New scoring system to distinguish deep invasive submucosal and muscularis propria colorectal cancer during colonoscopy: a development and global multicenter external validation study (e-T2 Score)

Yohei Koyama, MD, PhD,[1,2] Masayoshi Yamada, MD, PhD,[1,3] Mai Ego Makiguchi, MD,[1] Masau Sekiguchi, MD, PhD,[1] Hiroyuki Takamaru, MD, PhD,[1] Taku Sakamoto, MD, PhD,[1] Shin Kono, MD, PhD,[2] Masakatsu Fukuzawa, MD, PhD,[2] Shih Yea Sylvia Wu, MD,[4] Arjun Sugumaran, MD,[4] Takashi Kawai, MD, PhD,[5] Takahisa Matsuda, MD, PhD,[1] Takao Itoi, MD, PhD, FASGE,[2] Yutaka Saito, MD, PhD, FASGE[1]

Tokyo, Japan; Wellington, New Zealand

Background and Aims: Diagnostics to differentiate deep submucosal invasive (invasion depth ≥1000 μm [T1b]) colorectal cancer (CRC) from muscularis propria invasive (T2) CRC are limited. We aimed to establish and validate a scoring system that differentiates T1b from T2.

Methods: A multicenter retrospective cross-validation study was performed. Four hundred sixty-one consecutive pathologically confirmed T1b or T2 CRCs were divided into the development (T1b, 222; T2, 189) and internal validation (T1b, 31; T2, 19) cohorts. Eight potential endoscopic findings were evaluated using the development cohort: loss of lobulation, deep depression, demarcated depressed area, protuberance within the depression, expanding appearance, fold convergency, erosion or white plaque, and Borrmann type 2 or 3 tumor. A scoring system that differentiates T1b from T2 was developed, and diagnostic performance was tested using the internal validation cohort by 8 endoscopists. External validation was conducted using 50 CRC images by 4 endoscopists from other institutions, including outside of Japan.

Results: Multivariate analysis identified the following 5 independent predictive endoscopic findings of T2 CRC: deep depression (odds ratio [OR], 2.08; 95% confidence interval [CI], 1.07-4.04), demarcated depressed area (OR, 4.40; 95% CI, 1.39-13.9), 4-fold convergency or more (OR, 3.41; 95% CI, 1.90-6.11), erosion or white plaque (OR, 8.28; 95% CI, 2.77-24.7), and Borrmann type 2 or 3 tumor (OR, 8.76; 95% CI, 3.58-21.5). The area under the receiver-operating characteristic curve (AUROC) was .90 (95% CI, .87-.93) in the development cohort, .80 (95% CI, .76-.85) in the internal validation, and .76 (95% CI, .69-.83) in the external validation.

Conclusions: We established and validated a new scoring system to differentiate T1b from T2 CRC using 5 simple endoscopic findings. (Gastrointest Endosc 2022;96:321-9.)

(footnotes appear on last page of article)

Improvements in endoscopic technology and techniques, such as endoscopic submucosal dissection (ESD), have made it possible to remove large colorectal tumors without surgery.[1-3] Intramucosal colorectal cancer (CRC) has negligible risk of lymph node metastasis (LNM) and can be cured by ESD.[4,5] On the other hand, submucosal invasive (T1) CRC has an approximately 10% risk of LNM, and intestinal resection with lymph node dissection is recommended.[6,7] However, the risk of LNM in T1 CRC is different for T1a (invasion depth <1000 μm) and T1b (invasion depth ≥1000 μm) lesions. T1a CRC without lymphovascular invasion, grade 2 or 3 tumor budding, and histologic features of poorly differentiated adenocarcinoma/mucinous carcinoma has an extremely low risk of LNM and can be adequately managed by ESD alone.[8-13] Although T1b CRC has a more than 10% overall risk of LNM, studies have reported that the frequency of LNM of T1b CRC without the abovementioned risk factors is approximately 1% to 2%.[14-17] Furthermore, the frequency of LNM is similar to the surgical mortality rate of 1.7%.[18] Therefore, expanding the criteria of ESD for T1b CRC is under discussion.[19,20] In contrast, muscularis propria invasive (T2) CRC has an approximately 20% risk of LNM, making it difficult to manage with ESD alone, and surgery is usually recommended.[21,22]

To the best of our knowledge, reports of endoscopic findings that predict T2 CRC are scarce, with Borrmann type 2 or type 3 being one of the few recognized endoscopic findings.[23,24] However, T2 CRCs often do not display typical Borrmann type 2 or 3 ulceration endoscopically, resulting in difficulty in distinguishing between T1b and T2 with the current diagnostic methods. T-staging of rectal cancer by magnetic resonance imaging (MRI) is an established modality.[25] However, MRI cannot be used for the staging of colon cancer because bowel movements cause interference on imaging. A meta-analysis attempted to distinguish between T1 and T2 CRC using EUS, but the sensitivity for T2 lesions was low at 67%.[26]

To avoid the risk of underestimating T2 CRC as T1b CRC and then subsequently performing futile ESD, establishing an endoscopic diagnostic method to distinguish between T1b and T2 is essential. Therefore, in this study, we aimed to investigate the endoscopic findings that predict T2 CRC and to establish a scoring system that can effectively differentiate T1b CRC from T2 CRC.

METHODS

Study design

The study protocol conforms to the ethical guidelines of the Declaration of Helsinki and was approved by the institutional review boards of the National Cancer Center Hospital (NCCH) and Tokyo Medical University Hospital (registration nos. 2020-002 and T2020-0137). A multicenter retrospective cross-validation study at 2 referral institutions in Japan was performed using data of pathologically confirmed T1b or T2 CRC diagnosed between January 2015 and December 2020. In accordance with the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement on type 2b and type 4, this study consisted of 3 stages (Fig. 1).[27] The first stage was to develop a scoring system for predicting T2 CRC based on endoscopic findings (development study). The second stage was to validate the scoring system at the institution in which the score was developed (internal validation study). The final stage was validation of the scoring system by external institutions (external validation study).

Data collection

Endoscopic images of 527 consecutive CRCs that were endoscopically or surgically resected and had been pathologically diagnosed as T1b or T2 (T1b, 297; T2, 230) between January 2015 and December 2018 at the NCCH were collected. Seven pedunculated-type CRCs and 59 CRCs without detailed endoscopic information were excluded. Finally, 461 CRCs were enrolled (T1b, 253; T2, 208). The 461 CRCs were divided into 2 groups in a non-randomized fashion split by time period at a 1:9 ratio. Nine-tenths of the CRC cases became the development cohort of 411 CRCs collected between June 2015 to December 2018 (pathologically proved T1b [pT1b], 222; pathologically proved T2 [pT2], 189) and one-tenth of the cases became the internal validation cohort of 50 CRCs collected between January to May 2015 (pT1b, 31; pT2, 19) (Fig. 1).

For the external validation cohort, 271 consecutive CRCs that had been pathologically diagnosed as T1b or T2 (T1b, 148; T2, 123) at the Tokyo Medical University Hospital between January 2015 and December 2020 were collected. After excluding 5 pedunculated-type CRCs, 54 CRCs with no images, and 162 CRCs without indigo

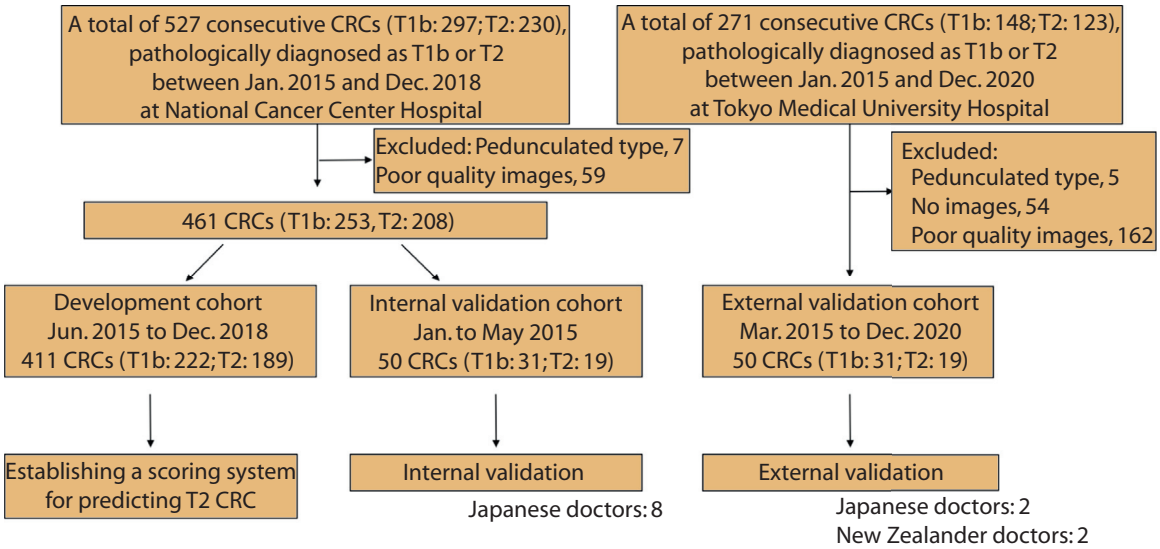


Figure 1. Study algorithm. *CRC*, Colorectal cancer; *T1b*, invasion depth ≥1000 μm; *T2*, muscularis propria invasive.

carmine dye images, 50 CRCs (T1b, 31; T2, 19) were included as the external validation cohort (Fig. 1).

Establishing a scoring system (development study)

Potential endoscopic findings. An exploratory meeting with 3 expert endoscopists (Y.K., M.Y., and Y.S.) was held, and 8 potential endoscopic findings of clinical T1b or T2 based on previous studies were determined as follows[28-30]: (1) loss of lobulation, without lobulation or fused nodules; (2) deep depression, described as >3-mm depression vertically with or without a demarcated area (the 3 mm was determined by evaluators' observations); (3) demarcated depressed area, defined as a definite depression with a circumferential margin; (4) protuberance within the depression, defined as an exposed neoplastic nodule within a depression resembling a submucosal tumor; (5) expanding appearance, described as the surface of the tumor appearing under tension, with a lustering redness because of expansive growth of the tumor; (6) fold convergency, which is a concentration of folds toward the tumor when observed under sufficient insufflation with full extension of the colorectal folds and until blood vessels surrounding the CRC can be clearly delineated (cutoff number of folds was determined statistically); (7) erosion or white plaque, defined as covering of tumor with a white material not easily removable after lavage; and (8) Borrmann type 2 or type 3 tumor, defined as ulcerated carcinomas with sharply demarcated and raised margins or ulcerated, infiltrating carcinomas without definite limits. Representative endoscopic findings are shown in Figure 2.

Establishment of a scoring system. The presence or absence of the 8 potential endoscopic findings were evaluated by an experienced endoscopist (Y.K.) using all endoscopic images of the 411 CRCs in the development cohort. With regard to fold convergency, the number of folds was recorded. A scoring system for predicting T2 CRC was established using the statistically significant endoscopic findings.

Interobserver and intraobserver reliability. To verify the reliability of the potential endoscopic findings, the initial endoscopist (Y.K.) and another experienced endoscopist (M.E.M.) independently evaluated the presence or absence of the statistically significant endoscopic findings for predicting T2 CRC, 6 months after the establishment of the scoring system, using endoscopic images of 100 randomly selected cases from the development cohort.

Internal validation study

To confirm the clinical practicality of the developed scoring system, 50 CRC images (T1b, 31; T2, 19) from the internal validation cohort were evaluated by 8 endoscopists at NCCH (experienced, 4; nonexperienced, 4). These images were independent of those from the

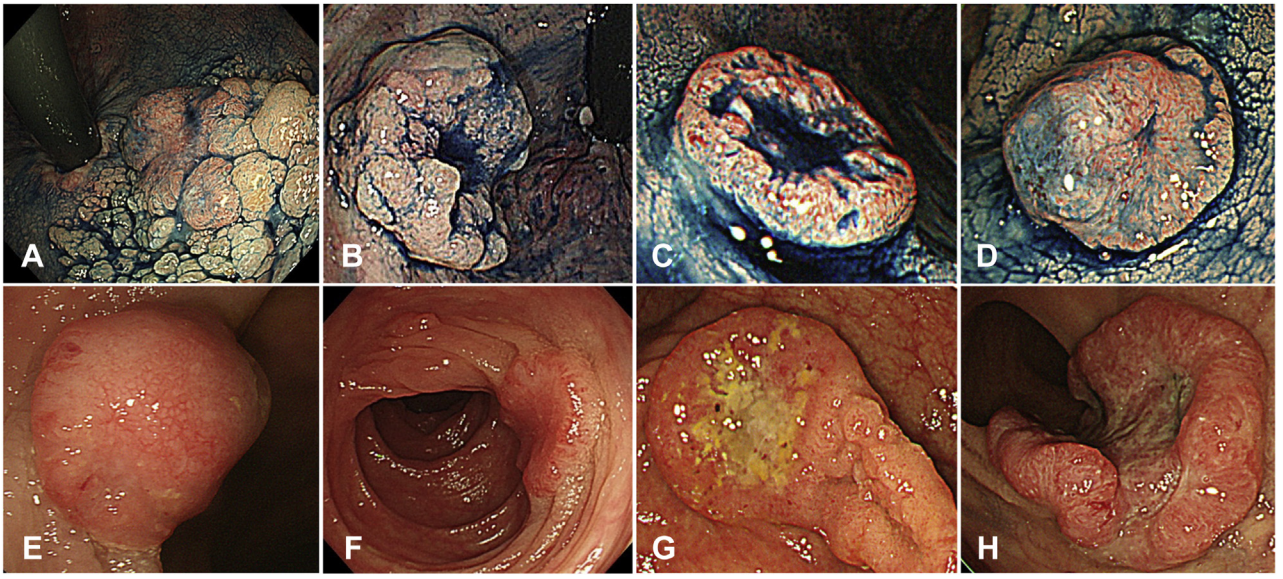


Figure 2. Eight representative potential endoscopic findings of clinical submucosal invasion depth ≥1000 μm or muscularis propria invasive colorectal cancer. **A,** Loss of lobulation: without lobulation, or fused nodules. **B,** Deep depression: >3-mm depression vertically with or without a demarcated area (the 3 mm was determined by evaluators' observation). **C,** Demarcated depressed area: definite depression with a circumferential margin. **D,** Protuberance within the depression: exposed neoplastic nodule within a depression resembling a submucosal tumor. **E,** Expanding appearance: surface of the tumor appearing under tension, with a lustering redness because of expansive growth of the tumor. **F,** Fold convergency: concentration of folds toward the tumor when observed under sufficient insufflation with full extension of the colorectal folds and until blood vessels surrounding the colorectal cancer can be clearly delineated. **G,** Erosion or white plaque: covering of tumor with a white material not easily removable after lavage. **H,** Borrmann type 2 or type 3 tumor: ulcerated carcinomas with sharply demarcated and raised margins or ulcerated, infiltrating carcinomas without definite limits.

development cohort. The evaluation test was conducted using a large projector monitor, on January 14, 2020, with a prior explanation for more than 30 minutes of representative endoscopic findings. Images were incorporated into a slideshow (Microsoft Power Point 2019; Microsoft, Redmond, Wash, USA). One white-light image and 1 indigo carmine dye image were simultaneously displayed for every CRC. The evaluators were blinded to both the histopathologic diagnosis and clinical information of each CRC.

External validation study

To validate the practicality of the scoring system, 50 CRCs from the external validation cohort were evaluated by 4 endoscopists outside of the NCCH (Japanese, 2; New Zealander, 2) between June and July 2021. The evaluation test was displayed on a personal computer monitor using a slideshow. Before the evaluation, the representative endoscopic findings were explained directly to the Japanese doctors and by web meeting to the New Zealander doctors for more than 30 minutes. One white-light image and 1 indigo carmine dye image were simultaneously displayed per CRC, and the evaluators independently assessed the 50 CRCs for the presence and absence of the endoscopic findings from the developed scoring system. The evaluators were blinded to both the histopathologic diagnosis and clinical information of each CRC.

Definition of the endoscopists and study evaluators

Experienced endoscopists were defined as those who were certified as gastroenterologists by the Japan Endoscopic Society and have performed more than 3000 colonoscopies at referral centers where colorectal ESD is performed. All other endoscopists were defined as nonexperienced endoscopists.

Statistical analysis

Because the present study was an exploratory study, as many cases as possible were collected during the study period. The criterion standard for the diagnosis of invasion depth is histopathologic diagnosis of resected specimens. The cutoff number of folds in fold convergency was determined based on the most accurate cutoff point in the receiver operating characteristic (ROC) curve.

In the development study, univariate and multivariate logistic regression analyses were performed for clarifying the independent endoscopic findings of T2 CRC. A $P < .05$ was considered to indicate a statistically significant difference between groups. For establishing a scoring system for predicting T2 CRC, weighted points proportional to the nearest integer of β-regression coefficient values for the endoscopic findings that were significant independent variables in the multivariate analysis were assigned. To assess the diagnostic value of the scoring system, the points were adapted to the development cohort, and the ROC-area under the curve (AUC) was then calculated. Similarly, for internal and external validation, the points were adapted based on image evaluation, and ROC-AUC was calculated.

Intraobserver and interobserver agreements were calculated based on kappa statistics.[31] The value of kappa was defined as follows: slight, 0 to .20; fair, .21 to .40; moderate, .41 to .60; substantial, .61 to .80; and almost perfect, .81 to 1.00. All statistical analyses were performed using SPSS version 27 software (IBM Corp, Armonk, NY, USA).

RESULTS

Clinicopathologic characteristics of CRC patients

The clinicopathologic characteristics of the CRC patients are shown in Table 1. There were no significant differences in age, sex, tumor location, and histopathologic type among the cohorts. In the development cohort, the mean tumor diameter was significantly larger at 29 mm for T2 compared with 23 mm for T1b tumors ($P < .01$). With regard to treatment, surgical resection was more likely to be performed for T2 CRCs than for T1b CRCs in the development and external cohorts ($P < .01$ and $P < .01$, respectively).

Development study (establishing the scoring system for predicting T2 CRC)

The ROC curve of the number of folds in the fold convergency is shown in Supplementary Figure 1 (available online at www.giejournal.org). The most accurate cutoff point was 4 folds, with a sensitivity and specificity of 70% and 74%, respectively. Based on these results, fold convergency was considered present when 4 folds or more were observed.

Univariate and multivariate analyses of the 8 endoscopic findings are shown in Table 2. In the univariate analysis, loss of lobulation, deep depression, demarcated depressed area, fold convergency, erosion or white plaque, and Borrmann type 2 or type 3 tumor were significantly associated with T2 CRC. Multivariate analysis demonstrated that independent predictive endoscopic findings of T2 CRC were deep depression, demarcated depressed area, fold convergency, erosion or white plaque, and Borrmann type 2 or type 3 tumor.

In developing the scoring system for predicting T2 CRC, points were assigned proportional to the β-regression coefficient values for each of the 5 independent endoscopic findings, as follows: 1 point for deep depression, 2 points each for demarcated depressed area and fold convergency, and 3 points each for erosion or white plaque and Borrmann type 2 or type 3 tumor. A total score ranging from 0 to 11 points was calculated for each CRC in the development cohort by adding together the points corresponding to the endoscopic findings (Table 2). In the development

TABLE 1. Characteristics of patients with T1b and T2 colorectal cancers

	Development cohort (n = 411)			Internal validation cohort (n = 50)			External validation cohort (n = 50)		
	pT1b (n = 222)	pT2 (n = 189)	*P* value	pT1b (n = 31)	pT2 (n = 19)	*P* value	pT1b (n = 31)	pT2 (n = 19)	*P* value
Mean age, y (SD)	66 (13)	67 (12)	.25	64 (12)	61 (14)	.52	69 (13)	69 (12)	.93
Sex			.28			.94			.27
Male	121 (55)	113 (60)		16 (52)	10 (53)		18 (58)	8 (42)	
Female	101 (45)	76 (40)		15 (48)	9 (47)		13 (42)	11 (58)	
Tumor location			.43			.76			.87
Proximal colon (cecum to transverse colon)	74 (33)	53 (28)		9 (29)	4 (21)		10 (32)	5 (26)	
Distal colon (descending sigmoid colon)	62 (28)	54 (29)		6 (19)	5 (26)		10 (32)	6 (32)	
Rectosigmoid colon and rectum	86 (39)	82 (43)		16 (52)	10 (53)		11 (36)	8 (42)	
Mean tumor diameter, mm (SD)	23 (16)	29 (11)	<.01	21 (13)	28 (11)	.08	22 (10)	27 (7)	.07
Treatment method			<.01			.07			<.01
Endoscopic resection	75 (34)	2 (1)		8 (26)	1 (5)		14 (45)	0 (0)	
Surgery	147 (66)	187 (99)		23 (74)	18 (95)		17 (55)	19 (100)	
Histopathologic type			.46			1.00			1.00
Differentiated	222 (100)	184 (99)		31 (100)	19 (100)		31 (100)	19 (100)	
Undifferentiated	0 (0)	1 (1)		0	0		0	0	

Values are n (%) unless otherwise defined. *T1b*, Invasion depth ≥1000 μm; *T2*, muscularis propria invasive; *pT1b*, pathologically proved T1b; *pT2*, pathologically proved T2; *SD*, standard deviation.

TABLE 2. Association of individual endoscopic findings and T1b and T2 colorectal cancers in the development cohort

			Univariate analysis		Multivariate analysis			
Endoscopic findings	pT1b (n = 222)	pT2 (n = 189)	Odds ratio (95% confidence interval)	*P* value	Odds ratio (95% confidence interval)	*P* value	β-regression coefficient	Score
Loss of lobulation, presence	108 (49)	159 (84)	6.07 (3.76-9.81)	<.001	1.28 (.66-2.52)	.46	.25	—
Deep depression, presence	59 (27)	150 (79)	10.6 (6.70-16.9)	<.001	2.08 (1.07-4.04)	.031	.73	1
Demarcated depressed area, presence	155 (70)	185 (98)	20.0 (7.13-56.1)	<.001	4.40 (1.39-13.9)	.012	1.48	2
Protuberance within the depression, presence	47 (21)	52 (28)	1.41 (.90-2.22)	.13	1.04 (.52-2.08)	.92	.04	—
Expanding appearance, presence	110 (50)	103 (55)	1.22 (.83-1.80)	.32	1.28 (.69-2.36)	.43	.25	—
Fold convergency, presence	54 (24)	126 (67)	6.78 (4.34-10.6)	<.001	3.41 (1.90-6.11)	<.001	1.23	2
Erosion or white plaque, presence	142 (64)	184 (97)	20.7 (8.18-52.5)	<.001	8.28 (2.77-24.7)	<.001	2.11	3
Borrmann type 2 or type 3 tumor, presence	10 (5)	111 (59)	30.2 (15.0-60.6)	<.001	8.76 (3.58-21.5)	<.001	2.17	3

Values are n (%) unless otherwise defined. *T1b*, Invasion depth ≥1000 μm; *T2*, muscularis propria invasive; *pT1b*, pathologically proved T1b; *pT2*, pathologically proved T2; —, not assessed.

cohort, the AUC of the scoring system for predicting T2 CRC was .90 (95% confidence interval [CI], .87-.93) (Fig. 3A). Using the most accurate cutoff value of 7 points, we found that the sensitivity and specificity were 82% and 83%, respectively (Table 3).

The kappa value for intraobserver and interobserver agreement of the 5 independent endoscopic findings in the development study were as follows: .75 and .68 respectively for deep depression, .70 and .65 respectively for demarcated depressed area, .30 and .51 respectively for fold convergency, .39 and .54 respectively for erosion or white plaque, and .68 and .60 respectively for ulcerating tumor (Table 4).

Internal validation of the scoring system for predicting T2 CRC

In the internal validation study using a total of 400 (50 × 8) CRCs, 19 CRCs were considered to be indeterminable, and

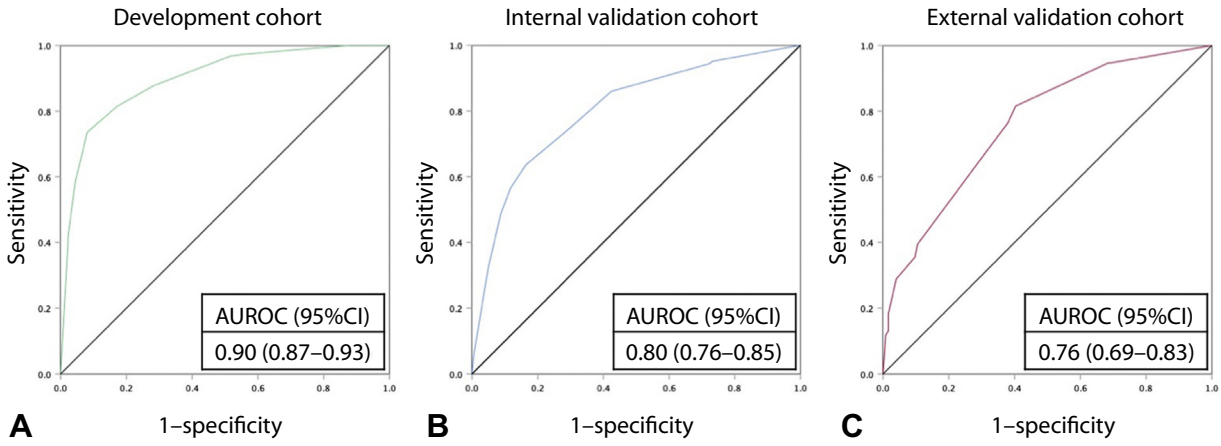


Figure 3. Receiver operating-characteristic curves for the endoscopic diagnosis of muscularis propria invasive colorectal cancer. **A,** Development cohort. **B,** Internal validation cohort. **C,** External validation cohort. *AUROC*, Area under receiver operating characteristic curve; *CI*, confidence interval.

ultimately 381 cases were analyzed. The proportion of individual endoscopic findings for T1b or T2 CRCs in the internal validation cohort are shown in Supplementary Table 1 (available online at www.giejournal.org). In the internal validation cohort, the AUC of the scoring system for predicting T2 CRC evaluated by 8 endoscopists was .80 (95% CI, .76-.85) (Fig. 3B). Given the most accurate cutoff value of 7 points in the development study, the sensitivity and specificity were 57% and 88%, respectively, in the internal validation cohort (Table 3).

External validation of the scoring system for predicting T2 CRC

The prevalence of individual endoscopic findings for T1b and T2 CRC in the external validation cohort are shown in Supplementary Table 2 (available online at www.giejournal.org). In the external validation cohort, the AUC of the scoring system for predicting T2 CRC evaluated by 4 endoscopists, including doctors from other countries, was .76 (95% CI, .69-.83) (Fig. 3C). Using the most accurate cutoff value of 7 points in the development study, we found that the sensitivity and specificity were 36% and 90%, respectively, in the external validation cohort (Table 3).

DISCUSSION

In this cross-validation study, we developed and validated a scoring system composed of 5 endoscopic findings to distinguish between T1b and T2 CRC. To the best of our knowledge, this is the first scoring system using endoscopic findings to distinguish between T1b and T2.

There have been reports using CT, MRI, and EUS as modalities to evaluate T1 and T2 lesions. In a study of preoperative T-stage diagnostic ability using CT colonography for 256 CRCs, the diagnostic accuracy of Tis/T1 versus T2 versus T3/T4 based on intestinal wall deformity was 77.6%.[32] A meta-analysis including 234 patients with rectal cancer reported that the pooled sensitivity and specificity of MRI for T-staging was 79% (95% CI, 72-85) and 85% (95% CI, 79-90), and the AUC of MRI was superior to that of EUS for T2 staging (.92 vs .82, $P < .01$).[33] In a meta-analysis including 208 cancers of the colon proximal to the rectum (T1, 39; T2, 35), the pooled sensitivity and specificity of T1 were good at 90% (95% CI, 66-98) and 98% (95% CI, 94-99), respectively. However, the 95% CI for sensitivity was broad, indicating that the accuracy was not satisfactory. Furthermore, the pooled sensitivity and specificity of T2 were 67% (95% CI, 50-80) and 96% (95% CI, 92-98), respectively, indicating the sensitivity was insufficient.[26]

These studies reported the diagnostic ability of not only T1 and T2 but also T3/T4 lesions. In fact, no reports to date have specifically analyzed the efficacy of diagnostic modalities to differentiate between T1b and T2, which is crucial for determining the suitability of ESD for the treatment of T1b CRCs, when considering expanding the indications of ESD. Although the diagnostic accuracy of our scoring system cannot be directly compared with that of CT, MRI, and EUS because of different study designs, the sensitivity and specificity of our scoring system with the cutoff value set at 7 points were 82% and 83%, respectively, and its high performance was also confirmed using internal and external validation. Considering the close invasion depth between T1b and T2 CRCs, our newly established scoring system can relatively effectively differentiate between these tumors. In addition, the scoring system uses endoscopic findings only, which is considered an advantage over the other modalities.

In this study, 5 independent endoscopic findings that effectively distinguish between T1b and T2 CRCs were clarified: deep depression, demarcated depressed area, fold convergency, erosion or white plaque, and Borrmann

TABLE 3. Risk score and diagnostic performance of the scoring system for predicting T2 colorectal cancer

	Development cohort (n = 411)				Internal validation cohort (n = 381)				External validation cohort (n = 200)			
Total score	T1b (n = 222)	T2 (n = 189)	Sensitivity % (95% CI)	Specificity % (95% CI)	T1b (n = 238)	T2 (n = 143)	Sensitivity % (95% CI)	Specificity % (95% CI)	T1b (n = 124)	T2 (n = 76)	Sensitivity % (95% CI)	Specificity % (95% CI)
0	27	0	100 (99-100)	0 (0-1)	64	7	100 (99-100)	0 (0-1)	38	4	100 (98-100)	0 (0-2)
1	0	0	—	—	2	1	95 (91-98)	27 (25-29)	1	0	95 (88-98)	31 (29-35)
2	30	2	100 (98-100)	12 (12-14)	71	12	94 (90-97)	28 (26-30)	35	10	95 (88-98)	32 (30-35)
3	41	3	99 (97-100)	26 (25-28)	28	15	86 (81-90)	58 (55-61)	3	4	82 (74-88)	60 (56-65)
4	9	1	97 (94-99)	44 (43-47)	4	2	76 (70-81)	69 (66-73)	11	9	76 (68-84)	62 (58-67)
5	52	17	97 (94-99)	48 (47-51)	30	15	74 (68-80)	71 (68-75)	23	19	65 (56-73)	71 (66-76)
6	25	12	88 (84-91)	72 (69-75)	11	10	64 (58-69)	84 (80-87)	1	3	40 (34-47)	90 (85-93)
7	20	15	82 (78-85)	83 (80-86)	7	11	57 (52-62)	88 (85-91)	7	5	36 (30-43)	90 (86-94)
8	8	28	74 (71-77)	92 (89-95)	9	23	49 (45-54)	91 (88-94)	3	8	29 (25-35)	96 (92-98)
9	5	32	59 (56-62)	95 (93-98)	10	36	33 (30-38)	95 (92-97)	0	4	18 (17-24)	98 (95-100)
10	0	3	42 (40-45)	98 (95-99)	1	4	8 (7-10)	99 (98-100)	1	1	13 (11-18)	98 (95-100)
11	5	76	40 (39-43)	98 (95-99)	1	7	5 (4-7)	100 (98-100)	1	9	12 (11-16)	99 (96-100)

T1b, Invasion depth ≥1000 μm; *T2*, muscularis propria invasive; *CI*, confidence interval; —, not assessed.

TABLE 4. Interobserver and intraobserver variability of the 5 independent endoscopic findings in the development cohort

Endoscopic findings	Intraobserver	Interobserver
Deep depression	.75	.68
Demarcated depressed area	.70	.65
Fold convergency	.30	.51
Erosion or white plaque	.39	.54
Borrmann type 2 or type 3 tumor	.68	.60

The value of kappa was defined as follows: slight, 0-.20; fair, .21-.40; moderate, .41-.60; substantial, .61-.80; and almost perfect, .81-1.00.

type 2 or type 3 tumor. In the development cohort, most pT2 CRCs showed deep depression. More than 70% of the T1b and T2 CRCs had a demarcated depressed area. Lesions with fold convergency of more than 4 folds had about 3 times the risk of being T2 CRC compared with T1b CRC. Almost all T2 CRCs had erosion or white plaques. These findings have long been widely known as characteristics of submucosal invasive cancers and were more often observed together with muscularis propria invasion.[34,35] However, approximately 40% of T2 CRCs did not show Borrmann type 2 or type 3 features and only 5% of the T1b CRCs were ulcerated (Table 2). Nevertheless, Borrmann type 2 or 3 tumors were highly specific for T2 CRCs and were able to diagnose pT2 CRCs with high reliability. Based on the above results, the 5 independent characteristic endoscopic findings were considered feasible for predicting T2 CRC. In the future, these characteristic endoscopic findings may be used to construct a rules-based artificial intelligence approach to differentiate T1b CRC from T2 CRC.

Because it is well known that the diagnosis of CRCs is affected by tumor size, tumor location, the doctor's experience, and so on, we performed a subgroup analysis and checked the diagnostic performance in various clinical subgroups.[24,36,37] In subgroup analysis, the present scoring system had high diagnostic performance in both the colon and rectum and when used by both experienced and nonexperienced endoscopists. However, the ROC-AUC was relatively low in tumors <25 mm compared with those >25 mm (Supplementary Fig. 2, available online at www.giejournal.org). One possible explanation is only a few patients had small T2 CRCs in our cohort. In any case, determination of the depth of small cancers is generally very difficult, particularly to distinguish between T1b and T2. Even when using EUS, it is difficult to clarify the microinvasion pattern. Therefore, the ROC-AUC at .69 (95% CI, .62-.77) in tumors <25 mm was considered sufficient for small T2 CRCs.

To expand the indications of ESD for T1b CRCs, evaluation of the risk of LNM of T1b CRCs in the clinical setting is also necessary. CT is routinely performed to rule out LNM, but its sensitivity is low when the lymph node is smaller than 10 mm. In recent years, there have been reports of the efficacy of artificial intelligence in measuring the risk of LNM after

resection of T1 CRC.[38,39] In the near future, endoscopic resection may become the primary management method for T1b CRCs coupled with postprocedural surveillance, and this approach may be concluded to be appropriate to avoid unnecessary surgical resection when the risk of LNM is low. In such a case, the strength of the scoring system is that the cutoff value can be altered according to the clinical purpose. For example, if endoscopic resection becomes the accepted primary treatment option for T1b CRC in the future, the cutoff value can be set to 2 or 3 points with a sensitivity of 86% to 94% to avoid ESD for T2 CRCs.

In addition, our scoring system was evaluated by endoscopists not only in Japan but also by endoscopists from other countries during the external validation. In the external validation cohort, the AUCs of Japanese endoscopists and endoscopists from other countries were .80 (95% CI, .72-.89) and .76 (95% CI, .64-.85), respectively (Supplementary Fig. 2). The diagnostic accuracy was slightly lower for endoscopists from other countries but was still high enough for the scoring system to be used around the world.

T1b is not an accepted T-stage for CRCs according to the American Joint Committee on Cancer. Intramucosal CRC (Tis in the American Joint Committee on Cancer staging) has negligible risk of LNM. On the other hand, T1 CRC (submucosal cancers) has an approximately 10% risk of LNM. The submucosa of the colon and rectum is believed to have lymphatic vessels and thus have risk of lymphatic spread of malignancy. Within T1 CRCs, because T1 CRCs with an invasion depth <1000 μm have an extremely low risk of LNM, T1 CRCs are divided into T1a and T1b in the Japanese Society for Cancer of the Colon and Rectum guidelines.[40,41] When discussing T1b and T2 CRCs, possibly patients may have T1N1-2, T2N1-2, and T3 CRCs that require surgical resection and possibly neoadjuvant therapy depending on the tumor location.[42,43] Our scoring system is a method to differentiate between T1b CRCs and T2 CRCs and does not predict T3 CRCs and LNM; hence, CT or MRI should be performed to determine the cancer staging before endoscopic resection. Furthermore, it is necessary to expand the scoring system to include T3 CRCs in the future.

There are some limitations to this study. First, this was a retrospective study. To reduce the selection bias, we enrolled as many consecutive CRCs as possible and evaluated all endoscopic images of the development cohort. Second, 2 images per CRC (1 white-light image and 1 indigo carmine dye image) were selected for evaluation for both the internal and external validation studies, and hence there may be selection bias of the images. However, the accuracy of this scoring system is comparable with that of MRI and EUS, and it is plausible that its accuracy will increase with its ongoing application in clinical practice. Third, fold convergency had a low concordance for both intraobserver and interobserver agreement. However, based on ROC analysis (Supplementary Fig. 1), as the number of folds increases, the specificity of T2 CRC increases and the sensitivity decreases, and hence there is a positive correlation between the number of folds and cancer depth. A further real-world prospective study is warranted to validate the efficacy of this scoring system for real-time diagnosis.

In conclusion, we established and validated a new scoring system to distinguish T1b and T2 CRCs using 5 simple endoscopic findings during colonoscopy. The new scoring system can be applied when determining the appropriate management of T1b and T2 CRCs.

ACKNOWLEDGMENTS

We thank the following participants for performing the internal validation: Hourin Cho, Yasuhiko Mizuguchi, Kengo Kasuga, and Kazumasa Kawashima. We are indebted to Helena Popiel, Instructor of the Department of International Medical Communications of Tokyo Medical University, for her editorial review of the manuscript. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Saito Y, Uraoka T, Yamaguchi Y, et al. A prospective, multicenter study of 1111 colorectal endoscopic submucosal dissections (with video). Gastrointest Endosc 2010;72:1217-25.
2. Yamada M, Saito Y, Takamaru H, et al. Long-term clinical outcomes of endoscopic submucosal dissection for colorectal neoplasms in 423 cases: a retrospective study. Endoscopy 2017;49:233-42.
3. Sakamoto T, Saito Y, Nakamura F, et al. Short-term outcomes following endoscopic submucosal dissection of large protruding colorectal neoplasms. Endoscopy 2018;50:606-12.
4. Yoshida N, Naito Y, Inada Y, et al. Cross-national analysis about the difference of histopathological management in Tis and T1 colorectal cancer between Japan and Korea. J Anus Rectum Colon 2019;3:18-26.
5. Bordet M, Bretagne JF, Piette C, et al. Reappraisal of the characteristics, management, and prognosis of intramucosal colorectal cancers and their comparison with T1 carcinomas. Gastrointest Endosc 2021;93:477-85.
6. Bosch SL, Teerenstra S, de Wilt JH, et al. Predicting lymph node metastasis in pT1 colorectal cancer: a systematic review of risk factors providing rationale for therapy decisions. Endoscopy 2013;45:827-34.
7. Wada H, Shiozawa M, Katayama K, et al. Systematic review and meta-analysis of histopathological predictive factors for lymph node metastasis in T1 colorectal cancer. J Gastroenterol 2015;50:727-34.
8. Kitajima K, Fujimori T, Fujii S, et al. Correlations between lymph node metastasis and depth of submucosal invasion in submucosal invasive colorectal carcinoma: a Japanese collaborative study. J Gastroenterol 2004;39:534-43.
9. Ueno H, Mochizuki H, Hashiguchi Y, et al. Risk factors for an adverse outcome in early invasive colorectal carcinoma. Gastroenterology 2004;127:385-94.
10. Yoda Y, Ikematsu H, Matsuda T, et al. A large-scale multicenter study of long-term outcomes after endoscopic resection for submucosal invasive colorectal cancer. Endoscopy 2013;45:718-24.
11. Ikematsu H, Yoda Y, Matsuda T, et al. Long-term outcomes after resection for submucosal invasive colorectal cancers. Gastroenterology 2013;144:551-9.
12. Kawachi H, Eishi Y, Ueno H, et al. A three-tier classification system based on the depth of submucosal invasion and budding/sprouting can improve the treatment strategy for T1 colorectal cancer: a retrospective multicenter study. Mod Pathol 2015;28:872-9.

13. Watanabe T, Muro K, Ajioka Y, et al. Japanese Society for Cancer of the Colon and Rectum (JSCCR) guidelines 2016 for the treatment of colorectal cancer. Int J Clin Oncol 2018;23:1-34.
14. Tanaka S, Haruma K, Oh EH, et al. Conditions of curability after endoscopic resection for colorectal carcinoma with submucosally massive invasion. Oncol Rep 2000;7:783-8.
15. Nakadoi K, Tanaka S, Kanao H, et al. Management of T1 colorectal carcinoma with special reference to criteria for curative endoscopic resection. J Gastroenterol Hepatol 2012;27:1057-62.
16. Oka S, Tanaka S, Nakadoi K, et al. Risk analysis of submucosal invasive rectal carcinomas for lymph node metastasis to expand indication criteria for endoscopic resection. Dig Endosc 2013;25:21-5.
17. Miyachi H, Kudo SE, Ichimasa K, et al. Management of T1 colorectal cancers after endoscopic treatment based on the risk stratification of lymph node metastasis. J Gastroenterol Hepatol 2016;31:1126-32.
18. Vermeer NCA, Backes Y, Snijders HS, et al. National cohort study on postoperative risks after surgery for submucosal invasive colorectal cancer. BJS Open 2019;3:210-7.
19. Tanaka S, Asayama N, Shigita K, et al. Towards safer and appropriate application of endoscopic submucosal dissection for T1 colorectal carcinoma as total excisional biopsy: future perspectives. Dig Endosc 2015;27:216-22.
20. Watanabe D, Toyonaga T, Ooi M, et al. Clinical outcomes of deep invasive submucosal colorectal cancer after ESD. Surg Endosc 2018;32:2123-30.
21. Huh JW, Kim HR, Kim YJ. Lymphovascular or perineural invasion may predict lymph node metastasis in patients with T1 and T2 colorectal cancer. J Gastrointest Surg 2010;14:1074-80.
22. Fields AC, Lu P, Hu F, et al. Lymph node positivity in T1/T2 rectal cancer: a word of caution in an era of increased incidence and changing biology for rectal cancer. J Gastrointest Surg 2021;25:1029-35.
23. Participants in the Paris Workshop. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. Gastrointest Endosc 2003;58:S3-43.
24. Endoscopic Classification Review Group. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. Endoscopy 2005;37:570-8.
25. Kijima S, Sasaki T, Nagata K, et al. Preoperative evaluation of colorectal cancer using CT colonography, MRI, and PET/CT. World J Gastroenterol 2014;20:16964-75.
26. Malmstrom ML, Saftoiu A, Vilmann P, et al. Endoscopic ultrasound for staging of colonic cancer proximal to the rectum: a systematic review and meta-analysis. Endosc Ultrasound 2016;5:307-14.
27. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-73.
28. Matsuda T, Fujii T, Saito Y, et al. Efficacy of the invasive/non-invasive pattern by magnifying chromoendoscopy to estimate the depth of invasion of early colorectal neoplasms. Am J Gastroenterol 2008;103:2700-6.
29. Ikehara H, Saito Y, Matsuda T, et al. Diagnosis of depth of invasion for early colorectal cancer using magnifying colonoscopy. J Gastroenterol Hepatol 2010;25:905-12.
30. Yasue C, Chino A, Takamatsu M, et al. Pathological risk factors and predictive endoscopic factors for lymph node metastasis of T1 colorectal cancer: a single-center study of 846 lesions. J Gastroenterol 2019;54:708-17.
31. Kundel HL, Polansky M. Measurement of observer agreement. Radiology 2003;228:303-8.
32. Utano K, Endo K, Togashi K, et al. Preoperative T staging of colorectal cancer by CT colonography. Dis Colon Rectum 2008;51:875-81.
33. Chan BPH, Patel R, Mbuagbaw L, et al. EUS versus magnetic resonance imaging in staging rectal adenocarcinoma: a diagnostic test accuracy meta-analysis. Gastrointest Endosc 2019;90:196-203.
34. Saitoh Y, Obara T, Watari J, et al. Invasion depth diagnosis of depressed type early colorectal cancers by combined use of videoendoscopy and chromoendoscopy. Gastrointest Endosc 1998;48:362-70.
35. Horie H, Togashi K, Kawamura YJ, et al. Colonoscopic stigmata of 1 mm or deeper submucosal invasion in colorectal cancer. Dis Colon Rectum 2008;51:1529-34.
36. Mukae M, Kobayashi K, Sada M, et al. Diagnostic performance of EUS for evaluating the invasion depth of early colorectal cancers. Gastrointest Endosc 2015;81:682-90.
37. Inoki K, Sakamoto T, Takamaru H, et al. The diagnostic performance for colorectal neoplasms using magnified endoscopy differs between experts and novice endoscopists: a post hoc analysis. Digestion 2020;101:590-7.
38. Ichimasa K, Kudo SE, Mori Y, et al. Artificial intelligence may help in predicting the need for additional surgery after endoscopic resection of T1 colorectal cancer. Endoscopy 2018;50:230-40.
39. Kudo SE, Ichimasa K, Villard B, et al. Artificial intelligence system to determine risk of T1 colorectal cancer metastasis to lymph node. Gastroenterology 2021;160:1075-84.
40. Weiser MR. AJCC 8th edition: colorectal cancer. Ann Surg Oncol 2018;25:1454-5.
41. Japanese Society for Cancer of the Colon and Rectum. Japanese classification of colorectal, appendiceal, and anal carcinoma: the 3rd English edition [secondary publication]. J Anus Rectum Colon 2019;3:175-95.
42. Rödel C, Graeven U, Fietkau R, et al. Oxaliplatin added to fluorouracil-based preoperative chemoradiotherapy and postoperative chemotherapy of locally advanced rectal cancer (the German CAO/ARO/AIO-04 study): final results of the multicentre, open-label, randomised, phase 3 trial. Lancet Oncol 2015;16:979-89.
43. Li Y, Wang J, Ma X, et al. A review of neoadjuvant chemoradiotherapy for locally advanced rectal cancer. Int J Biol Sci 2016;12:1022-31.

Abbreviations: AUROC, Area under receiver operating characteristic curve; CRC, colorectal cancer; ESD, endoscopic submucosal dissection; LNM, lymph node metastasis; NCCH, National Cancer Center Hospital; MRI, magnetic resonance imaging; ROC, receiver operating characteristic; T1, submucosal invasive; T1a, invasion depth <1000 μm; T1b, invasion depth ≥1000 μm; T2, muscularis propria invasive.

DISCLOSURE: *All authors disclosed no financial relationships.*

DIVERSITY, EQUITY, AND INCLUSION: We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. We worked to ensure that the language of the study questionnaires reflected inclusion. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

0016-5107/$36.00
https://doi.org/10.1016/j.gie.2022.03.002

Received October 16, 2021. Accepted March 1, 2022.

Current affiliations: Endoscopy Division (1), Department of Genetic Medicine and Services (3), National Cancer Center Hospital, Tokyo, Japan; Department of Gastroenterology and Hepatology (2), Department of Endoscopy (5), Tokyo Medical University Hospital, Tokyo, Japan; Endoscopy Unit, Hutt Hospital, Wellington, New Zealand (4).

Reprint requests: Masayoshi Yamada, MD, PhD, Endoscopy Division, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan.

APPENDIX

SUPPLEMENTARY TABLE 1. Proportion of individual endoscopic findings for T1b and T2 colorectal cancer in the internal validation cohort (n = 381)

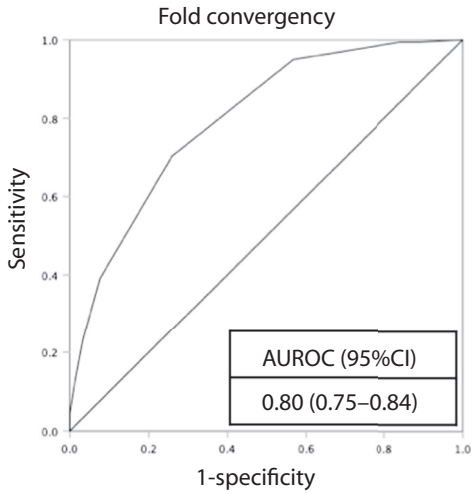
	Total		Tumor location		Lesion size		Experience	
	pT1b (n = 238)	pT2 (n = 143)	Colon (n = 183)	Rectum (n = 198)	<25 mm (n = 228)	≥25 mm (n = 153)	Nonexperienced (n = 200)	Experienced (n = 181)
Deep depression, presence	43 (18)	74 (52)	43 (23)	74 (37)	49 (21)	68 (44)	49 (25)	68 (38)
Demarcated depressed area, presence	147 (62)	121 (85)	134 (73)	134 (68)	165 (72)	103 (67)	132 (66)	136 (75)
Fold convergency, presence	29 (12)	43 (30)	48 (26)	24 (12)	51 (22)	21 (14)	32 (16)	40 (22)
Erosion or white plaque, presence	82 (34)	115 (80)	93 (51)	104 (53)	103 (45)	94 (61)	109 (55)	88 (49)
Borrmann type 2 or type 3 tumor, presence	17 (7)	56 (39)	21 (11)	52 (26)	26 (11)	47 (31)	39 (20)	34 (19)

Values are n (%). *T1b*, Invasion depth ≥1000 μm; *T2*, muscularis propria invasive; *pT1b*, pathologically proved T1b; *pT2*, pathologically proved T2.

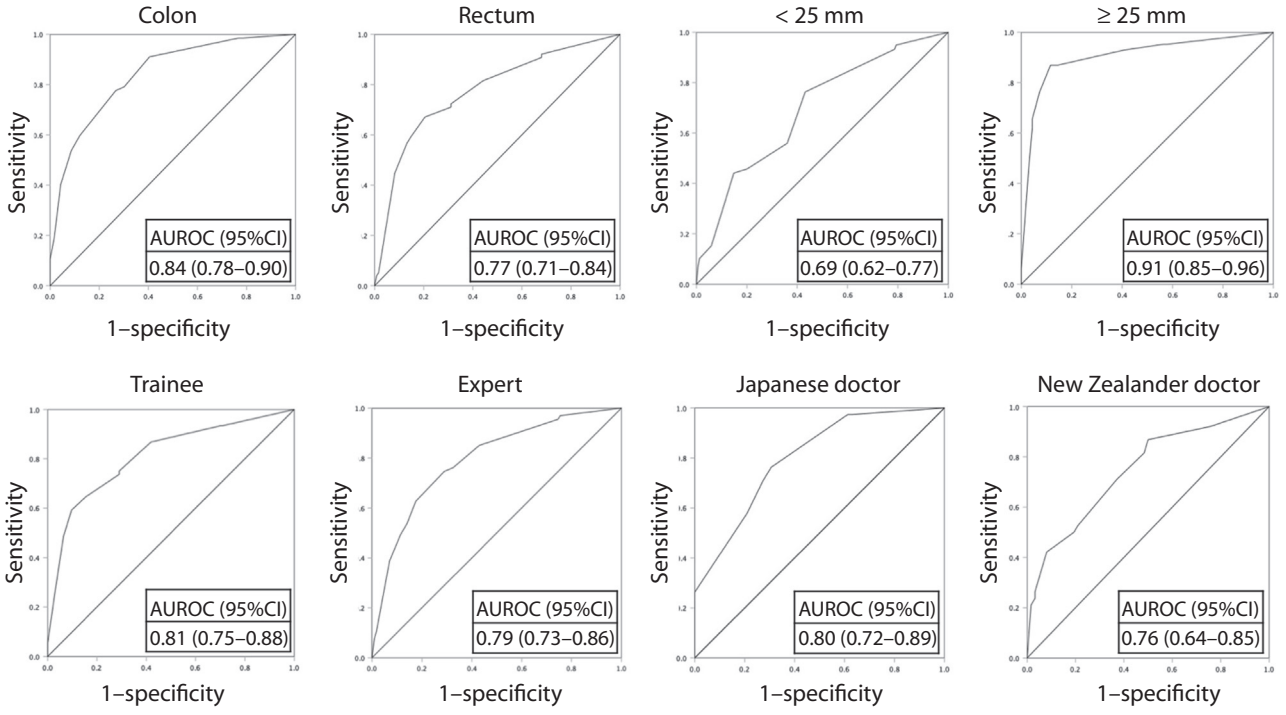
SUPPLEMENTARY TABLE 2. Proportion of individual endoscopic findings for T1b and T2 colorectal cancer in the external validation cohort (n = 200)

	Total		Nationality	
	pT1b (n = 124)	pT2 (n = 76)	Japanese (n = 100)	New Zealander (n = 100)
Deep depression, presence	5 (4)	21 (28)	8 (8)	18 (18)
Demarcated depressed area, presence	71 (57)	60 (79)	70 (70)	61 (61)
Fold convergency, presence	33 (27)	37 (49)	14 (14)	56 (56)
Erosion or white plaque, presence	34 (27)	50 (66)	38 (38)	46 (46)
Borrmann type 2 or type 3 tumor, presence	8 (6)	20 (26)	6 (6)	22 (22)

Values are n (%). *T1b*, Invasion depth ≥1000 μm; *T2*, muscularis propria invasive; *pT1b*, pathologically proved T1b; *pT2*, pathologically proved T2.



Supplementary Figure 1. Receiver operating-characteristic curve of fold convergency for the endoscopic diagnosis of muscularis propria invasive colorectal cancer. The most accurate is 4-fold convergency at 70% of sensitivity and 74% of specificity. *AUROC*, Area under receiver operating characteristic curve; *CI*, confidence interval.



Supplementary Figure 2. Receiver operating-characteristic curves for the endoscopic diagnosis of muscularis propria invasive colorectal cancer in subgroup analysis. Subgroup analysis shows the diagnostic performance in the following clinical settings: colon versus rectum, tumor size <25 mm versus ≥25 mm, trainee versus expert, and Japanese versus New Zealander doctor. All but <25-mm lesions show high AUC at more than about .8 and even the <25-mm lesions have the AUC of .69. *AUROC*, Area under receiver operating characteristic curve; *CI*, confidence interval.