



Characterization of liver nodules in patients with chronic liver disease by MRI: performance of the Liver Imaging Reporting and Data System (LI-RADS v.2018) scale and its comparison with the Likert scale

Andrea Esposito¹ · Valentina Buscarino² · Dario Raciti³ · Elena Casiraghi⁴ · Matteo Manini⁵ · Pietro Biondetti⁶ · Laura Forzenigo¹

Received: 3 February 2019 / Accepted: 25 September 2019 / Published online: 5 October 2019

© Italian Society of Medical Radiology 2019

Abstract

Objectives To evaluate the performance of the LI-RADS v.2018 scale by comparing it with the Likert scale, in the characterization of liver lesions.

Methods A total of 39 patients with chronic liver disease underwent MR examination for characterization of 44 liver lesions. Images were independently analyzed by two radiologists using the LI-RADS scale and by another two radiologists using the Likert scale. The reference standard used was either histopathological evaluation or a 4-year MRI follow-up. Receiver operating characteristic analysis was performed.

Results The LI-RADS scale obtained an accuracy of 80%, a sensitivity of 72%, a specificity of 93%, a positive predictive value (PPV) of 93% and a negative predictive value (NPV) of 70%, while the Likert scale achieved an accuracy of 79%, a sensitivity of 73%, a specificity of 87%, a PPV of 89% and a NPV of 70%. The area under the curve (AUC) was 85% for the LI-RADS scale and 83% for the Likert scale. The inter-observer agreement was strong ($k = 0.89$) between the LI-RADS evaluators and moderate ($k = 0.69$) between the Likert evaluators.

Conclusions There was no statistically significant difference between the performances of the two scales; nevertheless, we suggest that the LI-RADS scale be used, as it appeared more objective and consistent.

Keywords Liver cirrhosis · Carcinoma, hepatocellular · Magnetic resonance imaging · Early detection of cancer · Data interpretation, statistical

Abbreviations

LI-RADS	Liver Imaging Reporting and Data System
ROC	Receiver operating characteristic
PPV	Positive predictive value
NPV	Negative predictive value
AUC	Area under the curve
HCC	Hepatocellular carcinoma
ACC	Accuracy
SENS	Sensitivity
SPEC	Specificity
FNB	Fine-needle biopsy
HCV	Hepatitis C virus
HBV	Hepatitis B virus

✉ Andrea Esposito
andrea.esposito@policlinico.mi.it

¹ Division of Radiology, Foundation IRCCS Ca' Granda Maggiore Policlinico Hospital, Via F.Sforza 35, Milan, Italy

² Division of Radiology, Sant'Andrea Hospital, Corso M. Abbiate 21, Vercelli, Italy

³ Division of Radiology, Città di Sesto San Giovanni Hospital, Viale G. Matteotti 83, Sesto San Giovanni, Italy

⁴ Department of Computer Science (DI), University of Milan, Via G. Celoria 18, Milan, Italy

⁵ Department of Specialty and Transplant Medicine, Gastroenterology, Hepatology and Transplant Unit, Papa Giovanni XXIII Hospital, Piazza OMS 1, Bergamo, Italy

⁶ Division of Radiology, Abano Terme Policlinico Hospital, Piazza C. Colombo 1, Abano Terme, Italy

Introduction

The incidence of hepatocellular carcinoma (HCC) has risen rapidly over the last few years, especially in the USA [1, 2].

In Sweden, a long-linear model used to estimate a more correct incidence of hepatocarcinoma, showed that hepatitis C-associated liver cancer increased and constituted 20% of cases in 2010 [3].

HCC was the sixth most common cancer and the fourth leading cause of cancer-related deaths worldwide in 2018 [4].

The distribution of HCC varies according to geographic location, to viral hepatitis and to the age at which it was acquired. Chronic hepatitis B virus (HBV) and hepatitis C virus (HCV) infections represent the leading cause of HCC (60–70%). In most of Africa and Asia, HBV is the single leading risk factor for HCC, whereas in Japan, northern Europe and the USA, HCV is the major risk factor [5].

In 80–90% of cases, HCC occurs in the setting of cirrhosis [6]. However, a stepwise progression of hepatocarcinogenesis has been established [7], so when a hepatocellular nodule is detected, monitoring is recommended, in order to diagnose premalignant nodules and early HCC, when effective therapy can be applied.

Periodic serologic and imaging tests for patients known to be affected by chronic liver diseases are recommended and widely implemented in current clinical practice. However, the optimal surveillance interval and surveillance tools for HCC have not yet been standardized [8].

CT and MR are the imaging techniques that often allow making a definite diagnosis of HCC without a need to biopsy the lesion. HCC is the only tumor that can be diagnosed with imaging alone, without the need for histopathological confirmation [9].

MRI has been proposed as a sensitive (81%) and specific (85%) imaging modality for the evaluation of liver nodules in patients with cirrhosis [10, 11].

This is based on the unique properties of MR imaging resulting in a high intrinsic soft tissue contrast between normal liver parenchyma and liver lesions, which can be further enhanced with intravenous administration of non-specific (extracellular) and liver-specific (hepatobiliary) gadolinium-based contrast agents [12].

Several scientific organizations and societies have proposed diagnostic systems for the interpretation of liver examination, in order to reduce imaging interpretation variability and improve communication with clinicians [13].

Since March 2011, the Liver Imaging Reporting and Data System (LI-RADS) scale has been adopted by many clinical practices [14]. It is a system for standardizing the report, performance and diagnostic interpretation of liver nodules, using fixed criteria [15, 16]. The first version of the LI-RADS was launched in March 2011, an update was released in 2013 and 2014 [14], and the latest update was released in 2018 [15].

A different scale of diagnostic interpretation, adopted in many fields of research, is the Likert one [17, 18]. Our

purpose was to compare the performance of readers using a LI-RADS scale with that of readers using a Likert scale, in a cohort of patients with chronic liver disease who had undergone MRI for the discovery of a nodule in the sonography examination, using the histopathologic results from biopsy/liver transplant or an MRI follow-up over 4 years as reference standards.

Materials and methods

Patients

This retrospective study was approved by our institutional review board. We reviewed patients with cirrhosis, with no history of previous HCC, who underwent a MR examination, between February 2006 and March 2012, for the presence of new nodules, discovered with sonography. We identified 39 patients (*M/F*:24/15; mean age of 73.1 years; and age range of 54–91 years) with a total of 44 lesions.

For each patient, we registered the following data: age, date of the MR examination, number of nodules, segmental location of nodules, nodule size, radiologist evaluation for each nodule and the definitive diagnosis. This diagnosis was expressed as 0 to indicate “non-evidence of HCC” and as 1 for “histological confirmation of HCC”.

MRI technique

All MR examinations were performed on a 1.5-T system (Avanto; Siemens Medical Systems, Erlangen, Germany) using a phased-array coil for signal detection. All patients underwent axial T1-weighted and T2-weighted sequences and multiphase contrast-enhanced dynamic sequences of the whole liver with fat suppression. T1-weighted imaging included a breath-hold in-phase gradient echo sequence (175/5 TR/TE, 256 × 112 matrix, 70° flip angle) and an out-of-phase gradient echo sequence (175/2.38 TR/TE, 256 × 112 matrix, 70° flip angle). T2-weighted imaging included non-fat-suppressed (3945/66 TR/TE, 320 × 195 matrix) and fat-suppressed sequences (4185/53 TR/TE, 320 × 184 matrix). Dynamic sequences were performed with a T1 three-dimensional volumetric breath-hold examination using the following parameters: 4.7/2.3 TR/TE, 256 × 134 matrix, 10° flip angle, 3-mm slice thickness. Gadolinium (Gadobenate Dimeglumine; Multihance, Bracco, Milan, Italy) was injected at a dose of 0.2 mmol/kg at a rate of 2 mL/second, followed by 0.2 mL/kg of normal saline flush, with a second syringe (MedradStellant, Bayer, Germany) at the same injection rate. Arterial phase was acquired using a real-time bolus-tracking method: 9 s after the contrast entered the celiac axis, the liver was imaged. Additionally, a venous and delayed phase was obtained 60 and 180 s after

contrast administration. A breath-hold in-phase and out-of-phase T1-weighted sequence (175/2.3 TR/TE, 256 × 134 matrix) and a T1 three-dimensional volumetric breath-hold sequence were performed 2 h after contrast injection (hepatocyte phase).

Image interpretation

MRI studies were analyzed independently by two radiologists with, respectively, 10 and 20 years of experience in liver MRI, and they were independently analyzed by two radiologists with, respectively, 1 and 3 months of experience in liver MRI, using Synapse (PACS, Fujifilm Medical Systems, Japan). Two of these radiologists, with 1 month (IradioLIR) and 10 years (EradioLIR) of experience, evaluated the lesions using the LI-RADS scale v.2018, while the other two, respectively, with 3 months (IradioLik) and 20 years (EradioLik) of experience in liver MRI, evaluated the lesions using the Likert scale (scores 1–5). None of the radiologists were aware of the final diagnosis, official reports or clinical information. They only knew that in each liver, there was at least one lesion to characterize.

IradioLik obtained her 3 months MRI liver experience with EradioLik, without using the LI-RADS scale, while IradioLIR and EradioLIR had been training together for 1 month using the LI-RADS scale for MRI characterization of liver nodules.

LI-RADS scale v.2018

The final category of a liver nodule is determined by the following features: (1) the presence or absence of arterial phase hyper-enhancement; (2) the presence or absence of nonperipheral washout; (3) the presence or absence of an enhancing capsule; (4) the size of the nodule; and (5) the threshold growth of the lesion compared to previous examinations. Certain “ancillary features” and “tie-breaking” rules are applied to adjust category for a total of five categories [15, 16, 19, 20]: LR 1 = definitely benign, LR 2 = probably benign, LR 3 = intermediate probability for HCC, LR 4 = probably HCC, LR 5 = definitely HCC.

Likert scale

Nodular liver lesions were categorized as follows: 1 = HCC is highly unlikely to be present, 2 = HCC is unlikely to be present, 3 = the presence of HCC is equivocal, 4 = HCC is likely to be present and 5 = HCC is highly likely to be present.

According to international guidelines, the diagnosis of HCC was favored by the presence of these criteria: early post-contrast arterial enhancement and portal venous/

delayed phase washout, corona enhancement, capsule appearance and threshold growth [13].

Reference standard

Histopathological finding (72.7% of lesions) and MRI follow-up over 4 years (27.3% of lesions) were used as reference standards.

Histopathology was obtained from biopsy (56.8% of lesions) or liver transplantation specimens (15.9% of lesions). At MRI follow-up, a lesion was considered benign when it showed stable or reduced diameters.

Statistical analysis

Before choosing the optimal scale and score for HCC detection, the evaluation of inter-reader agreement was performed by computing linear weighted k coefficients and Pearson correlation coefficients. These analyses allowed us to understand which scale was the most objective and consistent. The k coefficient was interpreted as an indication of poor agreement when k was 0.40 or lower, as an indication of moderate/substantial agreement when k coefficient was higher than 0.40 and lower than 0.80 and as an indication of strong agreement when k coefficient was greater than 0.80. Pearson correlation coefficients were also employed to evaluate the inter-reader agreement and to confirm the results obtained by the k coefficient: Values lower or equal to 0.30 were an indication of poor agreement, values higher than 0.30 and lower or equal to 0.70 indicated moderate agreement, and values higher than 0.70 were evidence of strong inter-reader agreement.

Both for the LI-RADS v.2018 and Likert scales, receiver operating characteristic (ROC) curve analysis was performed to identify, for each evaluator, the optimal score for HCC detection, defined as the score that maximized the evaluator accuracy (ACC). Evaluations from each scale were classified as positive for HCC if the evaluation was equal to or higher than the derived optimal score; otherwise, they were classified as negative for HCC. Analysis of ROC curves allowed the comparison of Likert and LI-RADS v.2018 methods in terms of accuracy, sensitivity and specificity for HCC detection. With this aim, a logistic regression model (using the scores—from 1 to 5—of each evaluator as independent variable and the scores of the gold standard—0 or 1—as dependent variable) was used in each case to derive the probability of each lesion being HCC; fitted data were used to compute ROC curves. Furthermore, for both the LI-RADS and Likert scales, the scores from the two readers were pooled to derive a single optimal score between evaluators using the same scale. To simulate the use of pooled data, for each scale we employed a multiple regression model fitted to the scores of both radiologists as independent variable and the scores of

the gold standard as dependent variable. The z test for the difference between two proportions was applied to check the statistically significant difference among the performance ratios, achieved by the readers of the same scale. All p values were two-sided and were considered to indicate a significant difference at $p < 0.05$. The statistical analysis software was coded in MATLAB.

Results

Since five patients had two lesions to characterize, we evaluated a total of 44 lesions, 26 HCC and 18 non-HCC.

Using the LI-RADS scale, 34 lesions (19 HCC + 15 non-HCC = 77.27% of all the 44 lesions) obtained the same score. The k coefficient between the two evaluators of LI-RADS scale was 0.89, while the estimated Pearson correlation coefficient equaled 0.90.

Using the Likert scale, 22 lesions (11 HCC + 11 non-HCC = 50% of all lesions) were classified with equal

scores. The k coefficient and the Pearson correlation coefficient computed to evaluate the Likert scale inter-reader variability were much lower than those computed for the LI-RADS scale; they equaled, respectively, $k = 0.69$ and $\text{Pearson} = 0.63$.

The ROC curves computed to assess reader performance for HCC detection (Fig. 1) had the following area under the curves (AUCs):

- Using the LI-RADS scale, the AUCs were 0.87 (EradioLIR), 0.75 (IradioLIR), 0.91 (pooled data);
- Using the Likert scale, the AUCs were 0.79 (EradioLik), 0.83 (IradioLik), 0.87 (pooled data).

When pooled data were not considered, ROC analysis showed that the optimal threshold criterion, which allowed maximizing both the accuracy as well as the average of sensitivity (SENS) and specificity (SPEC) in the detection of HCC, was a score of 4 or higher for all the evaluators, using either the LI-RADS or the Likert scale. At this optimal

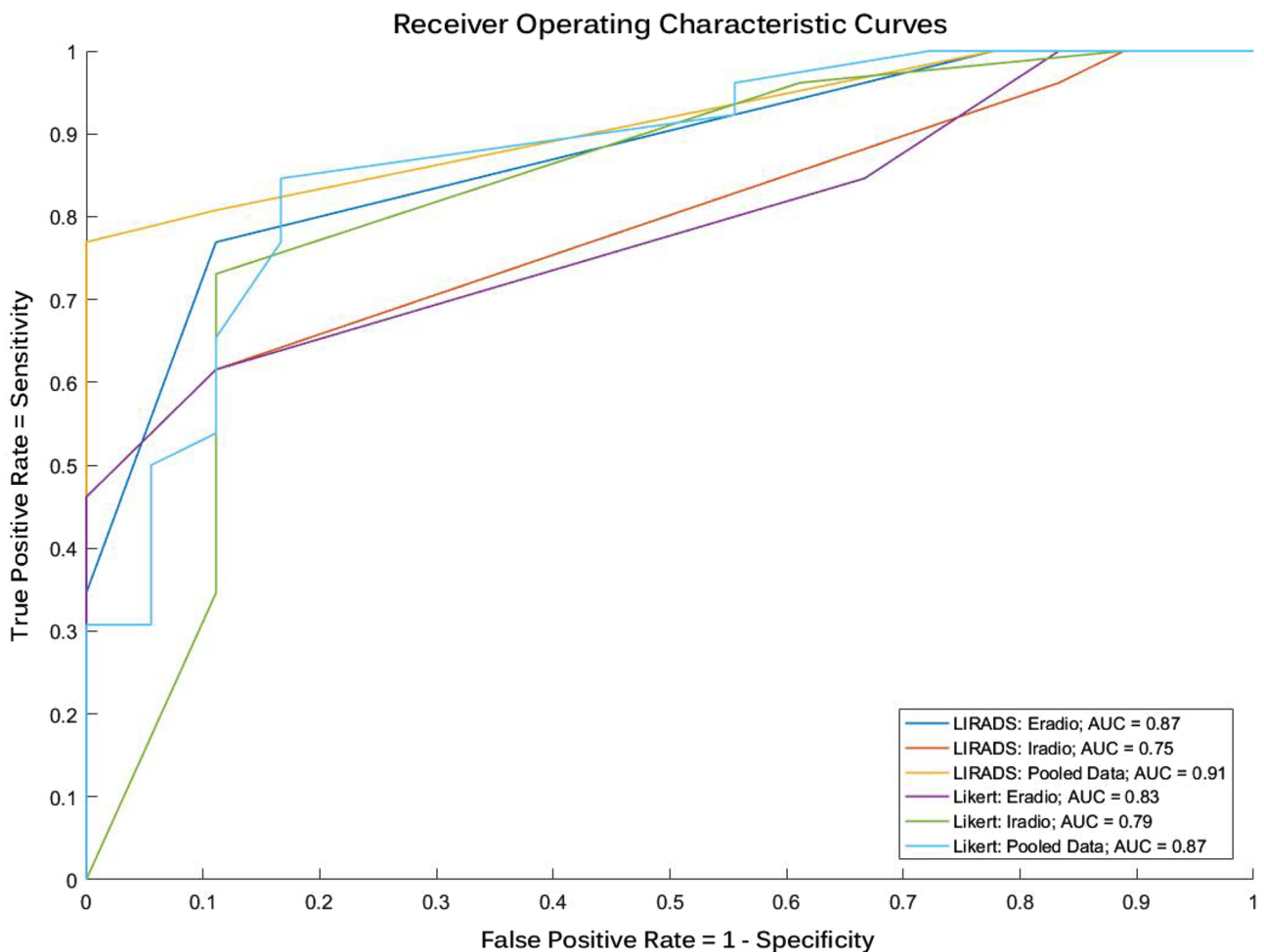


Fig. 1 ROC curves for each evaluator and for pooled data both for the LI-RADS and Likert scale

score, we calculated accuracy (ACC), sensitivity (SENS), specificity (SPEC), positive predictive value (PPV) and negative predictive value (NPV) (Table 1).

At the optimal score, all the evaluators and pooled data achieved also the maximum average of SENS and SPEC.

For both the scales, the reader performance improved when a cooperative diagnostic procedure was simulated by employing pooled data. Specifically, for the LI-RADS and Likert scales, the optimal pooled scores, which are composed by Eradio and Iradio scores, were, respectively, equal to (3, 2) and (2, 4) (Table 1).

To compare the two scales in terms of the achieved radiologist performances, for each scale the mean of the performance values (ACC, SENS, SPEC, PPV and NPV at the optimal scores, AUC) achieved by the two radiologists and the pooled data were computed.

The computed averages were similar, and the *z* test did not find any statistically significant difference between them. For the LI-RADS scale, we obtained: ACC = 0.80, SENS = 0.72, SPEC = 0.93, PPV = 0.93, NPV = 0.70, AUC = 0.85; for the Likert scale, the results were: ACC = 0.79, SENS = 0.73, SPEC = 0.87, PPV = 0.89, NPV = 0.70, AUC = 0.83.

The comparability of the achieved results suggests that neither scale is guaranteed to achieve a better performance than the other.

Discussion

MR is a noninvasive diagnostic modality with a high sensitivity (81%) and specificity (85%) for the detection and evaluation of liver nodules in patients with high risk of HCC [10]. When nodules are larger than 15 mm, the specificity becomes higher (100%) [21]. These characteristics make MRI relevant for the management of patients with suspicion of HCC, mostly for the detection of early HCC, usually composed of well-differentiated hepatocytes [22], thereby avoiding more invasive examinations such as fine-needle biopsy (FNB) samples [9, 21].

However, these results are influenced by the experience of the radiologists who examine the images and some clinicians complain of the lack of standardization in reporting liver MRI nodules.

In an effort to improve consistency of interpretations among radiologists, the introduction of a standardized scheme with fixed criteria, already applied for other organs such as breast, thyroid and prostate, has also been proposed for the liver.

The LI-RADS scale was created to standardize the reporting and data collected in patients with cirrhosis or other risk factors for developing HCC, both with CT and MRI, stratifying the risk of malignancy of a nodule with a scale of scores 1–5 (LR 1–LR 5) [13–16]. In particular, a nodular arterial

phase hyper-enhancement is a very important feature, as, if equal or superior to 2 cm or if in association with at least one of these three major features: the presence of nonperipheral “washout”, the presence of an enhancing “capsule” or a threshold growth, is to be considered either probably HCC (LR 4) or definitely HCC (LR 5) [15, 16, 23].

The Likert scale is a quicker and easier method, based on an analysis of items. In our study, we used five items (scores 1–5) to express a positive or a negative opinion to delineate the probability of a liver nodule to be HCC, based on dynamic enhancement and other features such as corona enhancement, capsule appearance and threshold growth [13].

To study the performance of the LI-RADS scale, we compared the diagnostic performance of readers using the latest version of the LI-RADS scale, updated in 2018, with that of readers using the Likert system.

With this aim, gastroenterologists gave us a homogeneous cohort of patients with chronic liver disease. They selected only patients with no previous HCC, with the first discovery of a nodule in the sonography examination, successively studied with MRI.

In both scales, radiologists performed well and accuracy was high.

Visual analysis of the ROC curves computed to assess reader performances for HCC would suggest that the LI-RADS scale provides the better performance values, since the mean AUC achieved by readers employing the LI-RADS scale equals 0.85, while the one achieved by readers employing the Likert scale is 0.83. Moreover, pooling data of readers employing the LI-RADS scale obtains AUC = 0.91, while the AUC achieved by evaluations made by pooled data of Likert scale equals 0.87.

A similar study was completed by Zhan et al. [24], using an earlier LI-RADS version (v2014).

These authors obtained substantial variations in liver observations between reporting by the LI-RADS and Likert methods. Particularly, there was only a slight agreement between the two methods in classification of probable HCC.

Zhang et al. demonstrated differences in diagnostic accuracy between LI-RADS and Likert (accuracy, 78.6% and 87.2%, $p < 0.001$), but did not evaluate potential differences in inter-reader agreement.

Moreover, they showed substantial discordance between computed tomography (CT) and magnetic resonance (MR) in stratification of hepatic nodules using LI-RADS.

We studied a more homogeneous cohort of patients than Zhang et al. did, because they selected from their database for liver CT or MRI reports in patients suspected of having a hepatic tumor, while we only evaluated subjects who underwent MRI for a nodule detected with sonography.

MRI has a higher per-lesion sensitivity than CT (80 vs. 68%) with a specificity of 94% and should be used as the

Table 1 Performance values for each observer score

Score	ACC	Sens	Spec	PPV	NPV	TP	TN	FP
<i>LI-RADS scale</i>								
Eradio LI-RADS	AUC	0.87						
Score > [5]	0.41	0.00	1.00	NaN*	0.41	0	18	0
Score ≥ [5]	0.61	0.35	1.00	1.00	0.51	9	18	0
Score ≥ [4]	0.82	0.77	0.89	0.91	0.73	20	16	2
Score ≥ [3]	0.68	1.00	0.22	0.65	1.00	26	4	14
Score ≥ [2]	0.64	1.00	0.11	0.62	1.00	26	2	16
Score ≥ [1]	0.59	1.00	0.00	0.59	NaN*	26	0	18
Iradio LI-RADS	AUC	0.75						
Score > [5]	0.41	0.00	1.00	NaN*	0.41	0	18	0
Score ≥ [5]	0.57	0.35	0.89	0.82	0.48	9	16	2
Score ≥ [4]	0.73	0.62	0.89	0.89	0.62	16	16	2
Score ≥ [3]	0.64	0.96	0.17	0.63	0.75	25	3	15
Score ≥ [2]	0.64	1.00	0.11	0.62	1.00	26	2	16
Score ≥ [1]	0.59	1.00	0.00	0.59	NaN*	26	0	18
Pooled data: Eradio + Iradio LI-RADS	AUC	0.91						
Score > [5, 4]	0.41	0.00	1.00	NaN*	0.41	0	18	0
Score ≥ [5, 4]	0.43	0.04	1.00	1.00	0.42	1	18	0
Score ≥ [5, 5]	0.61	0.35	1.00	1.00	0.51	9	18	0
Score ≥ [4, 3]	0.70	0.50	1.00	1.00	0.58	13	18	0
Score ≥ [4, 4]	0.84	0.73	1.00	1.00	0.72	19	18	0
Score ≥ [3, 2]	0.86	0.77	1.00	1.00	0.75	20	18	0
Score ≥ [4, 5]	0.84	0.81	0.89	0.91	0.76	21	16	2
Score ≥ [3, 3]	0.68	1.00	0.22	0.65	1.00	26	4	14
Score ≥ [2, 2]	0.66	1.00	0.17	0.63	1.00	26	3	15
Score ≥ [2, 3]	0.64	1.00	0.11	0.62	1.00	26	2	16
Score ≥ [1, 1]	0.59	1.00	0.00	0.59	NaN*	26	0	18
<i>Likert scale</i>								
EradioLikert	AUC	0.79						
Score > [5]	0.41	0.00	1.00	NaN*	0.41	0	18	0
Score ≥ [5]	0.68	0.46	1.00	1.00	0.56	12	18	0
Score ≥ [4]	0.73	0.62	0.89	0.89	0.62	16	16	2
Score ≥ [3]	0.64	0.85	0.33	0.65	0.60	22	6	12
Score ≥ [2]	0.66	1.00	0.17	0.63	1.00	26	3	15
Score ≥ [1]	0.59	1.00	0.00	0.59	NaN*	26	0	18
IradioLikert	AUC	0.83						
Score > [5]	0.41	0.00	1.00	NaN*	0.41	0	18	0
Score ≥ [5]	0.57	0.35	0.89	0.82	0.48	9	16	2
Score ≥ [4]	0.80	0.73	0.89	0.90	0.70	19	16	2
Score ≥ [3]	0.73	0.96	0.39	0.69	0.88	25	7	11
Score ≥ [2]	0.64	1.00	0.11	0.62	1.00	26	2	16
Score ≥ [1]	0.59	1.00	0.00	0.59	NaN*	26	0	18
Pooled data: Iradio + EradioLikert	AUC	0.87						
Score > [5, 5]	0.41	0.00	1.00	NaN*	0.41	0	18	0
Score ≥ [5, 5]	0.59	0.31	1.00	1.00	0.50	8	18	0
Score ≥ [4, 5]	0.57	0.31	0.94	0.89	0.49	8	17	1
Score ≥ [5, 4]	0.66	0.46	0.94	0.92	0.55	12	17	1
Score ≥ [4, 4]	0.68	0.50	0.94	0.93	0.57	13	17	1
Score ≥ [2, 5]	0.68	0.54	0.89	0.88	0.57	14	16	2
Score ≥ [3, 4]	0.75	0.65	0.89	0.89	0.64	17	16	2
Score ≥ [4, 3]	0.80	0.77	0.83	0.87	0.71	20	15	3

Table 1 (continued)

Score	ACC	Sens	Spec	PPV	NPV	TP	TN	FP
Score \geq [2, 4]	0.84	0.85	0.83	0.88	0.79	22	15	3
Score \geq [3, 3]	0.73	0.92	0.44	0.71	0.80	24	8	10
Score \geq [2, 3]	0.75	0.96	0.44	0.71	0.89	25	8	10
Score \geq [3, 2]	0.70	1.00	0.28	0.67	1.00	26	5	13
Score \geq [1, 3]	0.68	1.00	0.22	0.65	1.00	26	4	14
Score \geq [2, 2]	0.64	1.00	0.11	0.62	1.00	26	2	16
Score \geq [1, 1]	0.59	1.00	0.00	0.59	NaN*	26	0	18

NaN*not a number

preferred imaging modality for HCC diagnosis in patients at high risk [25].

MRI performed with a hepatobiliary agent has a significantly greater per-lesion sensitivity (87 vs. 74%) compared with MRI performed with an extracellular agent [26].

Choi et al. [27] demonstrated that LI-RADS v2014 LR-5 on gadoxetate disodium-enhanced MRI exhibited an excellent PPV for the diagnosis of HCC in patients with chronic liver disease and for some authors, it should be incorporated into LI-RADS as a major feature [9, 28, 29].

Evaluating readers using the same scale, both for LI-RADS v.2018 and Likert scale, we found that the differences between the SENS, SPEC and NPV values were statistically significant, but when we compared the performance achieved by employing different scales, we did not find any statistically significant difference between the aforementioned average performances, as obtained by Barth et al. [30].

The two radiologists who used LI-RADS v.2018 scale assigned the same score in 73.07% of HCC lesions and 83.33% of non-HCC lesions. Using Likert scale, the same score was assigned in 42.30% of HCC lesions and in 61.11% of non-HCC lesions.

This resulted in a strong agreement between the two evaluators of LI-RADS v.2018 scale (k coefficient = 0.89), while on the other hand, there was only a moderate agreement between the two evaluators of Likert scale (k coefficient = 0.69). These values, as stated by Barth et al. [30], were also confirmed by Pearson's correlation coefficient (LI-RADS v.2018 = 0.90; Likert = 0.63). These results suggest that the use of the LI-RADS v.2018 scale for HCC recognition provides a minor inter-reader difference compared with the Likert scale.

Ren et al. [26] compared the diagnostic performance of LI-RADS on MR for diagnosing HCC between v.2017 and v.2018, but they did not calculate the inter-reader agreement between the observers as we did to show the reproducibility of the method.

According to Petruzzi et al. [31], the LI-RADS scale produces strong reliability and validity, while aiming to

improve the clarity of clinical MRI reports [31]. Our findings confirmed those of Petruzzi et al.

A primary goal and motivation for the development of the LI-RADS scale is to reduce the inter-reader variability.

Clinicians have expressed serious concern over the inconsistency in liver lesion reporting among radiologists owing to the reader's experience level, differences in interpretation as well as differing personal preferences [32].

In our study, the LI-RADS scale has demonstrated the potential to reduce inter-reader variability and to enhance the communication with referring clinicians, as also observed by other authors [33].

There are several limitations in our study. A primary limitation is that this is a retrospective study with a small sample size of lesions. Second, the radiologists are from the same hospital and may have similar perspectives in terms of interpretation, and third, the experience of the radiologists involved in the study varied greatly. Fourth, we used a mixed reference standard of pathology and MRI follow-up over 4 years. Finally, as a contrast medium, for the evaluation of wash-in and washout and the hepatobiliary phase, we used only gadobenate dimeglumine and not gadoxetate disodium.

Conclusions

According to our results, there was not a statistically significant difference between LI-RADS v.2018 scale and Likert scale in the evaluation of liver nodules and detection of HCC; nevertheless, we suggest the use of LI-RADS v.2018 scale, because it appears more objective and consistent.

Our results reached the goal of LI-RADS which is to improve the consistency of radiology reports in imaging of high-risk patients for HCC.

Additional studies are warranted to evaluate the performance of these scales across radiologists from different institutions with different levels of experience.

Funding This study did not receive any funding.

Compliance with ethical standards

Conflict of interest All the authors declare that they have no conflict of interest.

Ethical standards This article does not contain any studies with animals performed by any of the authors. All procedures performed were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

- Sersté T, Barrau V, Ozenne V et al (2012) Accuracy and disagreement of computed tomography and magnetic resonance imaging for the diagnosis of small hepatocellular carcinoma and dysplastic nodules: role of biopsy. *Hepatology* 55(3):800–806. <https://doi.org/10.1002/hep.24746>
- Altekruse SF, Henley SJ, Cucinelli JE, McGlynn KA (2014) Changing hepatocellular carcinoma incidence and liver cancer mortality rates in the United States. *Am J Gastroenterol* 109(4):542–553. <https://doi.org/10.1038/ajg.2014.11>
- Törner A, Stokkeland K, Svensson Å et al (2017) The underreporting of hepatocellular carcinoma to the Cancer Register and a log-linear model to estimate a more correct incidence. *Hepatology* 65(3):885–892. <https://doi.org/10.1002/hep.28775>
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68(6):394–424. <https://doi.org/10.3322/caac.21492>
- Ringehan M, McKeating JA, Protzer U (2017) Viral hepatitis and liver cancer. *Philos Trans R Soc B* 372(1732):20160274. <https://doi.org/10.1098/rstb.2016.0274>
- Mittal S, El-Serag HB (2013) Epidemiology of hepatocellular carcinoma: consider the population. *J Clin Gastroenterol* 47(Suppl):S2–S6. <https://doi.org/10.1097/mcg.0b013e3182872f29>
- Jha RC, Mitchell DG, Weinreb JC et al (2014) LI-RADS categorization of benign and likely benign findings in patients at risk of hepatocellular carcinoma: a pictorial atlas. *AJR Am J Roentgenol* 203(1):48–69. <https://doi.org/10.2214/ajr.13.12169>
- Kim SY, An J, Lim YS et al (2017) MRI with liver-specific contrast for surveillance of patients with cirrhosis at high risk of hepatocellular carcinoma. *JAMA Oncol* 3(4):456–463. <https://doi.org/10.1001/jamaoncol.2016.3147>
- Rosiak G, Podgorska J, Rosiak E, Cieszanowski A (2018) Comparison of LI-RADS vol 2017 and ESGAR guidelines imaging criteria in HCC diagnosis using MRI with hepatobiliary contrast agents. *Biomed Res Int*. <https://doi.org/10.1155/2018/7465126>
- Marrero JA, Hussain HK, Nghiem HV, Umar R, Fontana RJ, Lok AS (2005) Improving the prediction of hepatocellular carcinoma in cirrhotic patients with an arterially-enhancing liver mass. *Liver Transpl* 11(3):281–289. <https://doi.org/10.1002/lt.20357>
- Lima PH, Fan B, Bérubé J et al (2019) Cost-utility analysis of imaging for surveillance and diagnosis of hepatocellular carcinoma. *AJR Am J Roentgenol* 17:1–9. <https://doi.org/10.2214/ajr.18.20341>
- Neri E, Bali MA, Ba-Ssalamah A et al (2016) ESGAR consensus statement on liver MR imaging and clinical use of liver-specific contrast agents. *Eur Radiol* 26(4):921–931. <https://doi.org/10.1007/s00330-015-3900-3>
- Cruite I, Tang A, Sirlin CB (2013) Imaging-based diagnostic systems for hepatocellular carcinoma. *AJR Am J Roentgenol* 201(1):41–55. <https://doi.org/10.2214/ajr.13.10570>
- Santillan CS, Tang A, Cruite I, Shah A, Sirlin CB (2014) Understanding LI-RADS. A primer for practical use. *Magn Reson Imaging Clin N Am* 22(3):337–352. <https://doi.org/10.1016/j.mric.2014.04.007>
- Chernyak V, Fowler KJ, Kamaya A, Kielar AZ, Elsavés KM, Bashir MR, Kono Y, Do RK, Mitchell DG, Singal AG, Tang A, Sirlin CB (2018) Liver Imaging Reporting and Data System (LI-RADS) version 2018: imaging of hepatocellular carcinoma in at-risk patients. *Radiology* 289:816–830. <https://doi.org/10.1148/radiol.2018181494>
- Chernyak V, Tang A, Flusberg M, Papadatos D, Bijan B, Kono Y, Santillan C (2018) LI-RADS® ancillary features on CT and MRI. *Abdom Radiol* 43(1):82–100. <https://doi.org/10.1007/s00261-017-1220-6>
- Likert R (1932) A technique for the measurement of attitudes. *Arch Psychol* 22(140):1–55
- Villers A, Puech P, Mouton D, Leroy X, Ballereau C, Lemaitre L (2006) Dynamic contrast enhanced, pelvic phased array magnetic resonance imaging of localized prostate cancer for predicting tumor volume: correlation with radical prostatectomy findings. *J Urol* 176(6 pt 1):2432–2437
- Mitchell DG, Bruix J, Sherman M, Sirlin CB (2015) LI-RADS (Liver Imaging Reporting and Data System): summary, discussion, and consensus of the LI-RADS management working group and future directions. *Hepatology* 61(3):1056–1065. <https://doi.org/10.1002/hep.27304>
- Sofue K, Burke LMB, Nilmini V et al (2017) Liver imaging reporting and data system category 4 observations in MRI: risk factors predicting upgrade to category 5. *J Magn Reson Imaging* 46(3):783–792. <https://doi.org/10.1002/jmri.25627>
- Forner A, Vilana R, Ayuso C et al (2008) Diagnosis of hepatic nodules 20 mm or smaller in cirrhosis: prospective validation of the noninvasive diagnostic criteria for hepatocellular carcinoma. *Hepatology* 47(1):97–104. Erratum in: *Hepatology* 47(2):769. <https://doi.org/10.1002/hep.21966>
- Kojiro M, Roskams T (2005) Early hepatocellular carcinoma and dysplastic nodules. *Semin Liver Dis* 25(2):133–142. <https://doi.org/10.1055/s-2005-871193>
- Fraum TJ, Tsai R, Rohe E et al (2018) 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* 286(1):158–172. <https://doi.org/10.1148/radiol.2017170114>
- Zhang YD, Zhu FP, Xu X, Wang Q, Wu CJ, Liu XS, Shi HB (2016) Classifying CT/MR findings in patients with suspicion of hepatocellular carcinoma: comparison of liver imaging reporting and data system and criteria-free Likert scale reporting models. *J Magn Reson Imaging* 43(2):373–383. <https://doi.org/10.1002/jmri.24987>
- Lee YJ, Lee JM, Lee JS, Lee HY, Park BH, Kim YH, Han JK, Choi BI (2015) Hepatocellular carcinoma: diagnostic performance of multidetector CT and MR imaging—a systematic review and meta-analysis. *Radiology* 275(1):97–109. <https://doi.org/10.1148/radiol.14140690>
- Ren AH, Zhao PF, Yang DW, Du JB, Wang ZC, Yang ZH (2019) Diagnostic performance of MR for hepatocellular carcinoma based on LI-RADS v2018, compared with v2017. *J Magn Reson Imaging*. <https://doi.org/10.1002/jmri.26640>
- Choi SH, Byun JH, Kim SY et al (2016) Liver Imaging Reporting and Data System v2014 with gadoxetate disodium-enhanced magnetic resonance imaging: validation of LI-RADS category 4 and 5 criteria. *Invest Radiol* 51(8):483–490. <https://doi.org/10.1097/rli.0000000000000258>

28. Cha DI, Jang KM, Kim SH et al (2017) Liver Imaging Reporting and Data System on CT and gadoxetic acid-enhanced MRI with diffusion-weighted imaging. *Eur Radiol* 27(10):4394–4405. <https://doi.org/10.1007/s00330-017-4804-1>
29. Joo I, Lee JM, Lee DH et al (2017) Liver Imaging Reporting and Data System v2014 categorization of hepatocellular carcinoma on gadoxetic acid-enhanced MRI: comparison with multiphase multidetector computed tomography. *J Magn Reson Imaging* 45(3):731–740. <https://doi.org/10.1002/jmri.25406>
30. Barth BK, Donati OF, Fischer MA et al (2016) Reliability, validity, and reader acceptance of LI-RADS-an in-depth analysis. *Acad Radiol* 23(9):1145–1153. <https://doi.org/10.1016/j.acra.2016.03.014>
31. Petruzzi N, Mitchell D, Guglielmo F et al (2013) Hepatocellular carcinoma likelihood on MRI exams: evaluation of a standardized categorization system. *Acad Radiol* 20(6):694–698. <https://doi.org/10.1016/j.acra.2013.01.016>
32. Davenport MS, Khalatbari S, Liu PS, Maturen KE, Kaza RK, Wasnik AP, Al-Hawary MM, Glazer DI, Stein EB, Patel J, Somashekar DK, Viglianti BL, Hussain HK (2014) Repeatability of diagnostic features and scoring systems for hepatocellular carcinoma by using MR imaging. *Radiology* 272(1):132–142. <https://doi.org/10.1148/radiol.14131963>
33. Sun HY, Lee JM, Shin CI, Lee DH, Moon SK, Kim KW, Han JK, Choi BI (2010) Gadoteric acid-enhanced magnetic resonance imaging for differentiating small hepatocellular carcinomas (< or = 2 cm in diameter) from arterial enhancing pseudolesions: special emphasis on hepatobiliary phase imaging. *Invest Radiol* 45(2):96–103. <https://doi.org/10.1097/rli.0b013e3181c5faf7>

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