and evaluation metrics.

MACHINE LEARNING

Machine learning, a subset of AI, refers to a set of computer algorithms that learn from the input data provided, adjust a model through a training process, and perform predictions in novel situations by using the trained model[18]. According to the type of learning strategy, machine learning algorithms can be classified into two categories: Supervised and unsupervised learning. In supervised learning, a training set that contains the input data with the correct response (targets) must be previously available. The model is trained using the training set until it is generalized to respond correctly to all possible inputs. In the case of unsupervised learning, the correct answers are not provided, and the model attempts to group the data into categories identifying similarities between such data[19,20]. In medicine, supervised learning is the most commonly applied strategy because the goal is predicting a known outcome by mimicking a physician or health professional.

World of artificial intelligence
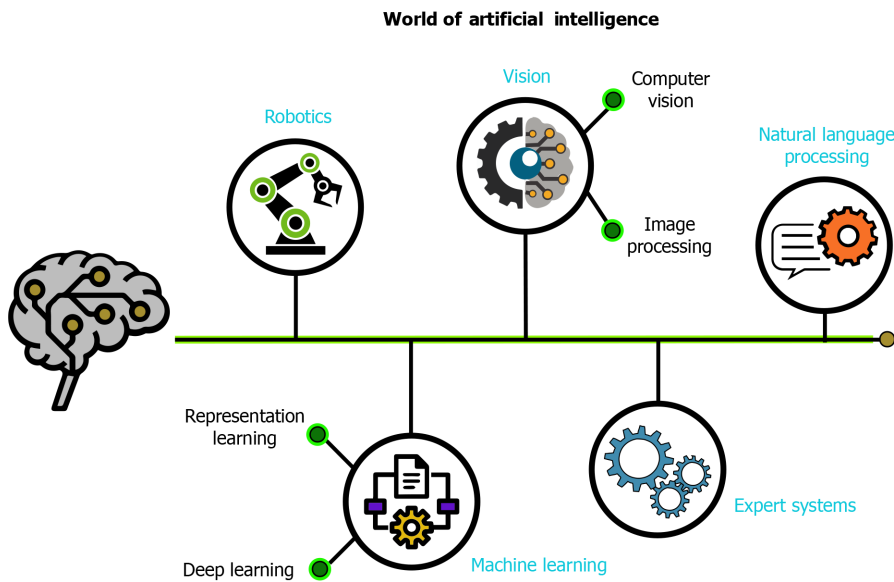
Figure 1 Artificial intelligence is a set of fields that are combining to improve tasks that involve human cognitive functions such as learning, reasoning and self-correction.

In the context of machine learning, healthcare data can be categorized as structured and unstructured. Imaging, genetic data, and electrophysiological data are some examples of structured data, whereas physical examination notes or clinical laboratory results that contain large portions of narrative texts are unstructured data[19]. The major digital data sources in medicine are images resulting from the development and improvement of different medical imaging techniques (*e.g.*, computer tomography, magnetic resonance imaging, ultrasound, X-ray and endoscopy)[10].

In gastroenterology, most computer-aided diagnosis/detection systems use images or videos, enabling the use of machine learning techniques to enhance their outcomes. In early works, CADe/x systems combined feature extractor methods and classical machine learning techniques such as random forest, decision trees, and support vector machines[21-23]. More recent works have shown applications that use deep learning algorithms such as convolutional neural networks (CNNs), in part due to the high performance and low latency of the systems[20].

The selection of one machine learning algorithm over another should be guided by analysing the available data as well as the task to be performed with it. Table 1 summarizes the main characteristics of the algorithms used in the last 5 years of state-of-the-art polyp detection and/or classification. We narrowed the analysis to the four most commonly used methods according to the Scopus and PubMed databases applied to medicine namely, support vector machines, random forest, decision trees and deep neural networks. Both support vector machines and random forests present high performance even if the data have high dimensionality. However, support vector machines are not recommended when the database is large because they increase training and inference time without improving performance[24]. Conversely, random forest presents high performance when working with large databases[25]. Deep neural networks outperform classical machine learning algorithms in almost all criteria but require large quantities of labelled data that may not be available, or the acquisition and labelling process may be very expensive or time consuming.

EVALUATION METRICS OF MACHINE LEARNING ALGORITHMS

Evaluation metrics are tied to the tasks (*e.g.*, classification, detection, localization, and segmentation) performed by the machine learning models. In gastroenterology in applications such as automatic polyp detection or classification, the evaluation metrics can be computed considering different levels: Video sequence, image, or region (pixel level).

Table 2 summarizes the terms and formulation of metrics commonly used for performance evaluation of machine learning models. In particular, those used in AI-based applications for colonoscopy. Some terms are key to understanding the

Table 1 Comparison between different types of machine learning approaches used in studies focused on polyp detection and classification

Characteristics	Support vector machine	Random forest	Decision trees	Deep neural networks	Context	Ref.
High dimensional data	High	High	Moderate	High	Performance	Shen *et al*[12]; Goodfellow *et al*[26]
Overlapped classes	Low	Low	Low	High		
Imbalance datasets	Moderate	High	Low	Moderate		
Non-linear data	Moderate	High	Moderate	High		
Larger dataset	Moderate[1]	High[1]	Low	High		
Outliers	Moderate	Moderate	Low	High	Robustness	Shen *et al*[12]; Yu *et al*[20]
Over-fitting	Moderate	High	Low	High		
Handling of missing values	Poor	Good	Good	Good		
Reproducibility	High	High	High	Moderate	Complexity	Yu *et al*[20]
Interpretability	Moderate	Moderate	High	Low		

[1]Consider that it leads to slow response time.

Table 2 Most common evaluation metrics found in the state of the art for detection, segmentation and classification tasks

Term	Symbol	Description
Positive	P	Number of real positive cases in the data
Negative	N	Number of real negative cases in the data
True positive	TP	Number of correct positive cases classified/detected
True negative	TN	Number of correct negative cases classified/detected
False positive	FP	Instances incorrectly classified/detected as positive
False negative	FN	Instances incorrectly classified/detected as negative
Area under curve	AUC	Area under the ROC plot
Term	**Task**	**Formulation**
Accuracy	C, D, S	$(TP + TN)/(TP + TN + FN + FP)$
Precision/PPV	C, D, S	$TP/(TP + FP)$
Sensitivity/Recall/TPR	C, D, S	$TP/(TP + FN)$
Specificity/TNR	C, D, S	$TN/(TN + FP)$
FPR	C, D, S	$FP/(TN + FP)$
FNR	C, D, S	$FN/(TP + FN)$
f1-score/DICE index	C, D, S	$2 \cdot (precision \cdot recall)/(precision + recall)$
f2-score	C, D, S	$4 \cdot (precision \cdot recall)/(4 \cdot precision + recall)$
IoU/Jaccard index	D, S	$(target \cap prediction)/(target \cup prediction)$
AAC	D, S	$(detected\ area \cap real\ area)/(real\ area)$

C: Classification; D: Detection; S: Segmentation. PPV: Positive predictive value; TPR: True positive rate; TNR: True negative rate. AAC: Annotated area covered.

evaluation metrics in algorithms used for automatic polyp detection and/or classification. There are two well-defined cases: Images with polyps (positive cases) and images without polyps (negative cases). In both cases, some authors[15,27,28] define a true positive (TP) when the algorithm output finds the correct region of the polyp

(detection) or labels the image as a polyp (classification). In the case of detection, only one TP is considered per polyp, avoiding over-detection. Any detection or classification as a positive case outside the region of polyp or images without polyps is considered false positive. The absence of positive output in detection or classification in images with a polyp is considered a false negative. If the algorithm does not provide any positive output in images without polyps, it is considered a true negative. Positive polyp detection is a common evaluation metric that can be computed as a true positive rate (see Table 2) or polyp-based analysis by defining a threshold of the positive frame-level predictions[29].

The most widely used evaluation metric is accuracy (see formulation in Table 2). It works well in datasets with an equal number of samples belonging to each class (*i.e.*, balanced dataset), but it is not recommended for imbalanced datasets[30]. Evaluation metrics such as sensitivity, specificity, and positive predictive value are not dependent on the class distribution; therefore, they are not biased by imbalanced datasets[31]. The use of evaluation metrics also depends on the task to be performed. In detection tasks, metrics such as the f1-score (or DICE index) and f2-score of the Jaccard index are widely used[17]. We analyse each evaluation metric below.

Accuracy represents the overall effectiveness of the algorithm when comparing the number of correctly classified/detected samples with the total number of samples[17].

Precision (positive predictive value) represents the proportion of predicted positive cases that are real positives[17].

Sensitivity (recall or true positive rate) measures the ability of the algorithm to correctly identify the positive cases[17].

Specificity (true negative rate) measures the ability of the algorithm to correctly identify the negative cases[17].

The false positive rate (FPR) represents the proportion of negative cases incorrectly identified as positive cases in the data. In statistics, the FPR is equivalent to the type I error[26].

The FNR represents the proportion of positive cases incorrectly identified as negative cases in the data. In statistics, the FPR is equivalent to the type II error[26].

The DICE index (f1-score) determines the similarity between two different areas whether the algorithm is performing segmentation or detection tasks. In classification, the f1-score is a metric to evaluate a trade-off between precision and recall[32].

The f2-score is a metric to evaluate a trade-off between precision and recall but lowers the importance of precision and increases the importance of recall[17].

The Jaccard index (IoU) is a metric mostly used in detection/segmentation algorithms and quantifies overlap between the target area and the area predicted by the algorithm[32].

Annotated area covered is an evaluation metric mostly used in detection or segmentation tasks. It represents the proportion of the real area detected/segmented by the algorithm[33].

The area under the curve (AUC) is a metric obtained from the receiver operating characteristic curve that relates to the sensitivity *vs* specificity of a binary classifier[31]. The best classifier is the one with the AUC closest to 1.

ENHANCING COLONOSCOPY OUTCOMES *VIA* AI

Colonoscopy exploration is performed through a flexible tube (endoscope) that contains a tiny video camera at the tip. The camera allows the physician to see the inside of the entire colon by displaying the image on a digital screen, as shown in Figure 2. During the process, the physician detects (or not) the presence of lesions on the colon and polyps, and then, depending on the shape, colour and texture of the polyp, determines whether to remove it[15]. The outcome of colonoscopy exploration depends on several factors. On the one hand, the procedure is intrinsically dependent on the technology used, such as the camera resolution, screen size and resolution, frame rate, and capability to deal with blurriness, among other issues[15]. On the other hand, the results can be affected by the cognitive capabilities of the physician (*e.g.*, tiredness, fatigue or concentration) during the exploration procedure[34]. Other factors, such as bowel preparation and the percentage of the colon explored, can also affect the outcome of the exploration[35].

Traditional colonoscopy has been shown to be successful when detecting polyps larger than 10 mm, which are easily detected by physicians during inspection. However, the miss rate of polyp detection increases with smaller sized and/or flat polyps[36]. There are both clinical and technical efforts to improve colonoscopy
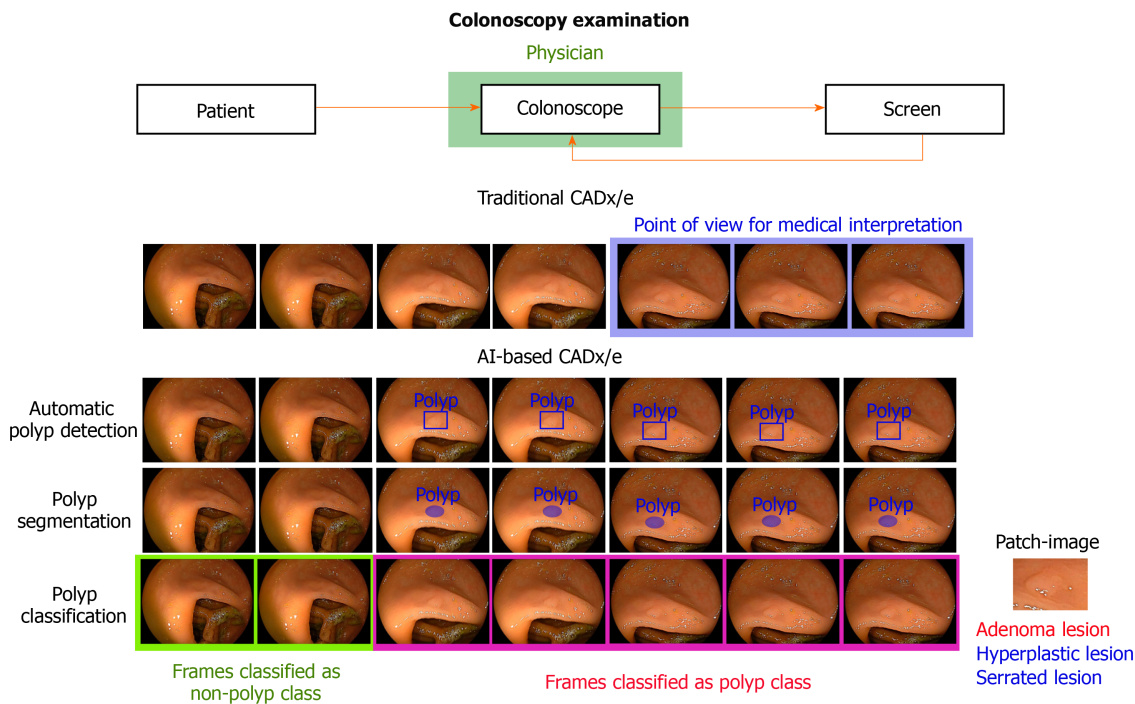
Figure 2 Comparison of traditional and AI-based computer-aided systems on colonoscopy examination.

results. For example, continuous improvement of the skills of physicians through training and practice[37] and the improvement of image/video acquisition devices combined with the development of clinically applicable CADx/e systems have been reported. Another technique used to make fine details of the mucosal surface more visible (evidencing small or slight lesions) on endoscopy tests is chromoscopy (also known as chromoendoscopy or chromocolonoscopy)[38].

Initially, chromoscopy consisted of spraying contrast dyes on the mucosa with the aim of outlining the mucosal morphology (dye-based chromoendoscopy DCE)[39]. The most frequent contrast dye used is indigo-carmine in concentrations varying from 0.2% to 2%[38]. DCE has been demonstrated to be a useful tool for endoscopists to detect and characterize lesions more accurately. A study presented by Brown *et al*[38] found that the rate of detection of small polyps was improved by DCE by approximately 90%. Such analysis was conducted on 2727 patients and showed that the detection of small polyps that could potentially develop into cancer was increased by 30% when chromoscopy was used. Although the DCE technique is simple to use and safe, it is labour intensive and time consuming, and the outcomes highly depend on bowel preparation[38,39]. Over the years, with the introduction of electronics and the improvement in technology, a new era of chromoscopy called virtual chromoendoscopy (VCE) has been adopted. VCE includes pre-processing optical imaging techniques, such as Olympus' narrow-band imaging (NBI) and autofluorescence imaging, as well as the post-processing techniques Pentax's i-SCAN and Fujinon intelligent chromoendoscopy[15]. Of all VCE techniques, NBI has been studied most frequently for assessing gastroenterology diseases[40]. This pre-processing technique uses light of specific wavelengths (green - 540 nm and blue -415 nm) to enhance detail on the mucosal surface. Although VCE techniques such as NBI can detect small-size or flat polyps, they suffer from some drawbacks, such as interobserver and intraobserver variability[41]. Such drawbacks refer to expertise, levels of distraction, or stress. However, the use of CADx/e systems may increase standardization in the process and, perhaps most importantly, more widespread adoption by non-experts in the field [41].

In this context, the new developments of CADx/e systems have focused on systems to assist in the detection and/or localization of polyps and the classification of the different types of polyps, both fundamental tasks to help clinicians at all stages of CRC diagnosis. AI has emerged as a powerful tool in two well-differentiated tasks: Polyp detection (including localization and segmentation) and polyp classification. By including an AI-based algorithm in the CADx/e systems that attend colonoscopy exploration, they can predict whether there are one (or more) polyps in a given video frame using white light alone, without the aid of advanced endoscopic imaging

modalities. If the purpose is also to locate the polyp, the algorithm predicts the position of the polyp in the image, as shown in Figure 2. If the physician requires a finer analysis, segmentation tools can allow isolation of the polyp region at the pixel level on the image. Once a polyp is detected, polyp classification aims to catalogue the type of polyp. The latter is particularly important because it allows the clinician to make a better decision as to whether to remove the polyp depending on whether it seems to be a benign, pre-malignant, or malignant polyp. Currently, to confirm if a polyp is malignant, the suspected polyp must be removed, and then a pathology test must be performed. However, an expected advance in the future is that AI can assist clinicians in differentiating polyps.

Colorectal polyps are anatomically pathologically classified as adenomatous, hyperplastic, or serrated and inflammatory, hamartomatous, juvenile, the latter being synthesized terminologically as miscellaneous, given their low prevalence (10%-20%) [42]. Adenomatous polyps are the most frequent (60%-70%). Depending on their histological characteristics, they can be tubular, villous, or tubulovillous. They can be of different degrees of dysplasia, which constitutes one of the elements for the diagnosis or the presumption of CRC[8,42]. Hyperplastic polyps have a prevalence between 10% and 30%. Although they are not usually neoplastic, there is a type of serrated polyp, the sessile serrated adenoma, which is considered a CRC precursor lesion through what is known as the serrated pathway of carcinogenesis[8,42,43].

ADVANCES IN AI FOR DETECTING AND CLASSIFYING COLORECTAL POLYPS

Automatic polyp detection—including classification and segmentation—in colonoscopy videos has been an active research topic during the last two decades. After analysing approaches reported in the literature, there are three well-defined methods: Hand-crafted, feature-based machine learning, and end-to-end learning approaches. Each method is discussed in more detail below. At the end of the analysis, we summarize the works in Table 3, showing the screening test, imaging modality and contribution for each one.

Hand-crafted approach

Hand-crafted methods refer to those based on exploiting low-level image processing techniques to obtain candidate polyp boundaries. This method considers the polyp as a protruding surface, with its boundaries detected using intensity valleys[37], Hessian filters[44], or the Hough transform[45].

Feature-based machine learning approach

Feature-based methods encompass the first era of machine learning: Designing a feature extractor and then training a classifier to predict a given class (*e.g.*, polyps or non-polyps). In early works, a texture descriptor was used to provide relevant features about the region of the image containing polyps using wavelet sub-band information in the work of Wimmer *et al*[46], Haralick co-occurrence matrix in the work of Hu *et al* [21], or Gaussian-kernel low pass filtering in the work of Mamanov *et al*[47]. Other features, such as shape, colour, and edge geometry, have also been used to create more robust detection systems that include polyp segmentation[32]. Glasmachers[33] proposed a CAD system that combines context-based image information to remove non-polyp information and shape features to reliably localize polyps.

After generating a feature vector using descriptors of texture, colour and/or shape, a classifier is required to predict whether polyps are present in the colonoscopy image, distinguishing between the different types of polyps or whether the region charac-terized on the image is a polyp (localization). The most commonly used classifiers are k-nearest neighbours[46], decision trees[22], random forests[22,32], and support vector machine[23].

End-to-end learning approach

The end-to-end (E2E) learning approach refers to training a learning system represented by a single model (generally a deep neural network)[33]. As the technology evolved and the computational capabilities increased, the use of convolu-tional neural networks as a key part of CADx/e systems is increasingly frequent in automatic polyp detection[48] and/or polyp classification tasks[49]. The advantage of the E2E approach is the possibility of designing more complex multitasking systems:

Table 3 Summary of studies focused on artificial intelligence applications for automatic polyp detection, classification, and segmentation

Study	Screening test	Imaging modality	Data type	AI-based algorithm	Contribution	Acc	Sen	Spe
Wimmer *et al* [46]	Colonoscopy	WL, NBI	Images	k-nearest neighbours	Polyp classification: non-neoplastic, neoplastic	80%	-	-
Tajbakhsh *et al* [22]	Colonoscopy	WL	Images	Decision trees; Random forest	Automatic polyp detection	-	88%	-
Hu *et al*[21]	CT Colonography	Greyscale	Images	Random forest	Polyp classification: non-neoplastic, neoplastic	-	-	-
Zhang *et al*[50]	Colonoscopy	WL, NBI	Images	CNN: Caffenet	Polyp detection and classification: benign from malignant	86%	88%	-
Shin *et al*[23]	Colonoscopy	WL	Images	Support vector machine	Whole image classification: polyps from non-polyps	96%	96%	96%
Sánchez-González *et al*[32]	Colonoscopy	WL	Images	Random forest; CNN: Bayesnet	Polyp segmentation	97%	76%	99%
Tan *et al*[52]	CT Colonography	Greyscale	Images	Customized CNN	Polyp classification: adenoma from adenocarcinoma	87%	90%	71%
Fonolla *et al*[51]	Colonoscopy	WL, NBI, LCI	Images	CNN: EfficientNet	Polyp classification: benign from pre-malignant	95%	96%	93%
Hwang *et al*[46]	Colonoscopy	WL	Images	Customized CNN	Polyp detection and segmentation	-	-	-
Park *et al*[53]	Colonoscopy	WL	Images	Customized CNN	Whole image classification: normal, adenoma and adenocarcinoma	94%	~94%	-
Viscaino *et al* [54]	Colonoscopy	Greyscale	Images	Support vector machine; Decision treesk-nearest neighbours; Random forest	Whole image classification: polyp and non-polyp	97%	98%	96%

WL: White light; NBI: Narrow-band imaging; LCI: Linked colour imaging; Acc: Accuracy; Sen: Sensitivity; Spe: Specificity.

Detecting polyps and then identifying whether the detected polyp is hyperplastic or adenomatous[50]. Information about whether a polyp can be malignant will assist the clinician in making a better clinical decision (to remove or not the polyp)[51].

There are other alternatives to colonoscopy; colonography or wireless capsule endoscopy are also used as screening techniques to detect polyps. Both alternatives are less invasive and do not present a risk of perforation to patients as colonoscopy does. AI-based algorithms have also been used to enhance the analysis *via* CT colonography. In particular, the use of greyscale information in the image (using a grey-level co-occurrence matrix in the work of Tan *et al*[52] or texture information in the work of Hu *et al*[21]) combined with CNN has been useful for differentiation of polyps: Adenoma from adenocarcinoma[53], non-neoplastic from neoplastic polyp[21], or images with polyps and without polyps[54].

MICCAI 2015 POLYP DETECTION CHALLENGE

MICCAI proposes a common validation and evaluation framework of new algorithms published in the field of biomedical image analysis (Bernal *et al*[15] and the references therein). Each year MICCAI launches international competitions (challenges) that allow for benchmarking algorithms on publicly released datasets and offers a basis to discuss validation strategies[22]. In 2015, the MICCAI sub-challenge on automatic polyps was launched and represented a significant advance in the area. As a result of this competition, three large endoscopic image databases were published, establishing a benchmark for new algorithms[22,37,55].

Colonoscopy datasets

To successfully train classical machine learning models, it is necessary to have reasonably sized databases[22]. However, to train deep learning models, large databases are required because the quantity of data is related to the network

performance. The most famous public databases used in computer science are ImageNet[56], with more than 14 million natural images hand-annotated in 20000 categories, or Microsoft's COCO[57], with more than 2500000 images. The state-of-the-art reports better than 90% accuracy in classification, object detection and localization tasks with deep neural networks pre-trained with these databases. In medicine, creating large databases represents a challenge because the data and the expertly annotated ground truth are required. In the case of colonoscopy, some publicly available datasets for polyp detection and classification have been released in the last few years. In particular, efforts such as the MICCA 2015 sub-challenge have prompted different groups to create and make available the databases summarized in Table 4.

Three datasets annotated for automatic polyp detection have been very popular in the scientific community: CVC-ClinicDB[37], ETIS-Larib[55], and ASU-Mayo Clinic Colonoscopy Video[22]. Both datasets CVC-ClinicDB and ETIS-Larib are composed of annotated frames, whereas the ASU-Mayo Clinic dataset is composed of 38 fully annotated videos selected to show maximum variation in colonoscopy procedures. All public databases are summarized in Table 4, as well as their characteristics.

MISCONCEPTIONS IN AI

Deep learning-based AI models offer promising results for medical image analysis. Nevertheless, a thorough understanding of the available data and its limitations and proficient curation of suitable training, testing, and validation subsets are required to successfully train these models responsibly and use them in a clinical setup, *e.g.*, as a diagnostic support tool. Following are some of the most common misconceptions.

Imbalanced datasets

In medicine, obtaining samples to create datasets can be a time consuming and expensive process[15]. This becomes even more complicated when samples are obtained from invasive procedures such as colonoscopy. Another important aspect is that as a result of data scarcity (*e.g.*, due to the low incidence of a medical condition), an intrinsic imbalance in the data can occur. Therefore, the available colonoscopy datasets mostly do not contain the same number of samples per class (also known as an imbalanced dataset)[29,63,64]. If a deep learning model is trained on such a dataset, the result will present a high risk of exhibiting bias against the minority classes and, in extreme cases, ignoring it altogether.

Moreover, the dataset structure needs to be considered when studying the performance metrics of deep learning models, such as accuracy and/or error rate, which are the most frequently used metrics when evaluating classification results. However, both are insufficient when working with an imbalanced dataset, as the relative contribution of the minority classes to these metrics is negligible[30]. Best practice is to be aware of the limitations of each metric and evaluate the performance of the algorithm with a set of complementary metrics (see Table 2).

The time and effort required to curate balanced datasets for intrinsically imbalanced problems have led researchers to develop techniques to enable AI models to be successfully trained on an imbalanced dataset[30]. Currently, the proposed methods can be grouped into data-level techniques and algorithm-level methods, which can be combined in hybrid approaches.

Data-level techniques aim to decrease the level of imbalance by modifying the class distribution within the available dataset. On the one hand, under-sampling methods voluntarily discard data from the majority classes, reducing the total information available to train the model. The simplistic approach to under-sampling is random under-sampling, which discards random samples from the majority classes. Notwithstanding, valuable information might be lost in the process. Intelligent under-sampling methods select removal candidates using more elaborated criteria, such as redundancy within each class in the majority group, known as one-sided selection[65], or their distance from minority samples, known as near-miss algorithms, as the several alternatives presented in Mani *et al*[66]. On the other hand, over-sampling methods artificially increase the quantity of available data in the minority classes. One technique, random over-sampling (ROS), which randomly duplicates samples from the minority classes, is the naive approach to over-sampling and is known to cause overfitting[67]. The model memorizes particular training samples instead of learning the underlying characteristics of the corresponding class and is then unable to generalize to novel data [26]. Several methods have been proposed to reduce over-fitting while over-sampling, such as the synthetic minority over-sampling technique introduced in Chawla *et al*[68]

Table 4 Summary of publicly available colonoscopy datasets

Dataset	Year	Description	Data type	Ground truth
CVC-ColonDB[29,58]	2012	380 sequential WL images from 15 videos	Images (574 × 500 pixels)	Binary mask to locate the polyp
CVC-PolypHD[58,59]	2012	56 WL images	Images (1920 × 1080 pixels)	Binary mask to locate the polyp
ETIS-Larib[55]	2014	196 WL images from 34 video sequences (44 different polyps)	Images (1125 × 966)	Binary mask to locate the polyp
CVC-ClinicDB[37]	2015	612 sequential WL images from 31 videos sequences (31 different polyps)	Images (388 × 284 pixels)	Binary mask to locate the polyp
ASU-Mayo[22]	2016	38 short video sequences (NBI, WL)	Video (SD and HD video)	Binary mask for 20 training videos
Colonoscopic dataset[49]	2016	76 short video sequences (NBI, WL)	Video	Labels: hyperplastic, adenoma and serrated
Kvasir-SEG[60]	2017	1000 images with polyps	Images	Binary mask to locate the polyp
CVC-ClinicVideoDB[61]	2017	18 sequences	Video (SD video)	Binary mask to locate the polyp
CP-CHILD-A, CP-CHILD-B[62]	2020	10000 images	Images (256 × 256)	Labels: polyp and non-polyp

WL: White light; NBI: Narrow-band imaging.

and its variants Han *et al*[69], Jo *et al*[70], or the cluster-based over-sampling method proposed in Jo *et al*[70].

Algorithm-level methods comprise cost-sensitive learning algorithms[71], which assign penalties to each majority class, increasing the importance of the minority classes, and decision process adjustments, which shift the decision threshold such that bias towards the minority classes is reduced.

Correlated data

Another dataset structure aspect to consider when inspecting performance metrics is the presence of correlation between dataset splits (most commonly training, testing and validation)[26]. We consider the task of analysing images obtained from a recorded colonoscopy to classify detected colorectal polyps as malignant or benign. If the training and validation dataset splits contain frames from the same video or patient, the correlation introduced by this situation will affect validation metrics, resulting in over-optimistic results and the risk of hidden generalization or over-fitting problems.

Interpretability

Machine learning, and more broadly AI, are essentially statistical models. During the training process, a set of parameters that define the specific behaviour of the base model is adjusted so that model predictions match expert annotations for elements in the database. Notwithstanding, commonly used models do not consider domain-specific expert knowledge in their predictions. Hence, it is possible for the trained model to learn features that are undesirable or incorrect, such as unintended patterns or visual artefacts present in the database, instead of constraining the feature space to medically relevant features only. To avoid this problem, manual assessment of the database elements is advised, along with internal feature visualization techniques[72]. See Deng *et al*[73] for a review of different strategies proposed to incorporate expert knowledge as prior information to machine learning (ML)/AI models.

FUTURE PROSPECTS

The results obtained from AI-based models are promising and establish an advantage compared to traditional methods. However, there are some limitations to be overcome by future research to propose clinically useful methods.

Overcoming real-time constraints: Videos in a colonoscopy exploration are generally acquired at 25 frames per second[15], which means that the maximum time

available to process each image (frame) must be less than 40 ms.

Increasing the variability of polyp cases, including studies with data from multiple medical centres if possible. However, given the scarcity of data on less common lesions (serrated adenomas) and knowing that deep learning approaches require vast numbers of labelled training samples, new research may include techniques such as few-shot learning introduced by Vinyals *et al*[74]. This technique focuses on learning a class from one or a few labelled samples and has been successfully applied in other medical areas, such as cervical cancer cell classification[75], breast cancer classification [76], and metastatic tumour classification[77].

Including in the polyp detection scheme the ability to detect other elements such as folds or blood vessels that can appear in a real exploration and can affect current methods' performance.

Tests were performed on complete video sequences to analyse the performance of the model under temporal consistency constraints and high variability in polyp appearance due to camera progression. Both conditions might impact the models' performance in a real clinical environment.

The ability to obtain uncertainty estimates from ML/AI model predictions is key to a responsible adoption of these techniques in clinical setups, as biased recommend-ations from CADx/e can have adverse effects on the final diagnosis. Bayesian deep learning has been proposed as a framework to address this problem, where deep learning models can deliver uncertainty information along with classification results [78] at the expense of an increased number of training parameters (and hence more training data required) or a more restricted model structure, *e.g.*, the need to incorporate dropout units within the mode architecture, as in Gal *et al*[79] 2016. Both of the abovementioned techniques have been successfully combined with active learning algorithms that enable incremental dataset labelling and/or training of the model parameters as new data become available (see Gal *et al*[80], 2017, and Woodward *et al* [81], 2016).

CONCLUSION

AI is a promising area in gastroenterology. With the processing power and high performance of algorithms such as deep learning, a new era of AI-based computer-aided systems can assist physicians in essential tasks such as colorectal polyp detection and classification. To achieve clinically useful systems, both clinicians and technicians must cooperate to mitigate AI drawbacks. Although most of the current technological effort has been focused on creating more precise polyp detection and classification tools, it remains a long path to be covered before adopting AI-based technology into the physician's daily work as an assistive tool for diagnosis decisions.

REFERENCES

1   **Ferlay J**, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. Global Cancer Observatory: Cancer Today [Internet]. Lyon, Francia: International Agency for Research on Cancer; 2020. Available from: https://gco.iarc.fr/today

2   **World Health Organization**.   Global Health Estimates 2019: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019. [Internet]. Geneva; 2020. Available from: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates

3   **Araghi M**, Soerjomataram I, Jenkins M, Brierley J, Morris E, Bray F, Arnold M. Global trends in colorectal cancer mortality: projections to the year 2035. *Int J Cancer* 2019; **144**: 2992-3000 [PMID: 30536395 DOI: 10.1002/ijc.32055]

4   **Winawer S**, Classen M, Lambert R, Fried M, Dite P, Goh KL, Guarner F, Lieberman D, Eliakim R, Levin B, Saenz R, Khan AG, Khalif I, Lanas A, Lindberg G, O'Brien MJ, Young G, Krabshuis J, Smith R, Schmiegel W, Rex D, Amrani N, Zauber A. Colorectal cancer screening World Gastroenterology Organisation/International Digestive Cancer Alliance Practice Guidelines. *South African Gastroenterol Rev* 2008; **6**: 13-20 [DOI: 10.4314/sagr.v6i1.30745]

5   **Levin B**, Lieberman DA, McFarland B, Smith RA, Brooks D, Andrews KS, Dash C, Giardiello FM, Glick S, Levin TR, Pickhardt P, Rex DK, Thorson A, Winawer SJ; American Cancer Society Colorectal Cancer Advisory Group; US Multi-Society Task Force; American College of Radiology Colon Cancer Committee. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin* 2008; **58**: 130-160 [PMID: 18322143 DOI: 10.3322/CA.2007.0018]

6   **Wolf AMD**, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, Etzioni R, McKenna

MT, Oeffinger KC, Shih YT, Walter LC, Andrews KS, Brawley OW, Brooks D, Fedewa SA, Manassaram-Baptiste D, Siegel RL, Wender RC, Smith RA. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin* 2018; **68**: 250-281 [PMID: 29846947 DOI: 10.3322/caac.21457]

7   **Walsh JM**, Terdiman JP. Colorectal cancer screening: scientific review. *JAMA* 2003; **289**: 1288-1296 [PMID: 12633191 DOI: 10.1001/jama.289.10.1288]

8   **Lin JS**, Piper MA, Perdue LA, Rutter C, Webber EM, O'Connor E, Smith N, Whitlock EP. Screening for Colorectal Cancer: A Systematic Review for the U.S. Preventive Services Task Force [Internet]. 2016. Rockville (MD): Agency for Healthcare Research and Quality (US); 2016 Jun. Report No.: 14-05203-EF-1 [PMID: 27441328]

9   **National Comprehensive Cancer Network**.   NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines): Colorectal cancer screening. National Comprehensive Cancer Network; 2013

10   **Giger ML**. Machine Learning in Medical Imaging. *J Am Coll Radiol* 2018; **15**: 512-520 [PMID: 29398494 DOI: 10.1016/j.jacr.2017.12.028]

11   **Wang KW**, Dong M. Potential applications of artificial intelligence in colorectal polyps and cancer: Recent advances and prospects. *World J Gastroenterol* 2020; **26**: 5090-5100 [PMID: 32982111 DOI: 10.3748/wjg.v26.i34.5090]

12   **Shen D**, Wu G, Suk HI. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng* 2017; **19**: 221-248 [PMID: 28301734 DOI: 10.1146/annurev-bioeng-071516-044442]

13   **Schootman M**, Hendren S, Ratnapradipa K, Stringer L, Davidson NO. Adoption of Robotic Technology for Treating Colorectal Cancer. *Dis Colon Rectum* 2016; **59**: 1011-1018 [PMID: 27749475 DOI: 10.1097/DCR.0000000000000688]

14   **Laique SN**, Hayat U, Sarvepalli S, Vaughn B, Ibrahim M, McMichael J, Qaiser KN, Burke C, Bhatt A, Rhodes C, Rizk MK. Application of optical character recognition with natural language processing for large-scale quality metric data extraction in colonoscopy reports. *Gastrointest Endosc* 2021; **93**: 750-757 [PMID: 32891620 DOI: 10.1016/j.gie.2020.08.038]

15   **Bernal J**, Tajbaksh N, Sanchez FJ, Matuszewski BJ, Hao Chen, Lequan Yu, Angermann Q, Romain O, Rustad B, Balasingham I, Pogorelov K, Sungbin Choi, Debard Q, Maier-Hein L, Speidel S, Stoyanov D, Brandao P, Cordova H, Sanchez-Montes C, Gurudu SR, Fernandez-Esparrach G, Dray X, Jianming Liang, Histace A. Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge. *IEEE Trans Med Imaging* 2017; **36**: 1231-1249 [PMID: 28182555 DOI: 10.1109/TMI.2017.2664042]

16   **Jiang F**, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H.   Artificial intelligence in healthcare: past, present, and future. Stroke and vascular neurology 2017; 2: 230-243 [DOI: 10.1136/svn-2017-000101]

17   **Nogueira-Rodríguez A**, Domínguez-Carbajales R, López-Fernández H, Iglesias Á, Cubiella J, Fdez-Riverola F, Reboiro-Jato M, Glez-Peña D. Deep Neural Networks approaches for detecting and classifying colorectal polyps. *Neurocomputing* 2021; **423**: 721-734 [DOI: 10.1016/j.neucom.2020.02.123]

18   **Deo RC**. Machine Learning in Medicine. *Circulation* 2015; **132**: 1920-1930 [PMID: 26572668 DOI: 10.1161/CIRCULATIONAHA.115.001593]

19   **Ravi D**, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ. Deep Learning for Health Informatics. *IEEE J Biomed Health Inform* 2017; **21**: 4-21 [PMID: 28055930 DOI: 10.1109/JBHI.2016.2636665]

20   **Yu KH**, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng* 2018; **2**: 719-731 [PMID: 31015651 DOI: 10.1038/s41551-018-0305-z]

21   **Hu Y**, Liang Z, Song B, Han H, Pickhardt PJ, Zhu W, Duan C, Zhang H, Barish MA, Lascarides CE. Texture Feature Extraction and Analysis for Polyp Differentiation *via* Computed Tomography Colonography. *IEEE Trans Med Imaging* 2016; **35**: 1522-1531 [PMID: 26800530 DOI: 10.1109/TMI.2016.2518958]

22   **Tajbakhsh N**, Gurudu SR, Liang J. Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information. *IEEE Trans Med Imaging* 2016; **35**: 630-644 [PMID: 26462083 DOI: 10.1109/TMI.2015.2487997]

23   **Shin Y**, Balasingham I. Automatic polyp frame screening using patch based combined feature and dictionary learning. *Comput Med Imaging Graph* 2018; **69**: 33-42 [PMID: 30172091 DOI: 10.1016/j.compmedimag.2018.08.001]

24   **Nalepa J**, Kawulok M. Selecting training sets for support vector machines: a review. *Artificial Intelligence Review* 2019; **52**: 857-900 [DOI: 10.1007/s10462-017-9611-1]

25   **Breiman L**. Random Forests. *Machine Learning* 2001; **45**: 5-32 [DOI: 10.1023/A:1010933404324]

26   **Goodfellow I**, Bengio Y, Courville A.   Chapter 5: Machine Learning Basics. In: Goodfellow I, Bengio Y, Courville A. Deep Learning. The MIT Press 2016: 96-161

27   **Qadir HA**, Balasingham I, Solhusvik J, Bergsland J, Aabakken L, Shin Y. Improving Automatic Polyp Detection Using CNN by Exploiting Temporal Dependency in Colonoscopy Video. *IEEE J Biomed Health Inform* 2020; **24**: 180-193 [PMID: 30946683 DOI: 10.1109/JBHI.2019.2907434]

28   **Patel K**, Li K, Tao K, Wang Q, Bansal A, Rastogi A, Wang G. A comparative study on polyp classification using convolutional neural networks. *PLoS One* 2020; **15**: e0236452 [PMID: 32730279 DOI: 10.1371/journal.pone.0236452]

29   **Misawa M**, Kudo SE, Mori Y, Cho T, Kataoka S, Yamauchi A, Ogawa Y, Maeda Y, Takeda K, Ichimasa K, Nakamura H, Yagawa Y, Toyoshima N, Ogata N, Kudo T, Hisayuki T, Hayashi T,

Wakamura K, Baba T, Ishida F, Itoh H, Roth H, Oda M, Mori K. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. *Gastroenterology* 2018; **154**: 2027-2029.e3 [PMID: 29653147 DOI: 10.1053/j.gastro.2018.04.003]

30   **Johnson JM**, Khoshgoftaar TM. Survey on deep learning with class imbalance. *J Big Data* 2019; **6**: 1-54 [DOI: 10.1186/s40537-019-0192-5]

31   **Viscaino M**, Maass JC, Delano PH, Torrente M, Stott C, Auat Cheein F. Computer-aided diagnosis of external and middle ear conditions: A machine learning approach. *PLoS One* 2020; **15**: e0229226 [PMID: 32163427 DOI: 10.1371/journal.pone.0229226]

32   **Sánchez-González A**, García-Zapirain B, Sierra-Sosa D, Elmaghraby A. Automatized colon polyp segmentation *via* contour region analysis. *Comput Biol Med* 2018; **100**: 152-164 [PMID: 30015012 DOI: 10.1016/j.compbiomed.2018.07.002]

33   **Glasmachers T**. Limits of End-to-End Learning. Proceedings of the Ninth Asian Conference on Machine Learning. *PMLR* 2017; **77**: 17-32

34   **Lee SH**, Chung IK, Kim SJ, Kim JO, Ko BM, Hwangbo Y, Kim WH, Park DH, Lee SK, Park CH, Baek IH, Park DI, Park SJ, Ji JS, Jang BI, Jeen YT, Shin JE, Byeon JS, Eun CS, Han DS. An adequate level of training for technical competence in screening and diagnostic colonoscopy: a prospective multicenter evaluation of the learning curve. *Gastrointest Endosc* 2008; **67**: 683-689 [PMID: 18279862 DOI: 10.1016/j.gie.2007.10.018]

35   **Lebwohl B**, Kastrinos F, Glick M, Rosenbaum AJ, Wang T, Neugut AI. The impact of suboptimal bowel preparation on adenoma miss rates and the factors associated with early repeat colonoscopy. *Gastrointest Endosc* 2011; **73**: 1207-1214 [PMID: 21481857 DOI: 10.1016/j.gie.2011.01.051]

36   **Leufkens AM**, van Oijen MG, Vleggaar FP, Siersema PD. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. *Endoscopy* 2012; **44**: 470-475 [PMID: 22441756 DOI: 10.1055/s-0031-1291666]

37   **Bernal J**, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Comput Med Imaging Graph* 2015; **43**: 99-111 [PMID: 25863519 DOI: 10.1016/j.compmedimag.2015.02.007]

38   **Brown SR**, Baraza W, Din S, Riley S. Chromoscopy *vs* conventional endoscopy for the detection of polyps in the colon and rectum. *Cochrane Database Syst Rev* 2016; **4**: CD006439 [PMID: 27056645 DOI: 10.1002/14651858.CD006439.pub4]

39   **Singh R**, Chiam KH, Leiria F, Pu LZCT, Choi KC, Militz M. Chromoendoscopy: role in modern endoscopic imaging. *Transl Gastroenterol Hepatol* 2020; **5**: 39 [PMID: 32632390 DOI: 10.21037/tgh.2019.12.06]

40   **van der Laan JJH**, van der Waaij AM, Gabriëls RY, Festen EAM, Dijkstra G, Nagengast WB. Endoscopic imaging in inflammatory bowel disease: current developments and emerging strategies. *Expert Rev Gastroenterol Hepatol* 2021; **15**: 115-126 [PMID: 33094654 DOI: 10.1080/17474124.2021.1840352]

41   **Alagappan M**, Brown JRG, Mori Y, Berzin TM. Artificial intelligence in gastrointestinal endoscopy: The future is almost here. *World J Gastrointest Endosc* 2018; **10**: 239-249 [PMID: 30364792 DOI: 10.4253/wjge.v10.i10.239]

42   **Moreira L**, Castells A, Castelvi S.   Pólipos y poliposis colorrectales. In: Montoro MA, García Pagán JC: Gastroenterología y Hepatología Problemas comunes en la práctica clínica. 2nd ed. Madrid: Asociación Española de Gastroenterología; 2012: 607-616

43   **East JE**, Saunders BP, Jass JR. Sporadic and syndromic hyperplastic polyps and serrated adenomas of the colon: classification, molecular genetics, natural history, and clinical management. *Gastroenterol Clin North Am* 2008; **37**: 25-46, v [PMID: 18313538 DOI: 10.1016/j.gtc.2007.12.014]

44   **Iwahori Y**, Hattori A, Adachi Y, Bhuyan MK, Woodham RJ, Kasugai K. Automatic Detection of Polyp Using Hessian Filter and HOG Features. *Procedia Computer Sci* 2015; **60**: 730-739 [DOI: 10.1016/j.procs.2015.08.226]

45   **Ruiz L**, Guayacán L, Martínez F.   Automatic polyp detection from a regional appearance model and a robust dense Hough coding. In: 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA) Bucaramanga, Colombia 2019: 1-5 [DOI: 10.1109/STSIVA.2019.8730270]

46   **Wimmer G**, Tamaki T, Tischendorf JJ, Häfner M, Yoshida S, Tanaka S, Uhl A. Directional wavelet based features for colonic polyp classification. *Med Image Anal* 2016; **31**: 16-36 [PMID: 26948110 DOI: 10.1016/j.media.2016.02.001]

47   **Mamonov AV**, Figueiredo IN, Figueiredo PN, Tsai YH. Automated polyp detection in colon capsule endoscopy. *IEEE Trans Med Imaging* 2014; **33**: 1488-1502 [PMID: 24710829 DOI: 10.1109/TMI.2014.2314959]

48   **Hwang M**, Wang D, Kong XX, Wang Z, Li J, Jiang WC, Hwang KS, Ding K. An automated detection system for colonoscopy images using a dual encoder-decoder model. *Comput Med Imaging Graph* 2020; **84**: 101763 [PMID: 32805673 DOI: 10.1016/j.compmedimag.2020.101763]

49   **Mesejo P**, Pizarro D, Abergel A, Rouquette O, Beorchia S, Poincloux L, Bartoli A. Computer-Aided Classification of Gastrointestinal Lesions in Regular Colonoscopy. *IEEE Trans Med Imaging* 2016; **35**: 2051-2063 [PMID: 28005009 DOI: 10.1109/TMI.2016.2547947]

50   **Zhang R**, Zheng Y, Mak TW, Yu R, Wong SH, Lau JY, Poon CC. Automatic Detection and Classification of Colorectal Polyps by Transferring Low-Level CNN Features From Nonmedical Domain. *IEEE J Biomed Health Inform* 2017; **21**: 41-47 [PMID: 28114040 DOI: 10.1109/JBHI.2016.2635662]

51   **Fonollà R**, van der Zander QWE, Schreuder RM, Masclee A, Schoon EJ, van der Sommen S, de With PHN. A CNN CADx System for Multimodal Classification of Colorectal Polyps Combining WL, BLI, and LCI Modalities. *Appl Sci* 2020; **10**: 5040 [DOI: 10.3390/app10155040]

52   **Tan J**, Gao Y, Liang Z, Cao W, Pomeroy MJ, Huo Y, Li L, Barish MA, Abbasi AF, Pickhardt PJ. 3D-GLCM CNN: A 3-Dimensional Gray-Level Co-Occurrence Matrix-Based CNN Model for Polyp Classification *via* CT Colonography. *IEEE Trans Med Imaging* 2020; **39**: 2013-2024 [PMID: 31899419 DOI: 10.1109/TMI.2019.2963177]

53   **Park HC**, Kim YJ, Lee SW. Adenocarcinoma recognition in endoscopy images using optimized convolutional neural networks. *Appl Sci* 2020; **10**: 1650 [DOI: 10.3390/app10051650]

54   **Viscaino M**, Cheein FA.   Machine learning for computer-aided polyp detection using wavelets and content-based image. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2019; 23: 961-965 [DOI: 10.1109/EMBC.2019.8857831]

55   **Silva J**, Histace A, Romain O, Dray X, Granado B. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. *Int J Comput Assist Radiol Surg* 2014; **9**: 283-293 [PMID: 24037504 DOI: 10.1007/s11548-013-0926-3]

56   **Deng J**, Dong W, Socher R, Li LJ, Li K, Li FF.   ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida. 2009: 248-255 [DOI: 10.1109/CVPR.2009.5206848]

57   **Lin TY**, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Zitnick CL, Dollár P.   Microsoft COCO: Common Objects in Context; 2015. [cited 10 January 2021] Available from: http://arxiv.org/abs/1405.0312

58   **Vázquez D**, Bernal J, Sánchez FJ, Fernández-Esparrach G, López AM, Romero A, Drozdzal M, Courville A. A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images. *J Healthc Eng* 2017; **2017**: 4037190 [PMID: 29065595 DOI: 10.1155/2017/4037190]

59   **Bernal J**, Sánchez J, Vilariño F. Towards automatic polyp detection with a polyp appearance model. *Pattern Recogn* 2012; **45**: 3166-3182 [DOI: 10.1016/j.patcog.2012.03.002]

60   **Pogorelov K**, Schmidt PT, Riegler M, Halvorsen P, Randel KR, Griwodz C.   KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. In: Proceedings of the 8th ACM on Multimedia Systems. 2017 [DOI: 10.1145/3083187.3083212]

61   **Angermann Q**, Bernal J, Sánchez-Montes C, Hammami M, Fernández-Esparrach G, Dray X, Romain O, Sánchez FJ, Histace A.   Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis. In: Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures, 2017: 29-41 [DOI: 10.1007/978-3-319-67543-5_3]

62   **Wang W**, Tian J, Zhang C, Luo Y, Wang X, Li J. An improved deep learning approach and its applications on colonic polyp images detection. *BMC Med Imaging* 2020; **20**: 83 [PMID: 32698839 DOI: 10.1186/s12880-020-00482-3]

63   **Chen PJ**, Lin MC, Lai MJ, Lin JC, Lu HH, Tseng VS. Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis. *Gastroenterology* 2018; **154**: 568-575 [PMID: 29042219 DOI: 10.1053/j.gastro.2017.10.010]

64   **Lui TKL**, Wong KKY, Mak LLY, Ko MKL, Tsao SKK, Leung WK. Endoscopic prediction of deeply submucosal invasive carcinoma with use of artificial intelligence. *Endosc Int Open* 2019; **7**: E514-E520 [PMID: 31041367 DOI: 10.1055/a-0849-9548]

65   **Kubat M**, Matwin S.   Addressing the Curse of Imbalanced Training Sets: One-Sided Selection. In: ICML; 1997; 97: 179-186

66   **Zhang J**, Mani I.   KNN Approach to Unbalanced Data Distributions: A Case Study Involving Information Extraction. In: Proceedings of workshop on learning from imbalanced datasets. 2003

67   **Van Hulse J**, Khoshgoftaar TM, Napolitano A.   Experimental perspectives on learning from imbalanced data. In: Proceedings of the 24th international conference on machine learning, New York, USA. 2007: 935-942 [DOI: 10.1145/1273496.1273614]

68   **Chawla NV**, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: synthetic minority over-sampling technique. *J Artif Int Res* 2002; 321-57 [DOI: 10.1613/jair.953]

69   **Han H**, Wang W-Y, Mao B-H.   Borderline-smote: a new over-sampling method in imbalanced data sets learning. In: Huang DS, Zhang XP, Huang GB, editors. Advances in Intelligent Computing. Berlin: Springer 2005; 878-887 [DOI: 10.1007/11538059_91]

70   **Jo T**, Japkowicz N.   Class imbalances vs small disjuncts. ACM SIGKDD Explorations Newsletter. 2004; 6: 40-49 [DOI: 10.1145/1007730.1007737]

71   **Thai-Nghe N**, Gantner Z, Schmidt-Thieme L.   Cost-sensitive learning methods for imbalanced data. The 2010 International Joint Conference on Neural Networks (IJCNN); 2010. Barcelona. Spain: 1-8 [DOI: 10.1109/IJCNN.2010.5596486]

72   **Mahendran A**, Vedaldi A.   Understanding deep image representations by inverting them. In: Proceedings of the IEEE conference on computer vision and pattern recognition; Boston, USA, 2015: 5188-5196 [DOI: 10.1109/CVPR.2015.7299155]

73   **Deng C**, Ji X, Rainey C, Zhang J, Lu W. Integrating Machine Learning with Human Knowledge. *iScience* 2020; **23**: 101656 [PMID: 33134890 DOI: 10.1016/j.isci.2020.101656]

74   **Vinyals O**, Blundell C, Lillicrap T, Kavukcuoglu K, Wierstra D.   Matching Networks for One Shot Learning. Part of Advances in Neural Information Processing Systems 29 (NIPS 2016)

75   **Yarlagadda DVK**, Rao P, Rao D, Tawfik O.   A system for one-shot learning of cervical cancer cell classification in histopathology images. Proc. SPIE 10956, Medical Imaging 2019: Digital Pathology;

2019: 1095611 2019 [DOI: 10.1117/12.2512963]

76   **Cano F**, Cruz-Roa A.   An exploratory study of one-shot learning using Siamese convolutional neural network for histopathology image classification in breast cancer from few data examples. In: 15th International Symposium on Medical Information Processing and Analysis 2020; 11330: 113300A [DOI: 10.1117/12.2546488]

77   **Mostavi M**, Chiu YC, Chen Y, Huang Y. CancerSiamese: one-shot learning for primary and metastatic tumor classification. *bioRxiv* 2020; Preprint [DOI: 10.1101/2020.09.07.286583]

78   **Kendall A**, Gal Y.   What uncertainties do we need in bayesian deep learning for computer vision? NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems 2017: 5580-5590

79   **Gal Y**, Ghahramani Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In: Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA. *PMLR* 2016; **48**: 1050-1059

80   **Gal Y**, Islam R, Ghahramani Z. Deep Bayesian Active Learning with Image Data. In: Proceedings of the 34th International Conference on Machine Learning. *PMLR* 2017; **70**: 1183-1192

81   **Woodward M**, Finn C.   Active One-shot Learning. In: NIPS 2016, Deep Reinforcement Learning Workshop, Barcelona, Spain. 2016