Abstract. The radiological diagnosis of Crohn's disease (CD)-related anorectal cancer is difficult; it is often found in advanced stages and has a poor prognosis because of the difficulty of curative surgery. However, there are no studies on predicting the diagnosis of CD-related cancer. The present study aimed to develop a predictive model to diagnose CD cancerous lesions more accurately in a way that can be interpreted by clinicians. Patients with CD who developed anorectal CD lesions at Hyogo Medical University (Nishinomiya, Japan) between March 2009 and June 2022 were included in the present study. T2-weighted and T1-weighted magnetic resonance (MR) images were utilized for our analysis. Images of anorectal lesions were segmented using open-source 3D Slicer software, and radiomic features were extracted using PyRadiomics. Six machine learning models were investigated and compared: i) Support vector machine; ii) naive Bayes; iii) random forest; iv) light gradient boosting machine; v) extremely randomized trees; vi) and regularized greedy forest (RGF). SHapley Additive exPlanations (SHAP) values were calculated to assess the extent to which each radiomic feature contributed to the model's predictions compared to baseline, represented as the average of the model's predictions for all test data. The T2-weighted images of 28 patients with anorectal cancer and 40 non-cancer patients were analyzed and the contrast-enhanced T1-weighted images of 22 cancer and 40 non-cancer patients. The model with the highest area under the curve (AUC) was the RGF-based model constructed using T2-weighted image features, achieving an AUC of 0.944 (accuracy, 0.862; recall, 0.830). The SHAP-based model explanation suggested a strong association between the diagnosis of CD-related anorectal cancer and features such as complex lesion texture; greater pixel separation within the same coronal cross-section; larger, randomly distributed clumps of pixels with the same signal intensity; and a more spherical lesion shape on T2-weighted images. The MRI radiomics-based RGF model demonstrated outstanding performance in predicting CD-related anorectal cancer. These results may affect the diagnosis and surveillance strategies of CD-related colorectal cancer.

Introduction

Anorectal cancer is the most prevalent form of Crohn's disease-related colorectal cancer (CD-CRC) in Japan and other Asian countries (1-3), and its incidence has been increasing since 2000 (4,5). In contrast, in Western countries, CD-CRC frequently develops in the right colon, and genome-wide association studies of CD patients have identified different associated mutations between patients of Asian and European descent (6,7). Nevertheless, some Western countries have reported anorectal cancer in CD patients, in which it ranks as the second most common form of CD-CRC, necessitating further research (7-11).

A large-cohort study from Japan reported that compared to sporadic CRC patients, CD-CRC patients were younger and had lower R0 resection rates, worse 5-year overall survival (OS) rates, and a poorer prognosis in stage II or III (12). The poor prognosis of patients may be related to the difficulty of early diagnosis, the complexity of radical surgery in advanced cancer, and the low efficacy of chemoradiotherapy due to the high incidence of mucinous carcinoma (2). Therefore, although periodic blood draw, magnetic resonance imaging (MRI), and endoscopic or transanal histology are recommended for early detection of these cancers, no effective surveillance program has been established (13). Typically, the diagnosis of anorectal cancer relies on the use of T2-weighted and post-contrast-enhanced T1-weighted MRI scans, with T2-weighted imaging (T2WI) playing a crucial role in assessing tumor depth (14). However, it has been reported that low anorectal carcinoma in CD arising in perianal fistulas can be difficult to diagnose clinically and radiologically (15,16).
To address these challenges, we performed radiomics analysis, an exhaustive analysis using high-dimensional quantitative features, on MR images of anorectal cancer and noncancerous lesions from CD patients to identify MRI features effective in discriminating benign and malignant anorectal cancer and predicting diagnosis. Common concerns regarding machine learning-based models, especially those using radiomics features, are that their outcomes are not easily explainable. Therefore, radiologists are hesitant to adopt such models (17). SHapley Additive exPlanations (SHAP) is a framework used to interpret predictions and define measures of additive feature importance and classes of theoretical outcomes (18). SHAP involves showing feature importance and its impact on the overall prediction model, with several reported benefits, including providing an understanding of the importance of individual features to model outputs (19,20). Therefore, the combination of SHAP and radiomics has the potential to allow clinicians to interpret predictions and definitions of measures of additive feature importance within a theoretical outcome framework (21).

We aimed to develop a predictive model to more accurately diagnose CD cancerous lesions and to combine the SHAP method with a predictive model to explain and visualize the model.

Materials and methods

Patient selection. We retrospectively analyzed the data of CD patients who developed anorectal CD lesions at Hyogo medical university between March 2009 and June 2022 in this study. The patients were divided into a cancer group or a noncancer group. The cancer group was diagnosed with anorectal cancer according to endoscopic biopsy or histopathology of the original operative specimens obtained at surgery. The noncancer group was diagnosed with anorectal CD lesions that were not identified as cancer, which should be those with anorectal lesions that had been present for at least 10 years. The data collected included sex, age at the time of the MRI scan, duration of disease, Montreal classification for CD, anorectal disease, prednisolone administration, biologic administration, histological findings, and cancer stage. In the present study, CD-CRC was classified using the 8th Union for International Cancer Control pathological TNM (pTNM) (pTNM) staging system.

We excluded patients in whom CD was not definitively diagnosed, those who did not undergo MRI, and those who underwent MRI with different imaging protocols, such as no TR/TE of 4,000-4,750/110-120 ms, 3 mm slice thickness/0.3 mm gap, 28x22 cm field of view (FOV), and 228x256-256x320 matrix. T1-weighted imaging (T1WI) was performed in the axial plane with a spin-echo TR/TE of 500-550/9-10 ms, 3 mm slice thickness/0.3 mm gap, 28x22 cm FOV, and 228x256-256x320 matrix. The MR images were retrieved from the facility's image archive and communication system and transferred to medical image merging software for image segmentation.

Image segmentation. Regions of interest (ROIs) within the acquired MR images were delineated using the open-source 3D Slicer program. The segmentation process involved the use of a threshold segmentation method, which defined the area within a fixed threshold range, followed by manual fine-tuning of the selected area as whole anorectal lesions, including fistula. The ROIs were delineated by a radiologist with over two decades of experience and subsequently reviewed by an equally seasoned radiologist with more than three decades of expertise.

Feature extraction and selection. Radiomic features were extracted with PyRadiomics, a Python library. A total of four feature extraction patterns were applied, including postfilter processing patterns. Pattern 1 involved no filter processing, Pattern 2 incorporated a wavelet filter, Pattern 3 employed a Laplacian of Gaussian filter with sigma values of 1 mm and 3 mm, and Pattern 4 amalgamated the features extracted from Patterns 1 to 3 while excluding common features. For each image, features were extracted in categories such as shape, first-order, gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level dependency matrix (GLDM), and gray level size zone matrix (GLSZM).

To prevent overfitting, seven feature selection methods were employed for the extracted features: Method 1 applied no feature selection, Method 2 applied feature selection based on the top 5 correlation coefficients with the outcome variable, Method 3 applied feature selection based on the top 50 significance was set at P<0.05. The odds ratio (OR) and 95% confidence intervals (CIs) were calculated for all variables in the univariate analysis. A multivariate logistic regression analysis was performed to examine the association between study variables and postoperative complications for factors with p values <0.20 in the univariate analysis. JMP ver. 16 (SAS Institute Inc. Cary, North Carolina, USA) was used to perform all analyses.
correlation coefficients with the outcome variable, Method 4 applied feature selection based on the top 5 mutual information content with the outcome variable, Method 5 applied feature selection based on the top 50 mutual information content with the outcome variable, Method 6 applied feature selection through least absolute shrinkage and selection operator (LASSO) regression, and Method 7 applied feature selection with L1 regularized logistic regression.

**Machine learning models and evaluation.** Six machine learning models were investigated and compared: support vector machine (SVM), naive Bayes (NB), random forest (RF), light gradient boosting machine (Light GBM), extremely randomized trees (ERT), and regularized greedy forest (RGF). A library called RGF-python was used for RGF, and Scikit-learn was used for the other models. The following hyperparameters were tried for each model. Default values were used for the other hyperparameters. The following is a list of parameters for which we have tried multiple values. Regularization parameters (C: 0.01, 0.1, 1.0, 10) for SVM. Parameters of variance smoothing for NB classifiers (var smoothing: 1e-09) for NB. Maximum depth of the decision tree (max_depth: 2, 3, 4, 5) for RF. Learning_rate (0.01, 0.05, 0.1) for LightGBM. Maximum depth of decision tree (max_depth: 2, 3, 4, 5) for ERT. Maximum number of leaves on the decision tree (max_leaf: 100, 500, 1000, 1500, 2000, 5000) for RGF.

A fourfold cross-validation score was employed to evaluate model performance. The data set was randomly partitioned into four equal subsets, with one serving as the test set and the remaining three as the training sets. This procedure was repeated four times, each time designating a different subset as the test set. Model performance was determined by averaging the scores obtained from the four evaluations, with the evaluation metric set as the AUC, accuracy and recall.

**Application of SHAP.** SHAP is a method for interpreting the contribution of each feature to the predictions of a machine learning model (SHAP value) in accordance with game theory principles. By analyzing the importance of features in the model's prediction process, one can gain insights into the relationship between these features and, in this case, malignant tumors associated with Crohn's disease, as well as their medical significance. The SHAP value is computed to quantify the extent to which each feature contributes to the model's predictions relative to a baseline value, represented as the average of the model's predictions for all test data. The predicted value for each data point was calculated as the base value plus the summation of the SHAP values.

**Comparison with a deep learning model.** Transfer learning, a type of deep learning, is widely used in medical image analysis, where the amount of data collected is often insufficient (24). In this method, a pre-trained model is used for a new task; the model is expected to produce good results with a small amount of data, and only features unique to the new task need to be learned.

One pre-trained model that can handle 3D images instead of 2D images is the 3D-CNN. The 3D-CNN was developed for video analysis, but it can also be applied to 3D medical images and has shown good performance in a variety of tasks (25).

**Comparison with other machine learning methods.** One machine learning technique is an ensemble learning method called stacking. Kwon et al reported that stacking exceeded the prediction results of a single model in a breast cancer prediction task (26). Stacking is a method in which the predictions of multiple models, called meta-features, for each dataset are used as features, and a new model is used to make the final prediction, which is called a meta-model.

We chose the predictions of a total pattern obtained from the predictions of the radiomics models as candidates for meta-features. If all features are used, there is a possibility of overfitting or duplication of similar data. Therefore, it is necessary to select meta-features appropriately based on specific criteria. Stacking is known to improve results by combining models that provide different prediction results (6). Therefore, we performed clustering using the k-means method based on the prediction patterns, and for each cluster, we selected the prediction closest to the center of the cluster. These selected predictions were used as meta-features. The number of clusters (equal to the number of meta-features) was 5, 10, 15, and 30, for a total of 4 patterns. For the meta-model, we tested three models: logistic regression, RF, and LightGBM.

**Results**

**Patient characteristics.** A total of 52 patients with CD-related anorectal cancer and 40 noncancer patients were initially included. Of these, 34 cancer patients and 40 noncancer patients had undergone MRI. Patients with different imaging conditions, such as not including fat suppression in T2-weighted and T1-weighted imaging sequences, were excluded. Finally, the T2-weighted images of 28 cancer and 40 noncancer patients and the contrast-enhanced T1-weighted images of 22 cancer and 40 noncancer patients were included in the study. Representative MRI image of CD-related anorectal cancer was shown in Fig. 1.

The characteristics of the patients in the cancer group and the noncancer group are shown in Table I. There were no significant differences in sex distribution, age at the time of the MRI scan, age at diagnosis, anoperineal disease, or pharmacotherapy between the groups. The duration of disease was significantly longer in the cancer group than in the noncancer group.
Although the proportion of L2 (colon type) was significantly higher in the cancer group than in the noncancer group (P<0.01), the proportion of L3 (ileocolonic type) was significantly lower in the cancer group than in the noncancer group (P<0.01). There were significantly more patients in the cancer group who had the structuring and penetrating (B2+B3) type (P=0.02). Although 100% of the patients had the penetrating type in both groups, two patients in the cancer group had no anoperineal disease.

The histological findings of anorectal cancer were well‑differentiated adenocarcinoma in 5 cases (17%), moderately differentiated adenocarcinoma in 2 cases (7%), poorly differentiated adenocarcinoma in 3 cases (10.7%), mucinous adenocarcinoma in 14 cases (50%), and signet‑ring cell carcinoma and squamous cell carcinoma in 2 cases (7%). There were stage I in 2 cases (7%), stage II in 11 cases (39.2%), stage III in 5 cases (17.8%), IV in 7 cases (25%), and X in 3 cases (10.7%). Mucinous adenocarcinoma was the most frequent histological finding, and there were few cases of early‑stage cancer.

### Table I. Characteristics of patients with CD with T2‑weighted images (cancer group and non‑cancer group).

<table>
<thead>
<tr>
<th>Factors</th>
<th>Overall (n=68)</th>
<th>Cancer group (n=28)</th>
<th>Non‑cancer group (n=40)</th>
<th>P‑value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex, n (%)</td>
<td>20 (29.4)</td>
<td>10 (35.7)</td>
<td>10 (25.0)</td>
<td>0.24</td>
</tr>
<tr>
<td>Age at the time of MRI scan, years, median (range)</td>
<td>45 (25‑81)</td>
<td>49.5 (34‑58)</td>
<td>41.5 (25‑81)</td>
<td>0.14</td>
</tr>
<tr>
<td>Duration of disease, months, median (range)</td>
<td>243 (129‑519)</td>
<td>288 (144‑516)</td>
<td>221.5 (129‑519)</td>
<td>0.01*</td>
</tr>
<tr>
<td>Montreal age at diagnosis classification, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1: ≤ 16 years old</td>
<td>13 (19.1)</td>
<td>4 (14.3)</td>
<td>9 (22.5)</td>
<td>0.87</td>
</tr>
<tr>
<td>A2: 17‑40 years old</td>
<td>53 (77.9)</td>
<td>24 (85.7)</td>
<td>29 (72.5)</td>
<td>0.15</td>
</tr>
<tr>
<td>A3: &gt; 40 years old</td>
<td>2 (2.9)</td>
<td>0 (0)</td>
<td>2 (5.0)</td>
<td>0.22</td>
</tr>
<tr>
<td>Montreal location of diagnosis classification, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1: ileum</td>
<td>6 (8.8)</td>
<td>3 (10.7)</td>
<td>3 (7.5)</td>
<td>0.64</td>
</tr>
<tr>
<td>L2: colon</td>
<td>15 (22.1)</td>
<td>1 (3.6)</td>
<td>14 (35.0)</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>L3: ileocolonic</td>
<td>47 (69.1)</td>
<td>24 (85.7)</td>
<td>23 (57.5)</td>
<td>0.01*</td>
</tr>
<tr>
<td>Montreal lesion behavior classification, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1: inflammatory</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.00</td>
</tr>
<tr>
<td>B2: stricturing</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.00</td>
</tr>
<tr>
<td>B3: penetrating</td>
<td>68 (100.0)</td>
<td>28 (100.0)</td>
<td>40 (100.0)</td>
<td>0.00</td>
</tr>
<tr>
<td>B2+B3</td>
<td>15 (22.1)</td>
<td>10 (35.7)</td>
<td>5 (12.5)</td>
<td>0.02*</td>
</tr>
<tr>
<td>Anoperineal disease, n (%)</td>
<td>66 (97.1)</td>
<td>26 (92.8)</td>
<td>40 (100.0)</td>
<td>0.08</td>
</tr>
<tr>
<td>PSL administration, n (%)</td>
<td>14 (20.6)</td>
<td>5 (17.9)</td>
<td>9 (22.5)</td>
<td>0.64</td>
</tr>
<tr>
<td>Biologic administration, n (%)</td>
<td>40 (58.8)</td>
<td>14 (50.0)</td>
<td>26 (65.0)</td>
<td>0.21</td>
</tr>
</tbody>
</table>

*P<0.05. The Mann‑Whitney U test was performed to compare continuous variables. The chi‑square test or Fisher's exact test was used to compare categorical variables. CD, Crohn's disease; PSL, prednisolone.

Radiomics models. The radiomics process is visually presented in the flowchart shown in Fig. 2. Prediction accuracy was calculated for a total of 616 patterns, which is the sum of 4 patterns for the filtering process, 7 patterns for feature selection, and various models (6 different models and hyperparameters for a total of 22 patterns).

The prediction models derived from the T2‑ and T1‑weighted images are described in Tables III and IV, respectively. The model with the highest AUC for the T2‑weighted image was RGF [pattern: no filter, using LASSO regression, max leaf was 100], with an AUC of 0.944 (accuracy: 0.862, recall: 0.830). For T1‑weighted images, the model with the highest AUC was based on SVM [pattern: no filter, no feature selection, regularization parameters was 0.1], with an AUC of 0.932 (accuracy: 0.877, recall: 0.747). The MRI‑based radiomic models showed higher AUCs than the patient characteristics model, and the RGF model based on T2‑weighted image features showed the highest AUC.

Explanation and visualization of the radiomics models. SHAP was used to quantitatively describe the output of the most accurate model (RGF) based on T2‑weighted images. SHAP summary plots were generated to provide a visually concise picture by representing the range and distribution of the feature importance values on the model output and relating the feature values to the feature impacts. The distributed importance plot in Fig. 3A lists the most important variables in descending order, with the top four being 1: GLCM_MCC (maximal correlation coefficient),
which reflects texture complexity; 2: SHAPE Maximum 2D Diameter Column, reflecting the maximum distance between pixels in the same coronal cross-section; 3: GLSZM _Zone Entropy, which indicates how randomly distributed the clump sizes of pixels of the same signal intensity are; and 4: SHAPE Surface Area to Volume Ratio, in which values closer to 0 indicate greater sphericity. Each dot in Fig. 3B is colored according to the value of the feature, with contributions ranging from negative (blue) to positive (red). Interpretation of these data indicates that if the signal intensity was heterogeneous in terms of the texture patterns (GLCM_MCC and GLSZM_Zone Entropy), the width was greater in the direction of the coronal cross-section (SHAPE_Maximum 2D Diameter Column) and the shape of the lesion was more spherical (SHAPE_Surface Area to Volume Ratio), the lesion was more likely to be judged as anorectal cancer.

A force plot (Fig. 4) was generated for each individual data set to show which and the degree to which features influenced the model to reach its predictions. Features with positive SHAP values are represented in red, features with negative SHAP values are represented in blue, and the length of the bar indicates the magnitude of the SHAP value for each feature. Two patients (Fig. 4A) for whom cancer could not be correctly predicted and one patient (Fig. 4B) for whom cancer could be correctly predicted are shown for comparison. GLCM_MCC in Case 1 and SHAPE_Maximum 2D Diameter Column and SHAPE_Surface Area to Volume Ratio in Case 2 had negative SHAP values. The two patients for whom cancer could not be correctly predicted had stage 1 early-stage cancer (Case 1) and squamous cell carcinoma (Case 2). There were two patients with stage 1 cancer and two with squamous cell carcinoma in all cases.

Comparison with a deep learning model: R(2+1)D. The prediction results using T2-weighted images are shown in Table V.
Comparison with other machine learning methods: prediction models-stacking. The prediction results using T2-weighted images are shown in Table VI. The pattern with the highest AUC among the predictions of each meta-model is listed. The stacking model (AUC: 0.912) exhibited a worse performance than the radiomics model (AUC: 0.947).

**Discussion**

To our knowledge, this is the first study to develop a machine learning-based radiomics model for predicting CD-related anorectal cancer. We demonstrate that the radiomics-based RGF model performed excellently in determining whether an anorectal lesion in CD was cancerous. Moreover, we also found that the extracted features could be clinically explained by applying SHAP.

Radiomics is derived from the terms radiology and omics (the science of systematically handling large amounts of information). It is a relatively new method for extracting large amounts of quantifiable image characteristics from conventional CT, MRI, and PET images and converting them into a form that can be data-mined (27). Visualization of tumor heterogeneity by extracting radiomics features is very important in assessing tumor grade and prognosis, and the ability of radiomics analysis to differentiate between cancer and benign tissue has already been demonstrated in various organs (28). In prostate cancer, a study of 147 biopsy-diagnosed patients reported that texture analysis was able to distinguish between cancerous and noncancerous prostate tissue on both T2-weighted MR images and diffusion coefficient maps obtained from diffusion-weighted MR images (28). Although data on the value of radiomics in colorectal cancer patients are limited, some studies have illustrated the usefulness of radiomics in staging colorectal cancer (29).

The models we developed in this study were able to determine the presence of anorectal cancer with a high degree of confidence, with all T2-weighted-based models having AUCs exceeding 0.9. Certainly, targeted endoscopic biopsy and transanal biopsy under anesthesia are important for the diagnosis of cancer, similar to MRI (13,30). However, a thorough examination of the rectum and perianal region is limited by pain in the perineal tissue and anal strictures. In addition, even examination under anesthesia by an experienced colorectal surgeon may miss early malignant lesions (31). The present results suggest that a predictive model using MRI may be particularly useful as a diagnostic aid in such cases. On the other hand, the T1WI-based model tended to have a lower overall recall and was not as effective at correctly predicting cancer as the T2WI models. Delayed postcontrast-enhanced axial T1-weighted fat-saturated images have been reported to show mild enhancement of internal nodular components and irregular internal walls (15). Therefore, it has been suggested that T1-weighted images are less suitable for radiomics analysis of overall volume and shape than T2-weighted images since only features of the surfaces showing contrast enhancement were extracted, which could explain the present results of the AUC for the prediction model.

The drawback of machine learning models is that it is difficult to interpret their predictive results, which hinders their use by clinicians. To date, SHAP has been used to identify patterns identified by complex machine learning algorithms that can differentiate adult pilocytic astrocytomas from high-grade gliomas and reveal early progression of nonmetastatic nasopharyngeal carcinoma and treatment response to whole-brain radiotherapy (32-34). In the present study, SHAP indicated that with the T2WI-based model, lesions with a complex signal...
intensity, a large width in the direction of the coronal section, and a more spherical shape could be interpreted as being more likely to be cancer. Complex signal intensity is a characteristic MRI finding of cancer, and a large width might be a characteristic advanced cancer or more complex fistula. Furthermore, T2-weighted images of mucinous adenocarcinoma have been reported to show lobular fluid accumulation similar to mosaic (15). The SHAP result that it was more nearly spherical may represent this mucinous lobular morphology. In fact, 50% of the histological types of cancer cases in the present results were mucinous adenocarcinoma.

Regarding the deep learning model, R(2+1)D had a considerably lower AUC than that of the radiomics models, and there was little effect of data augmentation.
the prediction may differ depending on the histological type.

Third, the lobular pattern is seen in mucinous carcinoma but not in squamous cell carcinoma, so the accuracy of the prediction results, and when combined with SHAP, the extracted radiomic features could be explained more effectively.

Furthermore, the model may not be suitable for detecting early-stage cancers, as it did not correctly predict patients with stage 1 cancer. However, good predictive results were obtained for stage 2 cancer patients, which may reduce total pelvic exenteration and missed surgical opportunities. Fourth, this study did not include genetic data. TP53 has been reported to be the most frequently mutated gene in CD-CRC (37). However, a previous multicenter study in which we participated failed to detect mutations in known oncogenes in half of the CD-CRC patients (38). In Japan, the most common types of CD-CRC are mucinous carcinoma of the anal canal and anal fistula, and unknown genomic alterations specific to Asia may affect this unique CD-CRC phenotype. Overall, there appears to be heterogeneity in the genetic events of CRC depending on background disease, histology, and ethnicity.

In the future, our findings should be confirmed by multicenter clinical trials with larger sample sizes. In addition, an automatic segmentation method needs to be developed. The proposed method requires image segmentation performed by a radiologist who is familiar with cancer associated with CD to produce predictive results. However, if an automatic segmentation method can be developed, it will be possible to make predictions without a radiologist. One method of automatic segmentation utilizes deep learning. By inputting images into a model that has been pre-trained on numerous medical images, such as U-Net, automatic segmentation results can be generated (39). However, accurate results cannot be obtained when the boundary between a malignant tumor and normal tissue is ambiguous, so it is necessary to develop a method to address this problem.

The MRI radiomics-based RGF model had an excellent ability to predict CD-related anorectal cancer. By applying SHAP, the extracted radiomic features could be explained clinically and radiologically. These results may affect the diagnosis and surveillance strategies of CD-CRC.

**Acknowledgements**

Not applicable.

**Funding**

No funding was received.

**Availability of data and materials**

The data generated in the present study may be requested from the corresponding author.

**Authors' contributions**

YH conceived and designed the study, acquired, analyzed and interpreted the data and drafted the manuscript. JI, KM and SO conceived and designed the study and acquired and analyzed the data. KN, KKu, RK, KKi, KKa and NB acquired and interpreted the data. MU, MI, TO, KY, HI analyzed and interpreted the data, critically revised the manuscript for important intellectual content and gave the final approval for publication.
the authenticity of all the raw data. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All study protocols were approved by the institutional review board at Hyogo Medical University (no. 4389), and written informed consent and agreement for the use of patient data were obtained before surgery.

Patient consent for publication

Written informed consent was obtained via an opt-out method.

Competing interests

The authors declare that they have no competing interests.

References