pISSN 2384-1095
eISSN 2384-1109

iMRI 2021;25:313-322

https://doi.org/10.13104/imri.2021.25.4.313

iMRI
Investigative Magnetic Resonance Imaging

Diagnostic Image Feature and Performance of CT and Gadoxetic Acid Disodium-Enhanced MRI in Distinction of Combined Hepatocellular-Cholangiocarcinoma from Hepatocellular Carcinoma

Hyunghu Kim[1], Seung-seob Kim[1], Sunyoung Lee[1], Myeongjee Lee[2], Myeong-Jin Kim[1]

[1]Department of Radiology, Research Institute of Radiological Science, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea
[2]Biostatistics Collaboration Unit, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea

Original Article

Received: January 7, 2021
Revised: May 10, 2021
Accepted: May 11, 2021

Correspondence to:
Myeong-Jin Kim, M.D., Ph.D.
Department of Radiology, Research Institute of Radiological Science, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.
Tel. +82-2-2228-2369
Fax. +82-2-2227-8337
E-mail: kimnex@yuhs.ac

Purpose: To find diagnostic image features, to compare diagnostic performance of multiphase CT versus gadoxetic acid disodium-enhanced MRI (GAD-MRI), and to evaluate the impact of analyzing Liver Imaging Reporting and Data System (LI-RADS) imaging features, for distinguishing combined hepatocellular-cholangiocarcinoma (CHC) from hepatocellular carcinoma (HCC).

Materials and Methods: Ninety-six patients with pathologically proven CHC (n = 48) or HCC (n = 48), diagnosed June 2008 to May 2018 were retrospectively analyzed in random order by three radiologists with different experience levels. In the first analysis, the readers independently determined the probability of CHC based on their own knowledge and experiences. In the second analysis, they evaluated imaging features defined in LI-RADS 2018. Area under the curve (AUC) values for CHC diagnosis were compared between CT and MRI, and between the first and second analyses. Interobserver agreement was assessed using Cohen's weighted κ values.

Results: Targetoid LR-M image features showed better specificities and positive predictive values (PPV) than the others. Among them, rim arterial phase hyperenhancement had the highest specificity and PPV. Average sensitivity, specificity, and AUC values were higher for MRI than for CT in both the first (P = 0.008, 0.005, 0.002, respectively) and second (P = 0.017, 0.026, 0.036) analyses. Interobserver agreements were higher for MRI in both analyses (κ = 0.307 for CT, κ = 0.332 for MRI in the first analysis; κ = 0.467 for CT, κ = 0.531 for MRI in the second analysis), with greater agreement in the second analysis for both CT (P = 0.001) and MRI (P < 0.001).

Conclusion: Rim arterial phase hyperenhancement on GAD-MRI can be a good indicator suggesting CHC more than HCC. GAD-MRI may provide greater accuracy than CT for distinguishing CHC from HCC. Interobserver agreement can be improved for both CT and MRI by analyzing LI-RADS imaging features.

Keywords: Liver neoplasms; Combined hepatocellular-cholangiocarcinoma; Differential diagnosis; Sensitivity and specificity; Contrast media

INTRODUCTION

Combined hepatocellular-cholangiocarcinoma (CHC) is an uncommon primary liver cancer, but recently has been increasingly reported in patients with liver cirrhosis (1). Because postsurgical outcome is generally poorer in CHC than in hepatocellular carcinoma (HCC), liver transplantation may not be advisable for treating CHC, for the sake of allocating organs to more appropriate conditions (2). Therefore, accurate differentiation of CHC from HCC is important in the treatment of patients suspected of primary liver cancer (3). In patients with cirrhosis or chronic B-viral hepatitis, it is acceptable to establish a diagnosis of HCC based on the characteristic imaging findings depicted on multiphase CT or MRI, without histologic evaluation, and proceed with treatment (4-6). Although imaging diagnosis of HCC has been reported to be more than 90% specific (4, 5), CHC is one of the main reasons for false positive diagnosis, especially in patients with chronic liver diseases (7, 8).

Most early studies that reported imaging findings of CHC were descriptive, and relatively few case-controlled studies have addressed the diagnostic accuracy of differentiating CHC from HCC (9-11). In an evaluation of CT and MRI diagnoses, Potretzke et al. (9) found that 88% of CHC cases had features of non-HCC malignancy, and could be correctly reclassified. With gadoxetic acid disodium-enhanced MRI (GAD-MRI), Lee et al. (11) demonstrated that 94% of CHC cases may demonstrate at least one feature of malignancy other than HCC, as defined in Liver Imaging Reporting and Data System (LI-RADS) and endorsed by the American College of Radiology (https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/LI-RADS). Contrarily, based on LI-RADS findings for GAD-MRI, Jeon et al. (10) reported that 37% of CHC occurrences were falsely categorized as LR-5 or LR-4. However, the diagnostic accuracies of CT and GAD-MRI have not been compared for differentiating CHC and HCC. Moreover, whether applying LI-RADS imaging criteria improves the diagnostic performance of radiologists is still unknown.

Therefore, the purpose of this study was to compare image features of CHC with HCC, to compare the diagnostic performance of multiphase CT and GAD-MRI for distinguishing CHC from HCC, and to evaluate the impact of analyzing LI-RADS imaging features.

MATERIALS AND METHODS

Patients

This retrospective study was approved by the Institutional Review Board of a tertiary medical center. The hospital ethics committee waived the requirement for written informed consent. From the pathology database, we retrieved 135 patients at our institution who were diagnosed with CHC histologically, through biopsy or surgery, between June 2008 and May 2018. Among them, 48 patients who had undergone both GAD-MRI and multiphase CT before treatment were included in this study. As a control group, 48 patients who had undergone both GAD-MRI and multiphase CT before treatment were randomly selected from among 2162 patients diagnosed with HCC by biopsy or surgery during the same period.

Among the 48 CHC patients, 28 had been included in previous studies at our institution (11, 12), which did not evaluate CT images. The current study included only patients who underwent both liver dynamic CT and GAD-MRI before histologic diagnosis. MRI images for the overlap patients were analyzed independently by investigators who were not involved in the previous studies.

CT and MRI Acquisition

Multiphase CT was performed with a 64- or more channel multidetector-row CT scanner. The CT images included dynamic sequences consisting of pre-contrast, late arterial phase, portal venous phase, and delayed phase. After pre-contrast scan, 600 mgI/kg of iodinated contrast material of 300 mgI/mL or 370 mgI/mL was injected intravenously to each patient. The late arterial phase scan was taken 18 seconds after the attenuation of abdominal aorta reached 100-HU above the baseline by using the bolus-tracking method. The portal venous phase images were acquired 30 seconds after obtaining late arterial phase, and the delayed phase images were acquired 150 seconds after obtaining the portal phase images (Supplementary Table 1).

Multiphase liver MRI was performed with a 1.5-T or 3.0-T MRI scanner. For pre-contrast scans, dual-echo, spoiled gradient-echo, T1-weighted in- and opposed-phase images, and fat-suppressed 3D spoiled gradient-echo images were obtained. For dynamic imaging, gadoxetic acid disodium (0.025 mmol/kg; Primovist, Bayer Pharma, Germany) was injected at 1 mL/s, followed by 20 mL saline at the same injection rate. The timing of late arterial phase was determined using a bolus-tracking method, then portal venous phase and transitional phase images were acquired.

Hepatobiliary images were obtained 15-20 min after the injection of contrast material. Multi- and single-shot T2-weighted turbo spin-echo images and diffusion-weighted images were acquired (Supplementary Table 2).

Image Analysis

Three abdominal radiologists, R1, R2, and R3, had clinical experience in abdominal imaging of over 25 years, more than 5 years, and less than 5 years, respectively. They independently evaluated CT and MRI images in random order, without knowledge of diagnosis or clinical characteristics, using a picture archiving and communication system (Centricity version 4, GE Healthcare, Barrington, IL, USA). The reviewers were informed that the patients had either CHC or HCC. The image analyses were conducted in two sessions. In the first analysis, each reviewer determined probability of CHC on CT and MRI using a five-point scale: 1, definitely not; 2, unlikely; 3, possibly; 4, probably; and 5, certainly CHC. In this analysis, the readers based their scores on their knowledge and experiences for the imaging findings of CHC, but without checking each individual LR-M (probably or definitely malignant, not necessarily hepatocellular carcinoma) and major imaging feature. In the second analysis conducted at least two months later, the same radiologists reviewed CT and MRI provided in random order, this time checking relevant imaging features defined

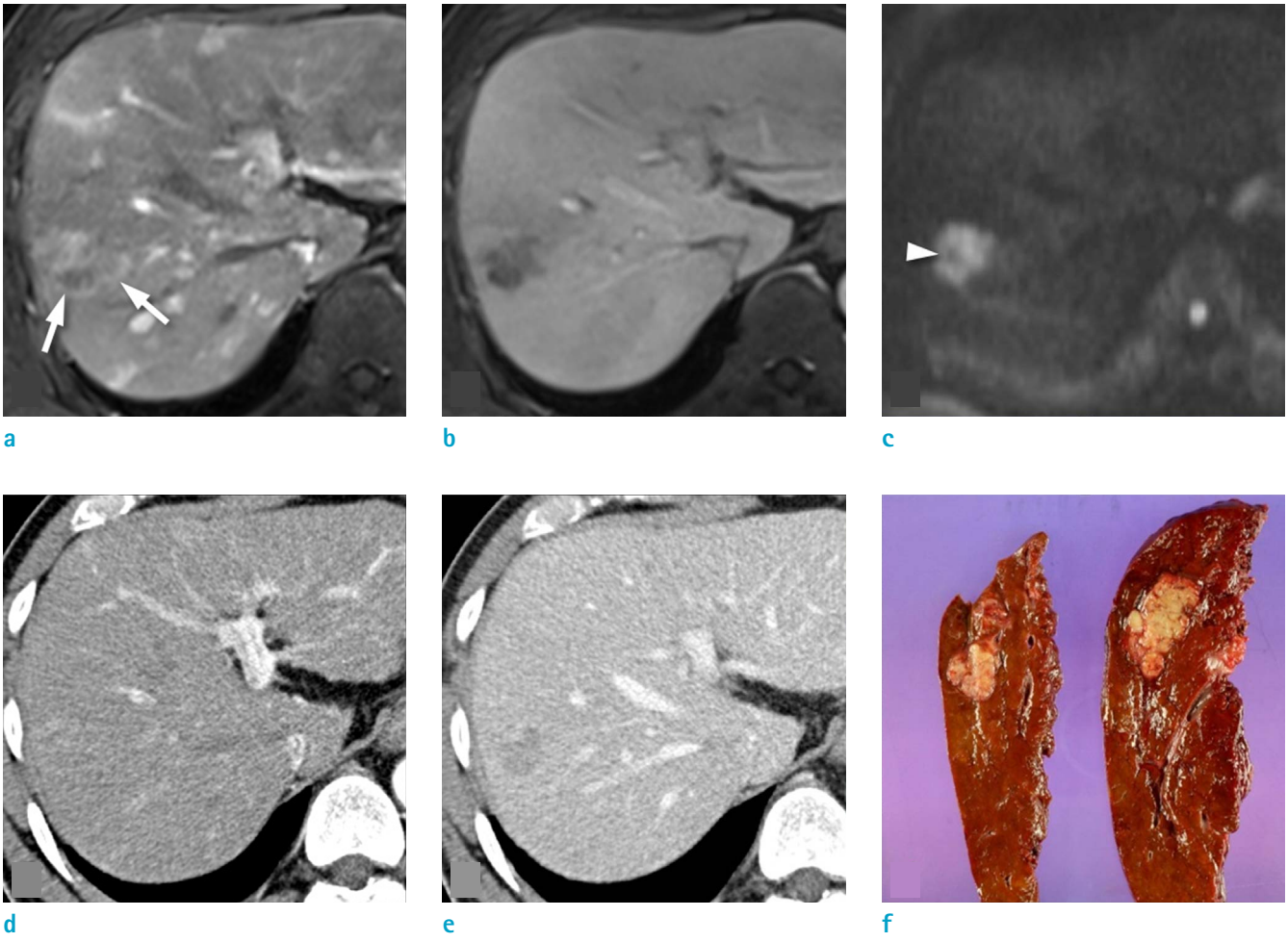

Fig. 1. Combined hepatocellular-cholangiocarcinoma in a 55-year-old man with chronic B-viral cirrhosis. (a-c) MRI of late arterial phase (a), portal phase (b), and diffusion-weighted image (c); (d-e) CT images of late arterial phase (d), and portal phase (e). Readers 1 and 2 considered the lesion to be hepatocellular carcinoma by CT, but combined hepatocellular-cholangiocarcinoma by MRI, because of rim arterial phase enhancement (a, arrows) and targetoid appearance in the diffusion-weighted image (c, arrowhead). The pathology was confirmed after right anterior sectionectomy. On the gross specimen (f), a 4.2 × 2.5 cm multinodular confluent mass is noted.

in LI-RADS version 2018 (Supplementary Table 3), and again determined probability of CHC for each lesion on the five-point scale.

Statistical Analysis

R version 3.6.1 was used for statistical analysis. Demographic data for the HCC and CHC groups were compared using an independent t-test for age, and a X^2 test for other categorical variables. For the first image analysis, and again for the second analysis, each reader's sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated separately for CT and MRI, as was area under the curve (AUC) of the reader's receiver operating characteristic (ROC) curve for diagnosing CHC (13). In the five-point scoring system, scores of 4 and 5 were considered to be positive results for CHC. Differences in the image findings between CHC and HCC were analyzed with the X^2 test. Intraobserver agreement and interobserver agreement between the first and second analyses were evaluated using Cohen's kappa coefficient, κ (14). The degree of agreement was graded as follows: $\kappa < 0$ as no agreement, $0 \leq \kappa < 0.20$ as slight, $0.20 \leq \kappa < 0.40$ as fair, $0.40 \leq \kappa < 0.60$ as moderate, $0.60 \leq \kappa < 0.80$ as substantial, and $0.80 \leq \kappa \leq 1$ as nearly complete agreement.

In multiple-reader, multiple-case ROC analysis to acquire average AUC, sensitivity, and specificity, readers were treated as fixed, and cases as random variables. The result was acquired with the Obuchowski-Rockette method, which

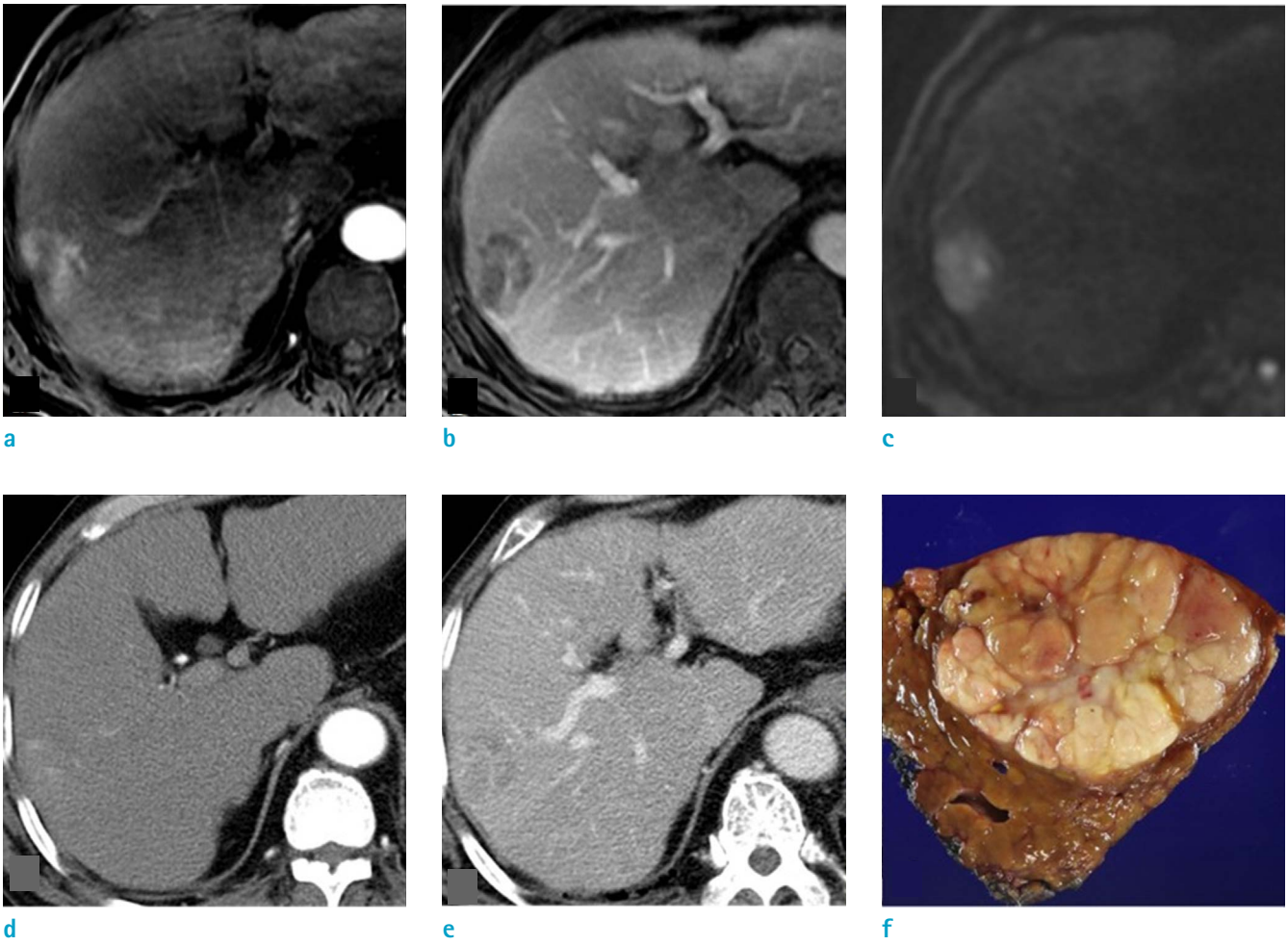

Fig. 2. Combined hepatocellular-cholangiocarcinoma in a 73-year-old man with alcoholic cirrhosis. (a-c) MRI of late arterial phase (a), portal phase (b), diffusion-weighted image (c); (d-e) CT images of late arterial phase (d), and portal phase (e). One reader correctly graded the lesion as probable combined hepatocellular-cholangiocarcinoma only in the second analysis of MRI, by recognizing rim enhancement (a) and delayed central enhancement (b). The diagnosis was confirmed by wedge resection. On the gross specimen (f), a nodular mass is seen, with small regions of necrosis and hemorrhage.

adopted the two-way, mixed effects ANOVA model. The 95% confidence interval was estimated with the jackknife method (15-17).

RESULTS

Patient Characteristics

The HCC and CHC groups did not differ significantly in age, sex, or prevalence of liver cirrhosis (Table 1). Hepatitis B virus (HBV) infection was the major cause of liver cirrhosis in both the HCC and the CHC group, respectively 82% (18/22) and 71% (15/21).

Table 1. Patient Characteristics

	Hepatocellular carcinoma (48 patients)	Combined hepatocellular carcinoma-cholangiocarcinoma (48 patients)	P-value
Age	59 ± 10	62 ± 11	0.544
Sex (Male:Female)	37:11	40:8	0.609
Etiology			
All patients			
HBV	42 (88%)	36 (75%)	0.190
HCV	3 (6%)	3 (6%)	>0.99
Alcohol	26 (54%)	27 (56%)	>0.99
NAFLD	8 (17%)	5 (10%)	0.552
NASH	2 (4%)	1 (2%)	>0.99
Liver cirrhosis	22 (46%)	21 (44%)	0.837
HBV	18 (82%)	15 (71%)	0.420
HCV	2 (9%)	2 (10%)	0.961
Alcohol	15 (68%)	13 (62%)	0.666
Child-Pugh classification of patients with cirrhosis			
A	16 (73%)	17 (81%)	
B	6 (27%)	4 (19%)	
C	0 (0%)	0 (0%)	
Diagnostic methods of tumors	0.537		
Resection	43 (90%)	41 (85%)	
Biopsy	5 (10%)	7 (15%)	

HBV = hepatitis B virus; HCV = hepatitis C virus; NAFLD = non-alcoholic fatty liver disease; NASH = non-alcoholic steatohepatitis
The data refers to the number of patients.

Image Findings of CHC

Rim arterial phase hyperenhancement (APHE) and infiltrative margin only showed significant difference between CHC and HCC on both CT and MRI. Peripheral washout, delayed central enhancement, and targetoid transitional phase (TP) or hepatobiliary phase (HBP) appearance were found significantly more on GAD-MRI analysis (Table 2, Figs. 1, 2). Among these features, rim APHE had the best sensitivity and PPV in reader 1 (sensitivity 72.9%, PPV 87.5%) and reader 2 (sensitivity 77.1%, PPV 90.2%), and second-best sensitivity and PPV in reader 3 (sensitivity 64.6%, PPV 81.6%). Specificities of targetoid LR-M features were quite higher than sensitivities of those, mostly near 90% or over in R1 and R2 (Table 3).

CT versus MRI in the First and Second Analyses

In the first analysis, averaged values of sensitivity (P = 0.008), specificity (P = 0.005), and AUC (P = 0.002) for the three readers were higher with MRI than CT. Individually, two of the three readers (R1 and R3) showed better AUC for diagnosing CHC with MRI than with CT (R1, 0.707 for CT vs. 0.864 for MRI, P < 0.001; R3, 0.709 for CT vs. 0.828 for MRI, P = 0.009) (Table 4). Sensitivities of R1 and R3 were also higher with MRI than with CT (P < 0.001). Specificities were comparable between MRI and CT for all readers.

In the second analysis, as in the first, average sensitivity (P = 0.017), specificity (P = 0.026), and AUC (P = 0.036) were all higher with MRI than CT. Individually, the least experienced reader (R3) showed better AUC with MRI (0.734 for CT vs. 0.847 for MRI, P = 0.007), while sensitivities of R1 and R2 were higher with MRI than CT (P < 0.001). Specificities were comparable between MRI and CT for all readers.

First versus Second Analyses of CT and MRI

Average sensitivity, specificity, and AUC were comparable between the first and second analyses for both CT and MRI. However, individually, AUC values increased in the second analysis of CT by R3 (0.709 vs. 0.734, P = 0.010), and MRI by R2 (0.848 vs. 0.879, P = 0.018). For CT, sensitivity also increased in the second analysis by R3 (38% vs. 54%, P = 0.005), while specificity increased for R1 (73% vs. 88%, P = 0.02). On MRI, sensitivity increased in the second analysis by R2 (65% vs. 77%, P = 0.034), but for R3, sensitivity decreased (81% vs. 69%, P = 0.014) and specificity increased (77% vs. 88%, P = 0.025).

Table 2. Analysis of Imaging Features Described on LI-RADS version 2018

	R1						R2						R3					
	CT			MRI			CT			MRI			CT			MRI		
	cHCC-CCA	HCC	P-value	cHCC-CCA	HCC	P-value	cHCC-CCA	HCC	P-value	cHCC-CCA	HCC	P-value	cHCC-CCA	HCC	P-value	cHCC-CCA	HCC	P-value
Major features																		
Nonrim APHE	23	43	<0.001	13	42	<0.001	22	44	<0.001	11	43	<0.001	22	39	<0.001	17	41	<0.001
Nonperipheral washout	30	44	<0.001	18	42	<0.001	30	43	0.002	18	44	<0.001	46	45	0.646	33	43	0.012
Enhancing capsule	8	38	<0.001	9	31	<0.001	10	27	<0.001	11	38	<0.001	10	24	0.003	8	27	<0.001
Ancillary features, favoring malignancy, HCC in particular																		
Nonenhancing capsule	0	6	0.011	5	9	0.247				0	6	0.011				4	22	<0.001
Nodule-in-nodule architecture	0	6	0.14	2	8	0.045	3	5	0.46	2	6	0.14	4	9	0.136	4	10	0.083
Mosaic architecture	12	20	0.001	6	19	0.003	10	13	0.473	6	20	0.001	9	17	0.066	11	20	0.049
Fat in mass	0	12	0.004	1	9	0.008	1	3	0.307	2	12	0.004	0	2	0.153	4	12	0.028
Blood products in mass	1	10	0.014	1	11	0.002	1	5	0.092	2	10	0.014	1	1	>0.999	4	13	0.016
LR-M targetoid																		
Rim APHE	21	4	<0.001	35	5	<0.001	23	2	<0.001	37	4	<0.001	19	6	0.003	31	7	<0.001
Peripheral washout	2	1	0.557	17	3	<0.001	1	1	>0.999	15	3	0.002	0	1	0.315	11	3	0.021

APHE = arterial phase hyperenhancement; CCA = cholangiocarcinoma; cHCC = combined hepatocellular-cholangiocarcinoma; HBV = hepatitis B virus; HCC = hepatocellular carcinoma; HCV = hepatitis C virus; NAFLD = non-alcoholic fatty liver disease; NASH = non-alcoholic steatohepatitis
The data refers to the number of patients.

Intraobserver and Interobserver Agreement

Interobserver agreement tended to be higher for MRI in both the first and second analyses, with greater agreement in the second analysis for both CT (P = 0.001) and MRI (P < 0.001) (Table 5). In the first analysis, agreement between any two readers was moderate (~0.428-0.552 for CT, ~0.566-0.596 for MRI), and overall agreement was fair for both CT (0.307) and MRI (0.332). Interobserver agreement was higher in the second analysis than in the first, with moderate to substantial agreement for both CT (~0.595-0.770) and MRI (~0.685-0.743), and greater overall agreement of 0.467 for CT and 0.531 for MRI.

DISCUSSION

Our study demonstrated that GAD-MRI may provide greater accuracy than CT for distinguishing CHC from hepatocellular carcinoma. Also, interobserver agreement can be improved for both CT and MRI by analyzing the imaging features defined in LI-RADS.

Better detection of HCC using GAD-MRI than CT has been reported previously (18-22). However, diagnostic accuracy in distinguishing CHC or other malignancy from HCC has not been specifically investigated. Although comparisons of GAD-MRI and conventional extracellular contrast-enhanced MRI are even more scarce, a study showed GAD-

Table 3. Sensitivity, Specificity and PPV for CHC of MRI on Each Readers

	R1			R2			R3		
	Sensitivity	Specificity	PPV	Sensitivity	Specificity	PPV	Sensitivity	Specificity	PPV
Rim APHE	72.92% (58.15-84.72)	89.58% (77.34-96.53)	87.50% (75.00-94.23)	77.08% (62.69-87.97%)	91.67% (80.02-97.68)	90.24% (78.14-95.99)	64.58% (49.46-77.84)	85.42% (72.24-93.93)	81.58% (68.40-90.06)
Peripheral washout	35.42% (22.16-50.54)	93.75% (82.80-98.69)	85.00% (63.97-94.76)	31.25% (18.66-46.25)	93.75% (82.80-98.69)	83.33% (60.73-94.17)	22.92% (12.03-37.31)	93.75% (82.80-98.69)	78.57% (52.17-92.50)
Delayed central enhancement	41.67% (27.61-56.79%)	89.58% (77.34-96.53)	80.00% (62.05-90.73)	45.83% (31.37-60.83%)	93.75% (82.80-98.69)	88.00% (70.15-95.81)	39.58% (25.77-54.73)	87.50% (74.75-95.27)	76.00% (58.09-87.85)
Targetoid restriction	33.33% (20.40-48.41%)	93.75% (82.80-98.69)	84.21% (62.42-94.48)	35.42% (22.16-50.54%)	93.75% (82.80-98.69)	85.00% (63.97-94.76)	10.42% (3.47-22.66)	97.92% (88.93-99.95)	83.33% (37.75-97.63)
Targetoid TP or HBP appearance	37.50% (23.95-52.65)	93.75% (82.80-98.69)	85.71% (65.40-95.01)	37.50% (23.95-52.65)	93.75% (82.80-98.69)	85.71% (65.40-95.01)	43.75% (29.48-58.82)	85.42% (72.24-93.93)	75.00% (58.48-86.47)
Infiltrative margin	56.25% (41.18-70.52)	85.42% (72.24-93.93)	79.41% (65.05-88.88)	52.08% (37.19-66.71)	89.58% (77.34-96.53)	83.33% (67.62-92.29)	45.83% (31.37-60.83)	89.58% (77.34-96.53)	81.48% (64.49-91.42)
Marked diffusion restriction	14.58% (6.07-27.76)	95.83% (85.75-99.49)	77.78% (43.37-94.12)	10.42% (3.47-22.66%)	95.83% (85.75-99.49)	71.43% (33.76-92.46)	18.75% (8.95-32.63)	93.75% (82.80-98.69)	75.00% (46.37-91.23)
Necrosis or severe ischemia	27.08% (15.28-41.85)	81.25% (67.37-91.05)	59.09% (40.56-75.35)	27.08% (15.28-41.85)	89.58% (77.34-96.53)	72.2% (50.12-87.06)	33.33% (20.40-48.41)	83.33% (69.78-92.52)	66.67% (48.62-80.87)

APHE = arterial phase hyperenhancement; CHC = combined hepatocellular-cholangiocarcinoma; HBP = hepatobiliary phase; PPV = positive predictive values; TP = transitional phase

MRI to perform potentially better in identifying non-HCC malignancy, by better demonstrating targetoid appearance (23, 24). On our study, nearly all targetoid LR-M features were found significant more with MRI in patients of CHC than in those of HCC, except targetoid restriction on R3. Considering that only infiltrative margin among non-targetoid LR-M features showed significant difference in patients with CHC and HCC, GAD-MRI can improve performance in identifying CHC by delineating targetoid appearance better.

Each image features' sensitivities, specificities and PPVs were also evaluated for each readers. Sensitivities for all evaluated image features were not that high, among which the highest sensitivity was less than 80%. However, targetoid LR-M image features evaluated with MRI had higher specificity and PPV than the others. Among these image features, rim APHE showed the highest PPV and specificity in R1 and R2, and second best PPV and specificity in R3. Considering the objective of distinguishing CHC from HCC, which is because treatment options such as liver transplantation for HCC are not identically appliable for CHC, these high specificities can be helpful clinical situations.

Comparing CT versus GAD-MRI, our study showed that average diagnostic performance values (sensitivity, specificity, and AUC) of three readers were higher on GAD-MRI than CT in both the first and second analyses. This result indicates that GAD-MRI can provide better diagnostic performance, with or without analyses of LR-M or major imaging features defined by LI-RADS. In both analyses, sensitivity of all readers was better ($P < 0.001$), or nearly so ($P = 0.052$), with MRI. Although specificity did not differ significantly between CT and MRI for any reader, average specificity was significantly higher with MRI. AUC values were also significantly better with GAD-MRI for two readers (the most and least experienced) in the first analysis, and one (the least experienced) in the second analysis. This result may indicate that readers may perform better using GAD-MRI regardless of their level of experience.

Between the first and second analyses, although overall diagnostic performance was not significantly different for either CT or MRI, AUC improved in the second analysis for at least one reader whether with CT (R2) or with MRI (R3). Sensitivity of one reader improved in the second analysis for CT, and two readers for MRI. Specificity of one reader improved in the second analysis for CT, and one reader for MRI. These results indicate that systematically analyzing the LI-RADS imaging features may improve the readers'

Table 4. Comparison of Sensitivity, Specificity, and AUC between CT and MRI and between the First and Second Analyses

	Sensitivity			Specificity			AUC		
	CT (%, 95% CI)	MRI (%, 95% CI)	P-value	CT (%, 95% CI)	MRI (%, 95% CI)	P-value	CT (%, 95% CI)	MRI (%, 95% CI)	P-value
R1									
Conventional	54% (40-68%)	81% (70-92%)	<0.001	73% (60-85%)	79% (68-91%)	0.366	64% (54-73%)	80% (72-88%)	<0.001
LR-M	42% (28-56%)	71% (58-84%)	<0.001	88% (78-97%)	85% (75-95%)	0.705	65% (55-74%)	78% (70-86%)	0.26
P-values	0.083	0.059		0.02	0.18		0.822	0.556	
R2									
Conventional	50% (36-64%)	65% (51-78%)	0.052	96% (90-101%)	88% (78-97%)	0.102	73% (64-82%)	76% (68-86%)	0.996
LR-M	50% (36-64%)	77% (65-89%)	<0.001	96% (90-101%)	92% (84-99%)	0.157	73% (64-82%)	84% (77-92%)	0.256
P-values	>0.999	0.034		>0.999	0.317		>0.999	0.018	
R3									
Conventional	38% (24-51%)	81% (70-92%)	<0.001	83% (73-94%)	77% (65-89%)	0.366	60% (51-70%)	79% (71-87%)	0.009
LR-M	54% (40-68%)	69% (56-82%)	0.052	88% (78-97%)	88% (78-97%)	1	71% (62-80%)	78% (70-86%)	0.007
P-values	0.005	0.014		0.48	0.025		0.01	0.751	
Average									
Conventional	50% (36-63%)	70% (54-85%)	0.008	77% (67-87%)	88% (81-96%)	0.005	75% (68-83%)	85% (78-91%)	0.002
LR-M	59% (44-74%)	75% (63-86%)	0.017	85% (76-95%)	93% (87-99%)	0.026	78% (70-86%)	85% (78-92%)	0.036
P-values	0.371	0.606		0.221	0.324		0.618	0.936	

AUC = area under the curve; CI = confidence interval; LR-M = Probably or definitely malignant, not necessarily hepatocellular carcinoma; R1 = radiologist with over 25 years of abdominal imaging experience, R2 = radiologist with more than 5 years of abdominal imaging experience, R3 = radiologist with less than 5 years of abdominal imaging experience

Table 5. Interobserver Agreements on CT and MRI Before and after Analyzing LI-RADS Imaging Features*

	R1 vs. R2	R1 vs. R3	R2 vs. R3	Overall agreements
First analysis				
CT	0.552 (0.059)	0.428 (0.064)	0.520 (0.059)	0.307 (0.032)
MRI	0.596 (0.051)	0.566 (0.053)	0.596 (0.053)	0.332 (0.032)
P-value	0.573	0.097	0.338	0.581
Second analysis				
CT	0.770 (0.045)	0.595 (0.059)	0.621 (0.056)	0.467 (0.034)
MRI	0.743 (0.046)	0.685 (0.049)	0.734 (0.048)	0.531 (0.038)
P-value	0.675	0.241	0.126	0.209
P-values for first vs. second analyses				
CT	0.003	0.055	0.214	0.001
MRI	0.032	0.099	0.054	<0.001

*: The degrees of agreement are displayed by kappa values (standard error).

determination, but its impact on diagnostic performance may depend upon the reader's experience level. In our study, diagnostic performance tended to improve in the second analysis for radiologists with moderate experience or less.

Our results also showed that interobserver agreement improved significantly in the second analysis for both CT and MRI, indicating that meticulous analysis of imaging features may help improve interobserver agreement. The tendency toward higher interobserver agreement, as well as better diagnostic performance with MRI than with CT, may be attributed to the more conspicuous presentation by MRI of imaging features that indicate non-HCC malignancy (LR-M features), such as targetoid appearance (i.e., rim arterial phase hyperenhancement, peripheral washout, targetoid diffusion restriction, and targetoid hypointensity on transitional or hepatobiliary phase images) (12, 25, 26).

Interobserver agreement in the present study was comparable with results of a recent study by Ludwig et al. (27), which pooled composite data of CT, extracellular contrast-enhanced MRI, and GAD-MRI. Kim et al. (12) reported fair interobserver agreement for identifying each LR-M feature in differentiating HCC or various other hepatic malignancies, including CHC. However, that study did not

assess the overall agreement for determining non-HCC malignancy. Also, our study included only patients with HCC and CHC. This binary choice may have contributed to higher interobserver agreement than was the case for individual imaging features.

Our study had limitations. First, it was a retrospective study from a single center. Second, we included only HCC and CHC patients, and therefore readers considered only these two diagnosis options. This binary choice may affect the accuracy of each reader in diagnosing CHC. Therefore, diagnostic accuracy obtained in our study may not be generalizable to clinical practice, because we should also consider other differential diagnosis for a lesion in our daily practice. However, the design of our study is justified by its purpose, which was to directly compare CT and MRI for distinguishing CHC from HCC under identical conditions, and to evaluate the impact of LI-RADS imaging criteria. Third, because each reader analyzed each patient's CT and MRI twice, the possibility of recall bias cannot be eliminated. To avoid it, the second analysis was conducted more than two months after the first, and CT and MRI were reviewed in random order as well.

In conclusion, our study demonstrated that targetoid LR-M features evaluated with GAD-MRI can distinguish CHC from HCC more specifically. Among these features, rim APHE can diagnose CHC with the highest specificity and PPV. GAD-MRI may provide greater accuracy and interobserver agreement than CT for distinguishing CHC from HCC. In addition, both diagnostic performance and interobserver agreement can be improved by analyzing imaging features defined in LI-RADS.

Supplementary Materials

The Table Supplements are available with this article at https://doi.org/10.13104/imri.2021.25.4.313

Acknowledgments

This study was supported by faculty research grant 6-2019-0062 of Yonsei University College of Medicine for 2019.

REFERENCES

1. Brunt E, Aishima S, Clavien PA, et al. cHCC-CCA: consensus terminology for primary liver carcinomas with both hepatocytic and cholangiocytic differentation. Hepatology 2018;68:113-126
2. Ramai D, Ofosu A, Lai JK, Reddy M, Adler DG. Combined hepatocellular cholangiocarcinoma: a population-based retrospective study. Am J Gastroenterol 2019;114:1496-1501
3. Cunha GM, Tamayo-Murillo DE, Fowler KJ. LI-RADS and transplantation: challenges and controversies. Abdom Radiol (NY) 2021;46:29-42
4. European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol 2018; 69:182-236
5. Marrero JA, Kulik LM, Sirlin CB, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the study of liver diseases. Hepatology 2018;68:723-750
6. Omata M, Cheng AL, Kokudo N, et al. Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. Hepatol Int 2017;11:317-370
7. Fraum TJ, Cannella R, Ludwig DR, et al. Assessment of primary liver carcinomas other than hepatocellular carcinoma (HCC) with LI-RADS v2018: comparison of the LI-RADS target population to patients without LI-RADS-defined HCC risk factors. Eur Radiol 2020;30:996-1007
8. Kim MJ, Lee S, An C. Problematic lesions in cirrhotic liver mimicking hepatocellular carcinoma. Eur Radiol 2019;29:5101-5110
9. Potretzke TA, Tan BR, Doyle MB, Brunt EM, Heiken JP, Fowler KJ. Imaging features of biphenotypic primary liver carcinoma (hepatocholangiocarcinoma) and the potential to mimic hepatocellular carcinoma: LI-RADS analysis of CT and MRI features in 61 cases. AJR Am J Roentgenol 2016;207:25-31
10. Jeon SK, Joo I, Lee DH, et al. Combined hepatocellular cholangiocarcinoma: LI-RADS v2017 categorisation for differential diagnosis and prognostication on gadoxetic acid-enhanced MR imaging. Eur Radiol 2019;29:373-382
11. Lee HS, Kim MJ, An C. How to utilize LR-M features of the LI-RADS to improve the diagnosis of combined hepatocellular-cholangiocarcinoma on gadoxetate-enhanced MRI? Eur Radiol 2019;29:2408-2416
12. Kim YY, Kim MJ, Kim EH, Roh YH, An C. Hepatocellular carcinoma versus other hepatic malignancy in cirrhosis: performance of LI-RADS version 2018. Radiology 2019;291:72-80
13. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. Radiology 2003;229:3-8
14. Gwet K. Computing inter-rater reliability with the SAS system. Stat Methods Inter-rater Reliability Assess 2002;3:1-16
15. Hillis SL, Berbaum KS, Metz CE. Recent developments in the Dorfman-Berbaum-Metz procedure for multireader

ROC study analysis. Acad Radiol 2008;15:647-661
16. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol 1992;27:723-731
17. Obuchowski NA Jr, Rockette HE Jr. Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests an anova approach with dependent observations. Commun Stat - Simul Comput 1995; 24:285-308
18. Baek CK, Choi JY, Kim KA, et al. Hepatocellular carcinoma in patients with chronic liver disease: a comparison of gadoxetic acid-enhanced MRI and multiphasic MDCT. Clin Radiol 2012;67:148-156
19. Semaan S, Vietti Violi N, Lewis S, et al. Hepatocellular carcinoma detection in liver cirrhosis: diagnostic performance of contrast-enhanced CT vs. MRI with extracellular contrast vs. gadoxetic acid. Eur Radiol 2020;30:1020-1030
20. Yoon JH, Lee JM, Lee YJ, Lee KB, Han JK. Added value of sequentially performed gadoxetic acid-enhanced liver MRI for the diagnosis of small (10-19 mm) or atypical hepatic observations at contrast-enhanced CT: a prospective comparison. J Magn Reson Imaging 2019;49:574-587
21. Chung YE, Park JY, Choi J-Y, Kim M-J, Park MS, Seong J. Noninvasive biomarker for predicting treatment response to concurrent chemoradiotherapy in patients with hepatocellular carcinoma. Investig Magn Reson Imaging 2019;23:351-360
22. Park IK, Yu J-S, Cho E-S, Kim JH, Chung J-J. Pseudoglandular formation in hepatocellular carcinoma determines apparent diffusion coefficient in diffusion-weighted MRI. Investig Magn Reson Imaging 2018;22:79-85
23. Min JH, Kim JM, Kim YK, et al. Prospective intraindividual comparison of magnetic resonance imaging with gadoxetic acid and extracellular contrast for diagnosis of hepatocellular carcinomas using the Liver Imaging Reporting and Data System. Hepatology 2018;68:2254-2266
24. Lee S, Kim MJ, Kim SS, et al. Retrospective comparison of EASL 2018 and LI-RADS 2018 for the noninvasive diagnosis of hepatocellular carcinoma using magnetic resonance imaging. Hepatol Int 2020;14:70-79
25. Fraum TJ, Tsai R, Rohe E, et al. Differentiation of hepatocellular carcinoma from other hepatic malignancies in patients at risk: diagnostic performance of the Liver Imaging Reporting and Data System Version 2014. Radiology 2018;286:158-172
26. Fowler KJ, Potretzke TA, Hope TA, Costa EA, Wilson SR. LI-RADS M (LR-M): definite or probable malignancy, not specific for hepatocellular carcinoma. Abdom Radiol (NY) 2018;43:149-157
27. Ludwig DR, Fraum TJ, Cannella R, et al. Hepatocellular carcinoma (HCC) versus non-HCC: accuracy and reliability of Liver Imaging Reporting and Data System v2018. Abdom Radiol (NY) 2019;44:2116-2132