



# Diagnostic Performance of Deep Learning-Based Lesion Detection Algorithm in CT for Detecting Hepatic Metastasis from Colorectal Cancer

Kiwook Kim, Sungwon Kim, Kyunghwa Han, Heejin Bae, Jaeseung Shin, Joon Seok Lim

All authors: Department of Radiology, Research Institute of Radiological Science and Center for Clinical Image Data Science, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

**Objective:** To compare the performance of the deep learning-based lesion detection algorithm (DLLD) in detecting liver metastasis with that of radiologists.

**Materials and Methods:** This clinical retrospective study used 4386-slice computed tomography (CT) images and labels from a training cohort (502 patients with colorectal cancer [CRC] from November 2005 to December 2010) to train the DLLD for detecting liver metastasis, and used CT images of a validation cohort (40 patients with 99 liver metastatic lesions and 45 patients without liver metastasis from January 2011 to December 2011) for comparing the performance of the DLLD with that of readers (three abdominal radiologists and three radiology residents). For per-lesion binary classification, the sensitivity and false positives per patient were measured.

**Results:** A total of 85 patients with CRC were included in the validation cohort. In the comparison based on per-lesion binary classification, the sensitivity of DLLD (81.82%, [81/99]) was comparable to that of abdominal radiologists (80.81%,  $p = 0.80$ ) and radiology residents (79.46%,  $p = 0.57$ ). However, the false positives per patient with DLLD (1.330) was higher than that of abdominal radiologists (0.357,  $p < 0.001$ ) and radiology residents (0.667,  $p < 0.001$ ).

**Conclusion:** DLLD showed a sensitivity comparable to that of radiologists when detecting liver metastasis in patients initially diagnosed with CRC. However, the false positives of DLLD were higher than those of radiologists. Therefore, DLLD could serve as an assistant tool for detecting liver metastasis instead of a standalone diagnostic tool.

**Keywords:** Artificial intelligence; Colorectal neoplasms; Neoplasm metastasis; X-ray computed tomography; Computer-assisted diagnosis

## INTRODUCTION

The liver is the most common site of distant metastasis

**Received:** April 14, 2020 **Revised:** November 23, 2020

**Accepted:** December 4, 2020

This study was supported by a Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07048179).

**Corresponding author:** Sungwon Kim, MD, Department of Radiology, Research Institute of Radiological Science and Center for Clinical Image Data Science, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.

• E-mail: DINBE@yuhs.ac

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

of colorectal cancer (CRC) [1], which is the second most common cause of cancer-related deaths worldwide [2]. Several population-based studies have detected the rate of cumulative liver metastasis to be 25–30% among patients diagnosed with CRC [3–5].

Computed tomography (CT) is a non-invasive and reliable method for liver assessment and is considered to be one of the standard imaging modalities for preoperative detection of liver metastasis and postoperative surveillance in patients with CRC [6,7]. It has been observed that performing curative hepatic resection in the early stages of cancer metastasis increases the chances of survival among patients with CRC [3,5,8]. Therefore, early detection of CRC liver metastasis using imaging modalities is an essential part of the preoperative cancer workup.

However, detection of metastatic lesions in the liver is an

arduous and time-consuming task owing to the small size of the early metastatic lesions and a variety of benign lesions that obstruct the radiologist's line of sight. Several studies have been conducted to develop computer-aided detection of hepatic focal lesions on CT images using machine-learning methods [9-11].

Deep convolutional neural networks (CNNs), a class of machine-learning methods, perform with high efficiency in the field of medical image analysis [12-14]. Additionally, several articles have been published on studies that used CNNs for the analysis of hepatic lesions [15-17]. Patients with CRC diagnosed by histologic examination through colonoscopy biopsy routinely undergo abdominopelvic CT for the preoperative staging of CRC [18], and the detection of liver metastasis in this CT staging necessitates a significant alteration of the treatment strategy. Therefore, attempts were made to examine the performance of a state-of-the-art CNN algorithm for lesion detection at this stage in the CRC workup and to evaluate the merits and demerits of the algorithm.

The purpose of this study was to evaluate the performance of the deep learning-based lesion detection algorithm (DLLD) in detecting liver metastasis in cancer workup settings and compare its performance with that of

radiologists.

## MATERIALS AND METHODS

### Ethical Approvals and Study Population Overlap

This retrospective study was approved by the Institutional Review Board. The requirement for informed consent was waived (IRB No. 4-2019-0187). There was a partial overlap in the study population between this study and a previous study [19]. These patients were exclusively included in the training cohort in our study.

### Study Population

The study population comprised a training cohort and temporally independent validation cohort (Fig. 1). For the training cohort, electronic medical records were retrospectively searched, and 4871 recently diagnosed colorectal adenocarcinoma patients were identified between November 2005 and December 2010. Among these patients, 624 consecutive patients who underwent pretreatment contrast-enhanced abdominopelvic CT followed by contrast-enhanced liver magnetic resonance imaging (MRI) to characterize the undetermined hepatic focal lesion were

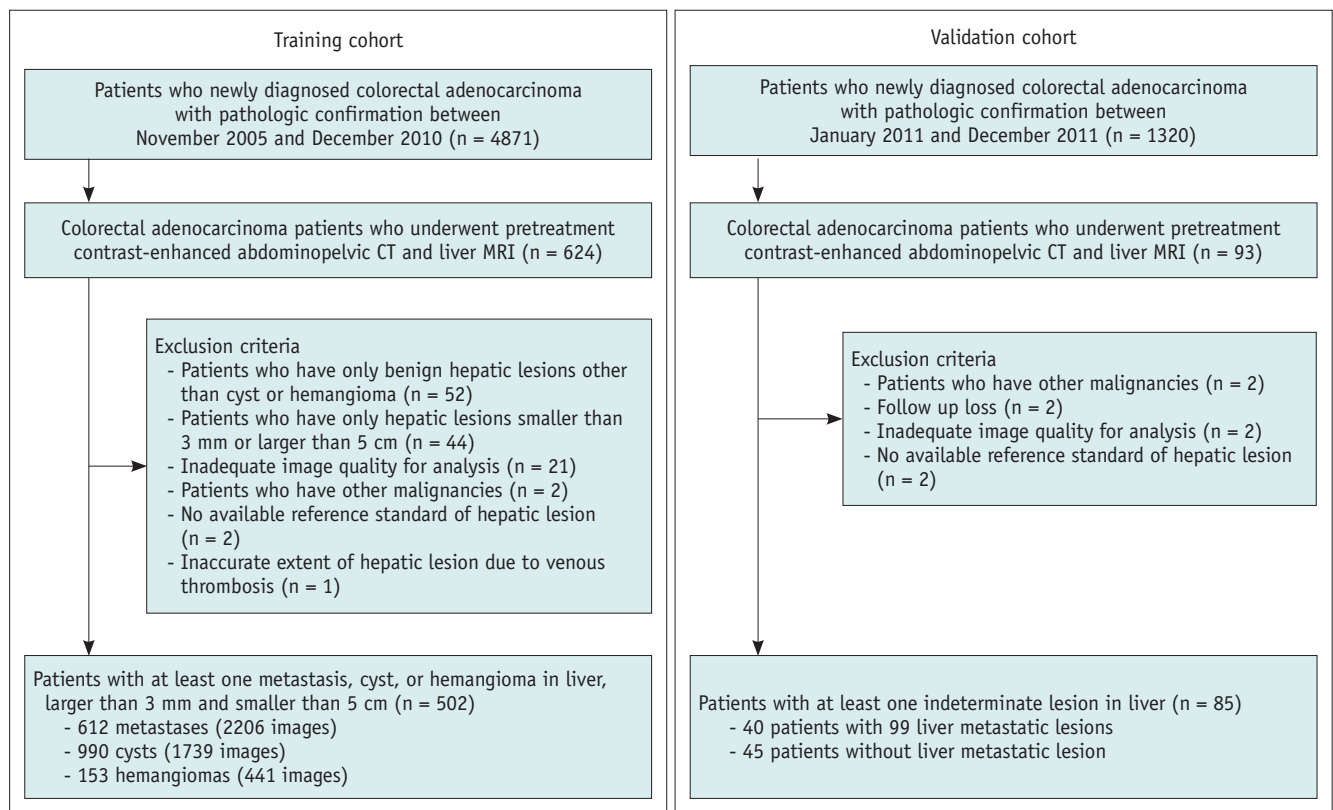


Fig. 1. Flow chart of the study population. CT = computed tomography, MRI = magnetic resonance imaging

identified. A total of 122 patients were excluded for the following reasons: 1) Fifty-two patients had only benign hepatic lesions other than cysts or hemangiomas; 2) Forty-four patients had hepatic lesions smaller than 3 mm or larger than 5 cm; 3) The CT image quality of 21 patients was inadequate for analysis; 4) Two patients had other malignancies; 5) The hepatic lesion in two patients did not meet the available reference standard; 6) The extent of hepatic lesion was inaccurate in one patient owing to venous thrombosis. Subsequently, 502 patients (a total of 4386 slice CT images in the portal phase) with 612 metastases (2206 images), 990 cysts (1739 images), and 153 hemangiomas (441 images) were included.

All lesions were confirmed using the following procedure. When surgical resection or biopsy was performed, the diagnosis was confirmed using pathology results. If pathological confirmation was unobtainable, typical MRI findings and imaging follow-up for a minimum of 1 year were used to characterize the hepatic lesions (Supplementary Materials 1).

Next, for the validation cohort, 1320 patients diagnosed with colorectal adenocarcinoma were identified for the first time between January 2011 and December 2011. Of these, 93 patients who underwent pretreatment contrast-enhanced abdominopelvic CT, followed by contrast-enhanced liver MRI owing to hepatic focal lesions and resectability evaluation, were identified. Eight patients were excluded for the following reasons: 1) two patients had other malignancies, 2) two patients were lost to follow-up studies, 3) the CT image quality of two patients was found to be inadequate for analysis; and 4) the hepatic lesions of two patients did not meet the available reference standard.

Finally, a group of 85 patients comprising 40 patients with 99 liver metastatic lesions and 45 patients without liver metastasis was selected for the validation cohort. Consequently, 229 cysts and 21 hemangiomas of these 85 patients were included in the validation cohort.

### Image Acquisition

The minimum requirement of the CT protocol is a portal venous phase with a thickness of 3 mm or 5 mm. Liver MRI was performed on patients suspected of having resectable liver metastases on CT scans, patients with an indeterminate lesion on CT, or high-risk patients at the discretion of the clinicians. Detailed technical information and imaging protocols for CT and MRI are provided in Supplementary Materials 2.

### Development of DLLD

CT images of the portal venous phase were used for the training and validation of DLLD. First, an abdominal radiologist with 1 year of experience detected all the metastases, cysts, and hemangiomas in the CT axial images of the patients included in the training cohort, drew a rectangular region-of-interest bounding each lesion, and recorded the type of each lesion. YOLOv3 [20], a state-of-the-art CNN object detection model, was used for DLLD architecture (<https://github.com/pjreddie/darknet>), and DLLD was initialized using the transfer learning method with a pre-trained volume in ImageNet. Detailed technical information regarding the YOLOv3 model is provided in Supplementary Materials 3, Supplementary Figure 1. DLLD was trained using a fully supervised learning method by inputting the converted CT images and labels indicating the type and location of the lesions. A false positive filtering method was applied to the post-processing step of lesion detection (Supplementary Materials 4, Supplementary Fig. 2). Three class-lesion training (metastasis, cyst, and hemangioma) was performed to teach the differences between different types of common hepatic lesions (Supplementary Materials 5, Supplementary Table 1).

### Performance Evaluation of DLLD

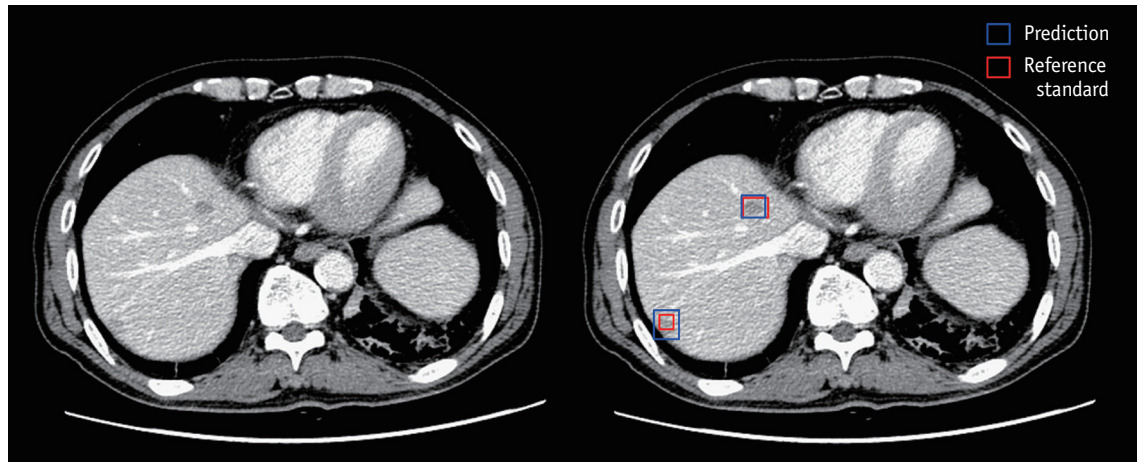
The trained DLLD received CT images of each patient in the validation cohort and derived a binary decision and a confidence score between 0 and 100 for each detected lesion (Fig. 2). Existing object detection models provide predictions per slice. Clinically, however, providing predictions per lesion is more practical. Thus, DLLD was designed to automatically derive predictions per lesion. First, lesions connected between adjacent slices were identified and recognized as one lesion. Next, the average of the metastasis scores of all the slices was defined as the metastasis confidence score of the corresponding lesion. For the binary decision, the class with the highest confidence score was considered to be the DLLD decision for the lesion (Fig. 3). As DLLD was designed for the detection and classification of liver lesions, the detection markings outside the liver were excluded from this evaluation.

### Observer Performance

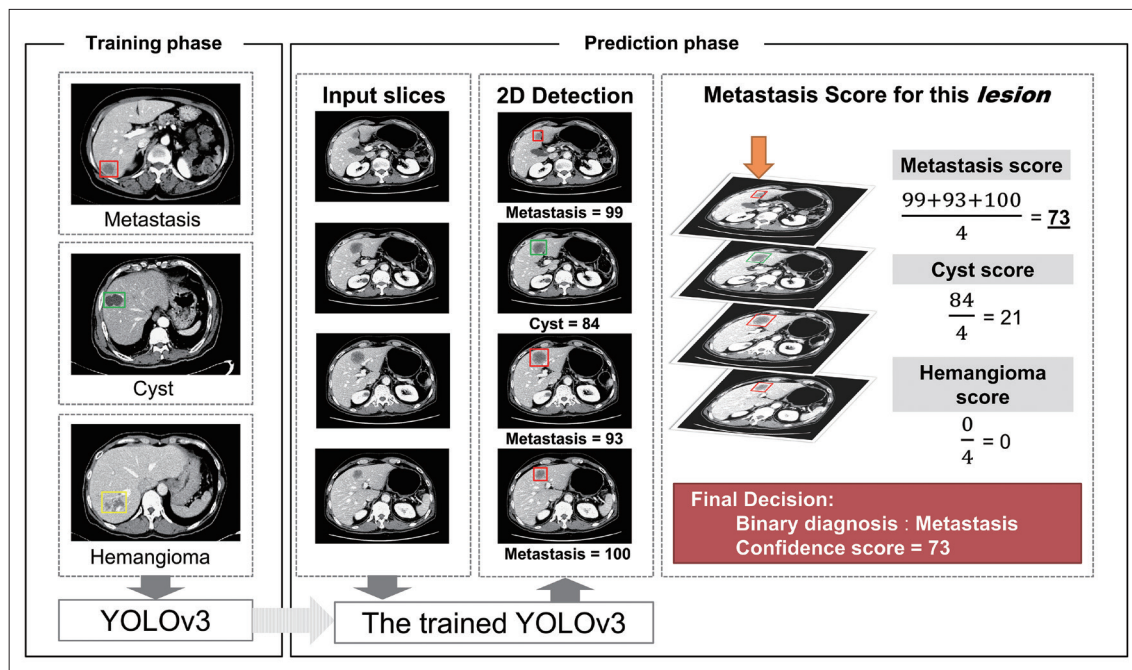
Observer performance tests were conducted to compare the performance of DLLD with that of the readers. The readers comprised two groups: three abdominal radiologists with 2, 3, and 20 years of experience in liver imaging, and

three second-year radiology residents. The readers were informed about the presence of CRC but not about the presence of hepatic metastasis and other clinicopathologic histories. The readers performed the hepatic metastasis

detection and grading task: detecting lesions with suspected hepatic metastases, marking the location of each suspected metastasis, and recording the confidence score of each marked lesion on a five-point scale (1 =



**Fig. 2. Example computed tomography images of a male patient aged 71 years diagnosed with rectal cancer.** There are two pathologically confirmed liver metastases. One 16-mm lesion is present in the right anterosuperior liver segment, and another 7-mm lesion in the right posterosuperior liver segment. The DLLD detects the lesion in the right anterosuperior liver segment and classifies it into a metastasis class with a 100-confidence score, and all six readers detect and properly classify it with a four- or five-point scale confidence score. The DLLD detects the lesion in the right posterosuperior liver segment and classifies it into a metastasis class with a 77-confidence score. However, none of the readers mark this lesion. DLLD = deep learning-based lesion detection algorithm



**Fig. 3. Schematic diagram of the DLLD architecture.** The YOLOv3 model was trained in a fully supervised manner using the 3-class CT images of the training cohort and the corresponding labels indicating the class and location of lesions. In the prediction phase, the trained YOLOv3 model finds hepatic focal lesions and predicts information about the detected lesions (location, class, and confidence score for each class) for each CT image in the validation set. Next, an abdominal radiologist manually identifies the lesions that are connected between the adjacent slices, recognizes them as one lesion, and calculates the confidence score for each class using the average of the confidence scores of all slices containing this lesion. If the confidence score of the metastasis is the highest among all classes, the binary classification result of the lesion is “metastasis.” Otherwise, the lesion is considered benign. CT = computed tomography, DLLD = deep learning-based lesion detection algorithm, 2D = two-dimensional

probably benign; 2 = indeterminate; 3 = possible metastasis with more than 50% confidence; 4 = probable metastasis with high confidence; 5 = definitely metastasis). They were then informed that lesions with a confidence score  $\geq 3$  were considered positive when analyzed using binary classification [13,21].

### Statistical Analysis

For evaluating the per-lesion diagnostic performance of DLLD and readers for the liver metastasis detection task based on the confidence score, the area under the alternative free-response receiver operating characteristic curve (AUA-FROC) was computed by performing an alternative free-response receiver-operating characteristic analysis [22]. Diagnostic performances of DLLD and the readers were compared using the 95% confidence interval (CI) for the difference in AUA-FROC values. The 95% CI was estimated using the bootstrap method with 1000 resamples.

For per-lesion binary classification, the sensitivity and false positives per patient (FPP) of DLLD and readers were measured and compared using generalized estimating equations. For the per-patient binary classification, patients with a minimum of one hepatic metastatic lesion were considered to exhibit a positive case of CRC metastasis, and the sensitivity and specificity of DLLD and readers were measured and compared using generalized estimating equations.

When performing the per-lesion binary classification, the Fleiss' Kappa statistic was used to analyze the intra-group inter-observer agreements among readers in each reader group. The kappa values were interpreted based on the guidelines provided by Landis and Koch [23].

A *p* value of less than 0.05 was considered statistically significant. All statistical analyses were performed using R (version 3.5.2; R Foundation for Statistical Computing) and SAS (version 9.4, SAS Institute Inc.).

## RESULTS

### Patient and Liver Metastatic Lesion Characteristics

The baseline characteristics of the patients and liver metastases in the validation cohort are listed in Table 1. Among the 99 CRC liver metastatic lesions from 40 patients, 63 were histologically confirmed after hepatic resection or percutaneous biopsy, and 36 lesions were diagnosed through MRI findings and upon follow-up imaging for a minimum of 1 year. The mean size of the metastatic lesions was  $2.2 \pm 2.2$

cm. Of the 99 metastatic lesions (27.3%), 27 were  $< 1$  cm in size. Nineteen patients (47.5%) had a solitary metastatic lesion, 11 patients (27.5%) had two or three metastatic lesions, and 10 patients (25%) had four or more metastatic lesions. The baseline characteristics of the patients in the training cohort are listed in Supplementary Table 2.

### Lesion-Based Diagnostic Performance of DLLD and Readers

Upon comparing the performance between DLLD and readers using the confidence score, it was found that the

**Table 1. Baseline Demographics and Clinical Characteristics of the Validation Cohort**

Variables*	Value
Mean age (year) ( $\pm$ SD)	67.5 ( $\pm$ 12.1), range 35–95
Sex, n (%)	
Male	54 (63.5)
Female	31 (36.5)
CRC location	
Colon	63 (74.1)
Rectum	22 (25.9)
CEA (ng/dL), median (IQR)	
CRLM group	31.8 (8.4–123.5)
Non-CRLM group	3.0 (1.9–8.3)
T stage (tumor invasion depth), n (%)	
T1–2 (confined to the bowel wall)	12 (14.1)
T3–4 (beyond the bowel wall)	73 (85.9)
N stage (nodal involvement), n (%)	
N- (node negative)	25 (29.4)
N+ (node positive)	60 (70.6)
M stage, n (%)	
M0	44 (51.8)
M1	41 (48.2)
Liver only	25 (61.0)
Lung only	1 (2.4)
Liver plus extrahepatic	15 (36.6)
Size of CRLM (cm), mean	2.2 ( $\pm$ 2.2), range 0.4–12.4
Number of CRLM per patient, mean ( $\pm$ SD)	2.5 ( $\pm$ 2.0), range 1–8
Number of CRLM $\leq 1$ cm, n (%)*	27 (27.3)
Confirmatory method for CRML, n (%)*	
Histopathology after hepatic resection or percutaneous biopsy	63 (63.6)
Suspicious MRI finding and follow-up imaging study	36 (36.4)

Values represent the number of subjects (%), median (IQR), or mean ( $\pm$  SD). \*The total number of CRLMs was 99. CEA = serum carcinoembryonic antigen, CRC = colorectal cancer, CRLM = colorectal cancer liver metastasis, IQR = interquartile range, MRI = magnetic resonance imaging, SD = standard deviation

AUAFROC value of DLLD (0.631, CI [0.520, 0.737]) was not significantly different from that of the abdominal radiologists (0.723, CI [0.574, 0.747],  $p = 0.085$ ) or that of the radiology residents (0.660, CI [0.640, 0.805],  $p = 0.584$ ) (Table 2).

In the comparison between DLLD and readers, based on the per-lesion binary classification, the sensitivity of DLLD was not significantly different from that of the abdominal radiologists and radiology residents (81.82%, CI [72.68, 88.39]; 80.81%, CI [73.03, 86.75],  $p = 0.795$ ; 79.46%, CI [70.76, 86.08],  $p = 0.569$ , respectively). The FPP of DLLD was higher than that of the abdominal radiologists and radiology residents (1.330 CI [1.052, 1.681]; 0.357, CI [0.275, 0.464],  $p < 0.001$ ; 0.667, CI [0.531, 0.838],  $p < 0.001$ , respectively) (Table 3).

Effect of threshold modification on diagnostic performance is provided in Supplementary Materials 6,

Supplementary Table 3 and Supplementary Table 4 and lesion-based diagnostic performance of DLLD in the detection and classification of cyst, hemangioma are listed in Supplementary Materials 7 and Supplementary Table 5.

**Sensitivities for Detecting Metastatic Lesions Less Than or More Than 1 cm**

A subgroup analysis for detecting metastatic lesions of < 1 cm ( $n = 23$ ) and > 1 cm ( $n = 76$ ) demonstrated that the sensitivity towards the former was statistically higher than that towards the latter, in DLLD (89.47%, CI [79.56, 94.89], 56.52%, CI [34.67, 76.10],  $p < 0.001$ ), and in abdominal radiologists and radiology residents (90.79%, CI [83.57, 95.03], 47.83%, CI [30.81, 65.36],  $p < 0.001$ ; 91.23%, CI [84.64, 95.15], 40.58%, CI [25.75, 57.35],  $p < 0.001$ , respectively) (Table 4).

**Table 2. Lesion-Based Diagnostic Performance of DLLD and Readers in the Detection of Colorectal Liver Metastasis**

Testee	AUAFROC (95% CI)	DLLD vs. Reader	
		Difference (95% CI)	P
DLLD	0.631 (0.520, 0.737)		
Radiologist		-0.092 (-0.202, 0.014)	0.085
Reader-averaged	0.723 (0.574, 0.747)		
Reader 1	0.738 (0.630, 0.841)		
Reader 2	0.730 (0.632, 0.812)		
Reader 3	0.702 (0.603, 0.798)		
Resident		-0.029 (-0.140, 0.08)	0.584
Reader-averaged	0.660 (0.640, 0.805)		
Reader 4	0.670 (0.558, 0.768)		
Reader 5	0.670 (0.568, 0.774)		
Reader 6	0.641 (0.531, 0.748)		

AUAFROC = area under the alternative free-response receiver operating characteristic curve, CI = confidence interval, DLLD = deep learning-based lesion detection algorithm

**Table 3. Lesion-Based Diagnostic Performance of DLLD and Readers in the Binary Classification of Colorectal Liver Metastasis**

Testee	Sensitivity (%) (95% CI)	DLLD vs. Reader (P)	FPP (95% CI)	DLLD vs. Reader (P)
DLLD	81.82 [81/99] (72.68, 88.39)		1.330 [113/85] (1.052, 1.681)	
Radiologist		0.795		< 0.001*
Reader-averaged	80.81 (73.03, 86.75)		0.357 (0.275, 0.464)	
Reader 1	77.78 [77/99]		0.118 [10/85]	
Reader 2	82.83 [82/99]		0.518 [44/85]	
Reader 3	81.82 [81/99]		0.435 [37/85]	
Resident		0.569		< 0.001*
Reader-averaged	79.46 (70.76, 86.08)		0.667 (0.531, 0.838)	
Reader 4	82.83 [82/99]		1.306 [111/85]	
Reader 5	80.81 [80/99]		0.471 [40/85]	
Reader 6	74.75 [74/99]		0.224 [19/85]	

\*Statistically significant. CI = confidence interval, DLLD = deep learning-based lesion detection algorithm, FPP = false positives per patient

### Patient-Based Diagnostic Performance of DLLD and Readers

In the per-patient binary classification, the sensitivity of DLLD (87.50%, CI [73.30, 94.70]) was not significantly different from that of the abdominal radiologists (85.80%, CI [78.40, 91.00],  $p = 0.79$ ) and radiology residents (85.00%, CI [77.40, 90.30] and  $p = 0.70$ ). However, the specificity of DLLD was 22.22% (CI [12.40, 36.60]), which was lower than that of the abdominal radiologists (66.67%, CI [58.31, 74.10],  $p < 0.001$ ) and radiology residents (55.56%, CI [47.09, 63.70],  $p < 0.001$ ) (Table 5).

### Intra-Group Inter-Observer Agreements among Readers in Two Groups

In the per-lesion binary classification task, diagnoses by three readers in the abdominal radiologist group were consistent in 82.8% of the metastases (82/99), and

the kappa value was 0.6458, indicating a substantial agreement. Similarly, diagnoses by three readers in the radiology residents' group were observed to be consistent in 80.8% of the metastases (80/99), and the kappa value was 0.6080, indicating a substantial agreement.

### DISCUSSION

For evaluating the diagnostic performance of DLLD, we developed a DLLD for detecting liver metastasis of CRC, based on the CNN model, and validated it using a temporally independent cohort of patients initially diagnosed with CRC. The results demonstrated that DLLD showed AUAFROC and sensitivity values comparable to those of abdominal radiologists. The sensitivity of DLLD was similar to that of radiologists in the detection of CRC liver metastasis in a CT scan, which is similar to the sensitivity of a recent meta-

**Table 4. Subgroup Analysis of Diagnostic Performance in the Binary Classification of Colorectal Liver Metastasis**

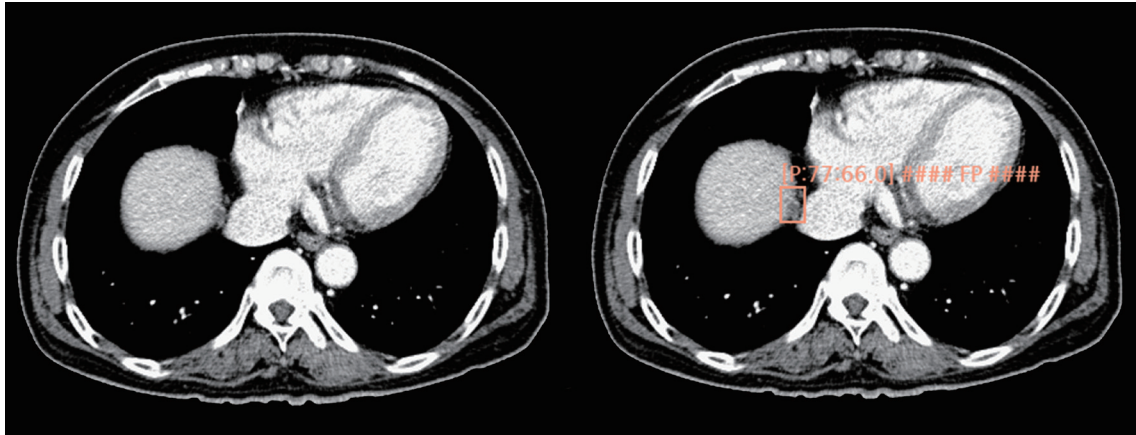
Testee	< 10 mm (n = 23)		≥ 10 mm (n = 76)		
	Sensitivity (%) (95% CI)	DLLD vs. Reader (P)	Sensitivity (%) (95% CI)	DLLD vs. Reader (P)	Sensitivity between Size Subgroup (P)
DLLD	56.52 [13/23] (34.67, 76.10)		89.47 [68/76] (79.56, 94.89)		< 0.001*
Radiologist		0.371		0.722	< 0.001*
Reader-averaged	47.83 (30.81, 65.36)		90.79 (83.57, 95.03)		
Reader 1	43.48 [10/23]		88.16 [67/76]		
Reader 2	47.83 [11/23]		93.42 [71/76]		
Reader 3	52.17 [12/23]		90.79 [69/76]		
Resident		0.132		0.593	< 0.001*
Reader-averaged	40.58 (25.75, 57.35)		91.23 (84.64, 95.15)		
Reader 4	56.52 [13/23]		90.79 [69/76]		
Reader 5	43.48 [10/23]		92.11 [70/76]		
Reader 6	21.74 [5/23]		90.79 [69/76]		

\*Statistically significant. CI = confidence interval, DLLD = deep learning-based lesion detection algorithm

**Table 5. Per-Patient Diagnostic Performance of DLLD and Readers in the Binary Classification of Colorectal Liver Metastasis**

Testee	Sensitivity (%) (95% CI)	DLLD vs. Reader (P)	Specificity (%) (95% CI)	DLLD vs. Reader (P)
DLLD	87.50 [35/40] (73.30, 94.70)		22.22 [10/45] (12.40, 36.60)	
Radiologist		0.790		< 0.001*
Reader-averaged	85.80 (78.40, 91.00)		66.67 (58.31, 74.10)	
Reader 1	82.50 [33/40]		91.11 [41/45]	
Reader 2	90.00 [36/40]		53.33 [24/45]	
Reader 3	85.00 [34/40]		55.56 [25/45]	
Resident		0.700		< 0.001*
Reader-averaged	85.00 (77.40, 90.30)		55.56 (47.09, 63.70)	
Reader 4	90.00 [36/40]		24.44 [11/45]	
Reader 5	87.50 [35/40]		57.78 [26/45]	
Reader 6	77.50 [31/40]		84.44 [38/45]	

\*Statistically significant. CI = confidence interval, DLLD = deep learning-based lesion detection algorithm



**Fig. 4. Example CT images of a female patient aged 78 years diagnosed with ascending colon cancer.** The deep learning-based lesion detection algorithm detects a nodule in the right anteriosuperior liver segment and classifies it into a metastasis class. CT image review reveals that it is a false-positive finding caused by the partial volume effect arising from the diaphragm. CT = computed tomography

analysis (82.1%) [24]. However, DLLD showed a statistically inferior FPP compared to that of the abdominal radiologists and radiology residents. In other words, although it can effectively detect hepatic focal lesions, DLLD may have limitations in its characterization.

This study was conducted on a retrospective cohort of 85 patients initially diagnosed with CRC, who displayed signs of suspected hepatic lesions on the initial staging of CT images and had to undergo further workup. The detection of liver metastasis at this stage has a significant impact on the development of the treatment strategy [3,25]. Thus, the results of the application of DLLD in the cohort helped in evaluating the merits and demerits of using the deep learning-based method. The results demonstrated that DLLD was as sensitive as a radiologist. However, for the characterization of hepatic lesions, verification by the radiologist is necessary owing to the excessive FPP of DLLD.

In the per-patient analysis, the sensitivities of DLLD and readers were similar; however, the specificity of the readers was much higher than that of DLLD. This result was presumed to be related to the deep background knowledge of the radiologists [26,27]. Owing to this difference in knowledge, the proposed DLLD seemed to have a relatively slow learning curve in terms of the characterization of the identified hepatic focal lesions. However, a level of sensitivity comparable to that of the radiologist was identified in this scale of the dataset, this is likely because the sensitivity is dominantly affected by image pattern recognition and therefore does not require prior medical knowledge. Additionally, in several cases, DLLD reported one or two false positives in patients without metastasis. These cases did not have a significant

impact on the per-lesion statistics; however, significantly impacted the per-patient decision. The standalone usage of DLLD without the supervision of a radiologist may lead to an increase in superfluous examinations and surgeries [28]. To prevent this, the DLLD should be used only as an assistant tool until its specificity is sufficiently verified. Notably, the performance comparison between the reader and a combination of reader and DLLD, which has often been used in several deep learning-based studies, may mask the disadvantage of the high FPP of deep learning-based detection methods. The reader can easily ignore the obvious false-positive lesions, such as partial volume artifact, definite cyst, and a cross-section of the bile duct, reported by DLLD (Fig. 4). Thus, the high FPP of DLLD could be obscured when the reader's evaluation is combined. Therefore, it is important to directly compare the FPP of DLLD with that of readers during evaluation.

In the subgroup analysis of lesions dichotomized by a maximum diameter of 1 cm, both DLLD and readers showed statistically significant lower sensitivity in finding metastatic lesions of less than 1 cm. Additionally, there was no significant difference in sensitivity between DLLD and readers in both size subgroups. This is consistent with the results of a previous study on the deep learning-based pulmonary nodule detection algorithm used in chest radiography [13]. The algorithm also cannot effectively detect small metastatic lesions ( $\leq 1$  cm). Therefore, DLLD should be used keeping this fact in mind.

The final diagnoses of the 113 false-positive findings of DLLD in the CRC liver metastasis detection task are summarized in Supplementary Table 6. The first and second causes of the false-positive findings were partial volume



artifacts arising from perihepatic normal structures (37, 32.7%) and the heterogeneity of liver parenchyma (22, 20.4%). These causes account for 53.1% of all false-positive findings of DLLD, and it was easy for radiologists to differentiate these findings from real metastasis. DLLD performed well in detecting all types of focal lesions; however, it could not determine well the class to which the lesion belonged. This may be owing to the difficulty in providing a training set that includes all morphologic forms of all focal lesions.

This study had several limitations. First, the number of patients for DLLD training was relatively small. However, in a recent study on the differential diagnosis of liver masses on CT, it was reported that a CNN trained for the study using only CT images of 460 patients could successfully classify the lesions into five categories [15]. In the present study, a larger number of images were used for training than the reported study. However, further multicenter studies are needed using a learning curve extrapolation method to estimate the required training data size. Second, there was no external validation of the DLLD performance. External validation through further studies is needed to confirm its general applicability. Third, the validation cohort group poses a risk of selection bias since this cohort comprised patients with both staging CT images and consecutive MR images. Most of these patients were those who showed an indeterminate lesion on CT and had taken an MRI scan. Although the validation cohort comprised patients with lesions that were more difficult to diagnose, the readers evaluated the same condition for fair performance comparison between DLLD and readers. Fourth, only the detection markings inside the liver were included in the performance evaluation. Since the detection markings outside the liver were excluded from the false-positive findings, the results may be exaggerated compared to other studies wherein the same procedure was not followed. However, DLLD was developed only for the liver and should only be used in this context. Therefore, extrahepatic detection can be ignored. In future studies, we plan to construct a DLLD that can automatically segment the liver and detect lesions within the segmented area. Fifth, all CT images were obtained more than 8 years back using CT scanners from only two vendors and using only the filtered back-projection algorithm. Further study is needed using CT images from up-to-date models and utilizing various image reconstruction methods from various vendors to overcome unintended time-dependent and vendor biases.

In conclusion, the sensitivity of DLLD was comparable to that of experienced radiologists when detecting liver metastasis in patients initially diagnosed with CRC. However, the FPP of DLLD was higher than that of radiologists. Therefore, DLLD could serve as an assistant tool for detecting liver metastasis instead of being utilized as a standalone diagnostic tool.

## Supplementary Materials

The Data Supplement is available with this article at <https://doi.org/10.3348/kjr.2020.0447>.

## Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

## Author Contributions

Conceptualization: Kiwook Kim, Sungwon Kim, Joon Seok Lim. Data curation: Kiwook Kim, Sungwon Kim, Heejin Bae, Jaeseung Shin. Formal analysis: Kiwook Kim, Sungwon Kim, Kyunghwa Han. Funding acquisition: Sungwon Kim. Investigation: Kiwook Kim, Sungwon Kim. Methodology: Sungwon Kim, Kyunghwa Han. Project administration: Sungwon Kim. Resources: Sungwon Kim. Software: Sungwon Kim. Supervision: Sungwon Kim, Joon Seok Lim. Validation: Sungwon Kim, Joon Seok Lim. Visualization: Kiwook Kim. Writing—original draft: Kiwook Kim, Sungwon Kim. Writing—review & editing: all authors.

## ORCID iDs

Kiwook Kim

<https://orcid.org/0000-0002-8798-5038>

Sungwon Kim

<https://orcid.org/0000-0001-5455-6926>

Kyunghwa Han

<https://orcid.org/0000-0002-5687-7237>

Heejin Bae

<https://orcid.org/0000-0002-1227-8646>

Jaeseung Shin

<https://orcid.org/0000-0002-6755-4732>

Joon Seok Lim

<https://orcid.org/0000-0002-0334-5042>

## REFERENCES

1. Qiu M, Hu J, Yang D, Cosgrove DP, Xu R. Pattern of distant

- metastases in colorectal cancer: a SEER based study. *Oncotarget* 2015;6:38658-38666
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424
  3. Manfredi S, Lepage C, Hatem C, Coatmeur O, Faivre J, Bouvier AM. Epidemiology and management of liver metastases from colorectal cancer. *Ann Surg* 2006;244:254-259
  4. Hackl C, Neumann P, Gerken M, Loss M, Klinkhammer-Schalke M, Schlitt HJ. Treatment of colorectal liver metastases in Germany: a ten-year population-based analysis of 5772 cases of primary colorectal adenocarcinoma. *BMC Cancer* 2014;14:810
  5. Engstrand J, Nilsson H, Strömberg C, Jonas E, Freedman J. Colorectal cancer liver metastases - a population-based study on incidence, management and survival. *BMC Cancer* 2018;18:78
  6. Valls C, Andía E, Sánchez A, Gumà A, Figueras J, Torras J, et al. Hepatic metastases from colorectal cancer: preoperative detection and assessment of resectability with helical CT. *Radiology* 2001;218:55-60
  7. Scheer A, Auer RA. Surveillance after curative resection of colorectal cancer. *Clin Colon Rectal Surg* 2009;22:242-250
  8. Dhir M, Sasson AR. Surgical management of liver metastases from colorectal cancer. *J Oncol Pract* 2016;12:33-39
  9. Bilello M, Gokturk SB, Desser T, Napel S, Jeffrey RB Jr, Beaulieu CF. Automatic detection and classification of hypodense hepatic lesions on contrast-enhanced venous-phase CT. *Med Phys* 2004;31:2584-2593
  10. Schwier M, Moltz JH, Peitgen HO. Object-based analysis of CT images for automatic detection and segmentation of hypodense liver lesions. *Int J Comput Assist Radiol Surg* 2011;6:737-747
  11. Chi Y, Zhou J, Venkatesh SK, Huang S, Tian Q, Henedige T, et al. Computer-aided focal liver lesion detection. *Int J Comput Assist Radiol Surg* 2013;8:511-525
  12. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316:2402-2410
  13. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology* 2019;290:218-228
  14. Wang J, Yang X, Cai H, Tan W, Jin C, Li L. Discrimination of breast cancer with microcalcifications on mammography by deep learning. *Sci Rep* 2016;6:27327
  15. 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;286:887-896
  16. Vivanti R, Joskowicz L, Lev-Cohain N, Ephrat A, Sosna J. Patient-specific and global convolutional neural networks for robust automatic liver tumor delineation in follow-up CT studies. *Med Biol Eng Comput* 2018;56:1699-1713
  17. Yan K, Bagheri M, Summers RM. 3D context enhanced region-based convolutional neural network for end-to-end lesion detection. International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2018:511-519
  18. Horton KM, Abrams RA, Fishman EK. Spiral CT of colon cancer: imaging features and role in management. *Radiographics* 2000;20:419-430
  19. Seo N, Park MS, Han K, Lee KH, Park SH, Choi GH, et al. Magnetic resonance imaging for colorectal cancer metastasis to the liver: comparative effectiveness research for the choice of contrast agents. *Cancer Res Treat* 2018;50:60-70
  20. Redmon J, Farhadi A. Yolov3: an incremental improvement. arXiv preprint 2018;arXiv:1804.02767
  21. de Hoop B, De Boo DW, Gietema HA, van Hoorn F, Mearadji B, Schijf L, et al. Computer-aided detection of lung cancer on chest radiographs: effect on observer performance. *Radiology* 2010;257:532-540
  22. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. *Med Phys* 2004;31:2313-2330
  23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174
  24. Choi SH, Kim SY, Park SH, Kim KW, Lee JY, Lee SS, et al. Diagnostic performance of CT, gadoxetate disodium-enhanced MRI, and PET/CT for the diagnosis of colorectal liver metastasis: systematic review and meta-analysis. *J Magn Reson Imaging* 2018;47:1237-1250
  25. Adam R, Delvart V, Pascal G, Valeanu A, Castaing D, Azoulay D, et al. Rescue surgery for unresectable colorectal liver metastases downstaged by chemotherapy: a model to predict long-term survival. *Ann Surg* 2004;240:644-657; discussion 657-648
  26. 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;210:71-74
  27. 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;158:535-539
  28. Garden OJ, Rees M, Poston GJ, Mirza D, Saunders M, Ledermann J, et al. Guidelines for resection of colorectal cancer liver metastases. *Gut* 2006;55 Suppl 3:iii1-8