"ARTIFICIAL HISTOLOGY" IN COLONIC NEOPLASIA:

A CRITICAL APPROACH.

Gavino FAA[1,2], Matteo FRASCHINI[3], Luca DIDACI[3], Luca SABA[4],

Mario SCARTOZZI[5], Enrico ORVIETO[6], Massimo RUGGE[7]

[1] Department of Medical Sciences and Public Health, Università degli Studi di Cagliari, 09123 Cagliari, Italy
[2] Department of Biology, College of Science and Technology, Temple University, Philadelphia, PA 19122, USA
[3] Department of Electrical and Electronic Engineering, Università degli Studi di Cagliari, 09123 Cagliari, Italy
[4] Department of Radiology, University Hospital, Università degli Studi di Cagliari, 40138 Cagliari, Italy
[5] Medical Oncology Unit, University Hospital of Cagliari, Università degli Studi di Cagliari, 09123 Cagliari, Italy
[6] Department of Pathology, ULSS 8 Berica, San Bortolo Hospital, 36100 Vicenza, Italy
[7] Department of Medicine – DIMED; General Anatomic Pathology and Cytopathology Unit, Università degli Studi di Padova, 35121 Padova, Italy

---

Word count (excluding abstract, table, and references): 3,076
Abstract words: 160
References: 67
Figures: 0
Tables: 1

Corresponding author:
Massimo RUGGE, MD, FACG
Department of Medicine - DIMED; University of Padova
Via A. Gabelli 61; 35121 Padova, Italy
E-mail: massimo.rugge@unipd.it

Gavino Faa gavino.faa@unica.it https://orcid.org/0000-0002-0189-8612
Matteo Fraschini fraschin@unica.it https://orcid.org/0000-0003-2784-6527
Luca Didaci didaci@unica.it https://orcid.org/0000-0003-4586-944X
Luca Saba luca.saba@unica.it https://orcid.org/0000-0003-2870-3771
Mario Scartozzi mario.scartozzi@unica.it https://orcid.org/0000-0001-5977-5546
Enrico Orvieto enrico.orvieto@aulss8.veneto.it https://orcid.org/0009-0001-3610-4149
Massimo Rugge massimo.rugge@unipd.it http://orcid.org/0000-0002-0679-0563

Abstract

Background: The histological assessment of colorectal precancer and cancer lesions is challenging and primarily impacts the clinical strategies of secondary colon cancer prevention. Artificial intelligence (AI) models may potentially assist in the histological diagnosis of this spectrum of phenotypical changes.

Objectives: To provide a current overview of the evidence on AI-based methods for histologically assessing colonic precancer and cancer lesions.

Methods: Based on the available studies, this review focuses on the reliability of AI-driven models in ranking the histological phenotypes included in colonic oncogenesis.

Results: This review acknowledges the efforts to shift from subjective pathologists-based to more objective AI-based histological phenotyping. However, it also points out significant limitations and areas that require improvement.

Conclusions: Current AI-driven methods have not yet achieved the expected level of clinical effectiveness, and there are still significant ethical concerns that need careful consideration. The integration of "artificial histology" into diagnostic practice requires further efforts to combine advancements in engineering techniques with the expertise of pathologists.

Keywords: colorectal dysplasia; colorectal cancer; artificial intelligence; deep learning; machine learning; gastrointestinal adenomas.

Introduction

In the last decades, the International Agency for Research on Cancer (IARC) has recorded a stepwise increase in colorectal cancer (CRC) incidence, prevalence, and mortality [1]. In high-income countries, CRC ranks third among the most incident malignancies and second among the most frequent causes of cancer-related deaths [1]. These epidemiological findings support the priority of implementing secondary CRC prevention strategies [2,3].

In high-income countries, the population-based CRC screening programs include fecal immunochemical testing, colonoscopy and histology. In this context, the histological assessment of the lesions belonging to the dysplasia spectrum (*i.e.* classical or serrated) may result in a personalized assessment of cancer risk which drives dedicated schedules of endoscopy follow-up [2,4–8].

The WHO International Agency defines dysplasia as intra-epithelial neoplastic transformation associated with the risk of neoplastic invasion. Based on the assessment of its cytoarchitectural phenotype, dysplasia (synonyms: intra-epithelial neoplasia [IEN], intraglandular neoplasia [IGN], non-invasive neoplasia [NIN]) is sub-typed in low-grade (LG-D) and high-grade dysplasia (HG-D), the latter almost always associated with invasive cancer progression.

International (cultural and educational) variability on the assessment of the histological phenotype of dysplastic lesions and pathologists' diagnostic experience result in significant intra- and inter-observer diagnostic inconsistency [9–13]. The clear-cut distinction among the lesions potentially included in the natural history of CRC (Hyperplastic phenotype-s, LG-D, HG-D, HG-D associated with invasive potential) is inconsistent [14]. The above-mentioned discrepancy deeply affects the strategies in secondary CRC prevention.

Several histological studies addressed the intra- and inter-observer consistency in dysplasia histologically grading. Unsatisfactory K-statistics values range from 0.31 to 0.55, with an estimated prevalence of misclassification in one out

of five cases [9,13,15–18]. Moreover, no substantial improvement has been achieved by applying molecular profiling techniques [19].

The current availability of digital technologies enabling to obtain high-resolution images from whole histological slides (WSI) has promoted the application of AI-based deep learning (DL) methods in histopathological diagnoses. [20,21]. Studies, conducted in heterogeneous clinical-pathological settings, have explored the reliability of the Artificial Intelligence (AI) in supporting endoscopy and histology in the assessment of gastrointestinal cancer precursors [22–25].

This review focuses on the advancements in AI-supported diagnostic models for assessing the histological phenotypes of colonic intraepithelial and invasive neoplastic lesions [11,26].

AI-driven models in the histological profiling of colorectal neoplastic lesions

In the last decade, numerous studies have explored the technical and clinical reliability of artificial intelligence in the histological profiling of colon precancer and cancer lesions [27]. Table 1 provides a summary of the machine learning (ML) and deep learning (DL) AI models that have been applied to whole slide digital images (WSIs).

TABLE 1 (provided in a separate file)

In 2017, Korbar and colleagues first applied deep learning models to digital whole slide images in the evaluation of colorectal precancerous lesions. [28]. In differentiating hyperplastic polyps, sessile serrated polyps, traditional serrated polyps, tubular adenomas, and tubulovillous adenomas, the results achieved an overall accuracy exceeding 90%. On this basis, the authors concluded that their approach “can reduce the cognitive burden on pathologists and improve their efficacy in histopathological characterization of colorectal polyps”. Upon critical reevaluation, these results reveal three significant limitations: i) the study was based

on a single internal dataset, which may result in low external reproducibility; ii) the study underlined a tendency to down-assess LG-D; iii) there was an error-prone distinction between sessile serrated and hyperplastic polyps.

Some years later, Wei *et al.* tested the accuracy of a new DL model in a representative data pool, including two large series of 508 internal and 1,182 external datasets [29]. The study involved hyperplastic polyps, sessile serrated adenomas and tubular or villous adenomas with LG-D. Unfortunately, no case of HG-D or CRC was included. The authors overtly recognized the significant differences in the diagnostic accuracy achieved by testing the internal (93.5% [95% CI, 89.6%-97.4%] and external (87.0% [95% CI, 82.7%-91.3%]) datasets.

Based on WSIs of colonic non-neoplastic lesions, adenomas with various grades of dysplasia, and CRC, a Japanese study by Iizuka and colleagues utilized convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to assess the diagnostic reliability of AI-driven models in the diagnosis of colorectal adenomas and CRC [30]. The study involved a dataset of 4,536 colon biopsies recruited from multiple centers. The accuracy values achieved by the digital model and expert pathologists were 95.6 and 85.89%, respectively. When the areas under the curve (AUCs) have been evaluated, the results showed promising values of 0.96 for CRC and 0.99 for colonic adenoma. However, it is noteworthy to observe that the translation of these results to non-Japanese contexts may be limited by the differences in diagnostic criteria applied in Eastern and Western cultural settings [31–33]. Moreover, the training set did not distinguish LG-D versus HG-D, which is a priority in pathology practice due to their major implication in histology-based clinical decision-making.

To test the “generalization ability” of their model based on DeepLab v2 with ResNet-34, Song and colleagues selected 156 "internal" histological specimens of colorectal adenomas and 111 adenomas gathered from two external institutions [34]. When tested on internal cases, the model achieved an AUC of 0.92, which was

comparable to the accuracy of over 90% on the histological slides obtained from the two external institutions. Additionally, the AI's diagnostic performance showed similar faults and learned rational reasoning to those of experienced pathologists.

In an international study involving the University of British Columbia, Canada and the Kaohsiung Medical University in Taiwan, Xu *et al*. developed a DL model to identify colorectal cancer cells in histology slides [35]. In their training dataset accounting for 275 cases, the patches containing more than 60% neoplastic cells were classified as "tumor positive." Patches with less than 40% cancer cells were deemed "tumor negative," while those containing 40-60% tumor cells were excluded from the analysis. The neural network showed impressive accuracy for non-cancer cases, achieving a median of 99.9%. However, its accuracy for colorectal cancer cases was below 95%. The authors attributed the lower accuracy in cancer predictions to the ranking choices made in the training dataset, which ultimately affected the model's performance.

In 2021, an international team including Chinese, USA, and German researchers, developed an innovative strategy for aggregating patches, focusing on the histological discrimination between CRC and non-cancer cases [36]. The model was trained and validated using a significant number of 170,099 patches and more than 14,680 WSIs. The dataset included significant variability in specimen size, shape, texture, and staining. The AI achieved an average diagnostic accuracy and AUC of 98.06% (95%CI: 97.36-98.75%) and 98.83% (95%CI 98.15- 99.51%), respectively. These values were significantly better than those obtained by pathologists, whose accuracy was 97.14% (95% CI: 96.12-98.15%). Despite the excellent results in assessing CRC, the study did not provide the expected information on the diagnostic performance in the histological spectrum of CRC precursors, which limits its usefulness in real-world clinical practice.

A subsequent study from the same group [37] evaluated the diagnostic performance of a semi-supervised learning method for assessing colorectal cancer based on whole slide images. The obtained results were comparable to those

achieved by expert pathologists. The authors concluded that labeling a limited number of image tiles to develop a baseline model could be a crucial step in integrating AI-driven models into clinical workflows.

In their study, Nasir-Moin *et al.* compared the diagnostic accuracy of 15 pathologists with that of an AI-augmented digital system in analyzing a series of 100 colorectal polypoid lesions, which included both tubular and tubulo-villous adenomas [38]. The focus of the authors was to determine whether an AI-driven model could enhance the accuracy of pathologists compared to “standard” (non-AI-supported) microscopic assessment. The results showed that classification accuracy improved from 73.9% with “traditional” microscopic analysis to 80.8% when using the AI model. The study's clinical impact is limited by the exclusion of dysplasia grading, which is a critical factor in clinical decision-making for patient management. Interestingly, the error analysis indicated that the digital system reduced the overall number of misclassifications made by the pathologists.

A study by Zhou, *et al.* tested the a deep learning model in identifying cancer cells without pixel-level annotations [39]. The study included 1,346 WSIs obtained from 1,212 CRCs and 134 non-cancer images, from The Cancer Genome Atlas (TCGA). A group of five pathologists fulfilled the annotations of cancer WSIs. The experimental results showed that this model could detect and locate cancer regions with image level only. The study had two significant limitations: first, blood vessels were misinterpreted as images of cancer, and second, blood cells were misidentified as cancer cells. Overall, the image-level framework outperformed the cell-level framework in classification tasks. Given these limitations, the authors suggest that future efforts should focus on automated feature selection and aggregation through adaptive learning. This approach would be more efficient and cost-effective, saving both time and resources.

A recent study by Ho, *et al*. examined the reliability of AI in identifying and grading precancerous lesions in the colon [40]. The study used a combination algorithm that included a DL model based on a Faster Region-Based Convolutional

Neural Network architecture, as well as a classical machine learning classifier. The initial training involved 66,191 image tiles extracted from 39 WSIs. The lesions were microscopically distinguished into “low risk" (benign, inflammatory) and "high risk" (dysplasia, malignancy). Despite the significant findings obtained in the detection of HG-D and CRC (AUC of 0.917) in the validation cohort, the relevance of the study is limited due to the histological dichotomic categorization, which does not align with the clinical priorities of diagnostic practice.

The study by Kim, *et al*. [41] is the first to really focus on using DL models to grade dysplasia on histological slides of CRCs. By assuming that dysplasia grading (LG-D *versus* HG-D) is a crucial clinical task in shaping the patient’s follow-up plans, this study evaluated the feasibility of a DL model for the detection of the following classes of colorectal lesions: i) benign; ii) low-grade dysplasia; iii) high-grade dysplasia; iv) adenocarcinoma. To this end, a deep neural network was trained with WSIs of colorectal resections. The model showed a high performance in adenocarcinoma detection, with accuracy ranging from 95.5% to 98.5%. These findings demonstrate a minimal missing rate for CRC, indicating promising opportunities for performance improvement Moreover, the model achieved a promising performance in dysplasia grading, with a sensitivity of 100% on low-grade dysplasia. These findings suggest that such a model might assist pathologists in the detection of adenocarcinoma on histological slides. Unfortunately, the error analysis evidenced the tendency of the model to overcall as adenocarcinoma cases of high-grade dysplasia. Eleven out of 63 cases labelled by the reference pathologists as high-grade dysplasia were predicted to be adenocarcinoma by the model. This effect can be seen in the sensitivity of HGD of 76.3%. When a senior gastrointestinal pathologist revised the 11 cases, the conclusion was that 6 cases should be eventually labelled adenocarcinoma. In contrast, 2 of the remaining cases should be considered as borderline lesions between high-grade dysplasia and adenocarcinoma. In the remaining 3 cases, the original diagnosis of high-grade dysplasia was confirmed by the expert gastrointestinal pathologist. Another “alarming” error regarded the prediction as a benign lesion by the model in one case of adenocarcinoma. The major limitation of this study is relative to the exclusive use of

WSIs from colorectal resections. The absence of colorectal polyps in this study indicates that the reported findings should be confirmed in further studies on colorectal polyps to better show this model's putative role in CRC screening and prevention.

In 2023, Bokhorst and coworkers used a DL model for digitally classifying histology images through a “multi-class semantic segmentation” (*i.e*.: digitalized segmentation of multiple tissue compartments) [42]. The study involved a data set of WSIs (H&E stain) of colon biopsies, including non-neoplastic samples, hyperplastic polyps (HP), LG-D and HG-D/CRC. In 1,000/1,054 tested cases, the diagnostic categorization was consistent between expert pathologists and the AI model (HG-D/CRC =265; LG-D= 679; HP= 29, non-neoplastic cases= 27). Significant inconsistencies were also documented: in 27 cases the diagnostic discrepancy involved CRC versus LG-D; in 14 cases LG-D versus HG-D, in 7 cases HP versus CRC or LG-D; in 6 cases non-neoplastic lesion versus LG-D or HP. A minor study weakness resulted from the inclusion of HG-D and CRC in the same diagnostic category. Despite the acceptable statistics of ROC/AUC values, the clinical dimension of the obtained results is still from being considered operationally and ethically acceptable in clinical practice [43].

An international group involving Portuguese and Swiss researchers developed a scalable WSI-based AI system for the diagnosis of colorectal cancer [44]. According to the authors’ definition, the model “learns from weak labels,” a strategy that is expected to lower the training step, without lowering the performance of the system. The study included an internal dataset of 10,500 WSIs of CRC cases and two external datasets. Two diagnostic classes were considered: i) dysplasia (collapsing LG-D and HG-D); ii) non-neoplastic samples. The accuracy in the distinction of the two histological phenotypes was 93.44%, with a sensitivity of 0.99%. Significant divergencies were observed in distinguishing LG-D from HG-D, with a discrimination power close to random. The capital impact of this distinction for any further clinical decision-making excludes the model from any clinical implementation.

Artificial Histology: Pitfalls and Recommendations.

In gastrointestinal settings, AI has been primarily tested as support for the endoscopy assessment of precancerous and cancerous lesions [45–48]. A recent meta-analysis, including 50 randomized controlled trials, found that the detection rates of colorectal neoplasia are higher when using computer-aided detection compared to other advanced techniques [47]. It's important to note that histology remains the reference standard for the conclusive assessment of the tested lesions, emphasizing the need of studies including both diagnostic procedures.

We currently lack comprehensive, evidence-based studies on the reliability of AI in assessing the full range of lesions involved in the natural history of CRC. The histological assessment of precancerous and CRC lesions is diagnostically challenging and is as clinically important as endoscopy profiling, both within and outside of colorectal cancer secondary prevention strategies [43,49,50].

The histological assessment and grading of precancer lesions lie on architectural and cytological variables. As a result, the dysplastic phenotype includes a combination of microscopic qualitative and quantitative variables that may (at least in part) explain the discrepancies in the pathologists' grading of the dysplastic phenotypes. The histological spectrum of dysplasia includes a range of morphological disarrangements involving the architecture of glands, cell phenotypes, and dysplasia-associated stroma. These changes increase in severity from low-grade to high-grade lesions and can often coexist in classical adenomas as well as in the precancerous progression of inflammatory bowel diseases [51–54].

Aimed to provide unequivocal diagnostic criteria, the International WHO agency provided detailed description of the diagnostic criteria distinguishing LG-D, HG-D, HG-D coexisting with intramucosal cancer. Despite this international effort, a marked interobserver variability in grading dysplasia has been reported in multiple studies [10,41]. An adjunctive drawback in the histology assessment may result from the piecemeal endoscopy removal of the dysplastic lesions. Unfortunately, such a

resection strategy increases the difficulties in assessing the histological architecture and ultimately limits the conclusive dysplasia grading.

The need for a reliable histological assessment of CRC precursors has prompted the investigation of artificial intelligence as a potential reference standard [11]. This review acknowledges the valuable efforts made to transition from subjective assessments by pathologists to objective, AI-based evaluations of dysplasia. However, this critical appraisal also reveals the current limitations of AI diagnostic models and highlights significant issues that require careful consideration.

1) The diagnostic criteria used in the pathologist-based assessment of dysplastic lesions are still controversial. Despite numerous authoritative efforts to combine the basic diagnostic criteria, significant discrepancies need to be properly solved [55];

2) The frequent piecemeal removal of raised and flat colonic mucosa lesions may significantly impact the consistent assessment of unevenly distributed architectural changes within the same biopsy set analyzed through WSIs [56,57];

3) The user-handling variabilities of technical procedures in managing tissue samples (tissue fixation, paraffin inclusion, sectioning, staining) may affect AI assessment of the lesions [58,59];

4) In most studies, the AI diagnostic accuracy depends on the microscopic evaluations of the reference pathologists in the training database. This means that any inconsistencies in the pathologists' assessments may be passed on to the "artificial eyes" when transitioning from internal/national to external/international testing databases. Assessing the reliability of AI diagnostic models and their application in real-world clinical settings emphasizes the importance of using internationally validated standardized classification systems [60];

5) In CRC precursors, as in other clinical settings, AI performance relates to the number of diagnostic classes, with dichotomous classifications performing better than multiclass/combined classification, the latter being those required in patient-centered clinical decision-making [61];

6) The role of AI in the diagnostic workflow must be clearly defined. It is essential to determine whether AI should eliminate cases from any further validation

or serve as a screening procedure to identify cases that require additional intervention from human experts. Both scenarios raise ethical concerns that need to be carefully addressed. Both scenarios raise ethical concerns that need to be thoroughly considered [62,63];

7) In multiple studies, AI has shown promise in distinguishing between different mucosal non-neoplastic and neoplastic phenotypes. However, in clinical practice, the established priority is to histologically distinguish lesions that require different patient management, such as LG-D *versus* HG-D [61,64].

As AI becomes more prevalent in pathology departments, it is still restricted to primarily implemented in advanced institutions. This emerging trend raises concerns about consolidating diagnostic expertise in only a few pathology departments. While centralizing the diagnostic procedures in large hospitals could improve the efficiency of the diagnostic workflow, it may also result in smaller institutions losing diagnostic competences, which could impact the training of future pathologists.

The diagnostic potential of machine learning (ML) and deep learning (DL) is compelling, and their current questionable performances are expected outcomes of initial exploratory efforts. Databases containing millions of images are, per se, imperfect, primarily depending on the quality of their training feeders [65]. In pathology (as in other clinical contexts), the performance of AI models depends on consistency between the training datasets and the algorithms applied in their real-world testing [66].

As critically considered through this review and most recently supported by a study testing ML and DL models focusing on CRC microsatellite instability detection, the current AI-driven methods did not still reach the expected level of clinical performance [67]. In this work-in-progress environment, we are committed to advancing our learning journey by blending rapid technical engineering progress with the traditional expertise of pathologists willing to embrace the challenge of "artificial histology" [62].

References.

[1] Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: An overview. International Journal of Cancer 2021;149:778–89. https://doi.org/10.1002/ijc.33588.

[2] Zauber AG, Winawer SJ, O'Brien MJ, et al. Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths. New England Journal of Medicine 2012;366:687–96. https://doi.org/10.1056/NEJMoa1100370.

[3] Zorzi M, Battagello J, Selby K, et al. Non-compliance with colonoscopy after a positive faecal immunochemical test doubles the risk of dying from colorectal cancer. Gut 2022;71:561–7. https://doi.org/10.1136/gutjnl-2020-322192.

[4] Mezzapesa M, Losurdo G, Celiberto F, et al. Serrated Colorectal Lesions: An Up-to-Date Review from Histological Pattern to Molecular Pathogenesis. Int J Mol Sci 2022;23:4461. https://doi.org/10.3390/ijms23084461.

[5] Rex DK, Boland CR, Dominitz JA, et al. Colorectal Cancer Screening: Recommendations for Physicians and Patients From the U.S. Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2017;153:307–23. https://doi.org/10.1053/j.gastro.2017.05.013.

[6] Wang J-D, Xu G-S, Hu X-L, et al. The histologic features, molecular features, detection and management of serrated polyps: a review. Front Oncol 2024;14. https://doi.org/10.3389/fonc.2024.1356250.

[7] Winawer SJ. Natural history of colorectal cancer. Am J Med 1999;106:3S-6S; discussion 50S-51S. https://doi.org/10.1016/s0002-9343(98)00338-6.

[8] Zorzi M, Battagello J, Amidei CB, et al. Low Colorectal Cancer Risk After Resection of High-Risk Pedunculated Polyps. Clin Gastroenterol Hepatol 2024;22:1518-1527.e7. https://doi.org/10.1016/j.cgh.2024.01.027.

[9] Denis B, Peters C, Chapelain C, et al. Diagnostic accuracy of community pathologists in the interpretation of colorectal polyps. Eur J Gastroenterol Hepatol 2009;21:1153–60. https://doi.org/10.1097/MEG.0b013e3283221124.

[10] Nagtegaal ID, Odze RD, Klimstra D, et al. The 2019 WHO classification of tumours of the digestive system. Histopathology 2020;76:182–8. https://doi.org/10.1111/his.13975.

[11] Oliveira SP, Neto PC, Fraga J, et al. CAD systems for colorectal cancer from WSI are still not ready for clinical acceptance. Sci Rep 2021;11:14358. https://doi.org/10.1038/s41598-021-93746-z.

[12] Schlemper RJ, Dawsey SM, Itabashi M, et al. Differences in diagnostic criteria for esophageal squamous cell carcinoma between Japanese and Western pathologists. Cancer 2000;88:996–1006. https://doi.org/10.1002/(sici)1097-0142(20000301)88:5<996::aid-cncr8>3.0.co;2-q.

[13] van Putten PG, Hol L, van Dekken H, et al. Inter-observer variation in the histological diagnosis of polyps in colorectal cancer screening. Histopathology 2011;58:974–81. https://doi.org/10.1111/j.1365-2559.2011.03822.x.

[14] Fujita R, Jass JR, Kaminishi M, Schlemper RJ, editors. Early Cancer of the Gastrointestinal Tract. Tokyo: Springer Japan; 2006. https://doi.org/10.1007/4-431-30173-9.

[15] Osmond A, Li-Chang H, Kirsch R, et al. Interobserver variability in assessing dysplasia and architecture in colorectal adenomas: a multicentre Canadian

study. J Clin Pathol 2014;67:781–6. https://doi.org/10.1136/jclinpath-2014-202177.
[16] Rex DK, Alikhan M, Cummings O, Ulbright TM. Accuracy of pathologic interpretation of colorectal polyps by general pathologists in community practice. Gastrointest Endosc 1999;50:468–74. https://doi.org/10.1016/s0016-5107(99)70067-2.
[17] Terry MB, Neugut AI, Bostick RM, Potter JD, Haile RW, Fenoglio-Preiser CM. Reliability in the Classification of Advanced Colorectal Adenomas. Cancer Epidemiology, Biomarkers & Prevention 2002;11:660–3.
[18] Yoon H, Martin A, Benamouzig R, et al. [Inter-observer agreement on histological diagnosis of colorectal polyps: the APACC study]. Gastroenterol Clin Biol 2002;26:220–4.
[19] Siskova A, Cervena K, Kral J, Hucl T, Vodicka P, Vymetalkova V. Colorectal Adenomas—Genetics and Searching for New Molecular Screening Biomarkers. Int J Mol Sci 2020;21:3260. https://doi.org/10.3390/ijms21093260.
[20] Faa G, Castagnola M, Didaci L, et al. The Quest for the Application of Artificial Intelligence to Whole Slide Imaging: Unique Prospective from New Advanced Tools. Algorithms 2024;17:254. https://doi.org/10.3390/a17060254.
[21] Kumar N, Gupta R, Gupta S. Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. J Digit Imaging 2020;33:1034–40. https://doi.org/10.1007/s10278-020-00351-z.
[22] Gupta S, Lieberman D, Anderson JC, et al. Recommendations for Follow-Up After Colonoscopy and Polypectomy: A Consensus Update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2020;158:1131-1153.e5. https://doi.org/10.1053/j.gastro.2019.10.026.
[23] Hassan C, Spadaccini M, Mori Y, et al. Real-Time Computer-Aided Detection of Colorectal Neoplasia During Colonoscopy. Ann Intern Med 2023;176:1209–20. https://doi.org/10.7326/M22-3678.
[24] Maeda Y, Kudo S-E, Ogata N, et al. Can artificial intelligence help to detect dysplasia in patients with ulcerative colitis? Endoscopy 2021;53:E273–4. https://doi.org/10.1055/a-1261-2944.
[25] Wang P, Liu X, Berzin TM, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. Lancet Gastroenterol Hepatol 2020;5:343–51. https://doi.org/10.1016/S2468-1253(19)30411-X.
[26] Rugge M, Genta RM, Malfertheiner P, et al. RE.GA.IN.: the Real-world Gastritis Initiative-updating the updates. Gut 2024;73:407–41. https://doi.org/10.1136/gutjnl-2023-331164.
[27] Davri A, Birbas E, Kanavos T, et al. Deep Learning on Histopathological Images for Colorectal Cancer Diagnosis: A Systematic Review. Diagnostics (Basel) 2022;12:837. https://doi.org/10.3390/diagnostics12040837.
[28] Korbar B, Olofson AM, Miraflor AP, et al. Deep Learning for Classification of Colorectal Polyps on Whole-slide Images. J Pathol Inform 2017;8:30. https://doi.org/10.4103/jpi.jpi_34_17.
[29] Wei JW, Suriawinata AA, Vaickus LJ, et al. Evaluation of a Deep Neural Network for Automated Classification of Colorectal Polyps on Histopathologic Slides.

JAMA Network Open 2020;3:e203398. https://doi.org/10.1001/jamanetworkopen.2020.3398.
[30] Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. Sci Rep 2020;10:1504. https://doi.org/10.1038/s41598-020-58467-9.
[31] Riddell RH, Iwafuchi M. Problems arising from Eastern and Western classification systems for gastrointestinal dysplasia and carcinoma: are they resolvable? Histopathology 1998;33:197–202. https://doi.org/10.1046/j.1365-2559.1998.00520.x.
[32] Rugge M. Big Data on Gastric Dysplasia Support Gastric Cancer Prevention. Clin Gastroenterol Hepatol 2022;20:1226–8. https://doi.org/10.1016/j.cgh.2022.01.038.
[33] Schlemper RJ, Itabashi M, Kato Y, et al. Differences in diagnostic criteria for gastric carcinoma between Japanese and western pathologists. Lancet 1997;349:1725–9. https://doi.org/10.1016/S0140-6736(96)12249-2.
[34] Song Z, Yu C, Zou S, et al. Automatic deep learning-based colorectal adenoma detection system and its similarities with pathologists. BMJ Open 2020;10:e036423. https://doi.org/10.1136/bmjopen-2019-036423.
[35] Xu L, Walker B, Liang P-I, et al. Colorectal Cancer Detection Based on Deep Learning. J Pathol Inform 2020;11:28. https://doi.org/10.4103/jpi.jpi_68_19.
[36] Wang KS, Yu G, Xu C, et al. Accurate diagnosis of colorectal cancer based on histopathology images using artificial intelligence. BMC Medicine 2021;19:76. https://doi.org/10.1186/s12916-021-01942-5.
[37] Yu G, Sun K, Xu C, et al. Accurate recognition of colorectal cancer with semi-supervised deep learning on pathological images. Nat Commun 2021;12:6311. https://doi.org/10.1038/s41467-021-26643-8.
[38] Nasir-Moin M, Suriawinata AA, Ren B, et al. Evaluation of an Artificial Intelligence–Augmented Digital System for Histologic Classification of Colorectal Polyps. JAMA Network Open 2021;4:e2135271. https://doi.org/10.1001/jamanetworkopen.2021.35271.
[39] Zhou C, Jin Y, Chen Y, et al. Histopathology classification and localization of colorectal cancer using global labels by weakly supervised deep learning. Comput Med Imaging Graph 2021;88:101861. https://doi.org/10.1016/j.compmedimag.2021.101861.
[40] Ho C, Zhao Z, Chen XF, et al. A promising deep learning-assistive algorithm for histopathological screening of colorectal cancer. Sci Rep 2022;12:2222. https://doi.org/10.1038/s41598-022-06264-x.
[41] Kim J, Tomita N, Suriawinata AA, Hassanpour S. Detection of Colorectal Adenocarcinoma and Grading Dysplasia on Histopathologic Slides Using Deep Learning. Am J Pathol 2023;193:332–40. https://doi.org/10.1016/j.ajpath.2022.12.003.
[42] Bokhorst J-M, Nagtegaal ID, Fraggetta F, et al. Deep learning for multi-class semantic segmentation enables colorectal cancer detection and classification in digital pathology images. Sci Rep 2023;13:8398. https://doi.org/10.1038/s41598-023-35491-z.

[43] O'Brien MJ, Winawer SJ, Zauber AG, et al. The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. Gastroenterology 1990;98:371–9.
[44] Neto PC, Montezuma D, Oliveira SP, et al. An interpretable machine learning system for colorectal cancer diagnosis from pathology slides. NPJ Precis Oncol 2024;8:56. https://doi.org/10.1038/s41698-024-00539-4.
[45] Maida M, Marasco G, Facciorusso A, et al. Effectiveness and application of artificial intelligence for endoscopic screening of colorectal cancer: the future is now. Expert Rev Anticancer Ther 2023;23:719–29. https://doi.org/10.1080/14737140.2023.2215436.
[46] Spadaccini M, Troya J, Khalaf K, et al. Artificial Intelligence-assisted colonoscopy and colorectal cancer screening: Where are we going? Dig Liver Dis 2024;56:1148–55. https://doi.org/10.1016/j.dld.2024.01.203.
[47] Spadaccini M, Iannone A, Maselli R, et al. Computer-aided detection versus advanced imaging for detection of colorectal neoplasia: a systematic review and network meta-analysis. Lancet Gastroenterol Hepatol 2021;6:793–802. https://doi.org/10.1016/S2468-1253(21)00215-6.
[48] Wallace MB, Sharma P, Bhandari P, et al. Impact of Artificial Intelligence on Miss Rate of Colorectal Neoplasia. Gastroenterology 2022;163:295-304.e5. https://doi.org/10.1053/j.gastro.2022.03.007.
[49] Shaukat A, Kahi CJ, Burke CA, Rabeneck L, Sauer BG, Rex DK. ACG Clinical Guidelines: Colorectal Cancer Screening 2021. Am J Gastroenterol 2021;116:458–79. https://doi.org/10.14309/ajg.0000000000001122.
[50] Winawer SJ, Zauber AG, O'Brien MJ, Geenen J, Waye JD. The National Polyp Study at 40: challenges then and now. Gastrointest Endosc 2021;93:720–6. https://doi.org/10.1016/j.gie.2020.09.044.
[51] Choi W-T. Non-conventional dysplastic subtypes in inflammatory bowel disease: a review of their diagnostic characteristics and potential clinical implications. J Pathol Transl Med 2021;55:83–93. https://doi.org/10.4132/jptm.2021.02.17.
[52] Harpaz N, Goldblum JR, Shepherd NA, et al. Colorectal dysplasia in chronic inflammatory bowel disease: a contemporary consensus classification and interobserver study. Hum Pathol 2023;138:49–61. https://doi.org/10.1016/j.humpath.2023.05.008.
[53] Mescoli C, Frego M, Rugge M. Pathology of dysplasia and cancer in inflammatory bowel disease. Annali Italiani Di Chirurgia 2011;82:11–8.
[54] Zhang R, Lauwers GY, Choi W-T. Increased Risk of Non-conventional and Invisible Dysplasias in Patients with Primary Sclerosing Cholangitis and Inflammatory Bowel Disease. Journal of Crohn's and Colitis 2022;16:1825–34. https://doi.org/10.1093/ecco-jcc/jjac090.
[55] Schlemper RJ, Kato Y, Stolte M. Review of histological classifications of gastrointestinal epithelial neoplasia: differences in diagnosis of early carcinomas between Japanese and Western pathologists. J Gastroenterol 2001;36:445–56. https://doi.org/10.1007/s005350170067.
[56] Belderbos TDG, Leenders M, Moons LMG, Siersema PD. Local recurrence after endoscopic mucosal resection of nonpedunculated colorectal lesions:

systematic review and meta-analysis. Endoscopy 2014;46:388–402. https://doi.org/10.1055/s-0034-1364970.
[57] Imai K, Hotta K, Ito S, Yamaguchi Y, Kishida Y, Ono H. Piecemeal resection of 20- to 25-mm colorectal polyps necessitates short-term surveillance colonoscopy to reduce local recurrence similar to en bloc removal. Gastrointest Endosc 2020;92:1276–7. https://doi.org/10.1016/j.gie.2020.06.074.
[58] Clarke EL, Treanor D. Colour in digital pathology: a review. Histopathology 2017;70:153–63. https://doi.org/10.1111/his.13079.
[59] Kim J, Choi W, Yoo D, et al. Solution-free and simplified H&E staining using a hydrogel-based stamping technology. Front Bioeng Biotechnol 2023;11:1292785. https://doi.org/10.3389/fbioe.2023.1292785.
[60] Xu H-L, Gong T-T, Liu F-H, et al. Artificial intelligence performance in image-based ovarian cancer identification: A systematic review and meta-analysis. EClinicalMedicine 2022;53:101662. https://doi.org/10.1016/j.eclinm.2022.101662.
[61] Salinas MP, Sepúlveda J, Hidalgo L, et al. A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis. Npj Digit Med 2024;7:1–23. https://doi.org/10.1038/s41746-024-01103-x.
[62] Jackson BR, Ye Y, Crawford JM, et al. The Ethics of Artificial Intelligence in Pathology and Laboratory Medicine: Principles and Practice. Acad Pathol 2021;8:2374289521990784. https://doi.org/10.1177/2374289521990784.
[63] Rugge M, Sacchi D, Cesaro S, Sbaraglia M, Locatelli F. Ethics in clinical autopsy. J Clin Pathol 2020:jclinpath-2020-206793. https://doi.org/10.1136/jclinpath-2020-206793.
[64] Xu H-L, Gong T-T, Liu F-H, et al. Artificial intelligence performance in image-based ovarian cancer identification: A systematic review and meta-analysis. EClinicalMedicine 2022;53:101662. https://doi.org/10.1016/j.eclinm.2022.101662.
[65] Schwabe D, Becker K, Seyferth M, Klaß A, Schaeffter T. The METRIC-framework for assessing data quality for trustworthy AI in medicine: a systematic review. Npj Digit Med 2024;7:1–30. https://doi.org/10.1038/s41746-024-01196-4.
[66] Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. Npj Digit Med 2022;5:1–8. https://doi.org/10.1038/s41746-022-00592-y.
[67] Faa G, Coghe F, Pretta A, et al. Artificial Intelligence Models for the Detection of Microsatellite Instability from Whole-Slide Imaging of Colorectal Cancer. Diagnostics 2024;14:1605. https://doi.org/10.3390/diagnostics14151605.

TABLE 1

Study (year)	Method	Dataset	Lesions included in the study	AI diagnostic performance	Main findings of the study	Limits of the study
Korbar, *et al.* (2017) [28]	Deep Learning (DL) on Whole Slide Images (WSIs)	Internal dataset	HP, Sessile Serrated lesions, T; T-V adenoma	Accuracy >90%	Reduced pathologists' cognitive burden	Single dataset. down-assessment of LGD;)
Wei, *et al.* (2020) [29]	DL on WSIs	Internal dataset= 508 External dataset= 1,182	HP, Sessile serrated Adenoma, T-V Adenoma	Accuracy (internal)=93.5% Accuracy (external)= 87.0%	Significant difference between internal and external datasets	No cases of HGD or CRC included
Iizuka, *et al.* (2020) [30]	CNNs & RNNs on WSIs	Multicenter study including 4,536 biopsies samples	Adenoma, CRC	95.6% (model), 85.89% (pathologists; AUCs: 0.96 (CRC), 0.99 (adenoma)	Promising AUCs for CRC/ adenoma	No distinction of LGD vs HGD
Song, *et al.* (2020) [34]	DeepLab v2 with ResNet-34	Internal histology samples = 156; external histology samples =111	Colorectal adenoma	AUC= 0.92	Model accuracy comparable to pathologists	Limited generalization testing
Xu, *et al.* (2020) [35]	DL for CRC detection on WSIs	CRC=199 histology specimens; non-cancer cases= 76	CRC detection	Accuracy (non-cancer) = 99.9% Accuracy (CRC)= 94.8%	High accuracy in non-cancer cases	Labelling bias in patch annotations, small sample size
Wang, *et al.* (2021) [36]	DL for patch aggregation on WSIs	170,099 patches from 14,680 WSIs	CRC *vs* non-cancer cases	Accuracy= 98.06%. AUC= 98.83%	Excellent CRC assessment	No data on precancer lesions
Yu, *et al.* (2021) [37]	Semi-supervised learning method	NA	CRC detection	AI's performance close to that expert pathologist	Limited details on training datasets	Limited details on training datasets
Nasir Moin, *et al.* (2021) [38]	AI-augmented digital system	100 colorectal polyps	T; T-V adenoma	Improved accuracy with AI: 73.9% to 80.8%	Reduced pathologist misclassification	No grading of dysplasia; limited clinical applicability
Zhou, *et al.* (2021) [39]	DL for CRC classification	CRC WSIs= 1,212. Non-cancer cases from TCGA= 134	CRC	Image-level model performance higher than cell-level	Blood vessels and blood cells misclassified as CRC	Blood vessels and blood cells misclassified as CRC
Ho, *et al.* (2022) [40]	DL (Faster Region-Based CNN + classical ML)	66,191 image tiles from 39 WSIs	Low-risk (benign, inflammatory) *vs* high-risk lesions (D, CRC)	AUC: 0.917 for HGD and CRC	Significant findings for HGD and CRC detection	Dichotomic lesion categorization misaligned with clinical practice
Kim, *et al.* (2023) [41]	DL model for dysplasia grading on WSIs	Colorectal resection slides	Non-cancer lesions, LGD, HGD, CRC	Diagnostic accuracy for CRC= 95.5%-98.5%; Sensitivity for LGD= 100%	High accuracy in CRC detection, promising dysplasia grading	Overassessment HGD as CRC; only colorectal resections included
Bokhorst, *et al.* (2023) [42]	DL for multi-class semantic segmentation on WSIs	1,054 WSIs (H&E) of colon biopsies	Non neoplastic samples, HP, LGD, HGD/CRC	Consistent diagnosis in 1,000/1,054 cases	High ROC/AUC values	HGD and CRC grouped together (concerns in clinical practice)
Neto, *et al.* (2024) [44]	WSI-based AI system with weak labels	10,500 Internal WSIs; 2 external datasets	No neoplasia, LGD, HGD	93.44% accuracy; 99% sensitivity	Fast training with weak labels	Weak discrimination of LGD *vs* HGD n; lacks clinical impact

Table 1.
Studies that applied artificial intelligence (AI) and deep learning (DL) models in the histological assessment of colorectal dysplasia and cancer lesions.
Each row refers to a study, detailing its key-methods, profile of the data set, Lesions included in the study, AI diagnostic performance, main study findings, and limits of the study.

Acronyms used in the table. AUC: Area under the curve; DL: Deep Learning; ML: Machine learning; WSI: whole slide image; T: tubular adenoma; V: villous adenoma; T-V: Tubulo-villous adenoma; HP: Hyperplastic Polyp; LGD: low grade dysplasia, HGD: High grade dysplasia; CRC: colorectal invasive cancer; NA: not available.