



| | |
|--------------|---|
| Title | The potential of an artificial intelligence for diagnosing MRI images in rectal cancer: multicenter collaborative trial |
| Author(s) | Hamabe, Atsushi; Takemasa, Ichiro; Ishii, Masayuki et al. |
| Citation | Journal of Gastroenterology. 2024, 59(10), p. 896-904 |
| Version Type | VoR |
| URL | https://hdl.handle.net/11094/98107 |
| rights | This article is licensed under a Creative Commons Attribution 4.0 International License. |
| Note | |

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka



The potential of an artificial intelligence for diagnosing MRI images in rectal cancer: multicenter collaborative trial

Atsushi Hamabe^{1,2} · Ichiro Takemasa¹  · Masayuki Ishii¹ · Koichi Okuya¹ · Koya Hida³ · Daisuke Nishizaki³ · Atsuhiko Sumii³ · Shigeki Arizono⁴ · Shigeshi Kohno⁴ · Koji Tokunaga⁵ · Hirotsugu Nakai⁶ · Yoshiharu Sakai⁷ · Masahiko Watanabe⁸

Received: 24 April 2024 / Accepted: 20 June 2024
© The Author(s) 2024

Abstract

Background An artificial intelligence-based algorithm we developed, mrAI, satisfactorily segmented the rectal tumor, rectum, and mesorectum from MRI data of rectal cancer patients in an initial study. Herein, we aimed to validate mrAI using an independent dataset.

Methods We utilized MRI images collected in another nationwide research project, "Open versus Laparoscopic Surgery for Advanced Low Rectal Cancer Patients". MRIs from 467 cases with upfront surgery were utilized; six radiologists centralized the MRI evaluations. The diagnostic accuracies of mrAI and the radiologists for tumor depth were compared using pathologic diagnosis as a reference.

Results For all cases, centralized diagnosis demonstrated 84.2% sensitivity, 37.7% specificity, and 73.7% accuracy; mrAI exhibited 70.6% sensitivity, 61.3% specificity, and 68.5% accuracy. After limiting MRIs to those acquired by a Philips scanner, with an inter-slice spacing of ≤ 6 mm—both conditions similar to those used in the development of mrAI—the performance of mrAI improved to 76.8% sensitivity, 76.7% specificity, and 76.7% accuracy, while the centralized diagnosis showed 81.8% sensitivity, 36.7% specificity, and 71.3% accuracy. Regarding relapse-free survival, the prognosis for tumors staged $\geq T3$ was significantly worse than for tumors staged $\leq T2$ ($P=0.0484$) in the pathologic diagnosis. While no significant difference was observed between $\geq T3$ and $\leq T2$ tumors in the centralized diagnosis ($P=0.1510$), the prognosis for $\geq T3$ was significantly worse in the mrAI diagnosis ($P=0.0318$).

Conclusion Proper imaging conditions for MRI can enhance the accuracy of mrAI, which has the potential to provide feedback to radiologists without overestimating tumor stage.

✉ Ichiro Takemasa
itakemasa@sapmed.ac.jp

Atsushi Hamabe
ahamabe@gesurg.med.osaka-u.ac.jp

- 1 Department of Surgery, Surgical Oncology and Science, Sapporo Medical University, S1 W16, Chuo-Ku, Sapporo 060-8543, Japan
- 2 Department of Gastroenterological Surgery, Graduate School of Medicine, Osaka University, 2-2-E2 Yamadaoka, Suita, Osaka 565-0871, Japan
- 3 Department of Surgery, Kyoto University Graduate School of Medicine, Kyoto, Japan
- 4 Department of Diagnostic Radiology, Kobe City Medical Center General Hospital, Kobe, Japan
- 5 Department of Diagnostic Radiology, Japanese Red Cross Wakayama Medical Center, Wakayama, Japan
- 6 Department of Radiology, Mayo Clinic, Rochester, MN, USA
- 7 Department of Surgery, Osaka Red-Cross Hospital, Osaka, Japan
- 8 Department of Surgery, Kitasato University Kitasato Institute Hospital, Tokyo, Japan

Keyword Rectal cancer · MRI · Artificial intelligence

Abbreviations

| | |
|-------|--|
| AI | Artificial intelligence |
| CRM | Circumferential resection margin |
| CRT | Chemoradiotherapy |
| DNA | Deoxyribonucleic acid |
| EMVI | Extramural vascular invasion |
| JSLCS | Japan Society of Laparoscopic Colorectal Surgery |
| LPND | Lateral pelvic lymph node dissection |
| MRI | Magnetic resonance imaging |
| TNT | Total neoadjuvant therapy |

Introduction

In the treatment of locally advanced rectal cancer, neoadjuvant chemoradiotherapy (CRT) has long been a standard of care in Western countries. More recently, total neoadjuvant therapy (TNT) has emerged as an intensified option proven to reduce the risk of distant metastasis [1–3]. However, there are patients who can achieve a cure without neoadjuvant treatment, for whom neoadjuvant CRT or TNT might be harmful [4, 5]. As multimodal therapies grow increasingly complex, accurate diagnosis of baseline tumor characteristics becomes vital for individualized decision-making. Among the preoperative assessments, MRI stands out as the most crucial tool, capable of revealing the tumor's various malignant features [6, 7]. Yet, even when interpreted by expert radiologists, the accuracy of MRI findings can be further refined. The development of technology to support radiologic interpretation of MRI is crucial in improving the prognosis of locally advanced rectal cancer and optimizing its treatment.

In our prior research, we developed an AI-based algorithm to segment the rectal tumor, rectum, and mesorectum from MRI data (referred to as mrAI). This can evaluate the T stage or identify areas at risk of circumferential resection margin (CRM) involvement [8]. While there have been recent developments in AI systems for diagnosing rectal cancer, our mrAI stands out, in that it was created using ground-truth data that aligns cancerous areas in pathologic specimens with corresponding regions on high-resolution MRI. Moreover, it segments three distinct areas—the tumor, rectum, and mesorectum—offering visual insights useful for selecting the optimal dissection layer during surgery and ensuring the preservation of the CRM. Although the precision of this technology was deemed satisfactory, it was developed using a ground-truth label based solely on MRI data from one institution (Sapporo Medical University) with a single data acquisition protocol. Consequently, additional validation using an independent dataset is essential to highlight the utility of mrAI.

The Japanese nationwide study titled “Open versus laparoscopic surgery for advanced low rectal cancer patients” was conducted across 69 institutes. This research was a project under the Japan Society of Laparoscopic Colorectal Surgery (JSLCS) spanning the period from January 2010 through December 2011 [9]. In a subsequent survey, MRI data were retrospectively collected to analyze correlations between MRI-related factors and clinicopathological outcomes. Here, MRIs underwent centralized review by expert radiologists. This study highlighted the result that lateral pelvic lymph node dissection (LPND) improved recurrence-free survival for cases with a lateral pelvic node short axis ≥ 5 mm [10], and a nomogram was devised to anticipate metastasis to the lateral pelvic node [11]. In the

present research, our aim was to validate mrAI using the MRI dataset collected in the JSLCS study.

Materials and methods

Study design and participants

This retrospective study utilized data from a multicenter cohort study spanning 69 institutions affiliated with the JSLCS. In the initial study, 1608 patients diagnosed with clinical stage II/III low rectal cancer below the peritoneal reflection and who underwent rectal resection between January 2010 and December 2011 were registered [9]. Clinical data were prospectively collected, demonstrating that laparoscopic surgery could be a feasible treatment option for advanced low rectal cancer (UMIN registration number: 000013919). In a subsequent study, 752 MRIs were retrospectively analyzed to identify the subset of patients who could benefit from LPND [10]. Six radiologists reviewed the MRI-related findings, such as tumor depth, lymph node enlargement, mrEMVI [12, 13], and mrCRM [14], revealing that LPND was advantageous for patients with a lateral pelvic node measuring between 5 and 10 mm. The study received approval from the institutional review boards of Kyoto University and all participating centers (UMIN registration number: 000026789). An opt-out method was used to obtain consent for study inclusion, as well as potential secondary data usage, in line with the Japanese Ethical Guidelines for Medical and Health Research Involving Human Subjects. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. In the current study, the validity of mrAI was assessed using these datasets after securing approval from the institutional review boards of Sapporo Medical University and Kyoto University. A total of 503 patients who did not receive preoperative treatment were selected to compare the mrAI-predicted T stage and the radiologic T stage diagnosed by the certified radiologists. The pathologic stage was used as a reference.

Interpretation of MRI

As detailed in a previous study, six radiologists, after a consensus meeting on MRI findings, reviewed the MRIs to assess tumor depth, mrCRM, and lymph-node size [10]. Tumor depth was categorized as T0–2, T3, or T4. It was classified as T3 if the tumor penetrated the muscularis propria and as T4 if it invaded adjacent organs. A detailed description of mrAI is available in a prior report [8]. In summary, the algorithm was developed using a deep neural

network. MRIs were annotated with ground-truth labels of tumors by verifying pathologically confirmed lesions on sections of circular resected specimens. The T stage was semi-automatically determined based on the positional relationships among the tumor, rectum, and mesorectum. Radiologists and mrAI evaluated the T2-weighted axial images of the collected MRIs. MRI data were loaded onto a laptop equipped with mrAI for analysis. The algorithm generated segmentation results for the tumor, rectum, and mesorectum (Fig. 1), from which the T stage and mesorectal fascia (MRF) involvement were automatically calculated. We categorized the MRI diagnosis into three groups, in which “local diagnosis” was the diagnosis made at each hospital,

“centralized diagnosis” was made by the above qualified radiologists, and “mrAI diagnosis” was made by mrAI.

Statistical analyses

All statistical analyses were conducted using JMP Pro 16 (SAS Institute, Cary, NC). Results are displayed as either the number of cases evaluated (for categorical data) or the median and range (for quantitative data). Univariate analyses utilized Fisher’s exact test. The Kaplan–Meier method estimated relapse-free survival, and the log-rank test determined statistical significance. All *P* values were two-tailed, with *P* values < 0.05 deemed statistically significant. A 2 × 2 table

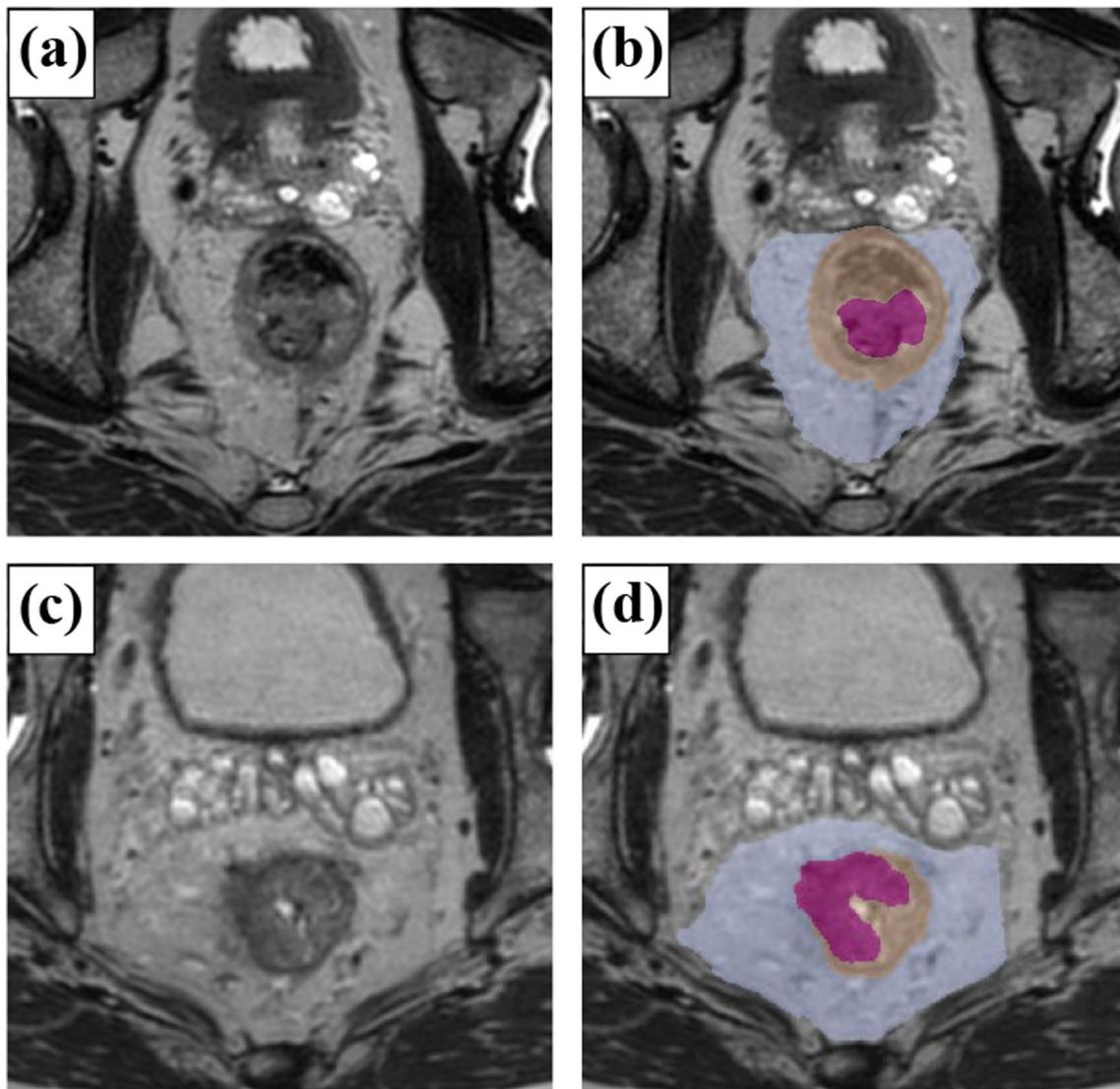


Fig. 1 Segmentation images based on mrAI. **a** MRI image of rectal cancer diagnosed as T2 by mrAI. **b** Identical case as (a) overlaid with segmentation results. Pink area indicates tumor, beige area indicates rectum, and blue area indicates mesorectum. **c** MRI image of rectal

cancer diagnosed as T3 by mrAI. **d** Identical case as (c) overlaid with segmentation results. Pink area indicates tumor, beige area indicates rectum, and blue area indicates mesorectum

was used to calculate the sensitivity, specificity, and accuracy of MRI. For example, in calculating the characteristics of the diagnostic test for $\geq T3$ staging, sensitivity is defined as the proportion of cases staged as $\geq T3$ on MRI among pathologic $\geq T3$ cases, specificity is the proportion of cases staged as $\leq T2$ on MRI among pathologic $\leq T2$ cases, and accuracy is defined as the proportion of cases in which the pathologic tumor depth ($\leq T2$ or $\geq T3$) was correctly diagnosed by MRI in all evaluated cases.

Results

Patients

From the enrolled cases, 11 with mucinous carcinoma were excluded because they were demonstrated to be unsuitable for analysis by mrAI due to characteristic MRI findings. In addition, 25 cases were excluded due to poor MRI quality, such as motion artifacts. Consequently, 467 cases were included in our analysis (Fig. 2).

The median duration of post-surgical follow-up was 2119 days (interquartile range: 1620–2418 days). Table 1 presents the patients' characteristics, MRI details, TNM classification, surgical procedures performed, and pathologic findings. The median age of the patients was 63 years (range: 26–86 years), and 317 of the 467 patients were male.

MRI images were primarily obtained using 1.5-T MRI scanners ($N=349$). The median inter-slice spacing was 6.1 mm (range: 3.0–10 mm), and the median in-plane resolution was 0.52×0.52 mm. The manufacturers of the MRI scanners included Philips (242 cases), GE Healthcare (106 cases), Siemens (90 cases), and others (29 cases). The

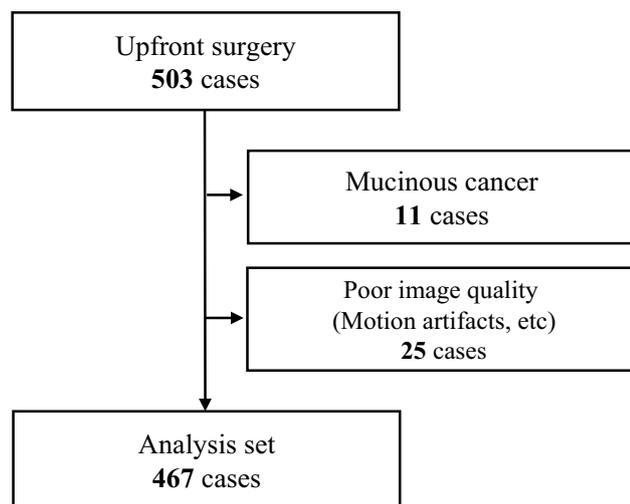


Fig. 2 Flowchart of patient inclusion in the study

relationship between manufacturers and magnetic field strength is shown in Supplementary Table 1.

Table 1 also displays the clinical T and N staging determined at each hospital. There were 431 cases staged as cT3/4. The median distance from the anal verge was 5 cm. Laparoscopic surgeries were performed in 138 cases, and the procedures included 233 anterior resections. LPND was performed in 281 cases.

Pathologic findings revealed that there were 106 cases of pT1/2, which was more frequent than the local diagnosis which was made at each hospital ($P < 0.0001$).

Precision of T staging

Following the method adopted in the previous studies that examined the performance of MRI diagnosis for rectal cancer [15–19], we initially assessed the diagnostic performance of mrAI for T staging, comparing it with that of radiologic experts (Table 2). For all cases, centralized diagnosis had an 84.2% sensitivity, 37.7% specificity, and 73.7% accuracy, whereas mrAI diagnosis exhibited a performance of 70.6% sensitivity, 61.3% specificity, and 68.5% accuracy. Local diagnosis showed 96.1% sensitivity, 20.8% specificity, and 79.0% accuracy. Given that the entire set comprised various MRIs acquired under different protocols or with different scanners, we hypothesized that the suboptimal results from mrAI might be due to discrepancies in image quality between the analyzed MRIs and those used to develop the algorithm. As noted in our previous report [8], we used high-resolution or 3D MRIs obtained with a Philips scanner. Thus, we evaluated the performance of mrAI using MRIs acquired under similar conditions. Out of all cases, 226 had an inter-slice spacing of 6 mm or less. Analysis of these cases indicated that mrAI diagnostic performance improved to 73.0% sensitivity, 67.3% specificity, and 71.7% accuracy, while the performance based on centralized diagnosis as well as local diagnosis remained similar to that in the first evaluation. Subsequently, MRIs acquired with a Philips scanner were isolated, and 129 cases were analyzed (selected group). The performance of mrAI furthermore improved to 76.8% sensitivity, 76.7% specificity, and 76.7% accuracy, while the performance of the centralized diagnosis was consistent with that in the first analysis. Local diagnosis yielded a performance of 99.0% sensitivity, 23.3% specificity, and 81.4% accuracy. These data suggested a tendency for local diagnosis to overestimate the tumor stage. As for diagnosis by mrAI, scanners from the other manufacturers did not yield results as favorable as those from Philips.

The diagnostic performance for stage T4 tumors was further detailed by categorizing tumor depth as $\leq T2$, T3, or T4 in the selected group. In the diagnosis of T4 tumors, the sensitivity and specificity were 85.7%/90.4% for local diagnosis,

Table 1 Patients' characteristics, MRI details, TNM classification, surgical procedures performed, and pathologic findings

| | |
|---|--|
| Patient characteristics | |
| Age in years, median (range) | 63 (26–86) |
| Biologic sex, male/female | 317/150 |
| BMI (kg/m ²), median (range) | 22.3 (15.0–40.9) |
| PS, 1/2/3/4 | 174/263/28/1 |
| MRI | |
| Magnetic field strength (T), 1.0/1.5/3.0/unknown | 1/349/92/25 |
| In-plane resolution (mm), median (range) | 0.52 × 0.52 (0.27 × 0.27– 1.02 × 1.02) |
| Inter-slice spacing (mm), median (range) | 6.1 (3.0–10) |
| Manufacture, P/G/S/T/H ^a | 242/106/90/25/4 |
| Tumor characteristics at initial examination at each hospital | |
| Distance from anal verge (cm), median (range) | 5 (0–12) |
| cT1/2/3/4 | 3/33/351/80 |
| cN, negative/positive/unknown | 173/292/2 |
| Suspected metastasis to lateral lymph nodes | 59 |
| Operative results | |
| Approach, open/laparoscopy | 329/138 |
| Procedure, anterior resection/ISR/APR/Hartmann/extended resection | 233/79/123/15/17 |
| LPND, yes/no | 281/186 |
| Pathologic findings | |
| Histology, tub1/tub2/pap/por/sig | 146/299/6/15/1 |
| pT1/2/3/4 | 11/95/309/52 |
| pN0/1/2/3 | 228/123/68/48 |
| pM0/1 | 462/5 |
| pStage I/II/IIIa/IIIb/IV | 66/162/122/112/5 |
| Metastasis to lateral lymph node, yes/no | 43/424 |

APR, abdominoperineal resection; BMI, body mass index; IQR, interquartile range; ISR, intersphincteric resection; LPND, lateral pelvic lymph node dissection; pap, papillary adenocarcinoma; por, poorly differentiated adenocarcinoma; PS, performance status.; sig, signet-ring cell carcinoma; tub1, well-differentiated adenocarcinoma; tub2, moderately-differentiated adenocarcinoma

^aP, Philips Medical Systems; G, GE Medical Systems; S, SIEMENS; T, Toshiba; H, Hitachi Medical Corporation

57.1%/94.8% for centralized diagnosis, and 28.6%/97.4% for mrAI-based diagnosis (Supplementary Table 2).

Precision of predicting MRF involvement

We aimed to validate the accuracy of mrAI in predicting MRF involvement. However, we identified two major challenges in verifying the presence of MRF involvement by pathology. First, at the time the analyzed cases underwent surgery, the CRM was not generally evaluated using a circumferential specimen in Japan. Second, even if MRF involvement was present, the resection of surrounding tissues such as nerves might result in the absence of cancer infiltration at the resection margin of the specimen. Therefore, we decided to evaluate the performance of mrAI by assessing the correlation of MRF involvement between centralized diagnosis and mrAI diagnosis. As

shown in Table 3, in the selected group, the mrAI evaluation significantly correlated with the centralized diagnosis ($P=0.0003$).

Long-term survival

Regarding relapse-free survival, we evaluated the association with long-term prognosis for each diagnostic modality by comparing $\geq T3$ with $\leq T2$ tumors. In local diagnosis, no significant difference was observed between $\geq T3$ and $\leq T2$ ($P=0.8717$) (Fig. 3a). While no significant difference was observed between $\geq T3$ and $\leq T2$ in the centralized diagnosis ($P=0.1510$), the prognosis for $\geq T3$ was significantly worse in the mrAI diagnosis than it was for $\leq T2$ ($P=0.0318$), consistent with the pathologic results in which the prognosis for $\geq T3$ was significantly worse compared with that for $\leq T2$ ($P=0.0484$) (Fig. 3b–d).

Table 2 Comparison of the diagnostic accuracy of the local, centralized, and mrAI diagnoses for $\leq T2$ and $\geq T3$, using pathological diagnosis as a reference

| | Inter-slice spacing | | | | | | | | | | | | | | |
|-----------------------|---------------------|-------------|-----------|-----|-------------|-------------|-----------------------|-------------|-------------|-----------|-----------|-------------|-------------|----------|-------|
| | Whole (N = 467) | | | | | | ≤ 6 mm (N = 226) | | | | | | | | |
| | Diagnosis | Tumor depth | Pathology | | Sensitivity | Specificity | Accuracy | Diagnosis | Tumor depth | Pathology | | Sensitivity | Specificity | Accuracy | |
| | | $\leq T2$ | $\geq T3$ | | | | | | | $\leq T2$ | $\geq T3$ | | | | |
| Manufacturers | Whole (N = 467) | Local | $\leq T2$ | 22 | 14 | 96.1% | 20.8% | 79.0% | Local | $\leq T2$ | 10 | 5 | 97.1% | 19.2% | 79.2% |
| | | | $\geq T3$ | 84 | 347 | | | $\geq T3$ | | 42 | 169 | | | | |
| | | Centralized | $\leq T2$ | 40 | 57 | 84.2% | 37.7% | 73.7% | Centralized | $\leq T2$ | 19 | 29 | 83.3% | 36.5% | 72.6% |
| | | | $\geq T3$ | 66 | 304 | | | $\geq T3$ | | 33 | 145 | | | | |
| | | mrAI | $\leq T2$ | 65 | 106 | 70.6% | 61.3% | 68.5% | mrAI | $\leq T2$ | 35 | 47 | 73.0% | 67.3% | 71.7% |
| | | | $\geq T3$ | 41 | 255 | | | $\geq T3$ | | 17 | 127 | | | | |
| | Philips (N = 242) | Local | $\leq T2$ | 15 | 3 | 98.4% | 25.4% | 80.6% | Local | $\leq T2$ | 7 | 1 | 99.0% | 23.3% | 81.4% |
| | | | $\geq T3$ | 44 | 180 | | | $\geq T3$ | | 23 | 98 | | | | |
| | | Centralized | $\leq T2$ | 22 | 28 | 84.7% | 37.3% | 73.1% | Centralized | $\leq T2$ | 11 | 18 | 81.8% | 36.7% | 71.3% |
| $\geq T3$ | | | 37 | 155 | | | $\geq T3$ | 19 | | 81 | | | | | |
| mrAI | | $\leq T2$ | 40 | 49 | 73.2% | 67.8% | 71.9% | mrAI | $\leq T2$ | 23 | 23 | 76.8% | 76.7% | 76.7% | |
| | | $\geq T3$ | 19 | 134 | | | $\geq T3$ | | 7 | 76 | | | | | |
| Non-Philips (N = 225) | Local | $\leq T2$ | 7 | 11 | 93.8% | 14.9% | 77.3% | Local | $\leq T2$ | 3 | 4 | 94.7% | 13.6% | 76.3% | |
| | | $\geq T3$ | 40 | 167 | | | $\geq T3$ | | 19 | 71 | | | | | |
| | Centralized | $\leq T2$ | 18 | 29 | 83.7% | 38.3% | 74.2% | Centralized | $\leq T2$ | 8 | 11 | 85.3% | 36.4% | 74.2% | |
| | | $\geq T3$ | 29 | 149 | | | $\geq T3$ | | 14 | 64 | | | | | |
| | mrAI | $\leq T2$ | 25 | 57 | 68.0% | 53.2% | 64.9% | mrAI | $\leq T2$ | 12 | 24 | 68.0% | 54.5% | 64.9% | |
| | | $\geq T3$ | 22 | 121 | | | $\geq T3$ | | 10 | 51 | | | | | |

Table 3 The correlation of mesorectal fascia involvement between centralized diagnosis and mrAI diagnosis

| | Centralized diagnosis | |
|----------------|-----------------------|----------|
| | Negative | Positive |
| mrAI diagnosis | | |
| Negative | 77 | 35 |
| Positive | 4 | 13 |

Discussion

In this study, we were able to verify the usefulness of mrAI, which we developed using a multi-institutional dataset. The performance of mrAI was enhanced when characteristics of the MRI imaging environment, such as the imaging protocol and the manufacturer, were close to those of the ground-truth data used for algorithm development. This result suggests how sensitive AI can be as a diagnostic support technology and has revealed that addressing this issue will be mandatory in making this technology generally available in future. The result that mrAI had higher diagnostic accuracy with thinner slice thickness suggests that image quality improvement of each cross-sectional image and the information embedded within the continuity of multiple sections are likely important in enhancing diagnostic accuracy.

The most standard approach for assessing the accuracy of preoperative diagnosis is comparison with pathologic findings. While such verification was previously possible, it has become challenging to use unmodified specimens now that preoperative treatment has been standardized [15–18]. Japan has a unique environment where unmodified pathologic

information from relatively recent times is easily accessible, as surgery-first treatment has been standardized for a long time, even after neoadjuvant CRT became prevalent in the West. In this study, to develop an AI-based algorithm, we created ground-truth label data by correlating pathologic sections of circular specimens with MRI images. To assess diagnostic accuracy, we utilized nationwide, multicentric data to compare pathologic findings with MRI on a 1:1 basis. In Japan too, the importance of preoperative treatment has been recognized in recent years, and the number of cases undergoing preoperative treatment is increasing [20, 21]. Conducting similar research methods will become more and more challenging in future. We consider this study to have significant value for having utilized this rare opportunity.

Regarding T staging, MRI is perceived to have a tendency to overestimate tumor stage rather than underestimate it [22]. In the MERCURY study, a comparison was made between MRI diagnosis and pathologic diagnosis among 311 individuals who underwent surgery first. Of the cases pathologically diagnosed as T1/2, 36% were diagnosed as T3/4 on MRI, while 31% of the cases pathologically diagnosed as T3/4 were diagnosed as T1/2 on MRI [15]. Long-term prognosis analysis of the MERCURY study showed a poor disease-free survival rate in cases with a high risk of positive CRM on preoperative MRI. In this report, 23.2% were diagnosed as Stage I on preoperative MRI, and eventually 26.7% were diagnosed as Stage I pathologically. This fact suggests that among the cases diagnosed pathologically as T2 or lower, there were not a few cases diagnosed as T3/4 at the time of diagnosis [23]. Although the MERCURY study had strict management with tightened MRI diagnostic criteria, it is believed that

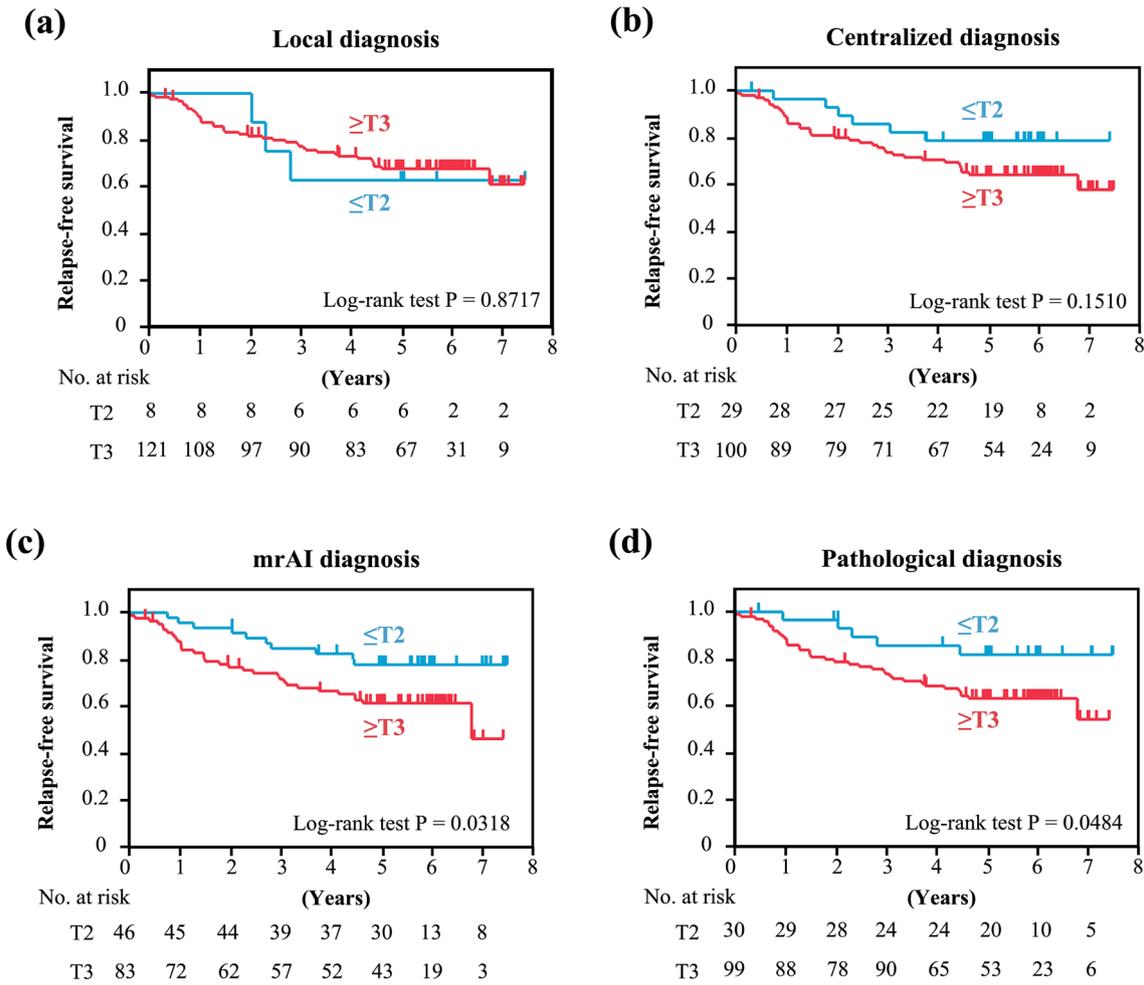


Fig. 3 Relapse-free survival based on depth classification per diagnostic method. **a** Local diagnosis. **b** Centralized diagnosis. **c** mrAI diagnosis. **d** Pathologic diagnosis

the tendency to overestimate the depth in actual clinical settings may be even more pronounced. Other retrospective studies examining the diagnostic accuracy of MRI for upfront surgery cases have also shown a tendency for MRI diagnosis to overestimate tumor stage [16–18], and a meta-analysis analyzing MRI diagnostic accuracy reported that the proportion of overestimation was 25% and of underestimation was 13% [19]. In not a few cases, it is difficult to distinguish between fibrotic and tumor tissues [22]. In radiologic diagnosis made by humans, intentions to avoid under-diagnosis by considering various clinical situations may significantly influence the diagnosis. In our study data, a similar trend was observed. Particularly in this study, which analyzed data from a time when MRI was not widespread in Japan, there was a tendency to extremely over-diagnose in real-world clinical settings. In centralized diagnosis involving experienced radiologists, this tendency was corrected compared with local diagnosis, but 63% of the cases pathologically diagnosed as

T2 or lower were over-diagnosed preoperatively as T3 or higher. While a central diagnostic setting is more conducive to neutral judgments, it would be even harder to avoid over-diagnosis in real-world clinical settings [24]. In this regard, mrAI’s diagnosis is always neutral, and even in cases that are difficult to judge, the algorithm can diagnose tumors without bias as T2 or below. This may contribute to optimizing each patient’s treatment. We do not assert that mrAI surpasses human diagnostic capabilities; rather, we emphasize that AI diagnosis should not be seen as a technology that confronts radiologic diagnosis. A comprehensive understanding of both the advantages and disadvantages of mrAI is imperative, and its application should be judiciously implemented to augment the diagnostic procedures conducted by radiologists. In addition, mrAI, when applied to research using MRI, is expected to be able to make unbiased decisions and also have the potential to reduce the workload of radiologists performing central diagnoses, since they can carry out bias-free diagnostics.

In this study, to evaluate the performance of mrAI, the depth of invasion was extracted based on segmentation and compared with centralized diagnosis. The depth of invasion is a well-known long-term prognostic factor, and in the MERCURY study the hazard ratio for DFS of pathologic Stage II was over five times that of Stage I [23]. Compared with centralized diagnosis, mrAI showed slightly lower sensitivity and higher specificity. To validate the clinical significance of this result, we compared the relapse-free survival of $\leq T2$ and $T3 \leq$ groups in MRI diagnosis. Patients with pathologic stage $\geq T3$ tumors had significantly poorer prognosis than those with tumors staged $\leq T2$, but a significant difference in relapse-free survival was shown only in the diagnosis derived by mrAI, not in the centralized diagnosis. Moreover, no association was found between the two groups in local diagnosis. Although we could not evaluate the performance of mrAI against pathologic assessment for MRF involvement, the observed correlation between centralized diagnosis and mrAI diagnosis suggests that mrAI may be useful for assessing MRF involvement.

An essential factor in deciding on preoperative treatment for locally advanced rectal cancer is the risk of recurrence for each case. The recently highlighted TNT has been proven to improve DFS in locally advanced rectal cancer, and its application is expanding, especially in Western countries [1–3]. However, for cases that can be cured without TNT, it represents overtreatment. One of the current challenges is to clearly define the indications for TNT [25]. We expect that mrAI, which can classify relapse-free survival similarly to pathologic diagnosis based on MRI findings at the time of diagnosis, will play a significant role in rectal cancer treatment as multidisciplinary treatments evolve. Especially as mentioned above, there might be a concern among radiologists about missing the opportunity for preoperative treatment, which tends to lead to overestimation. The importance of a tool to support neutral judgment is immense. While MRI can evaluate factors such as extramural vascular invasion (EMVI) [26], lymph node metastasis [27], and tumor deposits [28] that also affect long-term prognosis, they are not assessed by the current mrAI. We have not explored these in this study but plan to do so in future research. Furthermore, with recent advances in ctDNA and multiomics analysis, individualization of colorectal cancer treatment is becoming a reality [29–32]. With these technological innovations, establishing a method to comprehensively evaluate the individual risks of advanced rectal cancer will help pave the way to future individualized treatment of rectal cancer.

There are several limitations to this study. First, this study was retrospective, and therefore there is potential for bias. Because the MRIs were not obtained within a recent timeframe, in some cases the image quality was inferior compared with that of MRI obtained with standardized imaging protocols in Western countries [22, 33]. Second, our findings

could not prove that the current mrAI is universally applicable across different vendors and imaging environments. However, optimizing the imaging environment for MRI can enhance the applicability of mrAI, and we believe this challenge can be adequately addressed in future. Third, we did not individually review the segmentation results extracted by mrAI for each case. The accuracy of these segmentation results needs to be prospectively accumulated and verified, ensuring proper imaging conditions.

In conclusion, we were able to verify the performance of mrAI, which we developed using multi-institutional data. By ensuring proper imaging conditions for MRI, the accuracy of the results of the mrAI analysis can be enhanced, and mrAI has the potential to provide feedback to radiologists without leading to overestimation of tumor stage.

Acknowledgements This work was funded by the FUJIFILM Corporation. We are grateful to Japan Society of Laparoscopic Colorectal Surgery, for giving us the opportunity to utilize the data for secondary purposes.

Authors' contributions A. Hamabe, I. Takemasa, Y. Sakai, and M. Watanabe were involved in conceptualization and project administration. A. Hamabe, M. Ishii, K. Okuya, K. Hida, D. Nishizaki, A. Sumii, S. Arizono, S. Kohno, K. Tokunaga, and H. Nakai participated in data curation and investigation. A. Hamabe, I. Takemasa, M. Ishii, and K. Okuya participated in formal analysis. A. Hamabe and I. Takemasa obtained the funding. Y. Sakai, and M. Watanabe were involved in supervision. A. Hamabe, I. Takemasa, M. Ishii, K. Okuya, K. Hida, S. Arizono, Y. Sakai, and M. Watanabe participated in methodology and validation. A. Hamabe, M. Ishii, and K. Okuya were involved in visualization. A. Hamabe and I. Takemasa wrote the initial draft. Y. Sakai, and M. Watanabe reviewed and edited the draft. All authors critically revised the manuscript and approved the final version.

Funding Open Access funding provided by Osaka University. This study received funded support from the Fujifilm Corporation (<https://www.fujifilm.com>). They had no role in the study design, the gathering, analysis, or interpretation of data, or in the writing of the manuscript or decision to submit the paper for publication.

Declarations

Conflict of interests None of the authors has any relevant conflicts of interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Bahadoer RR, Dijkstra EA, van Etten B, et al. Short-course radiotherapy followed by chemotherapy before total mesorectal excision (TME) versus preoperative chemoradiotherapy, TME, and optional adjuvant chemotherapy in locally advanced rectal cancer (RAPIDO): a randomised, open-label, phase 3 trial. *Lancet Oncol.* 2021;22:29–42.
2. Conroy T, Bosset J-F, Etienne P-L, et al. Neoadjuvant chemotherapy with FOLFIRINOX and preoperative chemoradiotherapy for patients with locally advanced rectal cancer (UNICANCER-PRODIGE 23): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol.* 2021;22:702–15.
3. Jin J, Tang Y, Hu C, et al. Multicenter, randomized, phase III trial of short-term radiotherapy plus chemotherapy versus long-term chemoradiotherapy in locally advanced rectal cancer (STELLAR). *J Clin Oncol.* 2022;40:1681–92.
4. Birgisson H, Pahlman L, Gunnarsson U, Glimelius B. Swedish Rectal Cancer Trial G Adverse effects of preoperative radiation therapy for rectal cancer: long-term follow-up of the Swedish Rectal Cancer Trial. *J Clin Oncol.* 2005;23:8697–705.
5. Rombouts AJM, Hugen N, Elferink MAG, et al. Incidence of second tumors after treatment with or without radiation for rectal cancer. *Ann Oncol.* 2017;28:535–40.
6. Glynne-Jones R, Wyrwicz L, Tiret E, et al. Rectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up†. *Ann Oncol.* 2017;28:iv22–40.
7. Benson AB, Venook AP, Al-Hawary MM, et al. Rectal cancer, version 2.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw.* 2022;20:1139–67.
8. Hamabe A, Ishii M, Kamoda R, et al. Artificial intelligence–based technology for semi-automated segmentation of rectal cancer using high-resolution MRI. *PLoS ONE.* 2022;17:e0269931.
9. Hida K, Okamura R, Sakai Y, et al. Open versus laparoscopic surgery for advanced low rectal cancer: a large, multicenter, propensity score matched cohort study in Japan. *Ann Surg.* 2018;268:318–24.
10. Hida K, Nishizaki D, Sumii A, et al. Prognostic impact of lateral pelvic node dissection on the survival of patients in low rectal cancer subgroups based on lymph node size. *Ann Surg Oncol.* 2021;28:6179–88.
11. Sumii A, Hida K, Sakai Y, et al. Establishment and validation of a nomogram for predicting potential lateral pelvic lymph node metastasis in low rectal cancer. *Int J Clin Oncol.* 2022;27:1173–9.
12. Rouleau Fournier F, Motamedi MAK, Brown CJ, et al. Oncologic outcomes associated with MRI-detected extramural venous invasion (mrEMVI) in rectal cancer: a systematic review and meta-analysis. *Ann Surg.* 2022;275:303–14.
13. Chen S, Li N, Tang Y, et al. The prognostic value of MRI-detected extramural vascular invasion (mrEMVI) for rectal cancer patients treated with neoadjuvant therapy: a meta-analysis. *Eur Radiol.* 2021;31:8827–37.
14. Birbeck KF, Macklin CP, Tiffin NJ, et al. Rates of circumferential resection margin involvement vary between surgeons and predict outcomes in rectal cancer surgery. *Ann Surg.* 2002;235:449–57.
15. Group MS. Extramural depth of tumor invasion at thin-section MR in patients with rectal cancer: results of the MERCURY study. *Radiology.* 2007;243:132–9.
16. Burton S, Brown G, Daniels I, et al. MRI identified prognostic features of tumors in distal sigmoid, rectosigmoid, and upper rectum: treatment with radiotherapy and chemotherapy. *J Radiat Oncol Biol Phys.* 2006;65:445–51.
17. Kim H, Lim JS, Choi JY, et al. Rectal cancer: comparison of accuracy of local-regional staging with two- and three-dimensional preoperative 3-T MR imaging. *Radiology.* 2010;254:485–92.
18. Akasu T, Iinuma G, Takawa M, Yamamoto S, Muramatsu Y, Moriyama N. Accuracy of high-resolution magnetic resonance imaging in preoperative staging of rectal cancer. *Ann Surg Oncol.* 2009;16:2787–94.
19. Al-Sukhni E, Milot L, Fruitman M, et al. Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: a systematic review and meta-analysis. *Ann Surg Oncol.* 2012;19:2212–23.
20. Hashiguchi Y, Muro K, Saito Y, et al. Japanese society for cancer of the colon and rectum (JSCCR) guidelines 2019 for the treatment of colorectal cancer. *Int J Clin Oncol.* 2020;25:1–42.
21. Watanabe J, Kagawa Y, Chida K, et al. Phase III trial of short-course radiotherapy followed by CAPOXIRI versus CAPOX in locally advanced rectal cancer: the ENSEMBLE trial. *ESMO Gastrointest Oncol.* 2023;1:9–14.
22. Moreno CC, Sullivan PS, Mittal PK. MRI evaluation of rectal cancer: staging and restaging. *Curr Probl Diagn Radiol.* 2017;46:234–41.
23. Taylor FG, Quirke P, Heald RJ, et al. Preoperative magnetic resonance imaging assessment of circumferential resection margin predicts disease-free survival and local recurrence: 5-year follow-up results of the MERCURY study. *J Clin Oncol.* 2014;32:34–43.
24. Wong C, Fu Y, Li M, et al. MRI-based artificial intelligence in rectal cancer. *J Magn Reson Imaging.* 2023;57:45–56.
25. Takemasa I, Hamabe A, Miyo M, Akizuki E, Okuya K. Essential updates 2020/2021: advancing precision medicine for comprehensive rectal cancer treatment. *Ann Gastroenterol Surg.* 2022;7:198–215.
26. Smith NJ, Barbachano Y, Norman AR, Swift RI, Abulafi AM, Brown G. Prognostic significance of magnetic resonance imaging-detected extramural vascular invasion in rectal cancer. *Br J Surg.* 2008;95:229–36.
27. Brown G, Richards CJ, Bourne MW, et al. Morphologic predictors of lymph node status in rectal cancer with use of high-spatial-resolution MR imaging with histopathologic comparison. *Radiology.* 2003;227:371–7.
28. Lord AC, D’Souza N, Shaw A, et al. MRI-Diagnosed tumour deposits and EMVI status have superior prognostic accuracy to current clinical TNM staging in rectal cancer. *Ann Surg.* 2022;276:334–44.
29. Kotani D, Oki E, Nakamura Y, et al. Molecular residual disease and efficacy of adjuvant chemotherapy in patients with colorectal cancer. *Nat Med.* 2023;29:127–34.
30. Zhou J, Wang C, Lin G, et al. Serial circulating tumor DNA in predicting and monitoring the effect of neoadjuvant chemoradiotherapy in patients with rectal cancer: a prospective multicenter study. *Clin Cancer Res.* 2021;27:301–10.
31. Tie J, Cohen JD, Wang Y, et al. Serial circulating tumour DNA analysis during multimodality treatment of locally advanced rectal cancer: a prospective biomarker study. *Gut.* 2019;68:663–71.
32. Ullah I, Yang L, Yin FT, et al. Multi-omics approaches in colorectal cancer screening and diagnosis, recent updates and future perspectives. *Cancers.* 2022;14:5545.
33. Gormly. Rectal MRI: the importance of high resolution T2 technique. *Abdom Radiol.* 2021;46:4090–5.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.