Radiomics analysis of single-phase contrast-enhanced computed tomography for classification of hepatic focal lesions in colorectal cancer patients

Heejin Bae
Department of Medicine
The Graduate School, Yonsei University

Radiomics analysis of single-phase contrast-enhanced computed tomography for classification of hepatic focal lesions in colorectal cancer patients

Heejin Bae
Department of Medicine
The Graduate School, Yonsei University

Radiomics analysis of single-phase contrast-enhanced computed tomography for classification of hepatic focal lesions in colorectal cancer patients

Directed by Professor Joon Seok Lim

The Doctoral Dissertation
submitted to the Department of Medicine
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Heejin Bae

December 2020

This certifies that the Doctoral Dissertation of Heejin Bae is approved.

Thesis Supervisor: Joon Seok Lim

Thesis Committee Member#1: Sang Joon Shin

Thesis Committee Member#2: Helen Hong

Thesis Committee Member#3: Byung Soh Min

Thesis Committee Member#4: Jaemoon Yang

The Graduate School
Yonsei University

December 2020

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude and appreciation to my supervisor, Dr. Joon Seok Lim whose consistent guidance, support, and immense knowledge has been invaluable throughout this research project. I also wish to thank the rest of the research team, Dr. Helen Hong, Dr. Sungwon Kim, Dr. Han Sang Lee, and Dr. Kyunghwa Han, for their collaborative effort and valuable advice regarding this project.

Last but not least, I am incredibly grateful to my parents, my brother, my friends, and my lovely cat for their unconditional support throughout the completion of this dissertation and my life in general. Without their tremendous understanding and encouragement, it would be impossible for me to complete my study.

<TABLE OF CONTENTS>

LIST OF FIGURES

LIST OF TABLES

Abstract

Radiomics analysis of single-phase contrast-enhanced CT for classification of hepatic focal lesions in colorectal cancer patients

Heejin Bae

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Joon Seok Lim)

Objective: To evaluate the diagnostic performance of a radiomics model for classifying hepatic cysts, hemangiomas, and metastases in patients with colorectal cancer (CRC) from portal-phase abdominopelvic computed tomography (CT) images.

Methods: This retrospective study included 502 patients with CRC who underwent both contrast-enhanced abdominopelvic CT and contrast-enhanced liver magnetic resonance imaging between January 2005 and December 2010. Patients were divided into training (n = 386) and validation (n = 116) cohorts. Portal-phase contrast-enhanced CT images of 1290 liver lesions (size range, 3 mm–5 cm) were used to develop a radiomics model for differentiating three classes of liver lesions (cyst, hemangioma, and metastasis). Among multiple handcrafted features, the feature selection was performed using the ReliefF method, and random forest classifiers were used to train the selected features. The diagnostic performance of the developed model was evaluated and compared with that of four radiologists who classified liver lesions from the validation cohort (128 cysts, 30 hemangiomas, and 149 metastases). Additionally, a subgroup analysis was conducted based on lesion size (<10 mm or ≥10 mm).

Results: The radiomics model demonstrated significantly lower overall and hemangioma- and metastasis-specific polytomous discrimination index (PDI) (overall PDI, 0.8037; hemangioma-specific PDI, 0.6653; metastasis-specific PDI, 0.8027) than the radiologists' results, except that of the least-experienced radiologist (overall PDI, 0.9622–0.9680; hemangioma-specific PDI, 0.9452–0.9630; metastasis-specific PDI, 0.9511–0.9869). For differentiating subcentimeter lesions, the PDI of the radiomics model was different according to the lesion size (overall PDI of < 10 mm, 0.6486; overall PDI of ≥ 10 mm, 0.8264; p-value, 0.0692) while that of the radiologists was relatively maintained. For classifying benign lesions from metastasis, the radiomics model showed excellent diagnostic performance, with an accuracy of 84.36% (78.59–88.8) and an area under the receiver operating characteristic curve of 0.9426 (0.9149–0.9703). However, the three most experienced radiologists outperformed the radiomics model with an accuracy of 93.81–96.09% (p-value, 0.002–0.003).

Conclusion: The radiomics model achieved diagnostic accuracy comparable to that of radiologists when differentiating cysts, hemangiomas, and metastases from portal-phase CT images of patients with CRC and demonstrated potential for clinical use. However, this model was limited particularly to classifying hemangiomas and subcentimeter liver lesions, and therefore, unattended application of the system in daily clinical practice is not yet feasible.

Key Words: radiomics, computed tomography, colorectal cancer, liver metastasis, hepatic cyst, hemangioma, classification, diagnostic accuracy

Radiomics analysis of single-phase contrast-enhanced CT for classification of hepatic focal lesions in colorectal cancer patients

Heejin Bae

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Joon Seok Lim)

I. Introduction

Liver is the most common organ to metastasize in patients with colorectal cancer (CRC). Approximately 15% of these patients present with synchronous liver metastasis during diagnostic workup, and more than 50% will eventually develop liver metastasis, with a 5-year survival rate of less than 5% in cases of untreated liver metastasis[1-3]. With colorectal liver metastasis (CRLM) as an important prognostic factor, surgical resection remains a possible curative treatment with a reported 5-year overall survival rate as high as 58%, which is a better long-term survival benefit compared with palliative chemotherapy[4,5]. Therefore, accurate staging and prompt detection and characterization of liver lesions through imaging evaluations are essential for the selection of proper disease management. Primarily, differentiation of commonly encountered benign lesions such as hepatic cysts and hemangiomas is significant for the diagnosis of CRLM.

Radiomics is a rapidly emerging technique that extracts various qualitative data from medical images and uses them to support the detection and diagnosis of disease[6,7]. The development of a computer-aided classification system of liver lesions, particularly in cancer patients, is necessary to improve diagnostic accuracy and efficiency in daily clinical practice. The exclusion of benign liver lesions such

as cysts or hemangiomas would especially increase the work speed and efficiency of radiologists in reading initial and follow-up evaluations of cancer patients, considering that the use of computed tomography (CT) and the number of images acquired from a single examination are constantly increasing[8,9]. To date, several radiomics analyses have been conducted to classify liver lesions as benign or malignant based on various imaging modalities, and have achieved diagnostic accuracy comparable to radiologists[10-13]. However, due to the retrospective nature of these studies, researchers tended to adopt convenience sampling, the data collection method of sampling control and diseased subjects in a calculated ratio that can cause spectrum bias[14]. To achieve generalizability of the radiomics model, the study cohort should reflect the target population in terms of disease characteristics such as incidence or prevalence[14,15]. To overcome this sampling bias, we intended to register all patients and their liver lesions during a specific period.

Moreover, before the radiomics technique can be used in the clinical setting, its diagnostic performance must be verified in a setting similar to daily practice, wherein single-phase contrast-enhanced CT images are mainly performed for follow-up evaluations of cancer patients and radiologists review not only the axial but also coronal reformatted images. In practice, various imaging modalities, including ultrasonography, CT, gadolinium-enhanced magnetic resonance imaging (MRI), and fluorine-18 fluorodeoxyglucose (FDG) positron emission tomography fused with CT (PET/CT), are used for the initial diagnostic workup of CRC and detection of CRLM. Previous studies have shown that regardless of whether a patient received previous therapy, MRI is superior to CT and FDG PET/CT in detecting and characterizing hepatic lesions[16-21]. Particularly, the characterization of small hepatic lesions less than 10 mm in size can be challenging for CT due to volume averaging. Thus, further evaluation, including MRI, is routinely required to increase diagnostic confidence in identifying small indeterminate lesions with CT and to detect additional CRLM[22,23]. Despite the higher diagnostic accuracy of MRI, the accurate characterization and diagnosis of liver lesions on CT are significant

because CT is most commonly performed for both initial and post-treatment evaluation in patients with CRC due to its wide availability and cost-effectiveness. To the best of our knowledge, there is a lack of studies that have developed a radiomics model and compared its diagnostic performance with that of radiologists under circumstances similar to that of the clinical setting.

Therefore, in this study, we aimed to evaluate the performance of the radiomics model for the classification of CRLM, hepatic cysts, and hemangiomas using portal-phase abdominopelvic CT images, and compare its diagnostic accuracy with that of radiologists.

II. Materials and Methods

1. Study population

This retrospective study was approved by the institutional review board at our institution, and the requirement for informed consent was waived. A flow diagram of patient selection is shown in Figure 1.

A total of 676 CRC patients, who underwent both contrast-enhanced abdominopelvic CT and contrast-enhanced liver MRI between January 2005 and December 2010 were retrospectively registered in this study. Of these, 21 patients were excluded due to poor image quality (i.e., abdominopelvic CT without portal-phase or slice thickness larger than 5 mm). The following inclusion criteria were used for hepatic lesions in the study: (1) simple cyst, hemangioma, or metastasis larger than 3 mm and smaller than 5 cm; (2) in the case of multiple lesions, a maximum of five lesions per each lesion category (as many as 15 lesions per patient). Of the remaining 655 patients, 153 patients did not meet the inclusion criteria and were excluded for the following reasons: (1) had only a benign hepatic lesion other than a cyst or hemangioma (n = 52); (2) had a focal lesion, which was detected only on either of the imaging modalities, CT or MRI (n = 46); (3) had only a hepatic lesion smaller than 3 mm or larger than 5 cm (n = 44); (4) had a recurrent lesion after radiofrequency ablation (n = 6); (5) had other type of malignant tumor

(n = 2); (6) had no available reference standard for the hepatic lesion (n = 2); and (7) had images that revealed the inaccurate extent of the hepatic lesion due to venous thrombosis (n = 1). Subsequently, 502 CRC patients with 676 cysts, 130 hemangiomas, and 484 metastases were included in this study. In the final study population, training and validation cohorts were divided chronologically to assign 100 hemangiomas to a training cohort. Patients who underwent liver MRI after August 23, 2006 were included in the training cohort (n = 386; 214 men, 172 women; mean age, 60.8 ± 11.6 years) and those who received liver MRI before August 23, 2006 were allocated to the validation cohort (n = 116; 71 men, 45 women; mean age, 59.4 ± 10.3 years). The training set consisted of 548 cysts, 100 hemangiomas, and 335 metastases, and the validation set included 128 cysts, 30 hemangiomas, and 149 metastases.

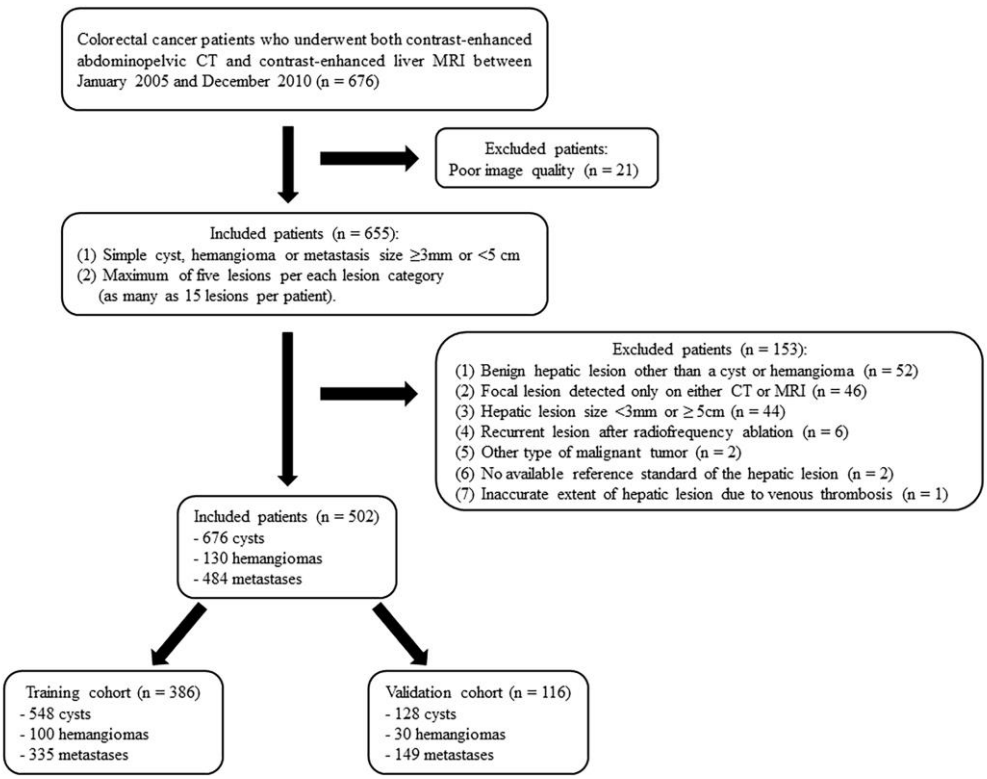

Figure 1. Flow diagram of patient selection.

Abbreviations: CT, computed tomography; MRI, magnetic resonance imaging.

2. Image acquisition

At our institution, single-phase contrast-enhanced CT with portal phase images was routinely performed to detect metastasis in both initial and follow-up evaluations of cancer patients. Following CT scans, liver MRI was usually performed to accurately evaluate the extent of metastases and to characterize the focal lesion when an indeterminate lesion was found on CT.

Various types of multidetector CT (MDCT) scanners were used in this study: a 4-channel MDCT scanner (Lightspeed Plus, GE Healthcare, Milwaukee, WI, USA), 16-channel MDCT scanners (Sensation 16, Siemens Healthcare, Erlangen, Germany; LightSpeed VCT, GE Healthcare, Milwaukee, WI, USA), and 64-channel MDCT scanners (Sensation 64, Somatom Definition Flash, Siemens Healthcare, Erlangen, Germany). The following CT parameters were used: tube voltage, 120–140 kVp; tube current, 100–280 mAs; beam pitch, 0.6–1.07; and reconstruction thickness, 3–5 mm. Portal-phase abdominopelvic CT images were acquired from the lower thorax to the lower pelvis approximately 70 seconds after the intravenous administration of 2.0 mL/kg of nonionic contrast material (Iopromide, Ultravist 300, Schering, Berlin, Germany) via an antecubital vein (flow velocity, 3–4 mL/s; fixed duration, 30 seconds) followed by a 20-mL saline bolus injection. The acquired images were reformatted in the coronal plane, and both axial and coronal image sets were transferred to our Picture Archiving and Communication System (PACS) for interpretation.

All liver MRI scans included in this study were obtained using a 1.5-T (Achieva 1.5-T, Philips Healthcare, Best, the Netherlands) or 3.0-T (Magnetom Trio a Tim, Siemens Healthcare, Forchheim, Germany; Achieva, Philips Healthcare, Best, the Netherlands) magnetic resonance scanners. Magnetic resonance sequences of both MRI with extracellular contrast agent (ECA-MRI) and MRI with hepatobiliary agent were as follows: dual-echo in-phase and opposed-phase spoiled gradient-echo T1-weighted images; multi-shot and single-shot turbo spin-echo T2-weighted images; and dynamic fat-suppressed spoiled gradient-echo T1-weighted images

conducted before and after the contrast agent injection, including arterial, portal venous, 3-minute delayed, and 5-minute delayed phases. Additionally, for dynamic scans with gadoxetic acid disodium (Primovist, Bayer Schering Pharma, Berlin, Germany), hepatobiliary phase images were obtained 15 or 20 minutes after 0.1 mL/kg (0.025 mmol/kg) contrast agent administration. For ECA-MRI, different types of gadolinium-based contrast agents including gadopentetate dimeglumine (Magnevist, Bayer Schering Pharma, Berlin, Germany) or gadoterate meglumine (Dotarem, Guerbet, Roissy, France) were injected intravenously with a volume of 0.1 mmol/kg.

3. Hepatic lesion confirmation

Hepatic lesions were pathologically confirmed by either surgery or percutaneous biopsy. If pathological results were not obtainable, typical MRI imaging findings of a cyst, hemangioma, or metastasis were used to characterize the focal hepatic lesion. For patients with focal hepatic lesions without pathological results and confirmatory imaging studies, the characterization of the lesion was determined through follow-up imaging for at least 1 year. An increase or decrease in lesion size during chemotherapy was considered as metastasis. On the other hand, lesions showing stability for at least 1 year were classified as benign.

4. Image analysis

Two board-certified abdominal radiologists and two radiology residents (second and fourth years of residency) who were unaware of outcomes of hepatic lesions independently reviewed the axial and coronal portal-phase images of each lesion in the validation cohort. Before the image review, each lesion was individually annotated by another radiologist on axial portal-phase images to indicate the location of the lesion. Radiologists evaluated each lesion and rated the probability of it being a cyst, hemangioma, or metastasis using a 6-point numerical rating scale with scores of 0, 2, 4, 6, 8, and 10 and the sum of the scale as 10.

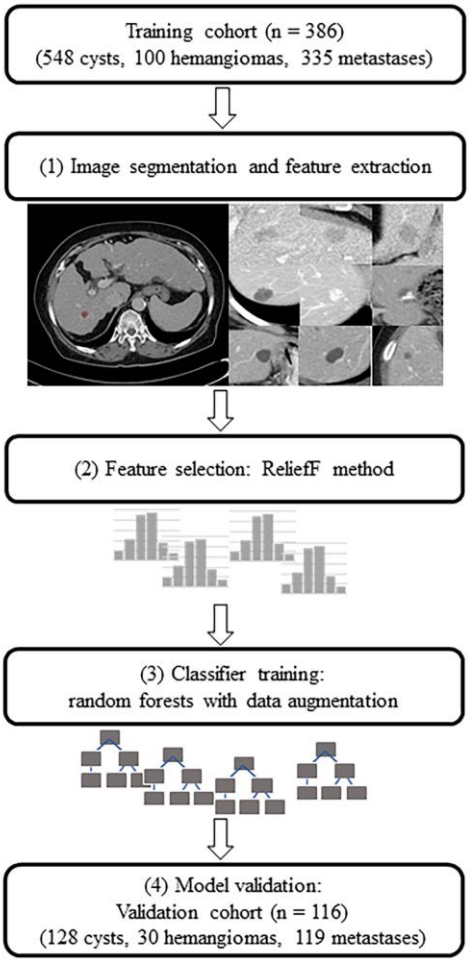

Figure 2. A flowchart of the radiomics model.

5. Radiomics feature analysis

A. Image segmentation

A flowchart of the radiomics analysis is summarized in Figure 2. From our PACS, axial portal-phase CT images were extracted in a Digital Imaging and Communications in Medicine (DICOM) format. One abdominal radiologist semiautomatically drew a region of interest (ROI) bounding the largest cross-sectional area of the hepatic lesion, using the medical image processing,

analysis, and visualization (MIPAV) software developed by the Center for Information Technology at the National Institutes of Health[24]. Another board-certified abdominal radiologist independently segmented the lesions of randomly selected 50 patients with 130 lesions to evaluate interobserver reproducibility. Among the extracted radiomics features, features with excellent interobserver reproducibility (i.e., intraclass correlation coefficient > 0.75)[25] were adopted for analysis.

B. Feature extraction

In feature extraction, the handcrafted features[26,27] representing the imaging characteristics of lesions, such as texture and shape, were computed. For each lesion, 119 dimensional texture features and 10 dimensional shape features were extracted from the segmented region. In texture features, the following features were included: 7 features of histogram characteristics, 6 features of the percentage of pixels below the thresholds, 6 features of the percentage of pixels above the thresholds, 5 features of histogram, 14 features of a gray-level co-occurrence matrix (GLCM), 22 features of a gray-level run-length matrix (GLRLM), and 59 features of local binary patterns (LBP). In shape features, size-related and roundness-related features such as area/perimeter ratio, eccentricity, or major axis length were included. A complete list of extracted features is summarized in Table 1.

Table 1. A complete list of hand-crafted features extracted from lesions

Categories (No. of features): Features (feature number)
Texture features (119)
Histogram features (7): Histogram mean (1), standard deviation (2), minimum (3) and maximum (4) intensities, skewness (5), kurtosis (6), and entropy (7)

GLRLM features (22):
Four direction mean and standard deviation of short run emphasis (39,40), long run emphasis (41,42), gray-level non-uniformity (43,44), run length non-uniformity (45,46), run percentage (47,48), low gray-level run emphasis (49,50), high gray-level run emphasis (51,52), short run low gray-level emphasis (53,54), short run high gray-level emphasis (55,56), long run low gray-level emphasis (57,58), long run high gray-level emphasis (59,60)
Percentages of pixels below the thresholds (6):
0 HU (8), 30 HU (9), 60 HU (10), 90 HU (11), 120 HU (12), 150 HU (13)
GLCM features (14):
Four direction mean and standard deviation of angular second moment (25,26), contrast (27,28), sum average (29,30), sum variance (31,32), sum entropy (33,34), entropy (35,36), and difference entropy (37,38)
Percentages of pixels above the thresholds (6):
150 HU (14), 170 HU (15), 190 HU (16), 210 HU (17), 230 HU (18), 250 HU (19)
LBP features (59):
10 uniform patterns in LBP histogram (61-119)
Percentile intensities at (5):
5% (20), 25% (21), 50% (22), 75% (23), 95% (24)
Shape features (10):
Area/perimeter ratio (120), convex area (121), eccentricity (122), Euler number (123), solidity (124), major-minor axis length ratio (125), major axis length (126), minor axis length (127), area (128), perimeter (129)

Abbreviations: GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run-length matrix; HU, Hounsfield unit; LBP, local binary patterns.

C. Feature selection and classifier training

Among the extracted 129 features, only 126 features showed excellent interobserver reproducibility, and therefore, were recruited for the analysis. Three excluded features were as follows: the GLRLM short-run low gray-level emphasis standard deviation, shape eccentricity, and shape major-minor axis length ratio. Using the ReliefF method[28], we selected 60 features in the order of high ReliefF scores (Fig. 3), which reflected the weights related to feature importance. The ReliefF method computed feature weights through iteratively updating them while emphasizing interclass margins and penalizing intraclass margins[26]. The random forest (RF) classifier[29] was trained on the selected features to identify the lesions as part of a cyst, hemangioma, or metastasis. For the classifier training, the number of decision trees for RF was experimentally determined and set to 100. To avoid overfitting and class imbalance, data augmentation was performed to increase the number of training data to 5000 for each class using random scaling and random angle rotation on the training images (Fig. 4).

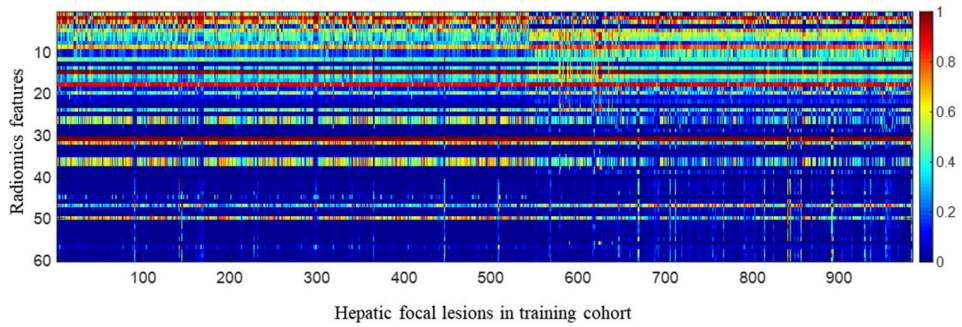

Figure 3. Heatmap of hepatic focal lesions in the training cohort. Each row represented hepatic focal lesions in the training cohort: cysts, 1-548; hemangiomas, 549-648; and metastases, 649-983. Each column displayed 60 different radiomics features in descending order by ReliefF scores. This heatmap demonstrated that uppermost radiomics features having high ReliefF scores could classify three categories of liver lesions.

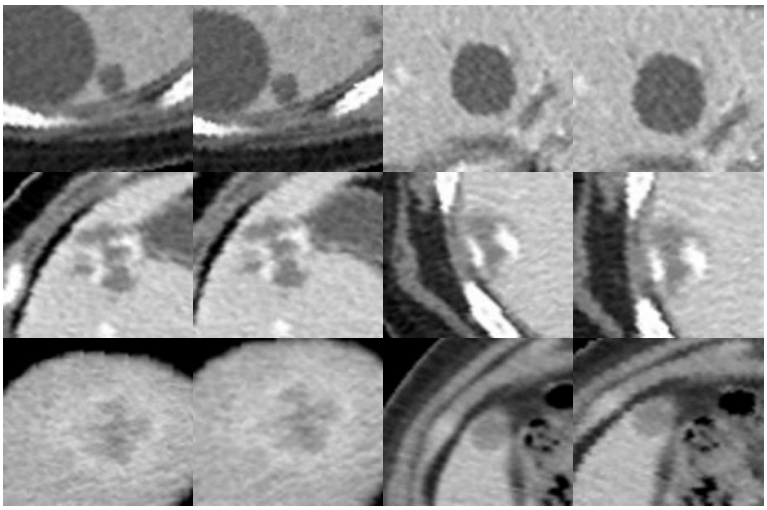

Figure 4. Representative images of augmented data of a cyst (top), hemangioma (middle), and metastasis (bottom). Using random scaling and random angle rotation, training data were augmented to avoid overfitting and class imbalance.

D. Performance evaluation

After the classifier training, we examined the accuracy of the trained model by distinguishing three categories of liver lesions in the validation cohort (Fig. 5). The model provided numbers that can be interpreted as the probability for each lesion category. The category with the highest output value was chosen as the diagnosis made by the radiomics model.

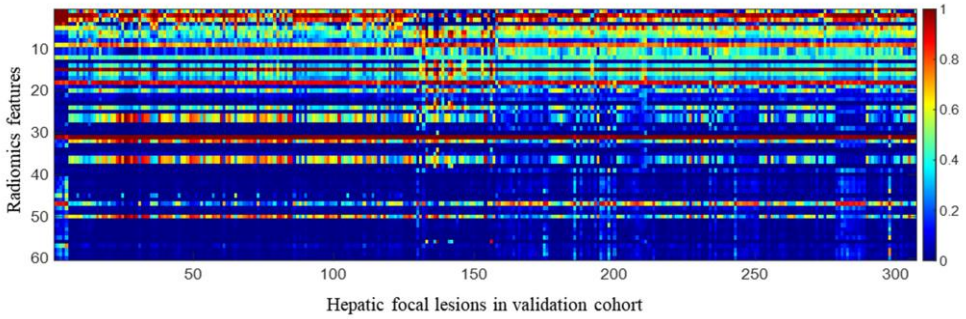

Figure 5. Heatmap of hepatic focal lesions in the validation cohort. Each row

represented hepatic focal lesions in the validation cohort: cysts, 1-128; hemangiomas, 129-158; and metastases, 159-307. Each column displayed 60 different radiomics features in descending order by ReliefF scores. Similar to the heatmap of the training cohort, that of the validation cohort revealed that uppermost radiomics features with high ReliefF scores could classify three categories of liver lesions.

6. Statistical analysis

The polytomous discrimination index (PDI)[30] and correct classification percentage (CCP)[31] were calculated to compare diagnostic performances in the classification of three classes of liver lesions. Both overall PDI and category-specific PDI were calculated and compared. In addition, 95% confidence intervals and p-values were computed using an adjusted bootstrap-corrected method with 1000 resampling.

For the two-class comparison of benign (hepatic cysts and hemangiomas) and metastatic lesions, the lesion was diagnosed as benign when the sum of the scale that radiologists assigned to the hepatic cyst and hemangioma exceeded 5 and when the sum of the number that the radiomics model assigned to the hepatic cyst and hemangioma was more than 0.5. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the accuracy of diagnosing liver metastases were analyzed using logistic regression in a generalized estimating equations model. Delong's method was used to calculate the area under the receiver operating characteristic curve (AUC).

Additionally, we conducted a subgroup analysis based on the lesion size (<10 mm group; ≥10 mm group). All statistical analyses were performed using the R software (version 3.6.2; R Foundation for Statistical Computing, Vienna, Austria). A p-value less than 0.05 was considered statistically significant. For a comparison between the radiomics model and average values of radiologists, a p-value less than 0.025 was set to be significant after applying Bonferroni correction.

III. Results

1. Patient characteristics

The patient characteristics of the training and validation cohorts are summarized in Table 2. Demographic characteristics among patients were not significantly different between the two cohorts. Mean size of the liver lesions in validation image sets was 16.0 ± 8.79 mm (cysts, 11.5 ± 6.50 mm; hemangiomas, 14.2 ± 6.52 mm; metastasis, 20.1 ± 8.95 mm). Image sets in the validation cohort were divided based on the size for subgroup analysis. The <10 mm group had 92 lesions with 76 cysts, 7 hemangiomas, and 9 metastases. The ≥10 mm group had 215 lesions, including 52 cysts, 23 hemangiomas, and 140 metastases.

Table 2. Patient characteristics in the training and validation cohorts

Parameter	Training (n = 386)	Validation (n = 116)	p-value
Mean age (y)	60.8 ± 11.6	59.4 ± 10.3	0.222
Sex			0.272
Men (%)	214/386 (55.4)	71/116 (61.2)	
Women (%)	172/386 (44.6)	45/116 (12.9)	
CEA level (ng/mL)	56.7 ± 642.1	20.9 ± 61.1	0.282
Primary cancer location			0.826
Colon (%)	234/386 (60.6)	69/116 (59.5)	
Rectum (%)	152/386 (39.4)	47/116 (40.5)	
Primary cancer T stage[1]			0.574
T1 (%)	18/383 (4.7)	6/113 (5.3)	
T2 (%)	34/383 (8.9)	11/113 (9.7)	
T3 (%)	266/383 (69.5)	83/113 (73.5)	
T4 (%)	65/383 (17.0)	13/113 (11.5)	
Primary cancer N stage[1]			0.173

N0 (%)	167/383 (43.6)	43/113 (38.1)
N1 (%)	134/383 (35.0)	36/113 (31.9)
N2 (%)	82/383 (21.4)	34/113 (30.1)

Abbreviations: CEA, carcinoembryonic antigen.

[1]Primary cancer stages of three patients in the validation cohort and three patients in the training cohort were unknown because they received colorectal surgeries at other institutions.

2. Diagnostic performance of three-class classification

PDI is the probability of correctly classifying a case within a set of cases[30]. Both overall PDI and category-specific PDI were analyzed to compare the ability of the model and radiologists to classify three categories of the liver lesions (Table 3, Fig. 6). Reviewer 1 and Reviewer 2 were board-certified abdominal radiologists, whereas Reviewer 3 and Reviewer 4 were fourth- and second-year residents, respectively. All reviewers except for the least-experienced radiologist demonstrated significantly higher overall (range, 0.9622–0.9680) and hemangioma- and metastasis-specific PDIs (hemangioma-specific PDI, 0.9452–0.9630; metastasis-specific PDI, 0.9511–0.9869) than the radiomics model (overall PDI, 0.8037; hemangioma-specific PDI, 0.6653; metastasis-specific PDI, 0.8027) with p-values ranging from <0.001 to 0.001. Particularly, the model showed the lowest category-specific PDI for hemangiomas (0.6653) among its category-specific PDIs; however, the radiologists maintained high category-specific PDIs regardless of the lesion type. For cysts, a statistically significant difference in category-specific PDI was observed between Reviewer 1 and the radiomics model (0.9889 and 0.9432, respectively; p-value, 0.007). Albeit statistically insignificant, Reviewer 2 and Reviewer 3 showed higher cyst-specific PDI (0.9741 and 0.9605, respectively) than the radiomics model (p-value, 0.103 and 0.385). Although the least-experienced reviewer did not show a significant difference, the overall and category-specific PDIs (overall PDI, 0.8458; cyst-specific PDI, 0.9478; hemangioma-specific PDI,

0.7305; metastasis-specific PDI, 0.8589) were numerically higher than those of the radiomics model.

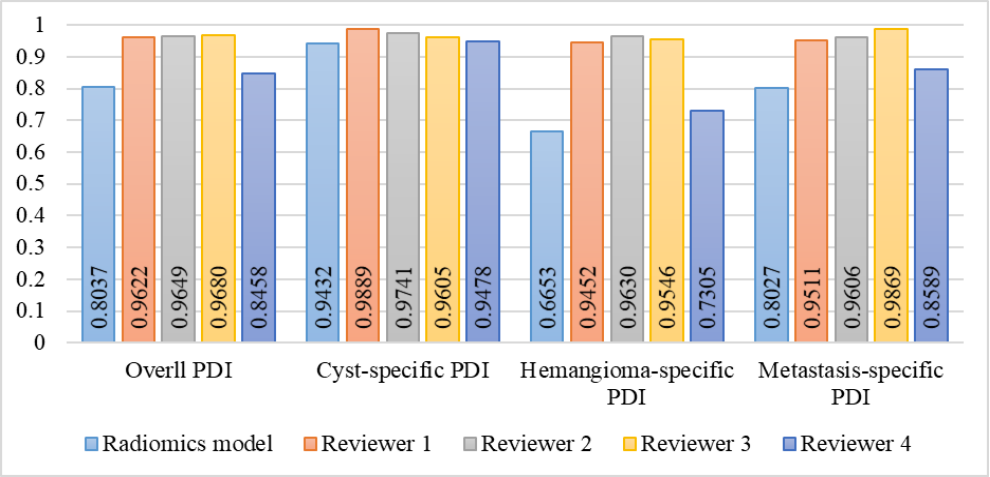

Figure 6. Comparison of the diagnostic performance between the radiomics model and radiologists measured by polytomous discrimination index (PDI). Reviewers 1 and 2 were board-certified abdominal radiologists, whereas Reviewers 3 and 4 were fourth- and second-year resident radiologists, respectively. All reviewers demonstrated higher overall and category-specific PDI than the radiomics model. The hemangioma-specific PDI of the radiomics model was the lowest among all values.

Table 3. Diagnostic performance to classify hepatic lesions as measured by polytomous discrimination index (PDI).

PDI	**Radiomics model**	**Radiologist (average)**	**Reviewer 1[1]**	**Reviewer 2[1]**
Overall (95% CI)	0.8037 (0.7291–0.8653)	0.9635 (0.9374–0.9799)	0.9622 (0.9268–0.9816)	0.9649 (0.9055–0.9846)
p-value	NA	<0.001	<0.001	<0.001
Cyst	0.9432	0.9815	0.9889	0.9741
p-value	NA	0.023	0.007	0.103
Hemangioma	0.6653	0.9541	0.9452	0.9630
p-value	NA	<0.001	<0.001	<0.001
Metastasis	0.8027	0.9558	0.9511	0.9606
p-value	NA	<0.001	<0.001	0.001
PDI	**Resident (average)**	**Reviewer 3[2]**	**Reviewer 4[2]**	
Overall (95% CI)	0.9069 (0.8731–0.9374)	0.9680 (0.9294-0.983)	0.8458 (0.778-0.902)	
p-value	0.003	<0.001	0.305	
Cyst	0.9541	0.9605	0.9478	
p-value	0.549	0.385	0.891	
Hemangioma	0.8425	0.9546	0.7305	
p-value	0.010	<0.001	0.418	
Metastasis	0.9229	0.9869	0.8589	
p-value	0.006	<0.001	0.253	

Abbreviations: CI, confidence interval; NA, nonapplicable.

[1]Reviewer 1 and Reviewer 2 were board-certified abdominal radiologists.

[2]Reviewer 3 was a fourth-year resident, and Reviewer 4 was a second-year resident.

Additionally, in the subgroup analysis (Table 4, Fig. 7), all the reviewers displayed consistently good performance in classifying the hepatic lesions regardless of lesion size, whereas the radiomics model showed numerically lower classification performance when evaluating subcentimeter lesions (overall PDI of <10mm group, 0.6486; overall PDI of ≥10 mm group, 0.8264; p-value, 0.069). Radiologists demonstrated overall PDIs of <10mm group ranging from 0.8855 to 0.9667, which was significantly higher than those of the radiomics model (p-value, <0.001–0.042). For ≥10 mm group, Reviewers 1, 2, and 3 showed significantly higher overall PDIs (0.9562, 0.9454 and 0.9621, respectively) than those of the radiomics model (p-value, 0.001–0.009). However, the least-experienced reviewer's overall PDI for classifying lesions ≥ 10 mm (0.8271; p-value, 0.993) was similar to that of the radiomics model.

All the reviewers demonstrated higher category-specific PDIs when assessing <10mm group (cyst-specific PDI, 0.9175–0.9910; hemangioma-specific PDI, 0.7898–0.9953; metastasis-specific PDI, 0.9131–0.9715) than the radiomics model (cyst-specific PDI, 0.8704; hemangioma-specific PDI, 0.4599; metastasis-specific PDI, 0.6155); however, statistically significant differences were only observed for cyst- and hemangioma-specific PDIs of Reviewers 1, 2, and 3 (p-value of cyst-specific PDI, 0.003-0.018; p-value of hemangioma-specific PDI, 0.002-0.006) and metastasis-specific PDIs of Reviewers 2 and 3 (p-value, 0.015-0.028). For ≥10 mm group, category-specific PDIs of reviewers, except for the least-experienced radiologist (cyst-specific PDI, 0.9421–0.9847; hemangioma-specific PDI, 0.9311–0.9505; metastasis-specific PDI, 0.9478–0.9872), were higher than those of the radiomics model (cyst-specific PDI, 0.9552; hemangioma-specific PDI, 0.7274; metastasis-specific PDI, 0.7966) with significant differences in hemangioma- and metastasis-specific PDIs (p-value of hemangioma-specific PDI, 0.004–0.006; p-value of metastasis-specific PDI, <0.001–0.011). Noticeably, in <10mm group, board-certified radiologists demonstrated significantly better performance in classifying benign lesions that

were cysts or hemangiomas (radiologist-averaged cyst-specific PDI, 0.9876; p-value, 0.003; radiologist-averaged hemangioma-specific PDI, 0.9649; p-value, 0.003), whereas resident radiologists showed statistically significantly higher accuracy in differentiating metastasis (resident-averaged metastasis PDI, 0.9602; p-value, 0.018).

Corresponding to the sensitivity and specificity in binary outcome evaluation, the CCP is the proportion of the diagnostic results that correctly match the gold standard among three classes of outcomes (Table 5). Similar to the PDI, the CCP was significantly higher for the results of the three most experienced radiologists (Reviewer 1, 0.9316; Reviewer 2, 0.9446; Reviewer 3, 0.9283) than that of the radiomics model (CCP, 0.8436; p-value, <0.001–0.001). In the subgroup analysis, differences in CCP were observed between the two subgroups, but the difference was not significant between the radiomics model and radiologists, except in the case of the least-experienced radiologist. Throughout the subgroup analysis, reviewers displayed better diagnostic performance (CCP of <10mm group range, 0.913–0.9457; CCP of ≥10 mm group range, 0.8465–0.9395) than the radiomics model (CCP of <10mm group, 0.8587; CCP of ≥10 mm group, 0.8372). Statistically significant differences were observed between Reviewers 2, and 4 for <10mm group (p-value, 0.009-0.030) and between Reviewers 1, 2, and 3 for ≥ 10 mm group (p-value, <0.001-0.001).

Table 4. Subgroup analysis of the polytomous discrimination index (PDI) based on hepatic lesion size

PDI	**Radiomics model**		**Radiologist (average)**		**Reviewer 1[2]**		**Reviewer 2[2]**	
	<10 mm[1] (n = 92)	≥10 mm[1] (n = 215)	<10 mm[1] (n = 92)	≥10 mm[1] (n = 215)	<10 mm[1] (n = 92)	≥10 mm[1] (n = 215)	<10 mm[1] (n = 92)	≥10 mm[1] (n = 215)
Overall	0.6486	0.8264	0.9565	0.9508	0.9462	0.9562	0.9667	0.9454
p-value	NA	NA	0.001	0.002	0.002	0.001	<0.001	0.009
Cyst	0.8704	0.9552	0.9876	0.9634	0.9910	0.9847	0.9841	0.9421
p-value	NA	NA	0.003	0.719	0.003	0.111	0.003	0.627
Hemangioma	0.4599	0.7274	0.9649	0.9399	0.9345	0.9311	0.9953	0.9487
p-value	NA	NA	0.003	0.003	0.005	0.004	0.002	0.006
Metastasis	0.6155	0.7966	0.9169	0.9493	0.9131	0.9507	0.9206	0.9478
p-value	NA	NA	0.051	0.003	0.107	0.002	0.028	0.011
p-value[4]		0.069		0.598		0.867		0.789

PDI	**Resident (average)**		**Reviewer 3[3]**		**Reviewer 4[3]**	
	<10 mm[1] (n = 92)	≥10 mm[1] (n = 215)	<10 mm[1] (n = 92)	≥10 mm[1] (n = 215)	<10 mm[1] (n = 92)	≥10 mm[1] (n = 215)
Overall	0.9216	0.8946	0.958	0.9621	0.8855	0.8271
p-value	0.005	0.079	0.001	0.002	0.042	0.993
Cyst	0.9417	0.9406	0.9659	0.9500	0.9175	0.9312
p-value	0.080	0.513	0.018	0.817	0.442	0.355
Hemangioma	0.8632	0.8391	0.9366	0.9505	0.7898	0.7277
p-value	0.036	0.111	0.006	0.005	0.164	0.951
Metastasis	0.9602	0.906	0.9715	0.9872	0.949	0.8248
p-value	0.018	0.031	0.015	<0.001	0.026	0.618
p-value[4]		0.589		0.511		0.841

Abbreviations: CI, confidence interval; NA, nonapplicable.

[1]The validation set (n = 307) was divided into two subgroups based on the lesion size (<10 mm or ≥10 mm).

[2]Reviewer 1 and Reviewer 2 were board-certified abdominal radiologists.

[3]Reviewer 3 was a fourth-year resident, and Reviewer 4 was a second-year resident radiologist.

[4]P-values for differences between <10mm group and ≥10 mm group.

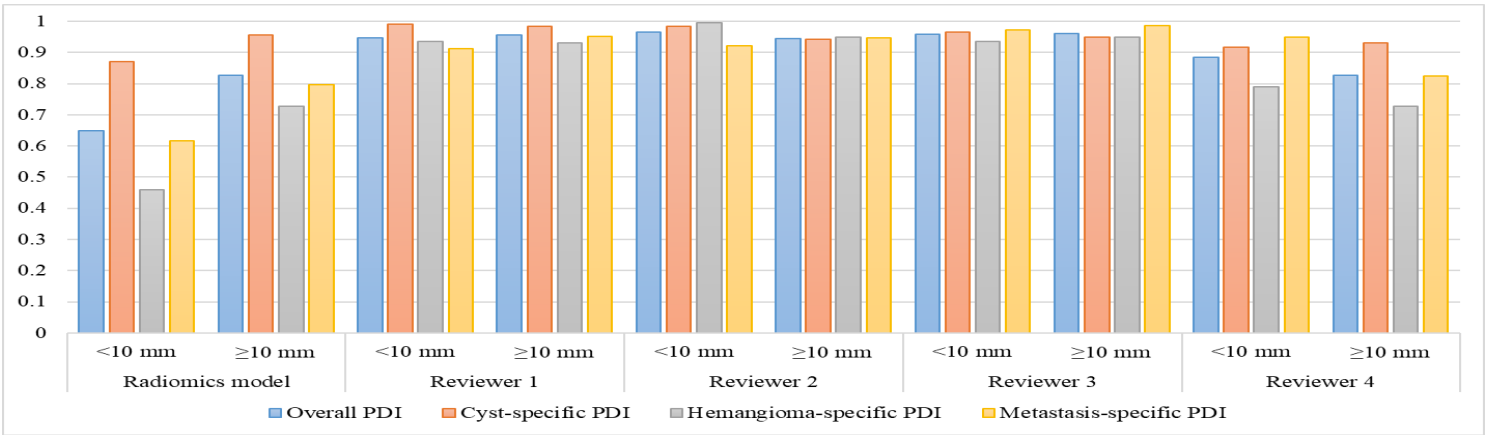

Figure 7. Comparisons of overall and category-specific polytomous discrimination index (PDI) between the radiomics model and radiologists classified by hepatic lesion size. Reviewers 1 and 2 were board-certified abdominal radiologists, whereas Reviewers 3 and 4 were fourth-year and second-year resident radiologists, respectively. All the reviewers, except for the least-experienced radiologist, displayed consistently good performance in classifying hepatic lesions, regardless of lesion size. However, the radiomics model demonstrated lower diagnostic performance for evaluating <10mm group than for classifying ≥ 10 mm group.

Table 5. Diagnostic performance to classify hepatic lesions as measured by correct classification percentage (CCP)

Subgroup[1]	**Radiomics model (95% CI)**	**Reviewer 1**[2] **(95% CI)**	**Reviewer 2**[2] **(95% CI)**	**Reviewer 3**[3] **(95% CI)**	**Reviewer 4**[3] **(95% CI)**
All (n = 307)	0.8436 (259/307) (0.8013–0.8827)	0.9316 (286/307) (0.9055–0.9642)	0.9446 (290/307) (0.9121–0.9642)	0.9283 (285/307) (0.9055–0.9593)	0.8697 (267/307) (0.8274–0.9023)
p-value	NA	<0.001	<0.001	0.001	0.271
<10 mm (n = 92)	0.8587 (79/92) (0.7717–0.913)	0.9348 (86/92) (0.8587–0.9674)	0.9457 (87/92) (0.8804–0.9674)	0.9130 (84/92) (0.8478–0.9733)	0.9348 (86/92) (0.8587–0.9674)
p-value	NA	0.050	0.009	0.285	0.030
≥ 10 mm (n = 215)	0.8372 (180/215) (0.788–0.8837)	0.9163 (197/215) (0.8681–0.9442)	0.9395 (202/215) (0.8977–0.9628)	0.9163 (197/215) (0.8698–0.9349)	0.8465 (182/215) (0.7953–0.893)
p-value	NA	0.001	<0.001	0.001	0.865
p-value[4]	0.612	0.581	0.684	0.488	0.010

Abbreviations: CI, confidence interval; NA, nonapplicable.

[1]The validation set was divided into two subgroups based on lesion size (<10 mm group or ≥10 mm group).

[2]Reviewer 1 and Reviewer 2 were board-certified abdominal radiologists.

[3]Reviewer 3 was a fourth-year resident, and Reviewer 4 was second-year resident.

[4]P-values for the differences between lesion size subgroups.

3. Diagnostic performance of binary classification

In addition to the three-class classification, the diagnostic performance of differentiating benign liver lesions (hepatic cysts and hemangiomas) from metastatic lesions was analyzed (Table 6). The radiomics model and all four radiologists showed high sensitivity (80.54%−95.97%) and specificity (86.71%−99.37%) for diagnosing liver metastases. Sensitivity, specificity, PPV, NPV, accuracy, and AUC of the radiomics model for diagnosing liver metastases were 80.54%, 87.97%, 86.33%, 82.74%, 84.36%, and 0.9426, respectively. Reviewers 1, 2, and 3 showed significantly higher accuracy (93.81–96.09%; p-value, 0.002–0.003) and AUC (0.9810–0.9928; p-value, <0.001–0.003) than those of the radiomics model. For a comparison between the radiomics model and the radiologists, a p-value less than 0.025 was set to be significant after applying Bonferroni correction. The average diagnostic performance of the board-certified radiologists was significantly higher than that of the radiomics model, with a sensitivity of 93.29% (p-value, 0.003), a specificity of 96.52% (p-value, 0.007), a PPV of 96.19% (p-value, 0.008), an NPV of 93.85% (p-value, 0.006), an accuracy of 94.95% (p-value, <0.001), and an AUC of 0.9814 (p-value, <0.001). Although the differences were not significant, the resident radiologists showed numerically higher sensitivity (90.60%; p-value, 0.030), specificity (93.04%; p-value, 0.187), PPV (92.47%; p-value, 0.166), and NPV (91.3%; p-value, 0.056) than those of the model. The less-experienced radiologists showed a significantly higher accuracy (91.86%; p-value, <0.001) and AUC (0.9623; p-value, <0.001) than those of the radiomics model. The fourth-year resident showed significantly higher specificity (p-value, 0.003), PPV (p-value, 0.005), accuracy (p-value, 0.026), and AUC (p-value, 0.001) than those of the second-year resident.

Table 6. Diagnostic performances of binary classification (hepatic cyst + hemangioma vs. metastasis)

Parameter	**Radiomics model**	**Radiologist (average)**	**Reviewer 1[1]**	**Reviewer 2[1]**
Sensitivity (%)	80.54	93.29	90.60	95.97
p-value	NA	0.003	0.047	0.012
Specificity (%)	87.97	96.52	96.84	96.20
p-value	NA	0.007	0.018	0.059
PPV (%)	86.33	96.19	96.43	95.97
p-value	NA	0.008	0.023	0.050
NPV (%)	82.74	93.85	91.62	96.20
p-value	NA	0.006	0.069	0.018
Accuracy (%)	84.36	94.95	93.81	96.09
(95% CI)	(78.59–88.8)	(92.15–96.79)	(89.78–96.32)	(91.28–98.3)
p-value	NA	<0.001	0.003	0.002
AUC	0.9426	0.9814	0.9819	0.981
(95% CI)	(0.9149–0.9703)	(0.9679–0.9901)	(0.9675–0.9964)	(0.9596–1.0000)
p-value	NA	<0.001	0.003	0.019

Parameter	**Resident (average)**	**Reviewer 3[2]**	**Reviewer 4[2]**
Sensitivity (%)	90.60	91.95	89.26
p-value	0.030	0.060	0.118
Specificity (%)	93.04	99.37	86.71
p-value	0.187	0.003	0.810
PPV (%)	92.47	99.28	86.36
p-value	0.166	0.004	0.996
NPV (%)	91.30	92.90	89.54
p-value	0.056	0.072	0.211
Accuracy (%) (95% CI)	91.86 (87.94–94.58)	95.77 (90.62–98.15)	87.95 (81.3–92.45)
p-value	0.013	0.003	0.359
AUC (95% CI)	0.9623 (0.9456–0.9760)	0.9928 (0.9842–1.0000)	0.9317 (0.8933–0.9701)
p-value	<0.001	<0.001	0.611

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; NA, notapplicable; NPV,

negative predictive value; PPV, positive predictive value.

[1]Reviewer 1 and Reviewer 2 were board-certified abdominal radiologists.

[2]Reviewer 3 was a fourth-year resident, and Reviewer 4 was a second-year resident.

IV. Discussion

We generated a computer-aided system to differentiate cysts, hemangiomas, and metastases from portal-phase CT images using handcrafted features and compared its diagnostic performance with radiologists. Our radiomics model demonstrated an overall PDI of 0.8037 with category-specific PDIs of 0.9432 for cysts, 0.6653 for hemangiomas, and 0.8027 for metastasis, indicating that the model showed good performance for discriminating cysts and metastasis among the three classes of liver lesions. Along with the diagnostic performance of three-class classification, the diagnostic accuracy of binary classification was also evaluated. The radiomics model of the current study showed 80.54% sensitivity, 87.97% specificity, 84.36% accuracy, and an AUC of 0.9426. In addition to substantial performance in binary classification, our classification system demonstrated accuracy for classifying three types of liver lesions with a CCP of 84.37% for all lesions, 85.87% for subcentimeter lesions, and 83.72% for lesions > 10 mm. Several previous studies that assessed the ability of machine learning for classifying liver lesions at various imaging modalities described good diagnostic performance, with an accuracy of 77.3%–97.2% and an AUC of 0.92–0.941[11-13, 32-36]. Allowing for varying performance of machine learning under different imaging acquisition conditions and imaging modalities, we expect our classification system to differentiate between the three classes of liver lesions with comparable accuracy.

However, between the radiomics model and radiologists, the latter outperformed the former throughout the analyses because the top three most experienced radiologists showed significantly higher overall, hemangioma-specific, and metastasis-specific PDIs, accuracy, and AUC than those of the radiomics model. Particularly, the model showed the lowest diagnostic performance of classifying hemangiomas compared to other values. This can be explained by the limitation of two-dimensional segmentation.

Although the ROI for image segmentation was drawn at the plane where the image was most suitable for diagnosis and represented the corresponding category of liver lesions the best, single portal-phase image may not be sufficient for identifying hepatic hemangioma due to its various imaging features. Depending on the acquisition time of portal phase images, the morphology of the hemangioma can be heterogeneous because the typical imaging finding of hepatic hemangioma on dynamic contrast-enhanced CT is early peripheral nodular enhancement with progressive centripetal fill-in[37,38]. Moreover, atypical hemangiomas with a flash-filling enhancement pattern or delayed fill-in enhancement pattern cause heterogeneity in image sets for radiomics model training[37,38]. Further study with three-dimensional (3-D) segmentation may increase the diagnostic accuracy of the radiomics model in classifying hepatic hemangiomas.

In subgroup analysis, albeit with insignificant difference, the radiomics model showed a lower diagnostic accuracy for classifying lesions < 10 mm than for classifying those ≥ 10 mm. A noticeable difference between radiologists and the radiomics model was that the radiologists maintained excellent diagnostic performance of discriminating subcentimeter lesions in the subgroup analysis. This discrepancy may be due to the bias of healthcare professionals who already know and have an experience that hepatic lesions that are too small to characterize are mostly benign even in cancer patients[39,40]. Particularly in colorectal cancer patients, the incidence of subcentimeter indeterminate lesions was reported to be 12.7% to 25.5% with a benign rate of 80% to 92.5%, which increased to as much as 97.8% in cases of small low attenuating hepatic lesions without larger lesions[41-43]. In accordance with this bias, in this study, more experienced board-certified radiologists were likely to demonstrate significantly better discrimination of subcentimeter benign lesions than metastatic lesion, but less-experienced resident radiologists showed lower category-specific PDIs for

cysts and hemangiomas than for metastases. Additionally, the diagnostic accuracy of the radiomics model for classifying lesions < 10 mm may decrease because precise drawing of ROI perfectly surrounding the subcentimeter-sized lesions is difficult and extractable information from these small lesions may be limited.

Contrary to our study, some studies showed that the diagnostic performance of radiologists was inferior to that of the classification models. Dankerl et al.[44] developed a retrieval-based computer-aided diagnosis system (accuracy for lesion type, 85.8%; accuracy for lesion histology, 75.1%) for differentiating 5 different types of liver lesions (cyst, hemangioma, hepatocellular carcinoma, metastasis and focal nodular hyperplasia) that provided diagnostic performance superior to three radiologists (accuracy for lesion type, 52–74%; accuracy for lesion histology, 22–51%). Similar to our study, only portal-phase CT images were used, and radiologists were allowed to review axial, coronal, and sagittal images. However, the liver lesion categories evaluated by Dankerl et al.[44] included hepatocellular carcinoma and focal nodular hyperplasia, for which multiphase CT is crucial for an accurate diagnosis. The inclusion of these liver lesions may have lowered the diagnostic accuracy of radiologists, which may explain the difference with our study results. Hamm et al.[45] proposed a deep learning system that classifies six categories of liver lesions (cyst, hemangioma, focal nodular hyperplasia, hepatocellular carcinoma, intrahepatic cholangiocarcinoma, and CRLM) on multiphase MRI with a diagnostic accuracy (90%) superior to that of two radiologists (80% and 85%). Other than a different imaging modality being used, the significant difference between our study and Hamm et al.[45] is the manner in which radiologists assessed liver lesions. While Hamm et al.[45] did not allow radiologists to scroll and view lesions other than the target liver lesion, radiologists from our study interpreted CT images similar to those seen in daily

clinical practice. In addition, we included every patient and liver lesion detected during the specific period in the study population and aimed to avoid convenience sampling and secure generalizability of the model. For implementation of the radiomics model in clinical practice, verification of its diagnostic performance and comparing the performance to that of radiologists in a clinical setting are crucial. Liver lesions that were both correctly and incorrectly diagnosed by the radiomics model are depicted in Figures 8 and 9.

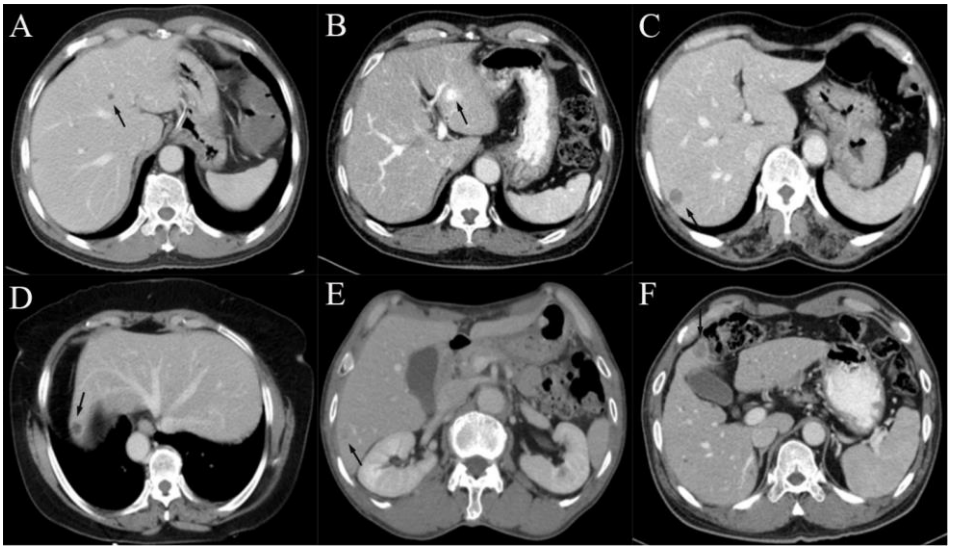

Figure 8. Representative images of a cyst, hemangioma, and metastasis that were correctly diagnosed by the radiomics model. (A) Cyst, (B) hemangioma, and (C) metastasis correctly diagnosed by both the radiomics model and radiologists. (D) Cyst correctly classified by the radiomics model but incorrectly diagnosed by all radiologists. (E) Hemangioma correctly diagnosed by the radiomics model but misclassified as metastasis by two radiologists. (F) Metastasis misdiagnosed as a hemangioma by resident radiologists.

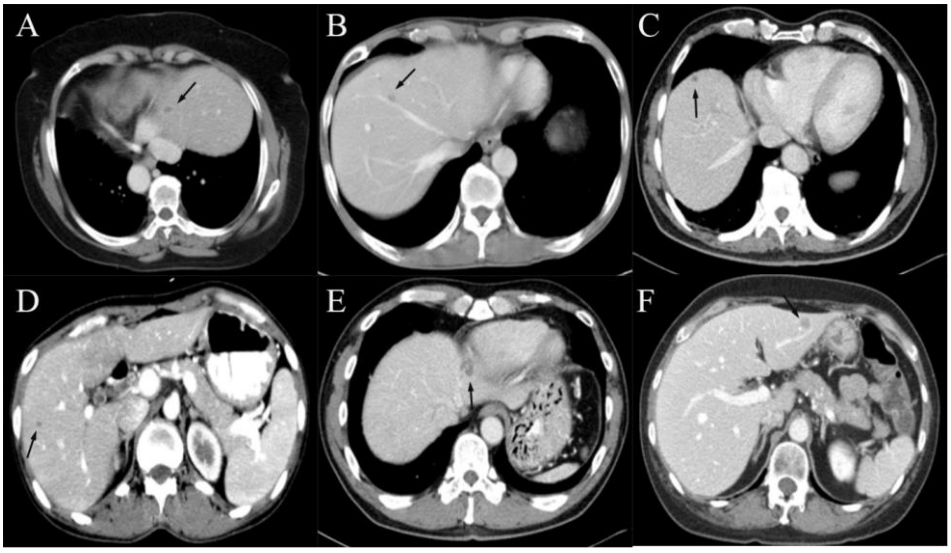

Figure 9. Representative images of a cyst, hemangioma, and metastasis that were incorrectly diagnosed by the radiomics model. (A) Cyst, (B) hemangioma, and (C) metastasis that were misclassified by both the radiomics model and radiologists. In the case of metastasis, there were no lesions that were misdiagnosed by all four radiologists. Board-certified radiologists misclassified this (C) metastasis as benign and the radiomics model misdiagnosed it as a cyst. (D) Cyst, (E) hemangioma, and (F) metastasis that were inaccurately classified by the radiomics model but correctly diagnosed by radiologists.

Some limitations exist in the current study. First, the sample size of hemangiomas was smaller than those of cysts and metastases. Although data augmentation was conducted to avoid an overfitting problem, this skewness in the dataset may have caused bias in the classifier training and induced decreased diagnostic performance for classifying hemangiomas. Second, the histopathological results of some of the liver lesions were not available. However, cysts and hemangiomas can be accurately diagnosed with typical MRI imaging features, and at least 1 year of follow-up images were reviewed to

judge the class of liver lesions. Finally, the subjective nature of semiautomatic segmentation might be another limitation of this study, particularly because the precise drawing of ROI of the subcentimeter lesion was challenging. To overcome this problem, two radiologists independently performed image segmentation and analyzed interobserver reproducibility.

V. Conclusion

We proposed a radiomics model that could differentiate cysts, hemangiomas, and metastases with substantial diagnostic performance using portal-phase abdominopelvic CT images from patients with CRC. Albeit inferior to the diagnostic accuracy of radiologists in this study, our classification model demonstrated high diagnostic accuracy comparable to that of previous studies and proved its potential to be used in clinical practice, where elimination of frequently detected benign liver lesions such as cysts or hemangiomas would help radiologists to review follow-up evaluations of cancer patients. However, the radiomics model demonstrated limitations especially in classifying hemangiomas and subcentimeter liver lesions. Therefore, at this stage, the unattended application of the radiomics model in the clinical setting may still be premature. Further studies using different techniques such as deep learning or 3-D segmentation are required to improve diagnostic accuracy and implement the system in daily practice.

References

1. Manfredi S, Lepage C, Hatem C, Coatmeur O, Faivre J, Bouvier A-M. Epidemiology and Management of Liver Metastases From Colorectal Cancer. Annals of Surgery 2006;244:254-9.
2. Zarour LR, Anand S, Billingsley KG, Bisson WH, Cercek A, Clarke MF, et al. Colorectal Cancer Liver Metastasis: Evolving Paradigms and Future Directions. Cell Mol Gastroenterol Hepatol 2017;3:163-73.
3. Bengtsson G, Carlsson G, Hafstrom L, Jonsson PE. Natural history of patients with untreated liver metastases from colorectal cancer. Am J Surg 1981;141:586-9.
4. Abdalla EK, Vauthey J-N, Ellis LM, Ellis V, Pollock R, Broglio KR, et al. Recurrence and Outcomes Following Hepatic Resection, Radiofrequency Ablation, and Combined Resection/Ablation for Colorectal Liver Metastases. Annals of Surgery 2004;239:818-27.
5. Jones RP, Kokudo N, Folprecht G, Mise Y, Unno M, Malik HZ, et al. Colorectal Liver Metastases: A Critical Review of State of the Art. Liver Cancer 2016;6:66-71.
6. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuze S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Ann Oncol 2017;28:1191-206.
7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278:563-77.
8. Siewert B, Sosna J, McNamara A, Raptopoulos V, Kruskal JB. Missed Lesions at Abdominal Oncologic CT: Lessons Learned from Quality Assurance. RadioGraphics 2008;28:623-38.
9. Brenner DJ, Hall EJ. Computed Tomography — An Increasing Source of Radiation Exposure. New England Journal of Medicine

2007;357:2277-84.

10. Alahmer H, Ahmed A. Computer-aided Classification of Liver Lesions from CT Images Based on Multiple ROI. Procedia Computer Science 2016;90:80-6.
11. Gletsos M, Mougiakakou SG, Matsopoulos GK, Nikita KS, Nikita AS, Kelekis D. A computer-aided diagnostic system to characterize CT focal liver lesions: design and optimization of a neural network classifier. IEEE Trans Inf Technol Biomed 2003;7:153-62.
12. Chang CC, Chen HH, Chang YC, Yang MY, Lo CM, Ko WC, et al. Computer-aided diagnosis of liver tumors on computed tomography images. Comput Methods Programs Biomed 2017;145:45-51.
13. Song S, Li Z, Niu L, Zhou X, Wang G, Gao Y, et al. Hypervascular hepatic focal lesions on dynamic contrast-enhanced CT: preliminary data from arterial phase scans texture analysis for classification. Clin Radiol 2019;74:653.e11-.e18.
14. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. Radiology 2018;286:800-9.
15. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and Generalizability in Radiomics Modeling: Possible Strategies in Radiologic and Statistical Perspectives. Korean J Radiol 2019;20:1124-37.
16. Tirumani SH, Kim KW, Nishino M, Howard SA, Krajewski KM, Jagannathan JP, et al. Update on the role of imaging in management of metastatic colorectal cancer. Radiographics 2014;34:1908-28.
17. Floriani I, Torri V, Rulli E, Garavaglia D, Compagnoni A, Salvolini L, et al. Performance of imaging modalities in diagnosis of liver metastases from colorectal cancer: a systematic review and

meta-analysis. J Magn Reson Imaging 2010;31:19-31.

18. Rojas Llimpe FL, Di Fabio F, Ercolani G, Giampalma E, Cappelli A, Serra C, et al. Imaging in resectable colorectal liver metastasis patients with or without preoperative chemotherapy: results of the PROMETEO-01 study. Br J Cancer 2014;111:667-73.
19. Sivesgaard K, Larsen LP, Sorensen M, Kramer S, Schlander S, Amanavicius N, et al. Diagnostic accuracy of CE-CT, MRI and FDG PET/CT for detecting colorectal cancer liver metastases in patients considered eligible for hepatic resection and/or local ablation. Eur Radiol 2018.
20. Kim HJ, Lee SS, Byun JH, Kim JC, Yu CS, Park SH, et al. Incremental value of liver MR imaging in patients with potentially curable colorectal hepatic metastasis detected at CT: a prospective comparison of diffusion-weighted imaging, gadoxetic acid-enhanced MR imaging, and a combination of both MR techniques. Radiology 2015;274:712-22.
21. Niekel MC, Bipat S, Stoker J. Diagnostic imaging of colorectal liver metastases with CT, MR imaging, FDG PET, and/or FDG PET/CT: a meta-analysis of prospective studies including patients who have not previously undergone treatment. Radiology 2010;257:674-84.
22. Zech CJ, Korpraphong P, Huppertz A, Denecke T, Kim MJ, Tanomkiat W, et al. Randomized multicentre trial of gadoxetic acid-enhanced MRI versus conventional MRI or CT in the staging of colorectal cancer liver metastases. Br J Surg 2014;101:613-21.
23. Jhaveri KS, Fischer SE, Hosseini-Nik H, Sreeharsha B, Menezes RJ, Gallinger S, et al. Prospective comparison of gadoxetic acid-enhanced liver MRI and contrast-enhanced CT with histopathological correlation for preoperative detection of colorectal liver metastases following chemotherapy and potential impact on surgical plan. HPB (Oxford)

2017;19:992-1000.

24. McAuliffe MJ, Lalonde FM, McGarry D, Gandler W, Csaky K, Trus BL. Medical Image Processing, Analysis and Visualization in clinical research. Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001; 2001. p.381-6.
25. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977;33:159-74.
26. Lee HS, Hong H, Jung DC, Park S, Kim J. Differentiation of fat-poor angiomyolipoma from clear cell renal cell carcinoma in contrast-enhanced MDCT images using quantitative feature classification. Med Phys 2017;44:3604-14.
27. Lee H, Hong H, Kim J, Jung DC. Deep feature classification of angiomyolipoma without visible fat and renal cell carcinoma in abdominal contrast-enhanced CT images with texture image patches and hand-crafted feature concatenation. Med Phys 2018;45:1550-61.
28. Robnik-Sikonja M, Kononenko I. An adaptation of Relief for attribute estimation in regression. ICML '97: Proceedings of the Fourteenth International Conference on Machine Learning 2000.
29. Ho TK. Random decision forests. Proceedings of 3rd international conference on document analysis and recognition: IEEE; 1995. p.278-82.
30. Van Calster B, Van Belle V, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW. Extending the c-statistic to nominal polytomous outcomes: the Polytomous Discrimination Index. Statistics in Medicine 2012;31:2610-26.
31. Dreižienė L, Ducinskas K, Paulionienė L. Correct Classification Rates in Multi-Category Discriminant Analysis of Spatial Gaussian Data. Open Journal of Statistics 2015;05:21-6.

32. Huang YL, Chen JH, Shen WC. Diagnosis of hepatic tumors with texture analysis in nonenhanced computed tomography images. Acad Radiol 2006;13:713-20.

33. Mougiakakou SG, Valavanis IK, Nikita A, Nikita KS. Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers. Artif Intell Med 2007;41:25-37.

34. Acharya UR, Koh JEW, Hagiwara Y, Tan JH, Gertych A, Vijayananthan A, et al. Automated diagnosis of focal liver lesions using bidirectional empirical mode decomposition features. Comput Biol Med 2018;94:11-8.

35. Yasaka K, Akai H, Abe O, Kiryu S. Deep Learning with Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-enhanced CT: A Preliminary Study. Radiology 2018;286:887-96.

36. Ye J, Sun Y, Wang S, Gu L, Qian L, Xu J. Multi-Phase CT Image Based Hepatic Lesion Diagnosis by SVM. 2009 2nd International Conference on Biomedical Engineering and Informatics; 2009. p.1-5.

37. Klotz T, Montoriol PF, Da Ines D, Petitcolin V, Joubert-Zakeyh J, Garcier JM. Hepatic haemangioma: common and uncommon imaging features. Diagn Interv Imaging 2013;94:849-59.

38. Caseiro-Alves F, Brito J, Araujo AE, Belo-Soares P, Rodrigues H, Cipriano A, et al. Liver haemangioma: common and uncommon findings and how to improve the differential diagnosis. European Radiology 2007;17:1544-54

39. Khalil HI, Patterson SA, Panicek DM. Hepatic lesions deemed too small to characterize at CT: prevalence and importance in women with breast cancer. Radiology 2005;235:872-8.

40. Jones EC, Chezmar JL, Nelson RC, Bernardino ME. The frequency and significance of small (less than or equal to 15 mm) hepatic lesions detected by CT. AJR Am J Roentgenol 1992;158:535-9.
41. Schwartz LH, Gandras EJ, Colangelo SM, Ercolani MC, Panicek DM. Prevalence and importance of small hepatic lesions found at CT in patients with cancer. Radiology 1999;210:71-4.
42. Lim GH, Koh DC, Cheong WK, Wong KS, Tsang CB. Natural history of small, "indeterminate" hepatic lesions in patients with colorectal cancer. Dis Colon Rectum 2009;52:1487-91.
43. Jang HJ, Lim HK, Lee WJ, Lee SJ, Yun JY, Choi D. Small hypoattenuating lesions in the liver on single-phase helical CT in preoperative patients with gastric and colorectal cancer: prevalence, significance, and differentiating features. J Comput Assist Tomogr 2002;26:718-24.
44. Dankerl P, Cavallaro A, Tsymbal A, Costa MJ, Suehling M, Janka R, et al. A retrieval-based computer-aided diagnosis system for the characterization of liver lesions in CT scans. Acad Radiol 2013;20:1526-34.
45. Hamm CA, Wang CJ, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. Eur Radiol 2019;29:3338-47.

ABSTRACT(IN KOREAN)

대장암 환자에서 발견되는 국소 간병변의 감별을 위한 조영증강 전산화 단층촬영에서의 라디오믹스 연구

<지도 교수 임준석>

연세대학교 대학원 의학과

배희진

목적: 대장암에서 간은 제일 흔하게 전이가 되는 장기로, 대장암 환자에서 간전이를 빠르게 진단하고, 간 병변을 정확하게 감별하는 것은 중요하다. 따라서, 본 연구에서는 조영증강 복부 전산화 단층촬영의 문맥기 영상에서 간 낭종, 혈관종 및 간 전이를 분류할 수 있는 라디오믹스 모델을 구축하고, 진단 정확도를 평가하고자 하였다.

방법: 2005년 1월부터 2010년 12월에 조영증강 복부 전산화 단층촬영과 조영증강 간 자기공명영상을 모두 시행한 502명의 대장암 환자를 후향적으로 분석하였으며, 2006년 8월 23일을 기준으로 그 이후에 조영증강 자기공명영상을 시행 받은 386명의 환자를 학습군으로 하였고, 나머지 116명의 환자를 검증군으로 포함하였다. 502명의 환자에서 병변의 크기가 3㎜ 이상, 5㎝ 이하인 간 낭종, 혈관종, 간 전이 병변이 포함이 되었으며, 총 1290개의 간 병변 (간 낭종 676개, 혈관종 130개, 간 전이 484개)의 문맥기 조영증강

전산화 단층촬영 영상을 이용하여 3 군의 간 병변을 분류할 수 있는 라디오믹스 모델을 구축하였다. 추출해낸 129개의 사람이 정의한 특징 (hand-crafted feature)중에서 관찰자간 재현성이 우수한 (intraclass correlation coefficient> 0.75) 126개의 특징만을 모델링에 사용하였다. ReliefF 방법을 통하여 특징 선택 (feature selection)을 시행하였으며, 선택된 특징들은 랜덤 포레스트를 이용하여 학습시킨 뒤에 최종적으로 라디오믹스 모델을 구축하였다. 구축한 모델의 진단능은 검증군에 포함된 간 병변을 이용하여 평가하였으며, 4명의 영상의학과 의사와 진단 정확도를 비교하였다. 이와 더불어 병변의 크기를 기준으로 10㎜ 미만과 10㎜ 이상으로 나누어서 추가적으로 분석을 시행하였다. 3군의 간 병변을 분류하는 진단능을 비교하기 위해서 polytomous discrimination index (PDI)와 correct classification percentage (CCP)를 구하였고, 양성 (간 낭종과 혈관종)과 악성의 이항 분류 진단능을 비교하기 위해서 민감도, 특이도, 양성 예측도, 음성 예측도, 진단 정확도 및 수신자 조작 특성 곡선 (receiver operating characteristic curve)의 곡선 아래 면적 (area under the curve)을 분석하였다.

결과: 라디오믹스 모델의 전체 PDI, 혈관종 PDI 및 간 전이 PDI는 각각 0.8037, 0.6653, 0.8027로 측정되었으며, 해당 수치는 가장 경험이 없는 영상의학과 의사를 제외한 모든 영상의학과 의사 (전체 PDI, 0.9622-0.9680; 혈관종 PDI, 0.9452-0.9630; 간 전이 PDI, 0.9511-0.9869)에 비해서 유의미하게 낮은 결과였다. 특히나, 병변에 상관없이 군 별 PDI가 비슷하게 높은 영상의학과 의사와 비교하여, 라디오믹스 모델의 혈관종 PDI는 0.6653으로 다른 군 별 PDI에 비해 낮은 수치로 측정되었다. 10㎜ 이상의 병변을 평가할 때와 비교하여

10㎜미만의 병변을 감별할 때 라디오믹스 모델의 PDI는 감소하였으나, 영상의학과 의사의 PDI는 병변의 크기에 상관이 없이 비교적 유지가 되는 결과를 보여주었다. 이항 분류에서는 라디오믹스 모델이 좋은 진단능을 보여주었으며, 간 전이를 감별하는 진단 정확도가 84.36% (95% 신뢰도, 78.59-88.8), 수신자 조작 특성 곡선의 곡선 아래 면적이 0.9426 (0.9149-0.9703)로 측정되었다. 그러나, 제일 경험이 적은 영상의학과 전공의를 제외하고 나머지 3명의 영상의학과 의사는 이항 분류에 있어서도 라디오믹스 모델을 뛰어 넘는 결과를 보여주었으며, 진단 정확도는 93.81%-96.09%로 측정되었고, 모델과의 차이는 통계학적으로 유의미한 차이였다 (p-value, 0.002-0.003).

결론: 라디오믹스 모델은 대장암 환자의 문맥기 조영증강 전산화 단층촬영 영상에서 간 낭종, 혈관종과 간 전이를 분류하는데 있어서 영상의학과 의사보다는 낮지만, 비교적 좋은 진단 정확도를 보여, 임상적으로 적용이 될 수 있는 잠재력을 보여주었다. 하지만, 라디오믹스 모델은 혈관종과 10㎜ 미만의 병변을 진단하는 것에 있어 한계를 보여 실제 진료 환경에서 사람의 개입 없이 적용이 되기에는 아직은 시기 상조로 판단된다.

핵심되는 말: 라디오믹스, 대장암, 간 전이, 간 낭종, 혈관종, 진단 정확도, 조영증강 전산화 단층촬영