INTRODUCTION

Colorectal cancer (CRC) is the second leading cause of cancer-related mortality worldwide, with the number of cases estimated to increase to 3.2 million by 2040.1,2 Population-based CRC screenings can improve patient outcomes through early diagnosis and treatment, but have led to higher incidences of T1 (early) CRC.3,4 T1 CRC can be grouped based on the invasion depth into the mucosa (Tis), superficial submucosa (T1a <1,000 µm submucosal invasion), and deep submucosa (T1b ≥1,000 µm submucosal invasion). Although endoscopic resection is the treatment modality of choice for superficial colorectal neoplasms,5,6 further surgical resection may be recommended based on the presence of risk factors after a full histological evaluation of the resected specimen.7-9 This is due to the risk of lymph node metastasis (LNM) in T1 CRC. The histological risk factors include lymphovascular invasion, tumor budding, and histological grade10-15 in addition to the depth of invasion. However, the risk of LNM in T1 CRC is estimated to be between 6% to 14%,16-19 which indicates that the postoperative morbidity and mortality associated with surgery for T1 CRC is avoidable.20,21 As such, accurately predicting the depth of invasion on the initial colonoscopy and consistent and precise histological specimen reports are crucial in patients with T1 CRC.

Artificial intelligence (AI) has been extensively studied in the context of polyp detection22-24 and, to a lesser extent, in the prediction of polyp histology during colonoscopy. Computer-aided diagnostic (CAD) systems that perform these functions are commercially available. However, predicting the risk of LNM is a complex task in CAD systems. Unlike CAD systems for the detection and prediction of polyp histology, determining the presence or absence of LNM in T1 CRC requires the input
of different forms of data from various sources. These include predicting the depth of invasion during colonoscopy, analyzing resected specimens for histology, and interpreting radiological images from cross-sectional imaging, which are sometimes performed in the context of rectal cancer.

This narrative review aimed to summarize the current evidence and clinical applications of AI in the prediction of LNM in T1 CRC. The role of AI in colonoscopy and histological examination will be examined, and the merits and limitations of its role in predicting LNM in T1 CRC will be discussed.

METHODS

A systematic search of the PubMed (Medline), Embase, and IEEE Xplore electronic databases was performed from the database inception up to November 18, 2022 (Fig. 1). The key search terms were AI, deep learning (DL), machine learning (ML), computer-aided diagnosis, T1 colon cancer, T1 rectal cancer, T1 CRC, LNM. Electronic searches were supplemented with manual searches of the references of all the retrieved studies to identify other relevant publications. Only studies published in English were included in this review.

Common terms and definitions of the clinical and technical endpoints used in studies evaluating AI in endoscopy have already been described in our earlier review and other published papers in this field.24-31

AI PREDICTION OF THE DEPTH OF INVASION IN T1 CRC DURING ENDOSCOPY

The depth of invasion is a known risk factor for LNM in T1 CRC. Traditionally, predicting the depth of invasion during colonoscopy depends largely on the availability and use of image-enhanced endoscopy (IEE),32 with33,34 or without magnification,35-38 to accurately classify the neoplastic potential of polyps based on the surface pattern and vessel appearance. The overall morphological appearance of colorectal tumors is also a known predictor of the depth of invasion, with features such as large size, pseudo-depressed or depressed areas, and the presence of large nodules indicative of a higher risk for deep and multifocal submucosal invasion.39,40 However, IEE systems may not be readily available at all centers. Furthermore, structured training41 and experience are required even when these resources are available, resulting in wide interobserver variability.42

Early studies incorporating AI for CAD in CRC focused on differentiating invasive cancers from the normal colonic mucosa or adenomas.43 Some of these studies utilized endocytoscopy44 and confocal laser endomicroscopy45 with encouraging results, but were limited in that they could not accurately assess the depth of invasion of T1 CRC. Endocytoscopy and confocal laser endomicroscopy may not be practical in wide-scale applications, as additional training and highly specialized equipment are required, even with a CAD function to alleviate the need for training. These imaging modalities require the endoscopist to focus on a very small area of the tumor at a time, making them time-consuming and labor-intensive to use in clinical settings. Furthermore, the CAD function in these studies was not trained to consider the macroscopic features of the tumor of interest. Lui et al.46 trained an AI image classifier that could predict curative endoscopic resection in large colonic tumors with an overall accuracy of 85.5% and an area under the receiver operating characteristic (AUROC) curve of 0.837, which was similar to that of a senior endoscopist who had performed more than 200 IEE colonoscopies. However, the image classifier was unsuitable for clinical use because it required a senior endoscopist to manually map the region of interest before the AI image classifier could make a prediction.

Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram of the literature search. AI, artificial intelligence; CRC, colorectal cancer.
To overcome these technical difficulties, Luo et al.\(^47\) added a tumor-localization branch to a deep convoluted neural network (CNN) model developed by modifying the GoogLeNet architecture. This enabled the CNN model to highlight the tumor area by exploiting the localization features of class activation maps while preserving useful information that lies outside the tumor area. The classification branch then predicts the histological invasiveness in the tumor area. The AI-enhanced attention-guided white-light colonoscopy (AEWL) model achieved an overall accuracy of 91.1% (95% confidence interval [CI], 89.6%–92.4%), with an AUROC curve of 0.970 (95% CI, 0.962–0.978) in predicting non-invasive and superficially invasive colorectal tumors, which in this study were defined as Tis and T1a lesions. The corresponding sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were 91.2, 91.0, 87.6, and 93.7%, respectively. The performance of the AEWL model was evaluated against that of experienced endoscopists using white-light and IEE with magnification. The results of this study showed that the accuracy of the AEWL model in estimating the depth of CRC invasion was comparable to that of experienced endoscopists (91.1% vs. 92.6%). However, when discriminating between T1b CRC and superficially invasive CRC, the sensitivity and AUROC of the AEWL model were 51.5% and 0.637, respectively. When images of advanced CRC were added to the training dataset, the sensitivity and AUROC improved by 65.3% and 0.729, respectively. The authors hypothesized that the surface signatures of T1b and advanced CRC may share certain similarities; hence, the addition of advanced CRC images to the training dataset could improve the performance of the DL model.

The AEWL model is a fully automated CAD system that utilizes non-magnified white-light images during colonoscopy, circumventing the need for IEE images in the training of CAD systems. This is arguably clinically more useful, as white light colonoscopy is the most widely available imaging modality compared with electronic or dye-based IEE. Tokunaga et al.\(^48\) developed a CAD system using non-magnified white-light colonoscopy images and a single-shot multibox detector to differentiate advanced CRC or CRC with submucosal invasion ≥1,000 µm, which are not amenable to endoscopic resection, from superficially invasive and mucosal lesions, which could be resected endoscopically. The accuracy and AUROC curves for predicting endoscopically resectable lesions in this study were 90.3% and 0.913, respectively. The CAD system had similar sensitivity, specificity, and accuracy as expert endoscopists and was found to be superior to trainee endoscopists. However, in a subgroup analysis of T1b CRC, the rate of correct diagnosis was only 51.2%, although this outperformed that of the trainees and experts (31.5% and 41.1%; \(p<0.01\), and \(p=0.047\), respectively). This drop-off in accuracy in T1b CRC was similar to that described earlier for the AEWL model.\(^47\)

In a retrospective study using non-magnified white-light images, Ito et al.\(^49\) built a CNN specifically to assist in the diagnosis of T1b CRC. The authors augmented the data by adding flipped and rotated images, with up to six times as many images as in the original being used as input for training the CNN, while excluding images deemed unsuitable for learning in each augmentation process. A 3-fold cross validation method was used, which excluded images that were altered for data augmentation. Using these methods, the study reported an accuracy of 81.2% and an AUROC of 0.871 for differentiating T1b from T1a and Tis CRC. The reported sensitivity, specificity were 89% and 68%, respectively. In a separate study by Nakajima et al.\(^50\) non-magnified white-light images of early stage CRC labelled only with the T-stage were used to train a CNN that could output a probability level for T1b CRC. Data augmentation was applied with rotation, resizing, saturation, and exposure adjustments to increase the number of training images from the original. The CNN model was assessed on an independent test dataset from an external hospital, with a threshold of 95% used to predict T1b CRC (at least one image with a probability score of >0.95 was considered a positive predictor for T1b CRC). The specificity, which was the main outcome of this study, was 87%. This was superior to the specificity of the two novice endoscopists (48% and 22%, respectively) but inferior to that of expert endoscopists (100% and 96%, respectively). The CAD system accuracy in predicting T1b CRC was 78% and 85% for CRC ≤20 mm and >20 mm, respectively. A major limitation of this study was that the test dataset did not include T1a lesions; thus, we were unable to demonstrate the CAD system effectiveness at differentiating the threshold depth of submucosal invasion, which defines T1a and T1b CRC.

In a study by Lu et al.,\(^51\) white-light and IEE images were combined into image pairs for training Endo-CRC, a 2-model neural network consisting of white-light and IEE convolution branches, along with a feature fusion convolution block and classifier. Testing of the Endo-CRC system was conducted on video clips that ranged from 10 to 19 seconds and comprised white-light and IEE images from an external test dataset. Based on the test results from 35 videos, the authors reported an ac-
curacy of 100% in differentiating unresectable deeply invasive T1 CRC from resectable colorectal tumors. The speed of the Endo-CRC system was at least 21 image pairs per second, based on a real-time video analysis. While the results were encouraging when tested on colonoscopy videos, the Endo-CRC system is likely not ready for routine clinical use, as our experience has shown that video output from high-definition colonoscopy systems requires a processing speed of approximately 50 frames per second.24 Table 1 summarizes the current studies using AI to predict the depth of invasion in T1-CRC.46-51

**AI IN PREDICTION OF LNM ON HISTOLOGY**

Following the endoscopic resection of T1 CRC, the histological specimen was carefully examined for risk factors indicating the possibility of LNM to determine if the endoscopic resection was curative. In a clinical setting, the risk of LNM may be considered at the point of diagnosis of T1 CRC on endoscopy, when a decision on endoscopic resectability needs to be made, or after endoscopic resection, when the clinician needs to decide whether the patient requires additional surgery based on the histological findings of the resected specimen. The histological factors predicting LNM after endoscopic resection include the depth of invasion, tumor budding, histological grade, and lymphovascular invasion.10-12,52 However, the interobserver agreement between pathologists in T1 CRC for lymphovascular invasion has been shown to vary. This is further exacerbated by the fact that immunostaining may not be routinely performed in all centers.53,54 Furthermore, interobserver agreement has been reported to be even lower in the assessment of the depth of invasion and tumor budding.56-58 There is also conflicting evidence on the magnitude of the risk of LNM posed by the depth of invasion, with studies59-61 suggesting that this may not be a crucial risk factor for LNM. DL models have been studied in this context to provide an objective “2nd reader” function and for automated processing of histology slides in T1 CRC for predicting LNM.

Conventional light microscopy is considered the “gold standard” in surgical pathology;62,63 however, progress and innovations in digital imaging inspired by telepathology have led to the development of whole-slide imaging (WSI).64,65 WSI enables the digitalization of hematoxylin and eosin (H&E) slides, which can be stored, shared, and viewed by different pathologists. The standardization of H&E staining into a uniform digital format also means that DL algorithms can be deployed in histological image analysis.66 This has led to the development of AI systems that can robustly process large amounts of WSI for the diagnosis and prediction of outcomes in CRC.67 One study reported an AUROC curve of 0.988 for accurately diagnosing CRC on WSI, which was higher than that of expert pathologists (0.970)68 and

| Table 1. Summary of studies using CAD during endoscopy to predict depth of invasion in CRC |
| Study                          | Year published | AI instrument                | Data set                                      | Sensitivity (%) | Specificity (%) | PPV (%)  | NPV (%)  | AUROC (%) | Accuracy (%) |
| Lui et al.45(a)                | 2019           | CNN                          | 8,567 NBI and WLI images                     | 94.6            | 92.3            | 98.8     | 72.0     | 0.934      | 94.3         |
| Luo et al.67                   | 2021           | CNN                          | 9,368 Images (WLI)                           | 91.2            | 91.0            | 87.6     | 93.7     | 0.970      | 91.1         |
| Tokunaga et al.41(b)           | 2021           | Single shot multi-box detector | 3,442 Images (WLI)                           | 96.7            | 75              | 90.2     | 90.5     | 0.913      | 90.3         |
| Ito et al.51(a)                | 2019           | CNN                          | 190 Conventional WLI images                  | 67.5            | 89.0            | -        | -        | 0.871      | 81.2         |
| Nakajima et al.51(a)           | 2020           | CNN                          | 1,917 Plain endoscopic images                | 81              | 87              | 85       | 83       | 0.888      | 84           |
| Lu et al.51(c)                 | 2022           | CNN                          | 820,348 WLI and IEE images, 35 videos        | 90              | 94.2            | 64.7     | 98.8     | 0.956      | 93.8         |

CAD, computer-aided diagnostic; CRC, colorectal cancer; AI, artificial intelligence; PPV, positive predictive value; NPV, negative predictive value; AUROC, area under the receiver operating characteristic; CNN, convoluted neural network; NBI, narrow-band imaging; WLI, white-light imaging; IEE, image-enhanced endoscopy.

(a)Prediction endoscopically curable lesions (includes sessile serrated adenomas, tubular adenoma with or without villous component, intramucosal adenocarcinoma, and T1a lesions). (b)Differentiation between endoscopically curable lesions (adenomas, high-grade dysplasia, CRC with submucosal invasion <1,000 µm) vs. CRC with submucosal invasion >1,000 µm or advanced CRC. (c)Prediction of Tis/T1a lesions vs T1b lesions. (d)Prediction of T1b CRC. (e)Prediction of lesions with low-grade dysplasia, high-grade dysplasia, intramucosal cancer and CRC with submucosal invasion <1,000 µm vs. CRC with submucosal invasion ≥1,000 µm and advanced CRC.
could potentially be generalized for clinical use. To overcome the tedious and time-consuming process of examining specimens for abnormal areas on histology, Gupta et al.\textsuperscript{69} examined the use of DL models for classification and localization to determine regions of interest for pathologists to focus on in CRC. The study reported an AUROC curve of 0.97 using a pretrained Inception-v3 model and an AUROC curve of 0.99 with a customized Inception-ResNet-v2 Type 5 (IR-v2 Type 5) model. The prediction of LNM in CRC using DL models and WSI has also been studied.\textsuperscript{70} In a German study of 2,431 patients from the German DACHS cohort, a slide-based artificial intelligence predictor (SBAIP) score was combined with a logistic regression analysis of clinical data and externally tested in a different cohort of patients. The SBAIP had an AUROC curve of 0.612 in predicting LNM in CRC; although, it must be noted that the study included different stages of CRC and the small number of T1 CRC precluded a subgroup analysis.

Kudo et al. conducted a multicenter study to evaluate the accuracy of an artificial neural network (ANN) model for predicting LNM in patients with T1 CRC.\textsuperscript{71} Demographic and clinical data, such as patient age, sex, tumor size, location, and morphology, were combined with pathological data, such as lymphovascular invasion and grade, from 3,134 patients who had undergone endoscopic or surgical resection for T1 CRC in Japan. These clinicopathological data were used to train the ANN model, which was assessed against the current United States (US)\textsuperscript{13,14,72} and the Japanese Society for Cancer of the Colon and Rectum (JSCCR) 10 guidelines during external validation on a test dataset. The ANN model identified patients with LNM after initial endoscopic resection with an AUROC curve of 0.84, which outperformed the US (AUROC curve 0.77, \(p=0.005\)) and Japanese (AUROC curve 0.61, \(p<0.001\)) guidelines. However, histological factors, such as the depth of invasion and tumor budding, were not included in the training of the ANN model.

In a retrospective study of 316 patients with T1 CRC, Kang et al.\textsuperscript{73} evaluated the performance of the least absolute shrinkage and selection operator (LASSO) model with the JSCCR guidelines\textsuperscript{10} for prediction LNM. The ML model incorporates information from immunohistochemical staining and tumor-infiltrating lymphocytes (TIL), which mediate local host antitumor immunity, with histological factors such as depth of submucosal invasion, tumor budding, histological grade, and lymphovascular invasion. The AUROC curve in the validation set showed better accuracy in predicting LNM using the LASSO model than using the Japanese guidelines (0.765 vs. 0.518, \(p=0.003\)). An earlier Dutch study identified histological factors of lymphovascular invasion, Haggitt level 4 invasion, muscularis mucosa type B, poorly differentiated clusters, and tumor budding as differentiating factors for predicting LNM in patients with pedunculated T1 CRC.\textsuperscript{74} Using these histological factors, the LASSO model was evaluated in a large multicenter Dutch cohort of 708 patients with pedunculated T1 CRC and showed an AUROC of 0.83, which was superior to conventional models based on American/European and Japanese guidelines (AUROC curves of 0.67 and 0.64, respectively). Takamatsu et al.\textsuperscript{75} conducted a retrospective single-center study in which histological images from 397 patients with T1 CRC were used for supervised ML. The AUROC curve for the prediction of LNM was 0.938, using an optimal cut-off sensitivity of 80.0% and specificity of 94.5% in the ML model. Cross validation was performed with repeated random subsampling to generate 12 validation datasets, with an average AUROC curve of 0.822 (95% CI, 0.767–0.938). More recently, an attention-based DL model by Song et al.\textsuperscript{76} achieved an AUROC of 0.844 for predicting LNM in the test set for patients with a submucosal invasion of 1,000 to 2,000 µm. When the performance of this model was compared against the prediction of LNM using the JSCCR guidelines,\textsuperscript{10} the DL model was able to avoid 16.1% of unnecessary additional surgeries in this group of patients while not missing any patients with LNM.

To date, most studies on DL for predicting the risk of LNM on WSI have analyzed full histological specimens post-endoscopic or surgical resection. However, in a study by Kasahara et al.\textsuperscript{77} ML was used to train a model to predict the risk of LNM in biopsy specimens. The investigators analyzed the morphological features of cell nuclei extracted from WSI to create an LNM risk model with the aim of directing patients with T1 CRC to appropriate treatments based on their risk of LNM determined from pre-treatment biopsy specimens. The study demonstrated an accuracy of 80% to 85% in predicting LNM on biopsy specimens. In a separate study conducted in two large population-based cohorts of patients with T1 and T2 CRC, a DL system was used to direct human pathology experts to areas deemed to contain features highly predictive of LNM in the WSI of the primary tumor and surrounding tissues.\textsuperscript{78} An interesting finding from this study was that the hybrid application of human observers and DL-identified inflamed adipose tissue was the highest predictor of LNM. This has not been described as a known histological risk factor for LNM in T1 CRC and highlights the potential for using AI to discover new biomark-
<table>
<thead>
<tr>
<th>Study</th>
<th>Year published</th>
<th>AI instrument</th>
<th>Type of data</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>AUROC</th>
<th>Accuracy (%)</th>
<th>Features used for training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kwak et al.(^{70})</td>
<td>2021</td>
<td>CNN</td>
<td>164 cases of stage I, II, and III CRC(^{a})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.677 for PTS score</td>
<td>PTS score (consisting of adipose tissue, lymphocytes, mucus, smooth muscle, normal colon mucosa, stroma, colon cancer epithelium)</td>
</tr>
<tr>
<td>Kudo et al.(^{71})</td>
<td>2021</td>
<td>ANN</td>
<td>4073 cases of T1 CRC(^{b})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.83</td>
<td>-</td>
<td>Age, sex, tumor size, location, morphology, lymphatic invasion, vascular invasion and histological grade</td>
</tr>
<tr>
<td>Kang et al.(^{73})</td>
<td>2021</td>
<td>LASSO</td>
<td>316 cases of T1 CRC(^{c})</td>
<td>56.1(^{d})</td>
<td>87.3(^{d})</td>
<td>39.7(^{d})</td>
<td>93.0(^{d})</td>
<td>0.765</td>
<td>0.518(^{d})</td>
<td>Histology grade, lymphovascular invasion, tumor budding, background adenoma, CD3_IM, CD3_TC, CD8_IM, CD8_TC, FOXP3_TC</td>
</tr>
<tr>
<td>Takamatsu et al.(^{75})</td>
<td>2019</td>
<td>RFC</td>
<td>397 cases of T1 CRC(^{c})</td>
<td>80.0</td>
<td>94.5</td>
<td>-</td>
<td>-</td>
<td>0.938</td>
<td>-</td>
<td>Cytokeratin IHC of slides</td>
</tr>
<tr>
<td>Song et al.(^{76})</td>
<td>2022</td>
<td>Deep convolution neural network</td>
<td>400 cases of T1 CRC(^{b})</td>
<td>100(^{b})</td>
<td>45(^{b})</td>
<td>32.6(^{b})</td>
<td>-</td>
<td>0.764</td>
<td>63.8(^{b})</td>
<td>Size of cancer, depth of submucosal invasion, lymphovascular invasion, tumor budding, positive resection margin, and microsatellite instability</td>
</tr>
<tr>
<td>Kasahara et al.(^{77})</td>
<td>2022</td>
<td>Support vector machine and random forest</td>
<td>146 cases of T1b CRC(^{c})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>91.0</td>
<td>Cancer cell nuclei and their heterogeneity</td>
</tr>
<tr>
<td>Brockmoeller et al.(^{78})</td>
<td>2022</td>
<td>ShuffleNet network model</td>
<td>203 cases of T1 and T2 CRC(^{c})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.567 (for T1 CRC), 0.711 (for T2 CRC)</td>
<td>Tumor infiltrating lymphocytes, inflamed fat, inflammatory cells at the invasive edge and deeper into the submucosa and into muscularis propria, mesenteric fat, poorly differentiated tumor areas, necrosis, papillary growth pattern</td>
<td></td>
</tr>
</tbody>
</table>

AI, artificial intelligence; LNM, lymph node metastasis; PPV, positive predictive value; NPV, negative predictive value; AUROC, area under the receiver operating characteristic; CNN, convolutional neural network; CRC, colorectal cancer; PTS, peritumoral stroma; ANN, artificial neural network; LASSO, least absolute shrinkage and selection operator; RFC, random forest classifier; IHC, immunohistochemistry; -, results not reported in study.

\(^{a}\)Study included only surgically resected cases. \(^{b}\)Study included endoscopically resected, with or without additional surgical resection and lymph node dissection. \(^{c}\)Prediction of lymph node metastasis based on US guidelines (National Comprehensive Cancer Network). \(^{d}\)Prediction of lymph node metastasis based on the Japanese Society for Cancer of the Colon and Rectum guidelines. \(^{e}\)Assuming predicted probability of LNM is 20%.
ers for CRC progression. Table 2 summarizes the current studies using AI in histopathology for predicting the risk of LNM in T1 CRC.\textsuperscript{70,71,73,75-78}

DL has also been studied for the detection of microsatellite instability, mismatch repair genes, and other genetic alterations in CRC.\textsuperscript{79-81} This may highlight CRC biomarkers that can predict LNM when integrated into clinical decision-making tools. However, these DL models are still in the early stages of development and require extensive external validation. They also do not directly address the issue of LNM prediction in T1 CRC and are thus beyond the scope of this review.

**CURRENT LIMITATIONS OF STUDIES ON AI IN T1 CRC**

Despite the advances and reported outcomes of AI studies in predicting depth of invasion and LNM to guide the management of T1 CRC, there are still major gaps that limit its generalizability and clinical application. DL, a subbranch of the ML field, is the most commonly used tool in the literature on AI and colonoscopy.\textsuperscript{25,82} In this method, multiple linear and non-linear processing units are arranged in a deep architecture to extract useful information automatically and construct a model that generates the required output. DL models perform these tasks without requiring predefined features, which is characteristic of conventional ML techniques.\textsuperscript{25} The DL studied during colonoscopy is well suited for simple tasks such as polyp detection or polyp histology, as data from a single source of input (colonoscopy image projected from the processor) are passed through multiple layers in a neural network to produce a narrow output that is often binary (polyp or no polyp; hyperplastic or neoplastic, respectively). However, clinical decision-making regarding LNM in T1 CRC depends on more than one factor. The analysis of endoscopy videos during colonoscopy for depth of invasion, histopathology slides or reports on risk factors of LNM, radiological images and clinical and demographic characteristics of the population, mean that more than one source of input is available in T1 CRC cases (Fig. 2). No single DL model can accommodate the processing of all these information sources, such as how a clinician processes information during decision-making to obtain the required output in the presence or absence of LNM. Moreover, DL models require a large number of cases to build.\textsuperscript{83} When the relevant specimens available for analyses are limited—for instance, in T1b CRC—

![Fig. 2. Schematic diagram illustrating varied sources of input and differing outputs to reach a clinical decision on lymph node metastasis (LNM) in T1 colorectal cancer (CRC). AI, artificial intelligence; ML, machine learning; NLP, natural language processing; WSI, whole slide imaging; CT, computed tomography; MRI, magnetic resonance imaging; LVI, lymphovascular invasion.](image-url)
results of the models may be inconclusive at best, and in some instances investigators may need to rely on an ML method instead because of its inherent limitations.

The avoidance of overfitting a DL model and its reliability is highly dependent on the quality, number, and variability of the images used for training, as well as the demographic and clinical features of the populations from which the data are gathered. Most published studies on DL in T1 CRC acknowledge the limitations of their datasets, as the number of high-quality images of T1 CRC datasets is smaller and may lack detailed annotations compared to polyp databases used in training DL models for polyp detection and characterization. This is also reflected in the studies that may not contain sufficient or even any T1a CRC in the datasets used for validation, which prevents subgroup analysis and comparisons from obtaining clinically meaningful data for differentiating T1a from T1b CRC. Furthermore, the training, validation, and test datasets are often derived from populations in the same geographical location and are sometimes split from the same overall dataset in a single institution, leading to a risk of selection bias and overfitting due to the probability of significant overlaps in clinicopathological features when the baseline population is identical.

In addition, most studies evaluating CAD systems in T1 CRC during endoscopy are retrospective and utilize still images, which may be difficult to translate into clinical practice when the real-time prediction of the LNM risk during colonoscopy is required. When video clips are used for validation, the speed of the DL model may be inadequate for routine clinical use in high-definition systems. Although recent studies have almost uniformly assessed DL models (as opposed to conventional ML and other statistical methods) for the prediction of LNM in T1 CRC, there remains a lack of standardization in reporting methodologies and results, which may make meaningful comparisons of different CAD systems and meta-analyses of the available data difficult. Studies on AI that address key questions regarding its use in T1 CRC, as well as a minimum reporting standard such as that required for randomized controlled trials, are needed to overcome this discrepancy.

Similarly, in the fields of DL and WSI, the quality of the WSI used as input for training DL models is crucial for its accuracy in predicting LNM in T1 CRC. Owing to the high dimensionality of the data, the original image may need to be downsized, where pixel information may be lost or broken down into multiple smaller patches for information extraction, which comes at the expense of spatial information. As highlighted in the section on histology, the variations in interobserver variability for tumor budding and depth of invasion among pathologists, coupled with the controversies surrounding the role of depth of invasion in predicting the actual risk of LNM, translate into uncertainty in the “ground truth” and weight assignment in the training of DL models for use in predicting LNM from T1 CRC samples.

In practice, determining the risk of LNM in T1 CRC depends on the demographic and clinical profile of the patient, predicted depth of invasion prior to resection, detailed pathological assessment after resection, and preoperative lymph node staging on CT or MRI—not on any of these factors in isolation. The available literature on DL in T1 CRC focuses mainly on one of the aforementioned factors, with statistical regression or conventional ML models used to combine additional patient information in some studies. For a DL model to be accurate and clinically relevant, at least two of these factors must be incorporated. This involves the insertion of additional branches into ANN algorithms and the use of natural language processing to extract information from endoscopy, histology, and radiology reports, which is computationally expensive and technically demanding.

**CONCLUSIONS**

The field of DL in the management of T1 CRC is developing rapidly, with results showing its potential to accurately predict the depth of invasion and risk of LNM during endoscopy and pathological assessment. However, more data from external validation of independent samples from different centers, as well as further enhancements to DL models to integrate clinically significant information, are necessary before DL can be applied for routine clinical use.

**Conflicts of Interest**

The authors have no potential conflicts of interest.

**Funding**

None.

**Author Contributions**

Conceptualization: JWL; Data curation: JWL, LMW, KI, TLA; Formal analysis: JWL, LMW, KI, KWL, JCYN; Methodology: JWL,
REFERENCES


25. Li JW, Ang TL. Colonoscopy and artificial intelligence: bridging the gap or a gap needing to be bridged? Artif Intell Gastrointest Endosc